Slice-mask based 3D Cardiac Shape Reconstruction from CT volume

Xiaohan Yuan[1], Cong Liu[1], Fu Feng[1], Yinsu Zhu[2], and Yangang Wang[1⋆]

[1] School of Automation, Southeast University
[2] Dept. of Radiology, the First Affiliated Hospital of Nanjing Medical University

Abstract. An accurate 3D ventricular model is essential for diagnosing and analyzing cardiovascular disease. It is challenging to obtain accurate patient-specific models on scarce data via widely accepted deep-learning methods. To fully use the characteristics of medical volume-based images, we present a slice-mask representation to better regress the parameters of the 3D model. A data synthesis strategy is proposed to alleviate the lack of training data by sampling in the constructed statistical shape model space and obtaining the corresponding slice-masks. We train the end-to-end structure by combining the segmentation and parametric regression modules. Furthermore, we establish a larger left ventricular CT dataset than before, which fills the gap in relevant data of the healthy population. Our method is evaluated on both synthetic data and real cardiac scans. Experiments demonstrate that our method can achieve advanced results in shape reconstruction and segmentation tasks. Code is publicly available at https://github.com/yuan-xiaohan/Slice-mask-based-3D-Cardiac-Shape-Reconstruction.

Keywords: 3D Reconstruction · Segmentation · Cardiac CT.

1 Introduction

The heart is one of the vital organs of our body, and cardiovascular disease is the leading cause of death and morbidity worldwide [23]. The left ventricle (LV) is the most important chamber of the heart and main source of blood flow. However, current disease diagnosis and assessment are always guided by slice-based 2D images, making it difficult for clinicians to obtain intuitive patient-specific visualizations and resulting in inaccurate estimates of clinical metrics such as volume and ejection fraction. Therefore, boosting the limited 2D images with the prior of 3D heart shapes, and using it to instantiate cardiac models has important applications for surgical planning, morphological assessment, and educational purposes. Although recent years have witnessed the rapid progress of

⋆ Corresponding author: Yangang Wang. 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.

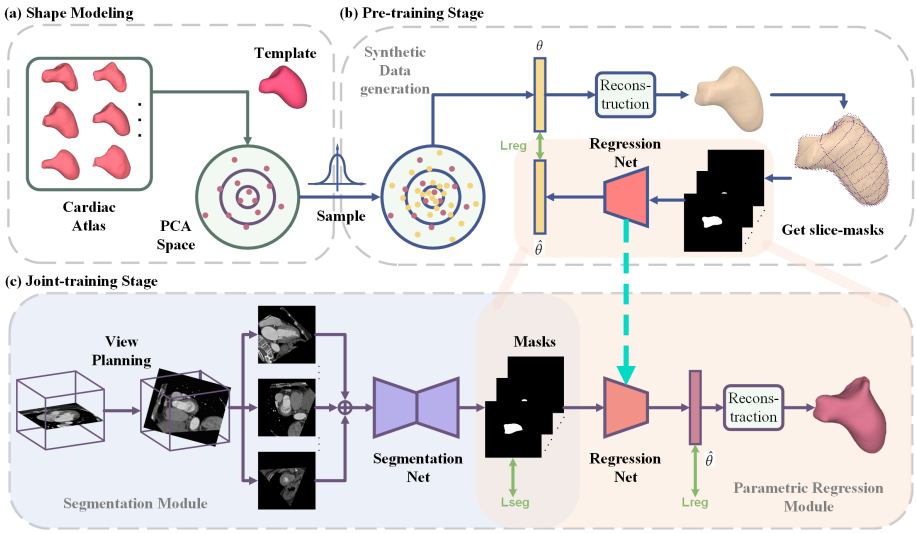

Fig. 1. The overall pipeline of the proposed framework. (a) Shape modeling. We first generate a cardiac atlas from the real meshes and construct a statistical shape model by PCA, resulting in low-dimensional parameter space. (b) In the pre-training stage, the parametric regression module is trained by generating massive synthetic data, consisting of reasonable model parameters and their corresponding slice masks. (c) Joint-training stage includes a segmentation module and a pre-trained parametric regression module. The input to the segmentation module is the set of view-planned slices, and the output is their masks, which are used as the input to the parametric regression module, finally obtaining the shapes.

deep learning in the field of medical image analysis, especially for the segmentation and reconstruction tasks, the lack of scale of annotated datasets impedes the generalization of trained models for 3D heart reconstruction. In this paper, we focus on the problem of accurate 3D reconstruction of the left ventricle in response to the data-poor dilemma.

Generally, existing 3D medical reconstruction methods often require fine-scale segmentation masks, mainly focusing on the precursor task of medical image segmentation [12]. The methods of directly generating a mesh from masks depend highly on the segmentation quality and often produce surfaces with stair-step artifacts that are affected by the low resolution of medical images. To avoid the lack of mesh integrity, a model-driven strategy, as a parametric method, is often addressed. The segmented contour is fitted by the constructed statistical shape model (SSM) to complete the shape prediction [8,10,13], but the separated multi-stage iterative strategy will greatly increase the computational cost. Recently, a few pioneers have devoted the organ geometry reconstruction of point clouds or meshes directly from images or volume with the deep neural network [6, 30, 32]. These works are often derived from the field of computer vision and lack consideration of the characteristics of medical image data.

In this paper, to fully exploit the volume characteristics of medical images, our key idea is to slice the three-dimensional SSM model from different views into a set of two-dimensional mask images via the technique of computer graphics, where the masks are named as **slice-mask**s. It is noted that the view we obtain is not a projection but a typical angle of slices in the volume, which is different from traditional multi-view reconstruction in the area of computer vision. In order to conveniently describe the anatomical cardiac model, the views of slices are predefined and chosen similar to the clinical strategy.

However, the scarcity of public datasets for cardiac CT images brings great difficulties to network training. To improve the variance of cardiac SSM parameters, we first construct the SSM space of the cardiac model and then sample the SSM parameters to synthesize massive 3D cardiac models that conform to the real anatomy, enlarging the size of the training set. The slice positions and 3D model landmarks can be easily defined in the SSM template model to obtain view-consistent slice-masks for all synthetic models. In such a manner, we can augment more slice-masks corresponding to the model without relying on the raw CT data.

With the obtained synthetic training data, the SSM parameter regression network is concatenated with a segmentation network to predict the SSM parameters. It is noted that these two modules might mismatch due to the imbalance of real and synthetic training data. To alleviate this circumstance, we propose a refining step to improve the accuracy of the segmentation module due to the anatomical constraints of the reconstruction task.

To further tackle the obstacle of data scarcity and fill the gap with normal human cardiac data, we increase the amount of real data by collecting a larger-scale healthy left ventricular CT dataset than most existing ones [14, 31, 35]. A more accurate low-dimensional parametric model is generated to facilitate the cardiac parameters regression and network training.

In summary, the main contributions of this work are as follows.

- Combining the characteristics of medical volume data with the anatomical knowledge of the heart, we use the proposed slice-mask representation to better regress the parameters of the 3D model. We synthesize massive models by sampling in a statistical shape model space and obtaining the corresponding slice-masks to alleviate the lack of training data.
- We design a training strategy for improving the accuracy of shape reconstruction and segmentation, where the end-to-end network consists of a segmentation module and a parameter regression module.
- We build a larger-scale CT atlas of the left ventricle than previous work, making up for the current scarcity of data, especially in healthy individuals.

2 Related work

Parametric Shape Reconstruction. Traditional 3D medical reconstruction methods often require fine-scale segmentation masks, mainly focusing on the precursor task of medical image segmentation. Marching Cube algorithm [19] is

then typically utilized to generate a mesh of segmented regions from the contours delineated from each layer of the image volume. Operations such as smoothing [3, 15] are often performed at the last step. Such methods cannot ensure the integrity of the mesh. Therefore, a model-driven strategy is often addressed, which usually completes the shape prediction by constructing a statistical shape model (SSM) to fit the initial contour of the segmentation [8,10,13]. Some methods for shape prediction based on segmentation graphs introduce shape priors into the segmentation tasks in the previous stage to ensure better results in the downstream reconstruction [5, 7, 18, 36]. For example, in the case of image artifacts, the constructed 3D cardiac shape can be ensured to have anatomical significance. [26] enforced robustness of shape prediction by simultaneously performing semantic segmentation, which is performed by regression of signed distance maps, trained using a loss function incorporating both distance and overlap measures. Other methods improve the reconstruction performance by directly introducing shape prior knowledge into the reconstruction task [1, 4, 27, 33]. Most recently, with the development of deep learning, methods combining SSM with a convolutional neural network(CNN) can achieve better results. Zhou et al. [33] borrowed the PointOutNet [9] to learn the relationship between a 2D image and a 3D SSM in a single stage for 3D shape prediction. Regression of shape coefficients using a CNN was performed by Bhalodia et al. [4] and Adams et al. [1] extended this work to a probabilistic approach to determine the credibility of the model output by quantifying uncertainty. Probabilistic surface prediction with a PCA shape prior was also performed by Tóthová et al. [27], the input of the network is three orthogonal standard MR views. Attar et al. [2] proposed a deep neural network using both CMR images and patient meta-data to directly predict 3D cardiac shape parameters instead of a pixel-wise classification across each 2D slice. This method uses the promising ability of SSM to simplify shape complexity. However, it needs over 3000 CMR image volumes with manual delineations to construct reference 3D shapes for training, and the patient metadata is often challenging to obtain.

Nonparametric Shape Reconstruction. Shape reconstruction methods in the form of a nonparametric model are generally based on deep learning. They can directly predict the surface mesh of cardiac structures from image or volume data. A series of works on recovering 3D shapes from 2D images have emerged in computer vision, usually consisting of an encoder that extracts image features and an encoder-decoder that generates grids [20, 29, 32]. PointOutNet [9] can generate unordered 3D points from a single RGB image cloud and 3D-LMNet [20] can utilize an image encoder to map a 2D image into a 3D point cloud latent space learned by an auto-encoder. Ye et al. [32] proposed a network that directly reconstructs LV from the volume of 2D CT slices and generates its segmentation masks from the predicted 3D point cloud. Wang et al. [29] used information extracted from single-view 2D medical images to predict the displacement of control points to learn the spatial deformation of lung organs. In recent years, graph convolutional neural networks have also shown promise for surface mesh reconstruction [16,30]. According to the volumetric properties of medical images,

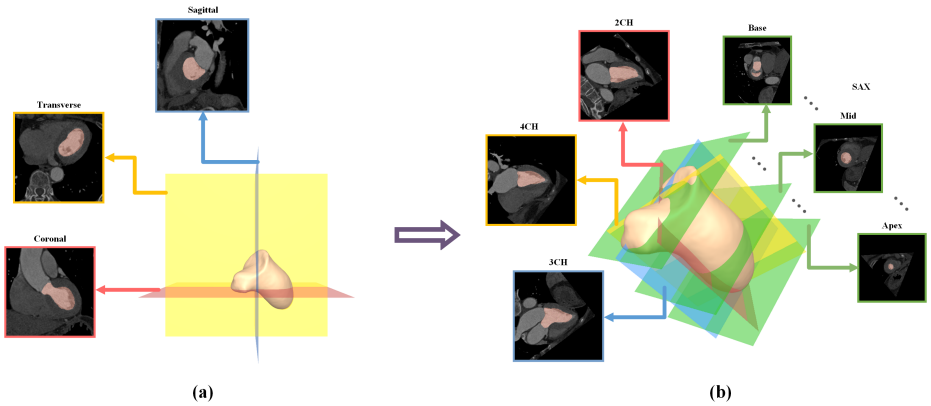

Fig. 2. Visual representation of the relative relationship between slice positions and ventricular model. Shapes with different color planes represent different views. On the left are the slices directly imaged by CT, and on the right are the slices we selected according to the view planning.

Wickramasinghe et al. [30] extended the Pixel2Mesh [28] from 2D images to 3D surface meshes, taking 3D volumes as input to solve the reconstruction problem of CT liver. Kong et al. [17] proposed a network that learns to deform a template grid into a volume of input image data by predicting the displacement of a multi-resolution control point grid.

3 Method

Figure 1 provides an overview of the proposed framework. First, we build a 3D model atlas of the left ventricle by manually segmenting and generating the meshes, then construct a low-dimensional parameter space using principal component analysis (PCA) (see Section 3.1). By sampling in the latent space, we make corresponding slice-masks for numerous reasonable samples generated, which are used to train the parametric regression network (see Section 3.2). After obtaining the pre-training weights of the parameter regression network, we add it to the joint architecture and optimize it with the segmentation network (see Section 3.3).

The joint architecture consists of two modules: the input CT volume is sliced into planes after view planning and used as the input to the segmentation module. The segmentation output will be used as the input to the parameter regression module, which directly estimates the model parameters.

3.1 Shape Modeling

Building an atlas is the first step in modeling the shape and constructing the parameter space. Benefiting from the high-resolution features of CT, we manually delineate each layer of the CT volume, so that the generated 3D model

has sufficiently high fidelity. Suppose that the heart model training set $\mathbb{D} = \{\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_M\}$ has M models, and different models have different vertices. The parametric model is mainly determined by the shape parameter $\alpha \in \mathbb{R}^K$. Since there is a large difference in the size of the heart at end-diastole and end-systole, to decouple it from shape, we add a size parameter $\beta \in \mathbb{R}^1$. As long as the overall parameter $\theta = [\alpha, \beta] \in \mathbb{R}^{K+1}$ is obtained, a heart model can be reconstructed.

We first use the Iterative Closest Point (ICP) algorithm to align each model with the template, and use the method proposed in [25] to deform the template onto each model to obtain a set of heart models with the same topology $\mathbb{S} = \{\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_M\}$, and $\mathbf{S}_i \in \mathbb{R}^{3\times N}, i \in 1 \cdots M$, where N is the number of vertices of the template. After the atlas is constructed, its parameter latent space can be obtained through statistical analysis. We use Principal Component Analysis (PCA) to construct the shape parameters of the model on the set $\mathbb{S}$ which is just gained by the registration. Map the shape $\mathbf{S}_i$ to vector $\mathbf{s}_i = [x_1, y_1, z_1, \cdots, x_N, y_N, z_N]^T \in \mathbb{R}^{3N}$, then let $\mathbf{S}_{map} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_M]^T \in \mathbb{R}^{M\times 3N}$ and the mean shape $\bar{\mathbf{s}} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{s_i}$. Through the singular value decomposition (SVD) of $\mathbf{S}_{map}$:

$$\mathbf{S}_{map} = \mathbf{U}\sum \mathbf{V}^T, \tag{1}$$

we get $\mathbf{V} \in \mathbb{R}^{3N\times K}$ that fully defines the linear function below:

$$\mathcal{B}(\theta) = \beta \cdot \mathcal{M}(\bar{\mathbf{s}} + \mathbf{V}\alpha), \tag{2}$$

where $\theta = [\alpha, \beta] \in \mathbb{R}^{K+1}$ denotes the shape coefficients, and the operator $M(\mathbf{s}) : \mathbb{R}^{3N} \mapsto \mathbb{R}^{\mathcal{3}\times N}$ maps the vector to the shape. The function $\mathcal{B}(\theta)$ produces the shapes of different hearts.

3.2 Data Synthesis Strategy

Based on the parametric models that have been proposed, we analyze the parameter distribution of real data and generate multivariate Gaussian distribution centered on real samples for sampling. The resampled parameters can obtain different hearts from the original training set while maintaining anatomical consistency and fidelity. We will introduce the parametric regression network and how to obtain the slice-masks for each generated model.

Parametric Regression Network. In this work, we regard the CT heart modeling problem as a multi-view reconstruction problem and hope to regress the parameters directly from the slice-masks.

Typically, raw CT images are scanned in standard planes of the body (transverse, sagittal, and coronal). However, due to the unique shape and location of the heart, it cannot be simply expressed by the standard axis of the human body (as shown in Figure 2(a)). Most existing methods do not consider the view problem but directly use the entire volume or images with the original view as input.

The former will bring the challenge of memory computing, and the latter cannot effectively contain the critical information for reconstructing the structure. Therefore, we propose to extract as much information as possible from a few slices from different perspectives. The process of obtaining these slices from the raw CT volume is called view planning, and the masks obtained from this series of slices are named slice-masks. The selection of slice locations is based on the anatomical structure of the heart: long-axis views (two-chamber (2CH), three-chamber (3CH), and four-chamber (4CH) heart) and short-axis views (SAX). We use the above-mentioned unique perspectives for the heart, equivalent to setting the "camera" in a place that can represent heart information more abundantly. The parameter regression network seeks the mapping relationship between slice-masks and parameter space, which significantly reduces the information redundancy caused by useless background. Figure 2(b) shows the relative position relationship between the position of each slice and the ventricular model. It can be seen visually that these perspective views cover the whole heart, and the segmented slices obtained in this way can comprehensively describe the shape of the ventricle.

For the parametric regression network, the input is 13 concatenated slice-masks (3 long-axis views and 10 short-axis views). Considering that the model parameters are generated by PCA operation, the coefficient corresponding to the eigenvector with a larger variance contribution is more important and can better reflect the topology of the model. Therefore, we use the weighted two-norm function to analyze the model parameters output by the network, which is,

$$\mathrm{L}_{reg} = \lambda_1 \cdot \sum_{i=1}^{K} (\alpha_i - \hat{\alpha_i})^2 \cdot \mathrm{w}_i + \lambda_2 \cdot (\beta - \hat{\beta})^2 \tag{3}$$

For supervision, where the weight w_i is positively related to the variance ratio of the feature vector, α_i and $\hat{\alpha_i}$ is the true value and predicted value of the shape parameter in the ith dimension, respectively. β and $\hat{\beta}$ is the true value and predicted value of the model size, respectively. λ_1 and λ_2 represent the influence of the shape term and the size term, respectively.

Acquisition of Slice-masks. According to the structural characteristics of the proposed regression network, the training data it needs is the pair of the slice-masks and the ground-truth parameters, so we propose a method to acquire the slice-masks of any generated model.

We will design the slices according to the given rules for each model to obtain the masks (See Supplementary Material for details). To plan the views directly, it is necessary to find the positional relationship between these slice-masks and the model. It is worth noting that the equations of slice planes are determined by several landmarks on the model, such as the apex (AP), the center of the mitral valve (MV), and the center of the aortic valve (AV). These anatomical landmarks are related to the vertices of the shape, and there is a one-to-one correspondence between the vertices of each parametric model. So

as long as there are three landmark indexes, one slice of the model can be directly determined. We ask professional doctors to manually determine a series of landmarks on the template model and then propagate it to any other generated shapes. Thus, we can obtain the slice-masks of the synthetic models and train the parameter regression network from these.

3.3 Training Strategy

Considering the scarcity of CT data and its annotations, direct training in the architecture of a joint network will cause specific difficulties. So we design a strategy for staged training. We first train the parameter regression network separately using the synthetic data to enhance its generalization performance. We place the pre-trained parametric regression network after the segmentation network and estimate the parameters directly from the former segmentation results. However, the pre-trained subnetwork uses perfect masks sliced from the ground-truth model during its training phase. The masks estimated from the segmentation network often contain artifacts that interfere with the parameter estimation. Therefore, combining the two networks and fine-tuning the parameter estimation network to adapt to this situation is necessary.

Here, the process of view planning for CT volume is similar to that of MR standard [24] (See Supplementary Material for details) to obtain the input of the segmentation network. We choose 2D UNet as the architecture, using cross-entropy loss L_{seg}. When two networks are coupled, exploiting their respective tasks to promote each other: adding segmentation guidance to reconstruction and adding reconstructed anatomical constraints to segmentation can lead to improved performance on both tasks. The network is optimized by a joint loss as follows:

$$\mathrm{L}_{joint} = \lambda \cdot \mathrm{L}_{seg} + \mu \cdot \mathrm{L}_{reg} \tag{4}$$

where L_{seg} is the segmentation loss and L_{reg} is the parametric regression loss. Their details will be further introduced in later sections, and λ and μ are the weights of them.

4 Experiments

4.1 Data

Table 1. A brief review of previous literature on CT cardiac atlases.

Methods	Size	Subjects
Ecabert et al. [8]	13	patients
Ordas et al. [22]	100	healthy/patients
Hoogendoorn et al. [13]	138	patients
Ours	**225**	**healthy**

Table 1 compares different studies built on CT cardiac atlases. The scale of the existing atlases is so small to directly build a parametric model and the statistical shape modeling method usually relies on the database of a healthy population. In order to fill the gap in relevant data and lay the foundation for in-depth research on the function of healthy hearts, we recruited more than 50 volunteers for cardiac CT scans.

Our raw cardiac CT images were collected in Jiangsu Province Hospital, and the whole volume of data covered the entire heart structure, ranging from the upper abdomen to the aortic arch. Five professional doctors were organized to label and review and finally obtained more than 200 left ventricular models of different phases. Different time phases reflect different states of heart movement, which increases the diversity of data.

4.2 Implementation Details and Evaluation Metrics

The segmentation module adopts the UNet structure [24], the input is 13 slices, and the size of each slice is scaled to 192×192. The parameter regression module adds fully connected layers to the encoder architecture of UNet, and finally outputs (50+1) parameters. To facilitate network regression, we normalize each dimension of the parameters. We use 3090Ti GPU for training, Adam optimizer with batch size 4, and initial learning rate set to 1e-4. In the joint loss, the weight of the segmentation loss λ is 1, and the weight of the parameter regression loss μ is 10.

For the segmentation task, we use the Dice Similarity Coefficient (DSC), intersection over union (IOU), and Hausdorff Distance (HD) for evaluation. For the reconstruction task, we use Mean Surface Distance (MSD) and Chamfer Distance (CD) for evaluation.

4.3 Ablation Study

We divided the data into a "real training set" (149 objects, which serves as a prior for building a parametric model, whose parameters are treated as ground-truth), and a "real test set" (76 objects, ground-truth without parameters). The synthetic data was also divided into a "synthetic training set" (3000 objects) and a "synthetic test set" (500 objects).

Data Synthesis Strategy Effectiveness. To demonstrate the data synthesis strategy, we trained the parameter regression network using only real data as the training set and adding synthetic data as the training set, respectively. As shown in Figure 2, both can achieve overfitting on the real training set, which shows that the idea of the network regressing parameters from a few segmentation masks is feasible. After adding the generated data to the training set, the performance on the synthetic test set is very obvious: the network after data augmentation shows better generalization performance.

Table 2. Results of the parameters MSE loss (Lreg) and mean surface distance (MSD) with or without synthetic data training, respectively. *R-Train* is to train on the real dataset, and *R+S-Train* is on both the real dataset and synthetic dataset; *R-Test* is the real test set, and *S-Test* is the synthetic test set.

Training set	Lreg		MSD	
	R-Test	S-Test	R-Test	S-Test
R-Train	0.001	0.015	0.927	1.557
R+S-Train	**0.0003**	**0.002**	**0.669**	**0.625**

Table 3. Results of different inputs to the regression network. *Tran* represents the raw transverse slices, *SAX* and *LAX* are the short-axis and long-axis slices after view planning, respectively. The number in parentheses indicates the number of slices of the input network.

Input	Tran(13)	SAX(10)	LAX(3)	SAX+LAX(13)
MSD	2.292±1.393	1.793±0.887	1.012±0.460	**0.926±0.354**
CD	4.693±2.92	3.585±1.669	2.083±0.775	**1.960±0.606**

Slice-masks Selection. We compared the effects of different slice-masks as inputs on the reconstruction results, as shown in Table 3. Using the raw transverse images obtained directly from CT is the choice of most existing work, but a small number of these views can not achieve ideal results. The selection of slice position is crucial, and LAX contains the most information because it is obtained according to the characteristics of heart structure. Taking both SAX and LAX perspectives as the input of the network can enhance the information obtained by the network.

Joint Optimization Effectiveness. We verified the effectiveness of the joint optimization of the segmentation module and the parametric regression module. As shown in Figure 3, whether the pre-training model is added or not affects the segmentation loss and parameter regression loss, respectively. In general, the loss of the joint network decreases with or without the addition of a pre-trained model. Adding one pre-training module will make the loss of another module drop more smoothly and speed up the training. Comparing the two loss images, it can be found that the regression loss is more volatile than the segmentation loss because our training of the parametric regression module alone is more "ideal", assuming that the inputs are all accurate segmentations. In practice, the segmentation estimated from the segmentation network is often not that imperfect. Therefore, it can be seen from Figure 4 (a) that if the segmentation module and the parameter regression module are entirely separated and optimized separately, the accuracy of the parameter regression will be significantly affected by the segmentation results, which reflects the advantages of the joint network. By optimizing simultaneously, this incongruity can be neatly balanced. The supervision of the parameter regression module will promote the improvement of segmentation accuracy, and the addition of the segmentation module

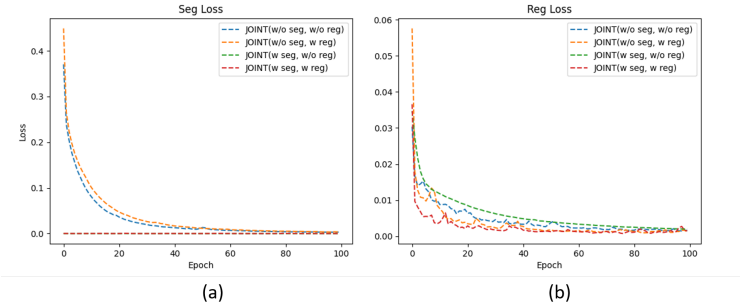

Fig. 3. Jointly optimized training. (a) and (b) are the curves of segmentation loss and parametric regression loss during training, respectively. Each figure shows the results of four training modes, where *seg* is the segmentation pre-training model, and *reg* is the parameter regression pre-training model.

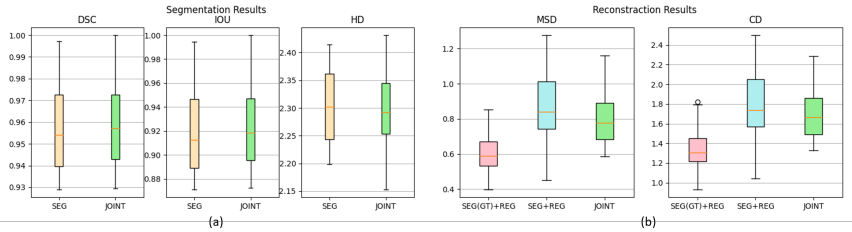

Fig. 4. The effectiveness of joint optimization. (a) and (b) are the performance results of different models on segmentation and reconstruction tasks, respectively. Among them, *SEG* is the segmentation model, *SEG(GT)+REG* represents that the input of the parameter regression module are the ground-truth masks, *SEG+REG* represents that the two modules are trained separately, and *JOINT* is the joint model.

can force the parameter regression module to learn parameters from imperfect segmentation, as shown in Figure 4 (b).

4.4 Comparison Experiments

- Raw images as input(**R-FCN**): We refer to the method from [2] and directly use the raw images as input to obtain shape parameters.
- Template fitting approach(**TF**): We use a two-stage method like [21] by minimizing the point-to-point distance between the obtained surface contour and the template mesh.
- Template fitting approach(**Voxel2Mesh**): A surface reconstruction method [30] based on GCN for 3D medical volumes.
- UNet and its variant versions: The classical medical image segmentation network (**UNet**) proposed by [24]. An encoder-decoder network (**UNet++**) with deep supervision and dense skip pathways [34]. A Context Encoder

Table 4. Results on the reconstruction task.

Methods	MSD(mm)	CD(mm)	Runtime(s)
R-FCN	2.004±0.920	4.014±1.693	**0.040**
TF	0.867±0.353	1.874±0.673	20.056
Voxel2Mesh	2.127±0.664	4.321±1.215	4.868
Ours-UNet++	1.021±0.577	2.135±1.192	0.113
Ours-CE-Net	0.937±0.328	1.936±0.564	0.085
Ours-UNet	**0.862±0.341**	**1.806±0.590**	0.106

Table 5. Results of the segmentation task on the specific views.

Methods	DSC(%)	IOU(%)	HD(mm)
R-FCN	0.876±0.063	0.787±0.097	2.687±0.237
TF	0.938±0.023	0.907±0.042	2.324±0.341
UNet++	0.917±0.114	0.864±0.129	2.416±0.482
Ours-UNet++	0.920±0.144	0.867±0.153	2.413±0.608
CE-Net	0.945±0.022	0.900±0.037	2.368±0.207
Ours-CE-Net	0.946±0.022	0.900±0.039	2.359±0.233
UNet	0.948±0.025	0.886±0.037	2.397±0.194
Ours-UNet	**0.952±0.025**	**0.907±0.041**	**2.321±0.347**

Network (**CE-Net**) to capture more high-level information and preserve spatial information for 2D medical image segmentation [11].

– **Ours**: Based on the above proposed UNet and its variants, our proposed parametric regression network is subsequently connected. To speed up the convergence, we train the parametric regression network with real and synthetic data and the segmentation network with real images and masks, respectively. The two networks are then concatenated to optimize the overall structure.

Results on the Reconstruction Task. Table 4 are the quantitative results of various methods on the reconstruction task, and our method is comparable to the TF method in accuracy. Combined with Table 5, it can be seen that the performance of the upstream segmentation task inevitably affects the results of the downstream reconstruction task. The segmentation network in our framework can be replaced with any SOTA structure; if the segmentation result is accurate enough, a better reconstruction effect can be obtained. The purpose of our joint optimization strategy is to minimize this effect.

Visualizations and error distributions reconstructed by several methods are presented in Figure 5. For the R-FCN and Voxel2Mesh, a large amount of training data must be used to achieve good results, and the lack of raw CT data greatly restricts it. Since, in the training phase, we use synthetic data that does not depend on the raw images, avoiding this dilemma and achieving better results under the same conditions. In addition, taking the overall volume as the input of the network requires down sampling to meet the memory needs, which

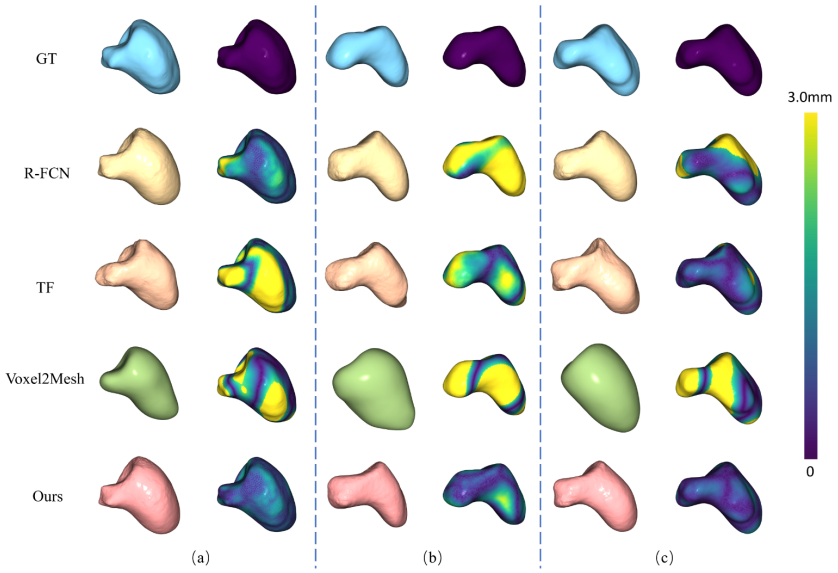

Fig. 5. Reconstruction results. Each alphabetical number represents an example, the first column of each example is the visualization, and the column row is the error distribution.

will lose detailed information. Although the error of the classical TF method is also tiny, it depends highly on the first step segmentation. The iterative process will take more time, reflecting the end-to-end efficiency of deep learning. In comparison, our inference speed is several times faster than iterative methods.

Results on the Segmentation Task. As shown in the table 5, the combination of Unet and regression module (Ours-UNet) achieves the best performance among all methods. Compared with the results of various optimization methods and their jointly optimized versions, the latter has a specific improvement in the segmentation effect. It shows that the supervision of the downstream parameter regression module can improve the performance of the upstream segmentation task. Figure 6 shows the visualization results of different methods. Our method can fit the contour of the ground-truth (due to the superposition of the two colors, it will appear almost white in the figure), while R-FCN and TF are pretty different from the ground-truth as a whole, sacrificing many details. Although the pixel classification-based method of UNet can also achieve high accuracy in numerical results, it is prone to some evident and non-physical flaws from the details shown in the orange box. Because a pure segmentation network only predicts from a few slices, it cannot incorporate the overall shape prior. In contrast, our method eliminates these flaws, balances the integrity of the model-based method with the fineness of classification-based segmentation, optimizes the overall contour, and performs numerically the best.

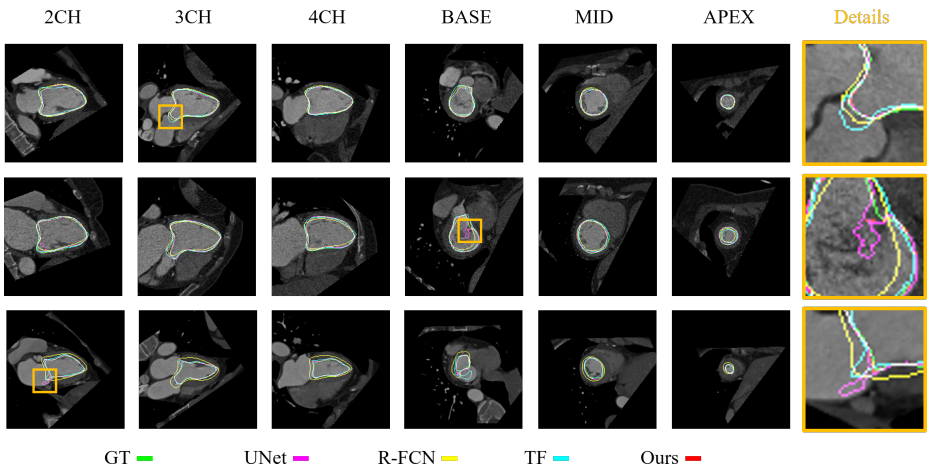

Fig. 6. Segmentation results on the specific views. Each row shows an example, and each column represents a different view from left to right: 2CH, 3CH, and 4CH slices on LAX; basal, middle, and apex slices on the SAX. The different colored outlines in the figure denote the results of different methods. The last column shows an enlarged version of the orange box in the figure.

5 Conclusion

Obtaining an accurate ventricular model from medical images is of great clinical value. Given the challenge that the network cannot be well-trained due to the lack of publicly labeled cardiac images, according to the characteristics of medical volume data, we propose a slice-mask representation. We alleviate the shortage of training data by sampling in the constructed parameter latent space and increasing the network's generalization performance. The segmentations network is connected with the pre-trained parameter regression network for end-to-end joint optimization to reduce the impact of the imperfection of the upstream task on the downstream task. We evaluate our proposed method on synthetic data and real cardiac scans. The results show that our method can achieve advanced shape reconstruction and segmentation tasks. We hope our dataset can support related fields, and we will expand our method to more organs in the future.

Acknowledgements This work was supported in part by the Natural Science Foundation of Jiangsu Province (No. BK20220127), the National Natural Science Foundation of China (No. 62076061), the "Young Elite Scientists Sponsorship Program by CAST" (No. YES20200025), and the "Zhishan Young Scholar" Program of Southeast University (No. 2242021R41083).

References

1. Adams, J., Bhalodia, R., Elhabian, S.: Uncertain-deepssm: From images to probabilistic shape models. In: International Workshop on Shape in Medical Imaging. pp. 57–72. Springer (2020)
2. Attar, R., Pereañez, M., Bowles, C., Piechnik, S.K., Neubauer, S., Petersen, S.E., Frangi, A.F.: 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. pp. 586–594. Springer (2019)
3. Augustin, C.M., Neic, A., Liebmann, M., Prassl, A.J., Niederer, S.A., Haase, G., Plank, G.: Anatomically accurate high resolution modeling of human whole heart electromechanics: a strongly scalable algebraic multigrid solver method for nonlinear deformation. Journal of computational physics **305**, 622–646 (2016)
4. Bhalodia, R., Elhabian, S.Y., Kavan, L., Whitaker, R.T.: Deepssm: a deep learning framework for statistical shape modeling from raw images. In: International Workshop on Shape in Medical Imaging. pp. 244–257. Springer (2018)
5. Boussaid, H., Rouet, L.: Shape feature loss for kidney segmentation in 3d ultrasound images. In: BMVC (2021)
6. Cai, J., Xia, Y., Yang, D., Xu, D., Yang, L., Roth, H.: End-to-end adversarial shape learning for abdomen organ deep segmentation. In: International Workshop on Machine Learning in Medical Imaging. pp. 124–132. Springer (2019)
7. Duan, J., Bello, G., Schlemper, J., Bai, W., Dawes, T.J., Biffi, C., de Marvao, A., Doumoud, G., O'Regan, D.P., Rueckert, D.: Automatic 3d bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. IEEE transactions on medical imaging **38**(9), 2151–2164 (2019)
8. Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M.J., Vembar, M., Olszewski, M.E., Subramanyan, K., Lavi, G., et al.: Automatic model-based segmentation of the heart in ct images. IEEE transactions on medical imaging **27**(9), 1189–1201 (2008)
9. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3d object reconstruction from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 605–613 (2017)
10. Frangi, A.F., Rueckert, D., Schnabel, J.A., Niessen, W.J.: Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. IEEE transactions on medical imaging **21**(9), 1151–1166 (2002)
11. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: Ce-net: Context encoder network for 2d medical image segmentation. IEEE transactions on medical imaging **38**(10), 2281–2292 (2019)
12. Habijan, M., Babin, D., Galić, I., Leventić, H., Romić, K., Velicki, L., Pižurica, A.: Overview of the whole heart and heart chamber segmentation methods. Cardiovascular Engineering and Technology **11**(6), 725–747 (2020)
13. Hoogendoorn, C., Duchateau, N., Sanchez-Quintana, D., Whitmarsh, T., Sukno, F.M., De Craene, M., Lekadir, K., Frangi, A.F.: A high-resolution atlas and statistical model of the human heart from multislice ct. IEEE transactions on medical imaging **32**(1), 28–44 (2012)
14. Karim, R., Blake, L.E., Inoue, J., Tao, Q., Jia, S., Housden, R.J., Bhagirath, P., Duval, J.L., Varela, M., Behar, J.M., et al.: Algorithms for left atrial wall segmentation and thickness–evaluation on an open-source ct and mri image database. Medical image analysis **50**, 36–53 (2018)

15. Kong, F., Shadden, S.C.: Automating model generation for image-based cardiac flow simulation. Journal of Biomechanical Engineering **142**(11) (2020)
16. Kong, F., Shadden, S.C.: 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. pp. 550–559. Springer (2021)
17. Kong, F., Wilson, N., Shadden, S.: A deep-learning approach for direct whole-heart mesh reconstruction. Medical image analysis **74**, 102222 (2021)
18. Lee, M.C.H., Petersen, K., Pawlowski, N., Glocker, B., Schaap, M.: Tetris: Template transformer networks for image segmentation with shape priors. IEEE transactions on medical imaging **38**(11), 2596–2606 (2019)
19. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. ACM siggraph computer graphics **21**(4), 163–169 (1987)
20. Mandikal, P., Navaneet, K., Agarwal, M., Babu, R.V.: 3d-lmnet: Latent embedding matching for accurate and diverse 3d point cloud reconstruction from a single image. arXiv preprint arXiv:1807.07796 (2018)
21. Medrano-Gracia, P., Cowan, B.R., Bluemke, D.A., Finn, J.P., Lima, J.A., Suinesiaputra, A., Young, A.A.: Large scale left ventricular shape atlas using automated model fitting to contours. In: International Conference on Functional Imaging and Modeling of the Heart. pp. 433–441. Springer (2013)
22. Ordas, S., Oubel, E., Sebastian, R., Frangi, A.F.: Computational anatomy atlas of the heart. In: 2007 5th International Symposium on Image and Signal Processing and Analysis. pp. 338–342. IEEE (2007)
23. Organization, W.H.: The world health report 2002: reducing risks, promoting healthy life. World Health Organization (2002)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
25. Sumner, R.W., Schmid, J., Pauly, M.: Embedded deformation for shape manipulation pp. 80–es (2007)
26. Tilborghs, S., Dresselaers, T., Claus, P., Bogaert, J., Maes, F.: Shape constrained cnn for cardiac mr segmentation with simultaneous prediction of shape and pose parameters. arXiv preprint arXiv:2010.08952 (2020)
27. Tóthová, K., Parisot, S., Lee, M., Puyol-Antón, E., King, A., Pollefeys, M., Konukoglu, E.: Probabilistic 3d surface reconstruction from sparse mri information. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 813–823. Springer (2020)
28. Wang, N., Zhang, Y., Li, Z., Fu, Y., Yu, H., Liu, W., Xue, X., Jiang, Y.G.: Pixel2mesh: 3d mesh model generation via image guided deformation. IEEE transactions on pattern analysis and machine intelligence **43**(10), 3600–3613 (2020)
29. Wang, Y., Zhong, Z., Hua, J.: 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). https://doi.org/10.1109/TVCG.2019.2934369
30. Wickramasinghe, U., Remelli, E., Knott, G., Fua, P.: Voxel2mesh: 3d mesh model generation from volumetric data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 299–308. Springer (2020)
31. Wolterink, J.M., Leiner, T., De Vos, B.D., Coatrieux, J.L., Kelm, B.M., Kondo, S., Salgado, R.A., Shahzad, R., Shu, H., Snoeren, M., et al.: An evaluation of automatic coronary artery calcium scoring methods with cardiac ct using the orcascore framework. Medical physics **43**(5), 2361–2373 (2016)

32. Ye, M., Huang, Q., Yang, D., Wu, P., Yi, J., Axel, L., Metaxas, D.: 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. pp. 117–126. Springer (2020)
33. Zhou, X.Y., Wang, Z.Y., Li, P., Zheng, J.Q., Yang, G.Z.: One-stage shape instantiation from a single 2d image to 3d point cloud. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 30–38. Springer (2019)
34. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 3–11. Springer (2018)
35. Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M.P., Oster, J., Wang, C., Smedby, Ö., Bian, C., et al.: Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. Medical image analysis **58**, 101537 (2019)
36. Zotti, C., Luo, Z., Lalande, A., Jodoin, P.M.: Convolutional neural network with shape prior applied to cardiac mri segmentation. IEEE journal of biomedical and health informatics **23**(3), 1119–1128 (2018)