Diffusion Models To Predict 3D Late Mechanical Activation From Sparse 2D Cardiac MRIs

Nivetha Jayakumar* VFB8ZV@VIRGINIA.EDU
Department of Electrical and Computer Engineering, University of Virginia, USA

Jiarui Xing* JX8FH@VIRGINIA.EDU
Department of Electrical and Computer Engineering, University of Virginia, USA

Tonmoy Hossain PWG7JB@VIRGINIA.EDU
Department of Computer Science, University of Virginia, USA

Fred Epstein FHE6B@UVAHEALTH.ORG
Department of Biomedical Engineering, University of Virginia, USA

Kenneth Bilchick KCB7F@UVAHEALTH.ORG
School of Medicine, University of Virginia, USA

Miaomiao Zhang MZ8RR@VIRGINIA.EDU
Departments of Electrical and Computer Engineering & Computer Science, University of Virginia, USA

Abstract

Identifying regions of late mechanical activation (LMA) of the left ventricular (LV) myocardium is critical in determining the optimal pacing site for cardiac resynchronization therapy in patients with heart failure. Several deep learning-based approaches have been developed to predict 3D LMA maps of LV myocardium from a stack of sparse 2D cardiac magnetic resonance imaging (MRIs). However, these models often loosely consider the geometric shape structure of the myocardium. This makes the reconstructed activation maps suboptimal; hence leading to a reduced accuracy of predicting the late activating regions of hearts. In this paper, we propose to use shape-constrained diffusion models to better reconstruct a 3D LMA map, given a limited number of 2D cardiac MRI slices. In contrast to previous methods that primarily rely on spatial correlations of image intensities for 3D reconstruction, our model leverages object shape as priors learned from the training data to guide the reconstruction process. To achieve this, we develop a joint learning network that simultaneously learns a mean shape under deformation models. Each reconstructed image is then considered as a deformed variant of the mean shape.

To validate the performance of our model, we train and test the proposed framework on a publicly available mesh dataset of 3D myocardium and compare it with state-of-the-art deep learning-based reconstruction models. Experimental results show that our model achieves superior performance in reconstructing the 3D LMA maps as compared to the state-of-the-art models.

1. Introduction

Cardiac resynchronization therapy (CRT) is widely used to treat cardiac conduction system disorders, such as left bundle branch block (LBBB) and intrinsic myocardial diseases (Abraham et al., 2002; Lindenfeld et al., 2007; Moss et al., 2009). However, standard CRT faces a substantial challenge with a high non-response rate, approximately around 40% (Chung et al., 2008; Exner et al., 2012). Implanting the CRT LV lead at an area with delayed activation may significantly decrease the non-response rate (Bilchick et al., 2014; Ramachandran et al., 2015). Cine displacement encoding with stimulated echoes (DENSE) MR imaging is an accurate and reproducible method for imaging regional myocardial displacement and strain (Aletras et al., 1999; Kim et al., 2004). Studies have shown that LMA can be effectively measured on circumferential strains of tis-

* These authors contributed equally

sue displacements imaged by DENSE (Bilchick et al., 2014; Ramachandran et al., 2015). The LMA in heart failure patients can be detected by the Time to the Onset of circumferential Shortening (TOS) (Kvåle et al., 2019; Xing et al., 2023, 2021). The 3D LMA map is then obtained by projecting the TOS values to a reconstructed 3D volume of the myocardium from a stack of sparse 2D slices. Despite the success of reconstructing the 3D LMA maps, current approaches (Kim et al., 2004; Kvåle et al., 2019; Xing et al., 2023, 2021) do not achieve satisfactory results since they rarely consider the shape variability among patients' hearts, resulting in unstable and poor performance of locating the late activating regions from reconstructed 3D LMA maps.

To address this challenge, recent advancements have leveraged deep learning techniques to substantially improve the performance of image reconstruction from a limited number of 2D slices (Cetin et al., 2023; Goodfellow et al., 2020; Maaløe et al., 2019). Extensive research has explored various network architectures for 3D image reconstruction, including UNets (Nguyen et al., 2021), transformers (Feng et al., 2021; Korkmaz et al., 2022), and state-of-the-art generative diffusion models (Waibel et al., 2023). These works have significantly improved the reconstruction efficiency by learning intricate mappings between stacks of 2D images and their corresponding 3D volumes. While the deep learning-based approaches have achieved impressive results in reconstructing detailed 3D images, they often lack explicit consideration of shape information during the learning process. Consequently, important geometric structures of objects depicted in the images may not be well preserved. This may lead to the occurrence of artifacts, such as discontinuities, holes, or mismatched connections between different parts, that break the topology of the reconstructed objects.

In this paper, we propose to utilize a recently developed work, shape-aware diffusion model (Jayakumar et al., 2023), to reconstruct 3D volumes of the myocardium from sparse 2D slices. We will introduce a jointly trained atlas-building network to determine the atlas (Wang and Zhang, 2022) (i.e., the mean image), along with a stack of sparse 2D slices, used as the input prior to train the diffusion model. In this framework, the predicted 3D volume is considered to be a deformed variant of the myocardium atlas. This reconstructed 3D volume can then be used to interpolate the predicted TOS values to obtain the cardiac mechanical activation maps. Due to the lack of 3D ground truth cine DENSE cardiac MRI scans, we train and validate our model on a public dataset of 3D myocardium mesh data and evaluate the results of the TOS prediction and LMA detection on the clinical test data. Our method is generally applicable to similar clinical applications, where collecting a complete 3D volume with high resolution is challenging. For example, reconstructing intra-operative 3D brain MRIs in real-time during brain surgeries.

2. Background

This section provides a brief background on the computation of strain matrices used as input to the neural network and an unbiased atlas-building framework trained to estimate the Fréchet mean of group-wise images.

2.1. Quantification of Late Mechanical Activation Through Strain Analysis

Circumferential strain has been demonstrated to depict myocardial contraction along the circular outline in the short axis, typically with a negative value (Budge et al., 2012). There are several advantages of analyzing strain values as it significantly reduces the computational complexity of the network input, shifting from vector-valued displacement data V to real-valued strain images represented as functions (S). Also, it provides more robustness by mitigating the impact of body motion artifacts commonly encountered in the original signals and it improves our ability to quantify dyssynchrony, as evidenced by studies such as (Budge et al., 2012). Following a recent work (Dandel et al., 2009), we calculate the activation time of the left ventricle using strains derived from the displacement maps. In Fig. 1a, we show examples of MR images overlaid with displacement maps over time.

Given a d-dimensional displacement map $u(x) : \Omega \to \mathbb{R}^d$, a Jacobian matrix $D^{d\times d}$ of u at each spatial location is

$$D = \begin{pmatrix} \Delta u_1^1 & \cdots & \Delta u_1^d \\ \vdots & \ddots & \vdots \\ \Delta u_d^1 & \cdots & \Delta u_d^d \end{pmatrix},$$

where $\Delta u_i^j = \frac{\mathbf{d}u_i}{\mathbf{d}x_j}$, with $i, j \in \{1, \cdots, d\}$ and $\Omega = \mathbb{R}^d/\mathbb{Z}^d$ be a d-dimensional torus domain with periodic boundary conditions, obtained from cine images. A strain matrix E is then computed as $E = \frac{1}{2}(D^T D - \mathrm{I})$,

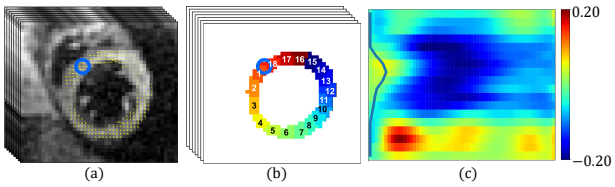

Figure 1: Example of (a) temporal DENSE magnitude MRIs overlaid with displacement maps; (b) myocardial sectors of left ventricles with strain values (blue color indicates contraction vs. red color indicates expansion); (c) strain matrix and its corresponding TOS curve. Blue circles in (a) and (b) indicate intersections of left and right ventricles (Xing et al., 2021).

where T denotes a matrix transpose and I is a $d \times d$ identity matrix. This paper focuses on the circumferential strain, which is the component of strain along the myocardium. In practical clinical applications, the muscle area of the left ventricle is often divided into a variable number of sectors (Cerqueira, 2002). In this work, we divide each of the 2D slices in the volumetric stack into 18 sectors which helps in understanding the motion of the myocardial wall in detail. Then we compute an $18-$*dimensional* strain vector from the full-resolution strain vector of the myocardium, where the strain value is extracted from the center of each of the 18 sectors for each slice. To facilitate the automatic estimation of TOS, we then rearrange the slices into a 2D matrix consisting of all strain vectors over a temporal dimension of the cardiac phases ((Xing et al., 2021)). An example of a spatiotemporal strain matrix overlaid with a circumferential TOS curve is shown in Fig. 1c.

2.2. Atlas Building Network

The deformation-based shape features of objects in an image can be utilized for image generation and registration tasks to ensure that the generated objects preserve the topological structures (Wang and Zhang, 2021; Zhang et al., 2013; Wu et al., 2022). Here, we briefly explain the notion of atlas-building ((Joshi et al., 2004)), which computes a Fréchet mean of a group of images as a deformation-based shape representation. For a given group of N images $\{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_N\}$, each image can be represented as a deformed version of the mean image. The problem of atlas building is to find the template or mean image $\mathcal{S}$ and deformation field $\{\phi_1, \phi_1, \cdots \phi_N\}$ with derived initial velocity fields $v_1, v_2, \cdots v_t$. This is achieved by minimizing the following energy function:

$$E(\mathcal{S}, \phi_n) = \sum_{n=1}^{N} \frac{1}{\sigma^2} \text{Dist}\left[\mathcal{S} \circ \phi_n(v_t),\, \mathcal{X}_n\right] + \text{Reg}\left(\phi_n(v_t)\right), \tag{1}$$

where σ^2 is the noise variance, $\text{Dist}[\cdot,\cdot]$ is a metric of dissimilarity such as local or global normalized cross-correlation, mean squared error, cosine distance, and mutual information. The $\circ$ denotes an interpolation operator that deforms image $\mathcal{X}_n$ with an estimated transformation ϕ_n. The $\text{Reg}(\cdot)$ is a regularizer that guarantees the smoothness of the transformation which can be denoted as

$$\text{Reg}[\phi_n(v_t)] = \int_0^1 (Lv_t, v_t)\, dt, \quad \text{with} \quad \frac{d\phi_n(t)}{dt} = -D\phi_n(t) \cdot v_n(t), \tag{2}$$

where $L : V \to \widetilde{V}$ is a symmetric, positive-definite differential operator that maps a tangent vector $v_t \in V$ into its dual space as a momentum vector $m_t \in \widetilde{V}$. This can be represented as $m_t = Lv_t$, or $v_t = Km_t$, with K being an inverse operator of L. The operator, D, denotes a Jacobian matrix and $\cdot$ represents the operation of element-wise matrix multiplication. For the metric L, we consider $L = (-\alpha\nabla^2 + \gamma \text{I})^3$ in this paper, where ∇^2 is the discrete Laplacian operator, α is a positive regularity parameter that controls the number of continuous derivates over the domain of the deformation fields, γ is a weighting coefficient, and I denotes an identity matrix.

With a given initial condition of velocity fields $v_n(0)$, the minimum of Eq. (2) is uniquely determined by solving a Euler-Poincaré differential equation (EPDiff) (Arnold, 1966; Miller et al., 2006) which is known as the *geodesic shooting* algorithm (Vialard et al., 2012) which can be represented as

$$\frac{\partial v_t}{\partial t} = -K\left[(Dv_t)^T \star m_t + \nabla \cdot (m_t \otimes v_t)\right], \tag{3}$$

where $\star$ is the truncated matrix-vector field auto-correlation. The operator $\nabla\cdot$ is the discrete divergence of a vector field and K is a smoothing operator with its inverse L. The computation of solving the EPDiff is time-consuming and expensive, particularly when images are with high dimensions. An

efficient reparameterization of the velocity fields was then developed to drastically decrease the computational cost of LDDMM with geodesic shooting without losing the model accuracy (Zhang and Fletcher, 2015, 2019).

We are now ready to minimize the atlas building energy function in Eq. (1) as

$$E(\mathcal{S},\, \phi_i) = \sum_{n=1}^{N} \frac{1}{\sigma^2} \text{Dist}[\, \mathcal{S} \circ \phi_n(v_n(t)),\, \mathcal{X}_n\,] + (Lv_n(0), v_n(0)),\ \text{s.t. Eq. (2) \& (3)}. \tag{4}$$

For the simplicity of notation, we will drop the temporal index in the upcoming equations.

3. Our Method

In this section, we present a novel LMA prediction framework based on SADIR, a recently developed diffusion models (Jayakumar et al., 2023), from a limited number of sparse 2D cardiac MRIs. First, we introduce shape-constrained diffusion models for myocardium reconstructions via atlas building. Here, we discuss two sub-modules, including an atlas building network, which enables us to learn shape priors from a given set of 3D myocardium images, and a forward-backward diffusion process leveraging the atlas or mean image generated from the atlas building network. Lastly, we discuss a pre-trained LMA quantification network to predict 3D LMA maps of the left ventricle using the reconstructed myocardium.

3.1. Diffusion Models For 3D Myocardium Reconstruction Via Atlas Building

Given a number of N myocardium training images $\{J_n, \mathcal{X}_n\}_{n=1}^{N}$ where J_n represents a collection of sparse 2D images paired with its corresponding full 3D volume $\mathcal{X}_n$. We design two sub-modules to reconstruct 3D myocardium from sparse 2D slices:

- *(i) An atlas building network (M_θ) that provides a mean image $\mathcal{S}$ of $\mathcal{X}_n$.* Followed by the network architecture of Geo-SIC (Wang and Zhang, 2022), we develop a learning-based atlas building network characterized by an initial velocity field ($v_n(0)$).

- *(ii) A diffusion reconstruction network G_θ, considering each reconstructed myocardium $\hat{\mathcal{X}}_n$ as a deformed variant of the obtained atlas, i.e., $\hat{\mathcal{X}}_n \triangleq \mathcal{S} \circ \phi_n(v_n(G_\theta))$.* We use the mean image from M_θ as the conditional shape feature for the myocardium reconstruction network G_θ to predict the velocity field v_n. Similar to the existing diffusion models (Ho et al., 2020; Waibel et al., 2023), we develop a forward process to generate an isotropic Gaussian distribution and a backward diffusion process to predict the velocity fields (Jayakumar et al., 2023) associated with the pair of input training myocardium images and an atlas image. For notation simplicity, we will omit the index n for each subject in the following sections.

An overview of our network architecture is shown in Fig. 2.

Forward diffusion process. For a given 3D volume of the myocardium x^0 and the time point of the diffusion process Υ, we assume the data distribution of x^Υ is a normal distribution which can be defined as $x^\Upsilon \sim \mathcal{N}(\mu, \eta)$, where μ is the mean and η is the variance. The forward diffusion process for time steps T is recursively defined via the posterior

$$q(x^\Upsilon | x^{\Upsilon-1}) = \mathcal{N}(x^\Upsilon; \sqrt{1-\eta^\Upsilon}x^{\Upsilon-1}, \eta^\Upsilon \mathbf{I}), \tag{5}$$

where $\mathbf{I}$ denotes an identity matrix and $\eta^\tau \in [0,1]$ denotes the variance scheduler increased along the time steps with $\eta^1 < \eta^2 < \cdots < \eta^\Upsilon$. We repeat the forward diffusion process for a fixed, predefined number of steps.

Following (Ho et al., 2020; Waibel et al., 2023), we obtain the result of the forward diffusion process as a one-step forward transition from x_0 to x_T as

$$q(x^\Upsilon \,|\, x^0) = \mathcal{N}(x^\Upsilon; \sqrt{\bar{\psi}^\Upsilon}x^0, (1-\bar{\psi}^\Upsilon)\,\mathbf{I}). \tag{6}$$

where $\psi^\Upsilon = 1 - \eta^\Upsilon$ and $\bar{\psi}^T = \prod_{i=1}^{\Upsilon} \psi^i$. Hence, we can write x^τ in terms of x^0 as

$$x^\Upsilon = \sqrt{\bar{\psi}^\Upsilon}x^0 + \sqrt{1-\bar{\psi}^\Upsilon}\epsilon,\ \text{with } \epsilon \sim \mathcal{N}(0, \mathbf{I}). \tag{7}$$

Reverse diffusion process. In the reverse process, we employ a conditional diffusion model, where the output mean 3D myocardium $\mathcal{S}$ from the atlas building network M_θ, along with a stack of sparse 2D slices of the volume to be reconstructed J_n is concatenated with x^τ from the forward diffusion process as the condition c to form the input of the network G_θ. This network is trained as the reverse process of the diffusion model.

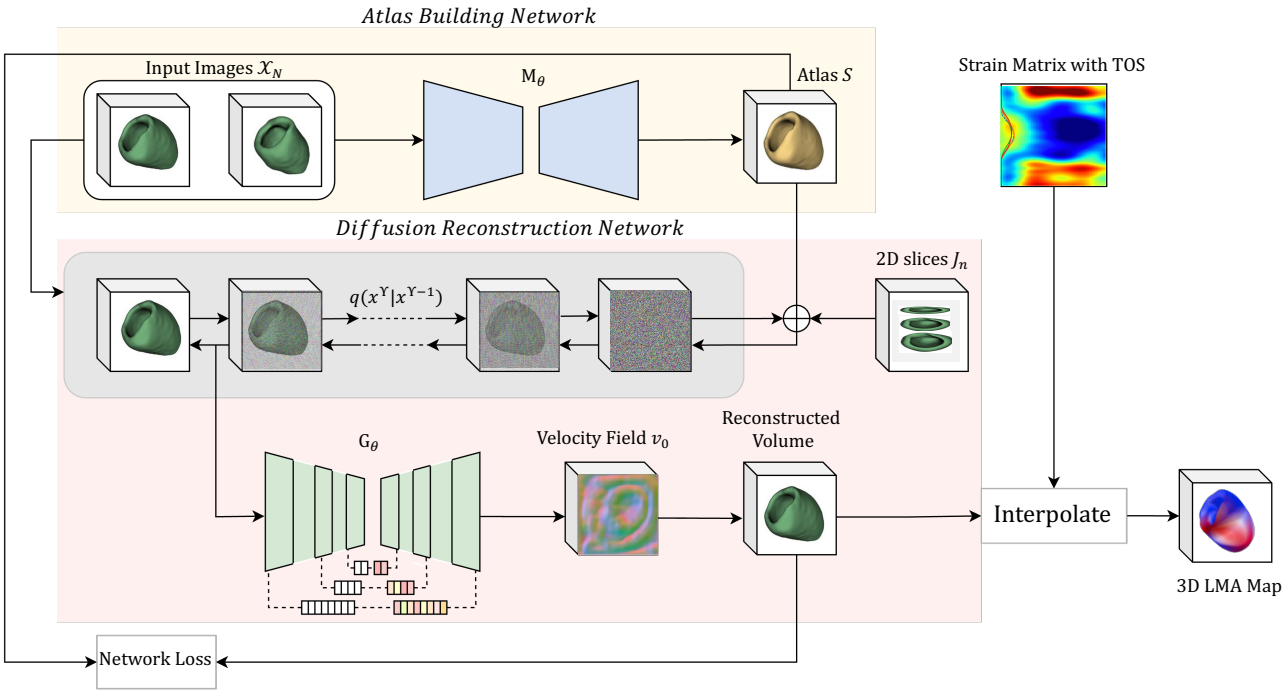

Figure 2: An overview of our proposed LMA prediction model from sparse 2D cardiac MRIs.

The joint probability distribution of the reverse process is given by $q(x^{\Upsilon-1}|x^{\Upsilon})$, that is learned by an Attention-UNet (Ronneberger et al., 2015), which is a type of Convolutional Neural Network with skip connections, residual blocks and excitation layers (Hu et al., 2019).

Defining the network parameter as G_θ, we can formulate the reverse process as

$$q_{G_\theta}(x^{\Upsilon-1}|x^{\Upsilon}) = \mathcal{N}(x^{\Upsilon-1}; \mu_{G_\theta}(x^{\Upsilon}, j_i, \mathrm{S}, \Upsilon), \Sigma_{G_\theta}(x^{\Upsilon}, j_i, \mathrm{S}, \Upsilon)) \quad (8)$$

As seen in (Jayakumar et al., 2023), the corresponding deformation field ϕ is then generated from the initial velocity field v_0, obtained from the network G_θ, to deform the atlas $\mathcal{S}$, thus giving us a reconstructed myocardium $\hat{\mathcal{X}} = \mathcal{S} \circ \phi(v_0(G_\theta))$ for a given stack of sparse 2D myocardium slices J_n.

3.2. LMA Prediction using Reconstructed Myocardium Volumes

To predict the 3D LMA map, we start by mapping the predicted TOS curves from each sparse 2D slice to their corresponding sectors in the myocardium segmentation of these slices. This produces a collection of sparse 3D vertices, each assigned a specific TOS value. To ensure consistency between the sparse slices and the overarching 3D model, we introduce an alignment procedure, refining the spatial positioning of these slices to fit well with the reconstructed 3D surface (Xing et al., 2021). Since the sparse slices are sampled from the dense slices, and the dense slices are the ground-truth of our model, we automatically know the spatial relationship between the sparse slices and reconstructed 3D model. The alignment procedure mainly builds a 3D coordinate system to hold both the sparse slices and the reconstructed 3D model.

The sparse slices in our datasets (particularly breath-hold MRIs) are set to be sampled at the same location of the subject's heart. A TOS curve com-

prises TOS values from specific sectors evenly spaced along the middle line of the left ventricle myocardium in the corresponding sparse slice. During the alignment process, the spatial positions of these TOS sectors adjust along with the sparse slices while maintaining their relative positions. For the alignment, we first build a 3D coordinate for the sparsely sampled 2D slices, based on their spatial information (including the slice spatial location and pixel resolution). The coordinates of the 3D mesh face vertices are automatically determined, as the 3D volume should occupy the same room as the sparsely sampled slices. With this data, we can conduct interpolation in the 3D space, enabling us to calculate the interpolated TOS values for each mesh face. For TOS prediction, we adopt a pre-trained model developed by (**?**). After the network predicts the TOS curve for all 2D sparse slices from the same subject, we create a 3D activation map by interpolating TOS values onto the left ventricles of the reconstructed 3D myocardium obtained G_θ.

3.3. Network Loss and Joint Training.

The network loss of the 3D myocardium reconstruction is a joint loss of the atlas building network M_θ and the reconstruction network G_θ. First, we can define the atlas building loss as

$$\mathcal{L}(M_\theta) = \sum_{n=1}^{N} \frac{1}{\sigma^2} \| \mathcal{S}(M_\theta) \circ \phi_n(v_n) - \mathcal{X}_n \|_2^2 + (Lv_n, v_n) + \mathrm{reg}(M_\theta), \tag{9}$$

where $\mathrm{reg}(\cdot)$ denotes a regularization on the atlas building network parameters.

Then we define the loss function of the diffusion reconstruction network from 2D sparse slices as a combination of sum-of-squared differences and Sørensen−Dice coefficient (Dice, 1945) loss between the predicted reconstruction and ground-truth in the following

$$\mathcal{L}(G_\theta) = \sum_{n=1}^{N} \|\mathcal{S} \circ \phi_n(v_n(G_\theta)) - \mathcal{X}_n\|_2^2 + \lambda\,[1 - \mathrm{Dice}(\mathcal{S} \circ \phi_n(v_n(G_\theta)), \mathcal{X}_n)] + \mathrm{reg}(G_\theta), \tag{10}$$

where λ is a weighting factor that obtains the extent of overlap between the predicted and ground-truth volumes. We can define the dice similarity score $\mathrm{Dice}(\cdot,\cdot)$ as

$$\mathrm{Dice}(\hat{\mathcal{X}}, \mathcal{X}_n) = 2(|\hat{\mathcal{X}}| \cap |\mathcal{X}_n|)/(|\hat{\mathcal{X}}| + |\mathcal{X}_n|), \tag{11}$$

where $\hat{\mathcal{X}}_n \triangleq \mathcal{S} \circ \phi_n(v_n(G_\theta))$. Defining α as the weighting parameter, we are now ready to define the joint loss of the reconstruction networks as

$$\mathcal{L} = \mathcal{L}(M_\theta) + \alpha\mathcal{L}(G_\theta). \tag{12}$$

Alternative optimization for the joint network learning. We employ an alternative optimization scheme (Nocedal and Wright, 1999) to minimize the network loss defined in Eq. (12). Here, we jointly optimize all network parameters by alternating between the atlas building network M_θ and reconstruction network G_θ without sharing weights. We summarize the joint training of our reconstruction model in Alg. 1.

Algorithm 1 Joint Training of the Diffusion Model for 3D Myocardium Reconstruction.

Input : A group of N input images with full 3D volumes $\{\mathcal{X}_n\}$ and a stack of sparse 2D images $\{J_n\}$.

Output: Generate mean shape or atlas $\mathcal{S}$, initial velocity fields v_n, and reconstructed volumes $\hat{\mathcal{X}}_n$

. **for** i = 1 to p **do**

  `/* Train geometric shape learning network */`

  Minimize the atlas building loss in Eq. (9)

  Output the atlas $\mathcal{S}$;

  `/* Train diffusion reconstruction network */`

  Minimize the diffusion reconstruction loss in Eq. (10);

  Output the initial velocity fields $\{v_n\}$ and the reconstructed images $\hat{\mathcal{X}}_n$;

end

Until convergence

4. Experimental Evaluation

We validate our proposed model to predict LMA through reconstructed 3D myocardium from 2D sparse cardiac MRIs.

Dataset. For 3D myocardium reconstruction network, we use 215 publicly available 3D myocardium mesh data from MedShapeNet data repository (Li,

2023). Using 3D slicer, we convert the mesh data to binary label maps (Fedorov et al., 2012). All the images were preprocessed as $222 \times 222 \times 222$ and pre-aligned with affine transformation.

For LMA prediction, we obtain cine DENSE volumes of 57 subjects along with 282 strain matrices. All the scans were acquired using a 1.5T MR scanner with a four-channel phased array radio-frequency coil. The cine DENSE imaging was performed in four short-axis planes at the basal, two ventricular, and apical regions with a resolution of $2.65 \times 2.65mm^2$, and the temporal resolution $17ms$. The field of view was $240 \times 240mm^2$, displacement encoding frequency $k_e = 0.1 \ \ cycles/mm$, echo time $= 1.08ms$, and flip angle 15°. Displacement was encoded in two orthogonal directions and a spiral k-space trajectory was used with six interleaves per image. The ground-truth TOS curves were manually labeled by experts.

We augment the dataset by shifting the strain matrices and their corresponding labels along the sector dimension. The last 5 frames of the strain matrices are trimmed due to noisy image signals. Also, we employ mixup (Zhang et al., 2017) to augment the dataset yielded 3844 strain matrices together with the ground truth. All datasets used in our work were publicly released or pre-collected. As such, no ethical considerations are necessary.

4.1. Experimental Settings.

We first validate our proposed model for reconstructing 3D myocardium from a sparse stack of eight 2D slices. We compare the performance of our model with three state-of-the-art deep learning-based reconstruction models: 3D-UNet (Çiçek et al., 2016b); DDPM, a probabilistic diffusion model (Ho et al., 2020); and DISPR, a diffusion model based shape reconstruction model with the consideration of geometric topology (Waibel et al., 2023). We visualize and compare the reconstructed 3D myocardium of all models. We evaluate all the reconstruction models using three evaluation metrics, including, the Sørensen–Dice coefficient (DSC) (Dice, 1945), Jaccard Similarity (Jaccard, 1908), and RHD95 score (Huttenlocher et al., 1993), validating the prediction accuracy of 3D myocardium for all methods. Lastly, we show the predicted 3D activation maps and compare them with the clinical reference model, concentrating on the LMA region in the reconstructed myocardium.

Parameter Setting. We set the mean and standard deviation as 0 and 0.1, respectively, for the forward diffusion process. We use linear scheduling for the noising process and is scaled to reach an isotropic Gaussian distribution irrespective of the value of T. In the reverse diffusion process, we set the depth of the 3D attention-UNet backbone as 6, introducing an attention mechanism via spatial excitation channels (Hu et al., 2019) with ReLU (Rectified Linear Unit) activation. For the atlas building network, we set the depth of the UNet architecture as 4. We set the number of time steps for Euler integration in EPDiff (Eq. (3)) as 10, and the noise variance $\sigma = 0.02$. We use a kernel map valued $[0.5, 0, 1.0]$ for the shooting operation. Besides, we set the parameter $\alpha = 3$ for the operator L. Similar to (Waibel et al., 2023), we set the batch size as 1 for all experiments.

We split the dataset into 70% training, 15% validation and 15% testing, utilizing cosine annealing learning rate scheduler that starts with a learning rate of $\eta = 1e^{-3}$. We run all models on training and validation images using the Adam optimizer and save the networks with the best validation performance. We conduct all experiments on one $80GB$ NVIDIA A100 GPU for $\sim$ 18 hours. For both training and testing, we downsample the volumes to $64 \times 64 \times 64$.

4.2. Experimental Results

Fig. 3 shows a comparison of the 3D myocardium between the ground truth and all reconstruction models, including such as 3D-UNet (Çiçek et al., 2016a), DDPM (Ho et al., 2020), DISPR (Waibel et al., 2023), and our model. It shows that our method outperforms other baselines in maintaining the topology of the myocardium. Besides, compared to the SOTA models, our reconstruction framework produces little to no artifacts, preserving the original shape of the organ. Our work uses sparse 2D binary label-maps of the myocardial walls to reconstruct the 3D volume by deforming the 3D atlas with the predicted deformation fields. This ensures that our model performs with similar scores when there are significant variations in image quality since it relies on the segmentations obtained from these scans. Moreover, the stack of sparse 2D slices provides sufficient information on patient-specific differences in the anatomy of the myocardium.

Tab. 1 reports the average DSC (Dice, 1945), Jaccard similarity (Jaccard, 1908), and Hausdorff distance (RHD95) (Huttenlocher et al., 1993) com-

Table 1: A comparison of 3D myocardium reconstruction for our framework compared to other SOTA models

Model	DSC ↑	Jaccard similarity ↑	RHD95 ↓
3D-Unet	0.870 ± 0.0158	0.771 ± 0.024	0.840 ± 0.202
DDPM	0.823 ± 0.014	0.668 ± 0.019	1.027 ± 0.093
DISPR	0.950 ± 0.017	0.906 ± 0.031	0.347 ± 0.032
Ours	**0.978 ± 0.016**	**0.957 ± 0.031**	**0.341 ± 0.023**

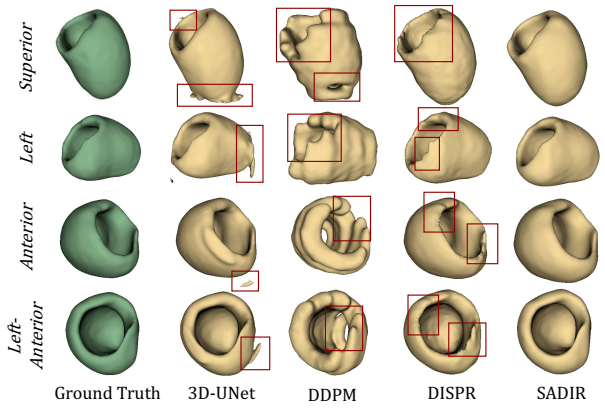

Figure 3: A comparison of reconstructed 3D myocardium between ground truth, 3D-UNet, DDPM, DISPR, and our model over four different views.

pared between the reconstructed myocardium from the ground truth and all models. Comparing the results of all models, our method outperforms all the baselines, achieving the best performance in DSC, Jaccard similarity, and RHD95 with the lowest standard deviations across all metrics.

Fig. 4 shows the generated 3D late mechanical activation maps using the reconstructed myocardial meshes for two subjects over three different views. The regions in red show larger TOS values, representing the late mechanical activation area while the regions in blue show areas containing lower TOS values, representing as the are normal regions. The 3D LMA maps provide intuitive visual information of LMA regions (highlighted in red in Fig. 4) location on the subject's left ventricle. Compared with previous methods, our models have much more accurate anatomical structures, which allows the doctors to more precisely identify the anatomical structure of the LMA regions. Additionally, it shows the differences between anatomical features of the myocardium interpolated using the current reference model used by clinicians, which does not provide information on patient-wise variability in the anatomy of the myocardium, and the myocardium reconstructed by our framework. It is evident that our model efficiently preserves the shape information of the organ. This plays a vital role in an improved and clinically significant detection of LMA regions, hence allowing for personalized surgical plans as opposed to the traditional one-size-fits-all model. While our method

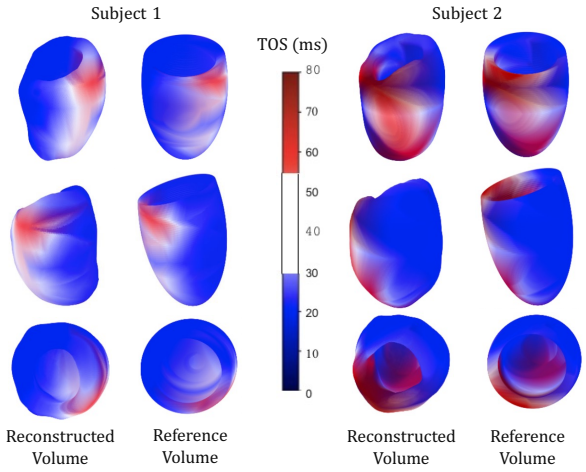

Figure 4: A comparison of 3D late mechanical activation maps using reconstructed myocardium volume of two subjects between our model and the clinical reference model.

effectively preserves the geometrical features of the myocardium with an improved performance of reconstruction, a thorough evaluation of the method on a variety of large datasets is needed. This future work could further improve the model's generalizabil-

ity across various image modalities with larger thicknesses of input slices.

5. Conclusion

This paper presents a shape-constrained diffusion model to obtain the 3D late mechanical activation of the myocardium given sparse 2D slices of DENSE MRI. The introduction of the mean image as a prior during the diffusion process helps the model to efficiently learn and preserve the inherent anatomical structure of the myocardium. We train and evaluate the performance of our reconstruction model on 3D myocardium mesh data. Using this pre-trained network, we predict the 3D volumes of the myocardium for two subjects and compare them to the current 3D reference model used in the clinical setting. Compared to the state-of-the-art models, our proposed network performed substantially well in predicting 3D late mechanical activations through myocardium reconstruction from sparse 2D slices. The clinical application of our framework would be pivotal since it can help in the accurate detection of LMA regions, thus allowing for effective CRT procedures.

6. Acknowledgement.

This work was supported by NSF CAREER Grant 2239977 and NIH 1R21EB032597.

References

William T Abraham, Westby G Fisher, Andrew L Smith, David B Delurgio, Angel R Leon, Evan Loh, Dusan Z Kocovic, Milton Packer, Alfredo L Clavell, David L Hayes, et al. Cardiac resynchronization in chronic heart failure. *New England Journal of Medicine*, 346(24):1845–1853, 2002.

Anthony H Aletras, Shujun Ding, Robert S Balaban, and Han Wen. Dense: displacement encoding with stimulated echoes in cardiac functional mri. *Journal of magnetic resonance (San Diego, Calif.: 1997)*, 137(1):247, 1999.

Vladimir Arnold. Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. In *Annales de l'institut Fourier*, volume 16, pages 319–361, 1966.

Kenneth C Bilchick, Sujith Kuruvilla, Yasmin S Hamirani, Raghav Ramachandran, Samantha A Clarke, Katherine M Parker, George J Stukenborg, Pamela Mason, John D Ferguson, J Randall Moorman, et al. Impact of mechanical activation, scar, and electrical timing on cardiac resynchronization therapy response and clinical outcomes. *Journal of the American College of Cardiology*, 63(16):1657–1666, 2014.

Loren P Budge, Adam S Helms, Michael Salerno, Christopher M Kramer, Frederick H Epstein, and Kenneth C Bilchick. Mr cine dense dyssynchrony parameters for the evaluation of heart failure: comparison with myocardial tissue tagging. *JACC: Cardiovascular Imaging*, 5(8):789–797, 2012.

Manuel D Cerqueira. American heart association writing group on myocardial segmentation and registration for cardiac imaging: Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. *Circulation*, 105: 539–542, 2002.

Irem Cetin, Maialen Stephens, Oscar Camara, and Miguel A González Ballester. Attri-vae: Attribute-based interpretable representations of medical images with variational autoencoders. *Computerized Medical Imaging and Graphics*, 104:102158, 2023.

Eugene S Chung, Angel R Leon, Luigi Tavazzi, Jing-Ping Sun, Petros Nihoyannopoulos, John Merlino, William T Abraham, Stefano Ghio, Christophe Leclercq, Jeroen J Bax, et al. Results of the predictors of response to crt (prospect) trial. *Echocardiography*, 2608:2616, 2008.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. *CoRR*, abs/1606.06650, 2016a. URL http://arxiv.org/abs/1606.06650.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016b.

Michael Dandel, Hans Lehmkuhl, Christoph Knosalla, Nino Suramelashvili, and Roland Hetzer. Strain and strain rate imaging by echocardiography-basic concepts and clinical applicability. *Current cardiology reviews*, 5(2):133–148, 2009.

Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

Derek V Exner, Angelo Auricchio, and Jagmeet P Singh. Contemporary and future trends in cardiac resynchronization therapy to enhance response. *Heart rhythm*, 9(8):S27–S35, 2012.

Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, and et al. 3d slicer as an image computing platform for the quantitative imaging network, Nov 2012. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3466397/.

Chun-Mei Feng, Yunlu Yan, Huazhu Fu, Li Chen, and Yong Xu. Task transformer network for joint mri reconstruction and super-resolution. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 307–317. Springer, 2021.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63 (11):139–144, 2020.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2019.

D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993. doi: 10.1109/34.232073.

Paul Jaccard. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, 44:223–270, 1908.

Nivetha Jayakumar, Tonmoy Hossain, and Miaomiao Zhang. Sadir: Shape-aware diffusion models for 3d image reconstruction. In *International Workshop on Shape in Medical Imaging*, pages 287–300. Springer, 2023.

Sarang Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.

Daniel Kim, Wesley D Gilson, Christopher M Kramer, and Frederick H Epstein. Myocardial tissue tracking with two-dimensional cine displacement-encoded mr imaging: development and initial evaluation. *Radiology*, 230(3):862–871, 2004.

Yilmaz Korkmaz, Salman UH Dar, Mahmut Yurt, Muzaffer Özbey, and Tolga Cukur. Unsupervised mri reconstruction via zero-shot learned adversarial transformers. *IEEE Transactions on Medical Imaging*, 41(7):1747–1763, 2022.

Kaja F Kvåle, Jørn Bersvendsen, Espen W Remme, Sebastien Salles, John M Aalen, Pål H Brekke, Thor Edvardsen, and Eigil Samset. Detection of regional mechanical activation of the left ventricular myocardium using high frame rate ultrasound imaging. *IEEE transactions on medical imaging*, 38(11):2665–2675, 2019.

Jianning Li. Medshapenet: A large-scale dataset of 3d medical shapes for computer vision, Mar 2023. URL https://medshapenet.ikim.nrw/.

JoAnn Lindenfeld, Arthur M. Feldman, Leslie Saxon, John Boehmer, Peter Carson, Jalal K. Ghali, Inder Anand, Steve Singh, Jonathan S. Steinberg, Brian Jaski, Teresa DeMarco, David Mann, Patrick Yong, Elizabeth Galle, Fred Ecklund, and Michael Bristow. Effects of Cardiac Resynchronization Therapy With or Without a Defibrillator on Survival and Hospitalizations in Patients With New York Heart Association Class IV Heart Failure. *Circulation*, 115(2): 204–212, January 2007. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIRCULATIONAHA.106.629261. URL https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.106.629261.

Lars Maaløe, Marco Fraccaro, Valentin Liévin, and Ole Winther. Biva: A very deep hierarchy of latent variables for generative modeling. *Advances in neural information processing systems*, 32, 2019.

Michael I Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2): 209–228, 2006.

Arthur J Moss, W Jackson Hall, David S Cannom, Helmut Klein, Mary W Brown, James P Daubert, NA Mark Estes III, Elyse Foster, Henry Greenberg, Steven L Higgins, et al. Cardiac-resynchronization therapy for the prevention of heart-failure events. *New England Journal of Medicine*, 361(14):1329–1338, 2009.

Tan Nguyen, Binh-Son Hua, and Ngan Le. 3d-ucaps: 3d capsules unet for volumetric image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 548–558. Springer, 2021.

Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

Raghav Ramachandran, Xiao Chen, Christopher M Kramer, Frederick H Epstein, and Kenneth C Bilchick. Singular value decomposition applied to cardiac strain from mr imaging for selection of optimal cardiac resynchronization therapy candidates. *Radiology*, 275(2):413–420, 2015.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Colin J Cotter. Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241, 2012.

Dominik J. E. Waibel, Ernst Röell, Bastian Rieck, Raja Giryes, and Carsten Marr. A diffusion model predicts 3d shapes from 2d microscopy images, 2023.

Jian Wang and Miaomiao Zhang. Bayesian atlas building with hierarchical priors for subject-specific regularization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 76–86. Springer, 2021.

Jian Wang and Miaomiao Zhang. Geo-sic: learning deformable geometric shapes in deep image classifiers. *Advances in Neural Information Processing Systems*, 35:27994–28007, 2022.

Nian Wu, Jian Wang, Miaomiao Zhang, Guixu Zhang, Yaxin Peng, and Chaomin Shen. Hybrid atlas building with deep registration priors. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022.

Jiarui Xing, Sona Ghadimi, Mohamad Abdi, Kenneth C. Bilchick, Frederick H. Epstein, and Miaomiao Zhang. Deep networks to automatically detect late-activating regions of the heart. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1902–1906, 2021. doi: 10.1109/ISBI48211.2021.9433796.

Jiarui Xing, Shuo Wang, Kenneth C. Bilchick, Frederick H. Epstein, Amit R. Patel, and Miaomiao Zhang. Multitask learning for improved late mechanical activation detection of heart from cine dense mri. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2023. doi: 10.1109/ISBI53787.2023.10230782.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

Miaomiao Zhang and P Thomas Fletcher. Finite-dimensional lie algebras for fast diffeomorphic image registration. In *International conference on information processing in medical imaging*, pages 249–260. Springer, 2015.

Miaomiao Zhang and P Thomas Fletcher. Fast diffeomorphic image registration via fourier-approximated lie algebras. *International Journal of Computer Vision*, 127(1):61–73, 2019.

Miaomiao Zhang, Nikhil Singh, and P Thomas Fletcher. Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In *International conference on information processing in medical imaging*, pages 37–48. Springer, 2013.