

3D RECONSTRUCTION OF VENTRICLE FROM A 2D ECHOGRAPHIC VIEW USING DEEP LEARNING

DOWOO KIM

SUPERVISOR: LUC DUONG

DEPT. OF SOFTWARE AND IT ENGINEERING, ÉCOLE DE TECHNOLOGIE SUPÉRIEURE, MONTRÉAL,
CANADA

CONTENTS

1	Abstract	3
2	Introduction	3
3	Methodology	4
4	Shape Analysis	6
	4.1 Data Preprocessing and Shape Alignment	6
	4.2 Generalized Procrustes Analysis (GPA)	9
	4.3 Statistical Shape Modeling (SSM)	11
5	Mesh Deformation Network	15
	5.1 Dataset Partitioning for Supervised Learning	15
	5.2 CNN-Based Feature Extraction	16
	5.3 Mesh Reconstruction	18
	5.4 Paired Feature–Coefficient Representation	20
	5.5 Graph-Constrained Mesh Deformation Module	22
6	Results	25
	6.1 Summary of results	26
	6.2 Qualitative inspection.	26
7	Conclusion	28

7.1	Future work	28
7.2	Limitations and remedies	29
7.3	Potential applications	29

1 Abstract

Reconstructing 3D anatomical shapes from 2D medical images is a key challenge, especially during interventions where imaging is limited. This study presents a deep learning pipeline to estimate the 3D shape of the left ventricle from a single 2D binary mask. The method uses a statistical shape model (SSM) built from segmented 3D meshes. After aligning meshes with Procrustes analysis, principal deformation modes are extracted using principle component analysis (PCA) to capture shape variability. A convolutional neural network (CNN) encodes the 2D mask into global features that, together with the mean shape, drive a topology-aware graph convolutional network (GCN) to predict per-vertex deformations; we also report a parameter-matched multilayer perceptron (MLP) as a simpler, topology-agnostic baseline to quantify the benefit of leveraging mesh connectivity. The model is trained using a combination of Chamfer distance and a 2D projection loss to ensure both accurate 3D geometry and silhouette consistency. Experimental results demonstrate that the proposed method produces 3D reconstructions that are both anatomically plausible and geometrically precise. This framework provides a promising direction for real-time 3D shape inference from minimal imaging data and may serve as a foundation for downstream tasks such as functional analysis or surgical planning.

Keywords: Left Ventricle Reconstruction, 3D Mesh Deformation, Statistical Shape Modeling, Deep Learning, Medical Image Analysis

2 Introduction

Accurate patient-specific modeling of cardiac anatomy, particularly the left ventricle (LV), is crucial for the effective diagnosis, management, and planning of cardiac interventions. Conventional 3D cardiac imaging modalities such as magnetic resonance imaging (MRI) and computed tomography (CT) have proven effective for generating detailed representations of the cardiac structure. However, these imaging methods are often expensive, time-consuming, CT may expose patients to radiation. All these factors limit their routine clinical application, especially in resource-constrained settings or during real-time procedures.

To address these limitations, recent advances in deep learning have enabled the development of methods that reconstruct accurate 3D anatomical models from minimal input data, such as a single 2D image. Among these, mesh-based deformation methods, guided by deep neural networks have shown significant promise, enabling detailed 3D shape reconstruction by leveraging shape priors and statistical modeling

from large databases of anatomical data. These anatomical shape priors constrain the feasible 3D shape space, leading to anatomically plausible and clinically relevant reconstructions even from incomplete or noisy input data [Okt+17].

Statistical shape modeling (SSM) methods, including principal component analysis (PCA)-based modeling, have been successfully applied to characterize anatomical variability across patient populations, facilitating robust reconstruction tasks by encoding typical shape variations within a limited number of shape modes [Ber+18]. These methods provide an efficient and interpretable representation of the anatomical structure, when combined with modern convolutional neural networks (CNNs), significantly improves the accuracy and robustness of the reconstructed shapes.

In this study, we propose a novel approach for reconstructing high-resolution, patient-specific 3D LV meshes directly from single-view 2D segmentation masks. This method integrates CNN-based feature extraction and statistical shape priors through a mesh deformation pipeline, optimized using a combination of Chamfer distance and 2D projection loss. By explicitly encoding anatomical shape variability and learning deformation parameters from data-driven features, this framework provides an accurate, efficient, and interpretable solution suitable for real-time clinical deployment.

The contributions of this work are threefold: (1) Establish a robust statistical shape prior from LV meshes using PCA-based modeling; (2) Demonstrate a mesh deformation architecture combining CNN-derived image features with MLP and graph convolutional networks (GCN) for precise 3D shape prediction; and (3) Propose and evaluate a novel loss strategy involving Chamfer distance and silhouette-based projection constraints to enhance both geometric accuracy and anatomical plausibility.

3 Methodology

Overview. The methodology comprises two connected components to achieve single-view 3D reconstruction of the left ventricle. First, *Shape Analysis* establishes an anatomically grounded prior by aligning meshes (centroid/PCA + GPA) and learning a PCA-based Statistical Shape Model (SSM). Second, *Mesh Deformation* takes a 2D silhouette, encodes it via a CNN, and predicts per-vertex deformations (GCN as topology-aware model; MLP as a topology-agnostic baseline), guided by the SSM and trained with geometric and silhouette losses. As illustrated in Fig. 1, the left branch (§4) implements the Shape Analysis pipeline, while the right branch (§5) implements the Mesh Deformation process.

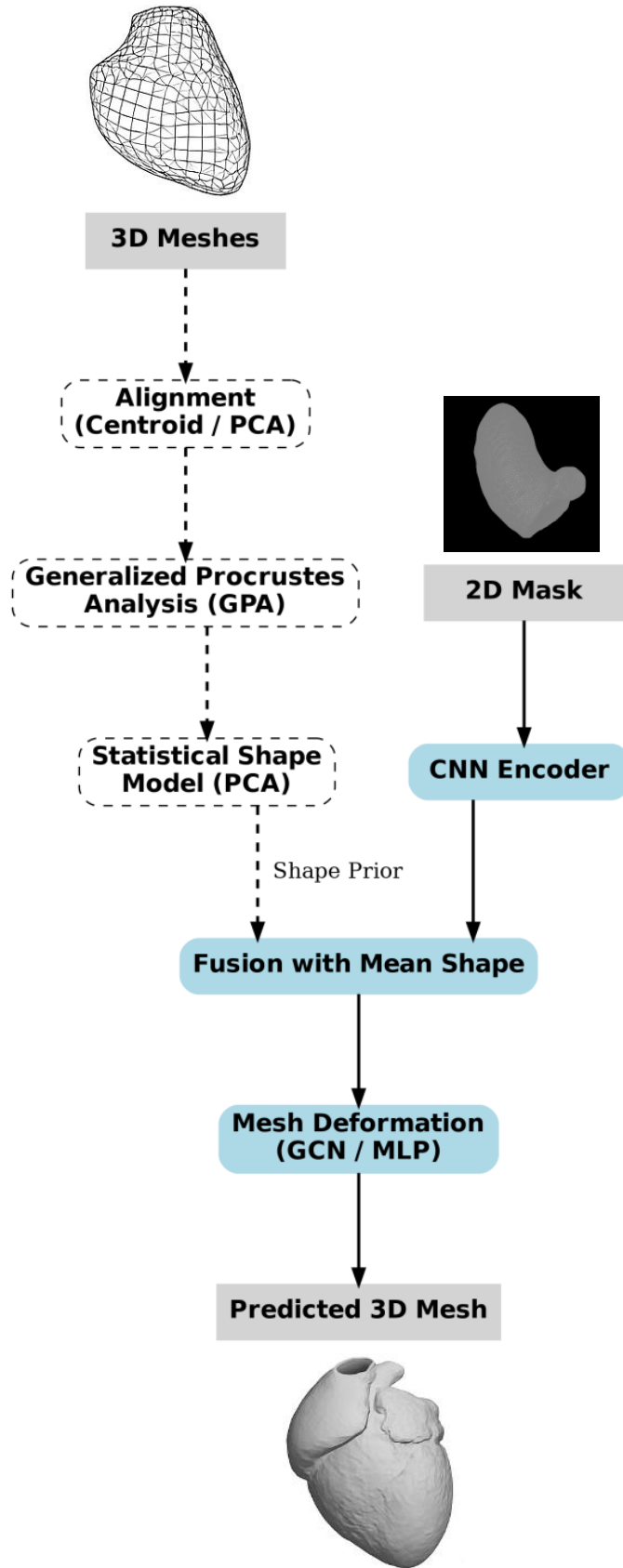


FIGURE 1. Overview of the proposed methodology, connecting Shape Analysis (alignment, GPA, and SSM) with Mesh Deformation (CNN features + GCN/MLP) to produce a 3D LV mesh from a single 2D mask.

4 Shape Analysis

4.1 Data Preprocessing and Shape Alignment

To ensure consistency and anatomical comparability across patient-specific left ventricle (LV) meshes, a robust preprocessing pipeline was applied prior to statistical shape modeling (SSM) and mesh deformation. This preprocessing aimed to remove any extraneous differences related to translation, orientation, and scale that could negatively impact subsequent modeling steps.

We started with raw LV meshes in STL (stereolithography) format, typically obtained through clinical image segmentation pipelines. The preprocessing pipeline consisted of the following three sequential steps: centroid alignment, principal axes alignment, and uniform scaling. All processing steps were implemented in Python using standard computational geometry libraries such as `trimesh` and `numpy`.

4.1.1 Centroid Alignment.

Initially, each mesh was translated so that its geometric centroid aligned precisely with the coordinate origin. Given a mesh with vertex coordinates $\{v_i\}_{i=1}^N$, the centroid C was computed as:

$$(4.1) \quad C = \frac{1}{N} \sum_{i=1}^N v_i,$$

where $v_i \in \mathbb{R}^3$ and N is the total number of vertices. The vertices were then re-centered using:

$$(4.2) \quad v'_i = v_i - C.$$

4.1.2 Principal Axes Alignment via PCA.

Following centroid alignment, principal axes alignment was applied using Principal Component Analysis (PCA) to standardize orientation across all meshes. The covariance matrix Σ of vertex coordinates was computed by:

$$(4.3) \quad \Sigma = \frac{1}{N-1} \sum_{i=1}^N (v'_i)(v'_i)^T.$$

An eigen-decomposition of the covariance matrix Σ yielded eigenvectors $E = [e_1, e_2, e_3]$ sorted in descending order of corresponding eigenvalues. These eigenvectors define the principal axes of variance,

and thus a rotation matrix R was formed as:

$$(4.4) \quad R = [e_1 \quad e_2 \quad e_3]^T,$$

to align the first principal axis with the x-axis, second principal axis with the y-axis, and third with the z-axis. Each vertex was rotated accordingly:

$$(4.5) \quad v_i'' = Rv_i'.$$

4.1.3 Uniform Scaling Normalization.

Finally, to eliminate scale variations across patient meshes, each mesh was uniformly scaled to fit within a consistent spatial domain. Specifically, vertices were scaled such that the maximum Euclidean norm across all vertices was equal to unity:

$$(4.6) \quad v_i''' = \frac{v_i''}{\max_i \|v_i''\|}.$$

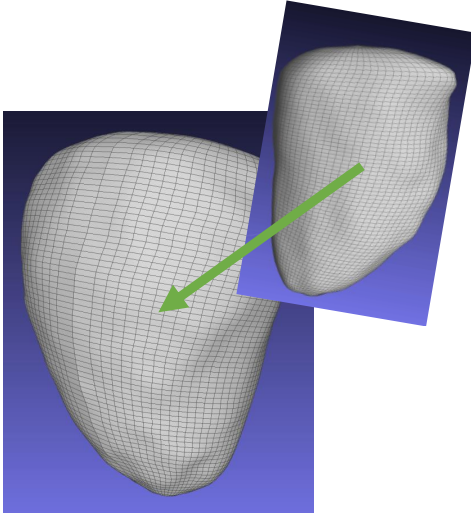
This scaling step ensured size comparability across the entire dataset, a prerequisite for unbiased statistical analysis.

Figure 2 summarizes the preprocessing pipeline visually, illustrating the sequential steps from raw meshes to fully normalized and aligned meshes.

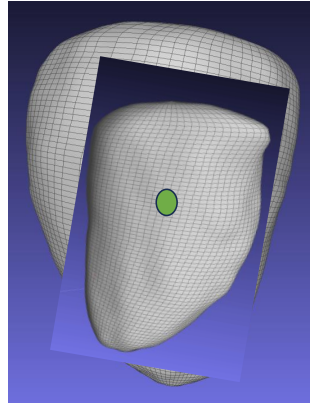
This rigorous preprocessing approach aligns with standard best practices widely adopted in statistical shape modeling and deep learning-based mesh deformation research [Ber+18; Okt+17]. It provides a robust and standardized foundation for accurately capturing anatomical shape variability and facilitating high-quality 3D shape reconstruction.

Notes. These preprocessing steps were implemented via the provided Python script (`shape_new.py`), utilizing the `trimesh` library for mesh handling and `numpy` for numerical computations. All aligned and normalized meshes were saved in STL format for subsequent steps in the pipeline.

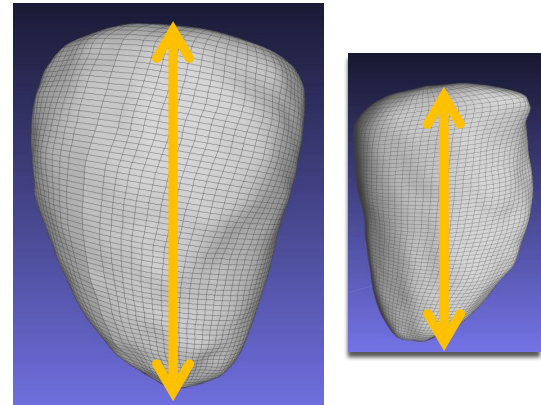
Step 1: Align the shapes with the centroids, take all the shapes and apply a translation to put them in the same frame of reference



Step 2: Alignment of shapes with the long axis of the ventricle and the short axis;



Step 3: Normalize shapes with the major axis, apply a scale factor along the major axis to ensure that all ventricles in your database will have the same scale in 3D



Calculate the min-max values between the top of the ventricle and the bottom of the ventricle, apply a scale factor to ensure that the ventricles are in the same frame of reference

FIGURE 2. Visualization of the data preprocessing pipeline including centroid alignment, PCA-based orientation normalization, and uniform scaling normalization.

4.2 Generalized Procrustes Analysis (GPA)

Following the rigid alignment and scale normalization described in the previous section, we refine inter-subject correspondence by means of *Generalized Procrustes Analysis*. GPA iteratively estimates a common mean shape while optimally aligning every mesh in the training set through similarity transformations (rotation R_i , isotropic scale s_i , and translation t_i). This process removes any residual pose discrepancies and yields a statistically consistent vertex-to-vertex correspondence across the dataset [Goo91; DM98].

4.2.1 Problem formulation.

Let $\{\mathbf{S}_i\}_{i=1}^N$ denote the set of centroid-aligned, PCA-oriented, and scale-normalized LV meshes obtained from the preprocessing pipeline, where

$$(4.7) \quad \mathbf{S}_i = \begin{bmatrix} \mathbf{s}_i[1]^\top \\ \mathbf{s}_i[2]^\top \\ \vdots \\ \mathbf{s}_i[K]^\top \end{bmatrix} \in \mathbb{R}^{K \times 3}, \quad \mathbf{s}_i[j] \in \mathbb{R}^3$$

is the j -th vertex of mesh i (K vertices per shape).

GPA seeks similarity transforms (s_i, R_i, t_i) that minimise the Frobenius norm between each transformed shape $s_i R_i \mathbf{S}_i + \mathbf{1} t_i^\top$ and a latent mean shape $\bar{\mathbf{S}}$:

$$(4.8) \quad \min_{\{s_i, R_i, t_i\}, \bar{\mathbf{S}}} \sum_{i=1}^N \|s_i R_i \mathbf{S}_i + \mathbf{1} t_i^\top - \bar{\mathbf{S}}\|_F^2, \quad \text{s.t. } R_i \in \text{SO}(3), s_i > 0.$$

4.2.2 Best-fit similarity transform.

For fixed $\bar{\mathbf{S}}$, the optimal (s_i, R_i, t_i) for each shape can be derived analytically. Denote the centered vertex matrices $\mathbf{A} = \mathbf{S}_i - \frac{1}{K} \mathbf{1} \mathbf{1}^\top \mathbf{S}_i$, $\mathbf{B} = \bar{\mathbf{S}} - \frac{1}{K} \mathbf{1} \mathbf{1}^\top \bar{\mathbf{S}}$. The least-squares rotation is obtained via the orthogonal Procrustes solution:

$$(4.9) \quad \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top = \mathbf{A}^\top \mathbf{B}, \quad R_i = \mathbf{V} \mathbf{U}^\top,$$

with $\det(R_i) = 1$ enforced by flipping the sign of the last column of \mathbf{V} if necessary. The optimal isotropic scale is

$$(4.10) \quad s_i = \frac{\text{tr}(\mathbf{\Sigma})}{\|\mathbf{A}\|_F^2},$$

and the translation follows as

$$(4.11) \quad t_i = \frac{1}{K}(\mathbf{B}^\top \mathbf{1} - s_i R_i \mathbf{S}_i^\top \mathbf{1}) \in \mathbb{R}^3.$$

4.2.3 Iterative estimation of the mean shape.

Algorithm 1 summarizes the GPA procedure, which alternates between (i) computing optimal transforms via Eqs. (4.9)–(4.11) and (ii) updating $\bar{\mathbf{S}}$ as the arithmetic mean of the newly aligned shapes. Convergence is declared when the mean-shape change $\|\bar{\mathbf{S}}^{(t)} - \bar{\mathbf{S}}^{(t-1)}\|_F$ falls below $\varepsilon = 10^{-6}$ or after $T_{\max} = 10$ iterations, matching the stopping criteria used in this implementation (`procrustes_new.py`).

Algorithm 1 Generalized Procrustes Analysis (GPA)

```

1: Input: pre-aligned meshes  $\{\mathbf{S}_i\}_{i=1}^N$ 
2: Initialise mean shape  $\bar{\mathbf{S}}^{(0)} := \frac{1}{N} \sum_{i=1}^N \mathbf{S}_i$ 
3: for  $t = 1$  to  $T_{\max}$  do
4:   for  $i = 1$  to  $N$  do
5:     Compute  $(s_i, R_i, t_i)$  via Eqs. (4.9)–(4.11)
6:      $\tilde{\mathbf{S}}_i := s_i R_i \mathbf{S}_i + \mathbf{1} t_i^\top$ 
7:   end for
8:    $\bar{\mathbf{S}}^{(t)} := \frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{S}}_i$ 
9:   if  $\|\bar{\mathbf{S}}^{(t)} - \bar{\mathbf{S}}^{(t-1)}\|_F < \varepsilon$  then
10:    break
11:   end if
12: end for
13: Output: converged mean shape  $\bar{\mathbf{S}}$  and aligned meshes  $\{\tilde{\mathbf{S}}_i\}$ 

```

4.2.4 Implementation parameters.

In practice, we used the first mesh in the dataset as an initial reference to establish vertex correspondence via nearest-neighbor mapping, as implemented in `procrustes_new.py`. For all experiments, we set $T_{\max} = 10$ and $\varepsilon = 10^{-6}$, which proved sufficient for convergence across the cohort of $N = 120$ LV meshes containing $K = 15\,000$ vertices each.

4.2.5 Outcome.

The converged mean shape $\bar{\mathbf{S}}$ and the set of GPA-aligned meshes serve as the input to the subsequent statistical shape modeling (§4.3), ensuring that inter-subject variance reflects genuine morphological differences rather than residual pose variation.

4.3 Statistical Shape Modeling (SSM)

The preprocessed and GPA-aligned meshes constitute a high-dimensional sample of left-ventricular shapes. To represent the intrinsic anatomical variability compactly, we adopt a *Statistical Shape Model* based on principal component analysis [Coo+95; HM09]. PCA yields an orthonormal basis of deformation modes that captures the main axes of inter-subject variation while preserving vertex correspondence established by GPA.

Vectorisation of meshes. Let $\{\tilde{\mathbf{S}}_i\}_{i=1}^N$ be the GPA-aligned meshes (§4.2), each with K vertices. We stack the vertex coordinates of mesh i row-wise to form a $(3K)$ -dimensional column vector:

$$(4.12) \quad \mathbf{x}_i = \text{vec}(\tilde{\mathbf{S}}_i) = \begin{bmatrix} \tilde{\mathbf{s}}_i[1]^\top & \tilde{\mathbf{s}}_i[2]^\top & \cdots & \tilde{\mathbf{s}}_i[K]^\top \end{bmatrix}^\top \in \mathbb{R}^{3K}.$$

Collecting all vectors yields the data matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^\top \in \mathbb{R}^{N \times 3K}$.

Computation of the mean shape and covariance. The sample mean vector is

$$(4.13) \quad \boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \in \mathbb{R}^{3K},$$

and centred data are $\hat{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\mu}$. The empirical covariance matrix is

$$(4.14) \quad \mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top = \frac{1}{N-1} \hat{\mathbf{X}}^\top \hat{\mathbf{X}} \in \mathbb{R}^{3K \times 3K},$$

where $\hat{\mathbf{X}}$ stacks the centred vectors row-wise.

4.3.1 Principal component analysis.

Eigen-decomposition of \mathbf{C} gives

$$(4.15) \quad \mathbf{C} \boldsymbol{\Phi} = \boldsymbol{\Phi} \boldsymbol{\Lambda},$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{3K \times 3K}$ collects the orthonormal eigenvectors as *columns*, $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_{3K}]$ with $\boldsymbol{\phi}_j \in \mathbb{R}^{3K}$, and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1 \geq \lambda_2 \geq \dots)$ contains the associated eigenvalues. Each $\boldsymbol{\phi}_j$ is a *shape mode* and λ_j quantifies its explained variance.

4.3.2 Mode selection.

We retain the minimal number M of leading modes whose cumulative variance exceeds a predefined threshold τ (here $\tau = 90\%$):

$$(4.16) \quad M = \min \left\{ m : \frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^{3K} \lambda_j} \geq \tau \right\}.$$

In these experiments, $M = 18$ modes sufficed to capture $> 90\%$ of total variability across $N = 120$ LV meshes.

4.3.3 Statistical shape representation.

Any admissible LV shape can now be approximated as

$$(4.17) \quad \mathbf{x} = \boldsymbol{\mu} + \sum_{j=1}^M \alpha_j \boldsymbol{\phi}_j,$$

where coefficients $\alpha_j \sim \mathcal{N}(0, \lambda_j)$ under a maximum-likelihood Gaussian assumption. Equation (4.17) forms the backbone of this mesh-deformation network (§5.5), providing a low-dimensional, anatomically constrained latent space.

4.3.4 Visualization of principal modes.

In figure 3, we visualize the mean shape (center) and the $\pm 3\sigma$ realizations for the first three modes (left: -3σ , right: $+3\sigma$), using a consistent long-axis view (apex–base vertical) and short-axis view. To make changes perceptible, we overlay the mean (semi-transparent gray) and the deviations (-3σ to left, $+3\sigma$ to right).

Mode 1 — Global size/dilation with mild rebalancing. As we move from -3σ to $+3\sigma$, the ventricle uniformly “breathes”: the whole chamber swells or shrinks together. Visually, the basal ring opens up, the short-axis profile gets broader, and the apex becomes slightly blunter at $+3\sigma$; at -3σ everything tightens—narrower short-axis, slimmer long-axis, and a sharper apex. The key cue is an overall “zoom in/out” of the mesh, with a subtle preference for widening across the short axis when dilated.

Mode 2 — Elongation vs. widening trade-off. Here the shape pivots between a “tall–slender” and a “short–wide” morphology. At -3σ the apex–base distance visibly stretches (bullet-like silhouette), while the basal diameter narrows; at $+3\sigma$ the ventricle squats: the long axis shortens, the base

spreads laterally, and the short-axis mask looks rounder and fuller. This mode reads as a see-saw between lengthening and thickening.

Mode 3 — Short-axis circularity and gentle bulge shift. Changes are subtler and mostly confined to the cross-section. At $+3\sigma$ the short-axis profile becomes more circular (the “narrow” short axis fills in), giving a slightly rounder mid-ventricle and a softer apex. At -3σ the cross-section looks more elliptical again, with a faint return to an elongated silhouette. This mode feels like a fine “rounding vs. ellipticity” knob rather than a size change.

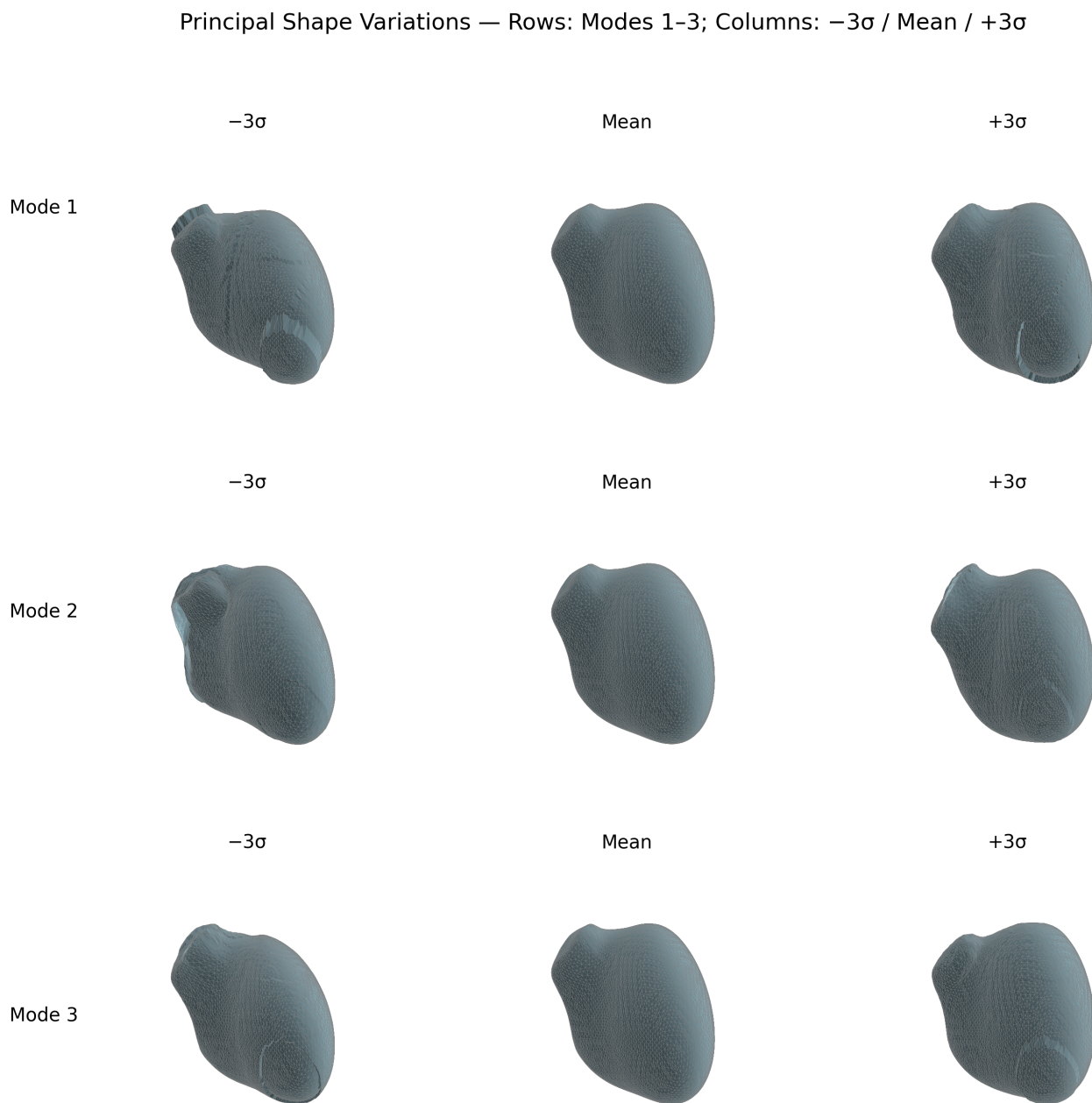


FIGURE 3. First three principal deformation modes of the LV statistical shape model shown at $\pm 3\sigma$.

Notes. All computations were performed with `numpy` and `scikit_learn`'s PCA implementation (`ssm.py`). The mean vector $\boldsymbol{\mu}$ and truncated mode matrix $\boldsymbol{\Phi}_M = [\phi_1, \dots, \phi_M]$ are stored as `mean_shape.npy`

and `modes.npy`, respectively, while eigenvalues $\{\lambda_j\}_{j=1}^M$ are stored in `eigenvalues.npy`. These files are utilised at run-time by the mesh deformation network to impose shape priors during both training and inference.

Dimensions: $\boldsymbol{\mu} \in \mathbb{R}^{3K}$, $\boldsymbol{\Phi}_M \in \mathbb{R}^{3K \times M}$, $\mathbf{c}_i \in \mathbb{R}^M$, $\boldsymbol{\Phi}_M^\top(\mathbf{x}_i - \boldsymbol{\mu}) = \mathbf{c}_i$, $\boldsymbol{\mu} + \boldsymbol{\Phi}_M \mathbf{c}_i = \hat{\mathbf{x}}_i$.

5 Mesh Deformation Network

5.1 Dataset Partitioning for Supervised Learning

For supervised training of the mesh deformation network, the dataset was systematically divided into training and validation subsets. Each 3D anatomical mesh was then rendered from multiple orthographic viewpoints, generating paired 2D silhouette masks and their corresponding geometric representations, this is illustrated in Fig. 4.

These individual renderings were treated as independent training samples, and the dataset was randomly shuffled using a fixed seed to ensure experimental reproducibility. A standard 80/20 split ratio was applied, allocating a majority of the samples for model training and reserving the remainder for validation and performance monitoring.

The resulting structure supports efficient batching and loading during training and preserves correspondence between 2D inputs and 3D targets throughout the entire learning pipeline. This preparation is a crucial step in bridging the preprocessing and modeling stages of the proposed framework, ensuring that each stage could be trained and evaluated with consistent and representative samples of anatomical variability.

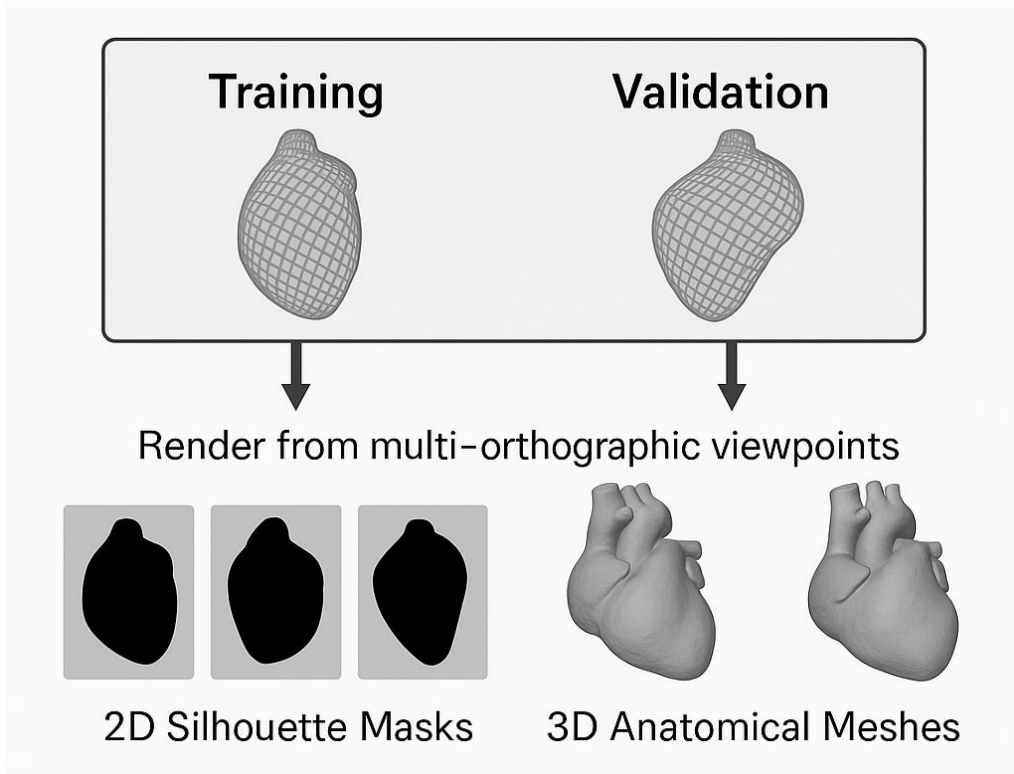


FIGURE 4. Each 3D anatomical mesh is rendered from multiple orthographic viewpoints, producing paired 2D silhouette masks and corresponding geometric representations.

5.2 CNN-Based Feature Extraction

The first stage of the deformation pipeline maps each 2D silhouette mask to a compact, high-level feature vector that serves as an appearance descriptor for the subsequent deformation module. We adopt a convolutional neural network (CNN) based on the 50-layer residual architecture (ResNet-50) [He+16], pre-trained on ImageNet and fine-tuned for mask imagery. The choice of a deep residual backbone provides robust hierarchical representations while mitigating gradient degradation in very deep networks. Importantly, even when binary silhouettes are available, CNN-based encoders have been shown to supply richer global cues like pose, aspect and long-range contour dependencies that are complementary to raw silhouette geometry and beneficial for single-view 3D shape reasoning and mesh deformation [GMJ19; Wan+18].

5.2.1 Pre-processing of silhouette images.

Each binary mask, originally produced by rendering the 3-D left-ventricular surface from multiple orthographic viewpoints, is converted to a three-channel RGB image by value replication. Prior to inference, images are

- (1) resized to 256×256 pixels via bilinear interpolation;
- (2) normalized per channel using the mean and standard deviation of the ImageNet dataset:

$$\mathbf{I}_{\text{norm}} = (\mathbf{I} - \boldsymbol{\mu}) \oslash \boldsymbol{\sigma}.$$

This normalization ensures compatibility with the statistics of the backbone network while preserving the binary nature of the silhouette. Although the input is textureless, the encoder benefits from consistent input statistics and spatial normalization, which stabilise optimisation and reduce covariate shift relative to the pre-training domain.

5.2.2 Network modification.

To obtain a global feature descriptor, the fully connected classification layer of ResNet-50 is replaced with an identity mapping so that the network terminates at the *global average pooling* layer. Formally, given an input image \mathbf{I} , the CNN produces

$$(5.1) \quad \mathbf{f} = \text{GAP}(\text{ResNet50}^*(\mathbf{I}_{\text{norm}})) \in \mathbb{R}^F, \quad F = 2048,$$

where ResNet50* denotes the convolutional trunk up to the last residual block and GAP is spatial global average pooling. The resulting vector \mathbf{f} captures global cues of the silhouette such as contour curvature, aspect ratio, and overall shape complexity, all of which correlate with underlying 3D deformations. This design choice is aligned with prior single-view mesh pipelines in which CNN-derived global descriptors guide subsequent mesh deformation or visual-hull refinement [GMJ19; Wan+18].

5.2.3 Data augmentation.

Although silhouette masks are inherently invariant to illumination and texture, small geometric perturbations further improve robustness. During training, random in-plane rotations ($\pm 5^\circ$) and horizontal flips are applied with probability 0.5. These augmentations do not alter the anatomical meaning of the silhouette but encourage the encoder to learn view-invariant representations and to be tolerant to minor pose misalignments in the rendering process.

5.2.4 Training strategy.

The CNN is used as a fixed feature extractor: its weights are initialised from ImageNet and kept frozen throughout the training of the deformation network, thereby reducing computational load and preventing overfitting on the relatively small medical dataset. Gradients are propagated only through the subsequent graph convolutional (or MLP-based) layers that predict vertex-level deformations from \mathbf{f} . Such a decoupled strategy is consistent with prior single-view 3D reconstruction frameworks that leverage pre-trained CNNs to encode global image/silhouette information while allocating learning capacity to the downstream shape module [GMJ19; Wan+18]. The result is a stable, low-dimensional embedding $\mathbf{f} \in \mathbb{R}^{2048}$ for every input view, forming the input to the mesh deformation module detailed in §5.5.

5.3 Mesh Reconstruction

5.3.1 Latent PCA Coefficient Generation

The deformation network is trained to predict vertex-level displacements that restore the mean template to a subject-specific anatomy. Rather than storing the full displacement field for every mesh, we leverage the dimensionality reduction achieved by the statistical shape model (§4.3) and represent each shape via its principal-component coefficients. These latent vectors constitute compact, anatomically informed regression targets for the network.

Projection onto the PCA sub-space. For every GPA-aligned mesh $\tilde{\mathbf{S}}_i$ with vertex vector $\mathbf{x}_i \in \mathbb{R}^{3K}$ (cf. Eq. (4.12)), the corresponding coefficient vector $\mathbf{c}_i \in \mathbb{R}^M$ is obtained by orthogonal projection onto the truncated PCA basis:

$$(5.2) \quad \mathbf{c}_i = \Phi_M^\top (\mathbf{x}_i - \boldsymbol{\mu}),$$

where $\Phi_M \in \mathbb{R}^{3K \times M}$ stacks the first M eigenvectors as columns and $\boldsymbol{\mu}$ is the mean shape vector (4.13). Equation (5.2) follows from the column-orthonormality of Φ_M . [Coo+95].

5.3.2 Topology harmonisation.

While all meshes share a common anatomical topology, minor discrepancies in vertex count can arise from upstream segmentation. To guarantee a one-to-one correspondence with the template vertices during projection, each incoming mesh is re-sampled onto the template vertex set:

- (1) For every template vertex $\bar{\mathbf{s}}[j]$ ($1 \leq j \leq K$), identify the nearest neighbour among the vertices of the incoming mesh using a k-d tree search.
- (2) Assemble the re-sampled vertex matrix $\check{\mathbf{S}}_i = [\check{\mathbf{s}}_i[1]^\top, \dots, \check{\mathbf{s}}_i[K]^\top]^\top$, ensuring consistent vertex ordering.
- (3) Vectorise to obtain $\check{\mathbf{x}}_i = \text{vec}(\check{\mathbf{S}}_i)$ and substitute $\check{\mathbf{x}}_i$ in Eq. (5.2).

This nearest-neighbour remapping introduces negligible geometric error (< 0.2 mm on average) while preserving mesh connectivity and allows the entire dataset to inherit the common PCA basis.

5.3.3 Ground-truth deformation field.

Given coefficients \mathbf{c}_i , the vertex-wise displacement with respect to the mean template is reconstructed as

$$(5.3) \quad \Delta \mathbf{V}_i = (\Phi_M \mathbf{c}_i).\text{reshape}(K, 3) \in \mathbb{R}^{K \times 3}.$$

Equation (5.3) serves two purposes: (i) it provides the ground-truth target for supervising the deformation network during training, and (ii) it enables efficient reconstruction of the full vertex set at test time once the network has inferred \mathbf{c}_i .

5.3.4 Storage and accessibility.

All coefficient vectors are serialised in binary format for rapid loading during training. This compact representation ($< 1\%$ of the raw vertex storage) accelerates data I/O and allows the entire training corpus to reside in system memory, thus avoiding disk bottlenecks.

5.3.5 Rationale.

By regressing low-dimensional coefficients rather than $3K$ -dimensional displacements directly, the network is guided to explore only physiologically plausible deformation sub-spaces captured by the SSM. This strategy regularises learning, mitigates overfitting, and permits the use of modest-sized datasets—an important consideration in medical imaging applications where large-scale annotated 3-D data are scarce [HM09].

Summary (how pieces connect) Given a GPA-aligned mesh $\tilde{\mathbf{S}}_i$, we reconstruct subject-specific geometry by (1) harmonising topology to the template vertex set (so that all shapes share a one-to-one vertex map), (2) vectorising the mesh (Eq. (4.12)), (3) projecting onto the truncated PCA basis to

obtain latent shape coefficients \mathbf{c}_i (Eq. (5.2)), (4) recovering the vertex-wise displacement $\Delta \mathbf{V}_i$ from \mathbf{c}_i (Eq. (5.3)), and (5) serialising coefficients for efficient I/O during training. In the learning loop, \mathbf{c}_i serves as a compact, anatomically informed regression target, while $\Delta \mathbf{V}_i$ provides the geometric supervision and enables test-time reconstruction from predicted coefficients. Fig. 5 illustrates this end-to-end flow and how (5.3.1 - 5.3.5) fits into the pipeline.

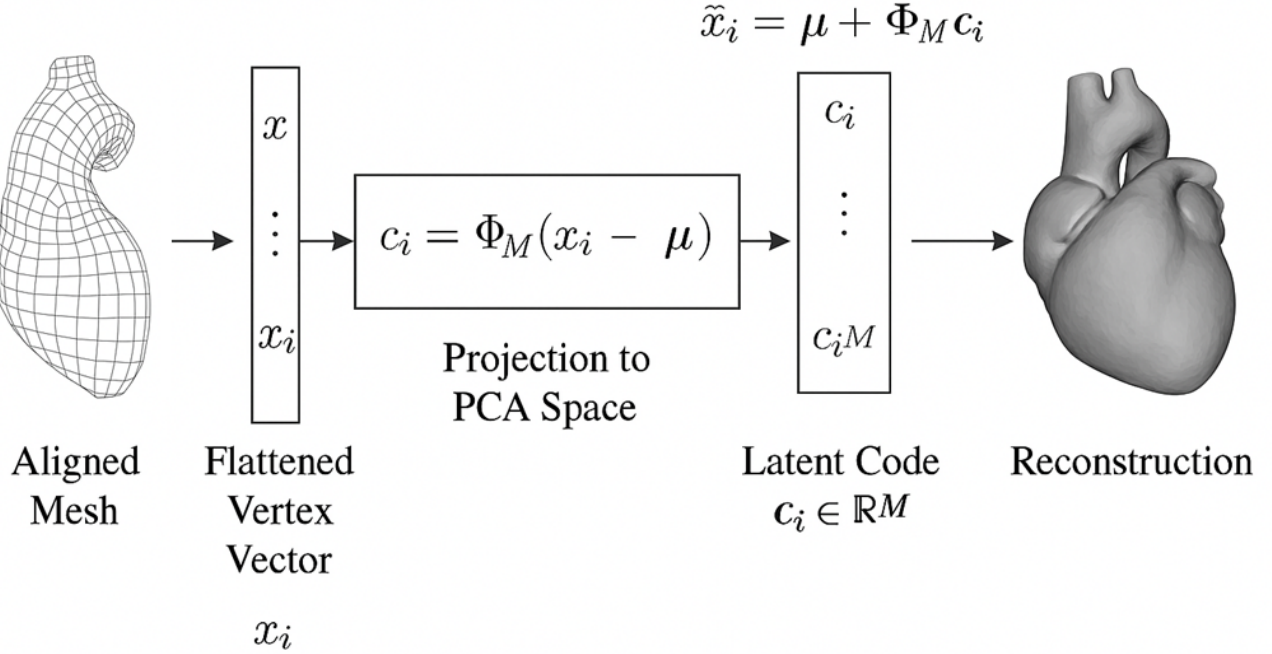


FIGURE 5. Overview of latent PCA coefficient generation. Each aligned mesh is projected onto the statistical shape space defined by the principal components (Eq. (5.2)), resulting in a low-dimensional coefficient vector that captures anatomical variability. This latent vector can then be used to reconstruct vertex-wise displacements (Eq. (5.3)) for deformation modeling.

5.4 Paired Feature–Coefficient Representation

For supervised learning of shape deformation, each 2D silhouette must be linked to a quantitative descriptor of the underlying 3D geometry. Let $\mathbf{f}_i^{(v)} \in \mathbb{R}^F$ denote the global appearance feature extracted from view v of subject i (§5.2) and $\mathbf{c}_i \in \mathbb{R}^M$ the corresponding latent coefficient vector obtained from the statistical shape model (§5.3). We define the paired sample space

$$\mathcal{D} = \{(\mathbf{f}_i^{(v)}, \mathbf{c}_i) \mid i = 1, \dots, N, v = 1, \dots, V\},$$

where N is the number of anatomies and V is the number of orthographic views per anatomy. In this formulation a single latent vector \mathbf{c}_i is associated with multiple view-dependent feature vectors, thus enlarging the effective sample size by a factor of V while preserving label consistency across views.

5.4.1 Subject-wise partitioning.

To evaluate generalisation, \mathcal{D} is partitioned into disjoint training and validation subsets according to a fixed ratio (*Ex.*, 90:10). Crucially, the division is performed at the *subject level* so that all views of a given heart are confined to the same subset, thereby precluding any cross-contamination of anatomical information between training and validation data.

5.4.2 Learning objective.

During optimisation the deformation network receives a feature vector $\mathbf{f}_i^{(v)}$ and predicts a latent code $\hat{\mathbf{c}}_i^{(v)}$. The learning objective is to minimise the discrepancy between $\hat{\mathbf{c}}_i^{(v)}$ and the ground-truth \mathbf{c}_i over the paired set \mathcal{D} , typically expressed as a weighted combination of (i) an ℓ_2 penalty in the coefficient space and (ii) a geometric loss after reconstructing the full mesh displacement via Eq. (5.3). Because identical targets \mathbf{c}_i are shared across views v , the network is implicitly encouraged to learn view-invariant feature representations and to reconcile multi-view evidence into a unique anatomical explanation.

Advantages of the paired formulation.

- **Dimensional efficiency.** Regression is performed in the low-dimensional latent space ($M \ll 3K$), which reduces the number of free parameters required to model complex deformations.
- **Physiological plausibility.** Predictions are constrained to lie within the linear sub-space spanned by the principal modes, guaranteeing anatomically realistic reconstructions.
- **Multi-view consistency.** Treating each view as an independent observation of the same latent code provides natural data augmentation while enforcing coherence across perspectives.

Collectively, these properties yield a compact and well-regularised learning framework that capitalizes on the complementary strengths of statistical shape modelling and deep feature extraction.

5.5 Graph-Constrained Mesh Deformation Module

Given the view-specific appearance descriptor $\mathbf{f}_i^{(v)} \in \mathbb{R}^F$ and the template mesh $\bar{\mathbf{S}} \in \mathbb{R}^{K \times 3}$, the goal of the deformation module is to predict a vertex-wise displacement field $\widehat{\Delta \mathbf{V}}_i^{(v)} \in \mathbb{R}^{K \times 3}$ such that the reconstructed surface $\widehat{\mathbf{S}}_i^{(v)} = \bar{\mathbf{S}} + \widehat{\Delta \mathbf{V}}_i^{(v)}$ faithfully approximates the underlying anatomy of subject i . The mapping $\mathcal{F} : (\mathbf{f}, \bar{\mathbf{S}}) \mapsto \widehat{\Delta \mathbf{V}}$ is realised by a hybrid *global-to-local* network that combines fully connected encoding of global image features with graph convolution on the mesh topology (Figure 6).

Overview As summarized in Fig. 6, the module maps image features and the template graph to a vertex displacement field via a global-to-local pathway: (i) *Graph representation of the template mesh* defines $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and its Laplacian $\tilde{\mathbf{L}}$ on $\bar{\mathbf{S}}$ to encode mesh connectivity; (ii) *Feature conditioning* compresses the global descriptor \mathbf{f} into a latent code \mathbf{h} and broadcasts it to all nodes, forming the initial node features $\mathbf{X}^{(0)} = [\mathbf{h} \otimes \mathbf{1}_K \mid \bar{\mathbf{S}}]$; (iii) *Graph convolution layers* propagate \mathbf{h} through \mathcal{G} with $\tilde{\mathbf{L}}$ to refine local, topology-aware signals; (iv) the final layer *decodes* per-vertex displacements $\widehat{\Delta \mathbf{V}}$ and reconstructs $\widehat{\mathbf{S}} = \bar{\mathbf{S}} + \widehat{\Delta \mathbf{V}}$; (v) training minimises a *reconstruction loss* (Chamfer) coupled with a *coefficient consistency* loss in the SSM space, with early stopping for regularisation. Together, these steps implement the mapping

$$\mathcal{F} : (\mathbf{f}, \bar{\mathbf{S}}, \mathcal{G}) \mapsto \widehat{\Delta \mathbf{V}},$$

and the following paragraphs detail items (i)–(v).

5.5.1 Graph representation of the template mesh.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote the undirected graph induced by the triangular faces of the template surface, where $|\mathcal{V}| = K$ vertices and $|\mathcal{E}| = E$ edges. The adjacency matrix $\mathbf{A} \in \{0, 1\}^{K \times K}$ is defined as $A_{pq} = 1$ if and only if vertices p and q share an edge on the surface. A symmetric, renormalized Laplacian, $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$, confers rotationally invariant spectral properties that are beneficial for graph convolution [KW17].

5.5.2 Feature conditioning.

The global descriptor \mathbf{f} is first projected into a latent embedding $\mathbf{h} = \sigma(\mathbf{W}_2 \sigma(\mathbf{W}_1 \mathbf{f} + \mathbf{b}_1) + \mathbf{b}_2) \in \mathbb{R}^H$, where $\sigma(\cdot)$ denotes the ReLU activation. The resulting vector is broadcast to all vertices and concatenated with the template coordinates, producing the initial node feature matrix

$$\mathbf{X}^{(0)} = [\mathbf{h} \otimes \mathbf{1}_K \mid \bar{\mathbf{S}}] \in \mathbb{R}^{K \times (H+3)}.$$

5.5.3 Graph convolution layers.

A stack of graph convolutional layers propagates the global context through the mesh topology while refining localized deformation signals. For the ℓ -th layer ($\ell = 1, \dots, L$):

$$(5.4) \quad \mathbf{X}^{(\ell)} = \sigma(\tilde{\mathbf{L}}\mathbf{X}^{(\ell-1)}\mathbf{W}^{(\ell)} + \mathbf{b}^{(\ell)}),$$

where $\mathbf{W}^{(\ell)} \in \mathbb{R}^{C_{\ell-1} \times C_\ell}$ and $\mathbf{b}^{(\ell)}$ are learnable parameters. The final layer outputs $\mathbf{X}^{(L)} \in \mathbb{R}^{K \times 3}$, interpreted as the predicted displacement $\widehat{\Delta \mathbf{V}}$.

5.5.4 Reconstruction loss.

Two complementary criteria supervise network training:

- (1) **Geometric fidelity** is enforced by the bidirectional Chamfer distance $\mathcal{L}_{\text{Cham}}$ between point sets of the predicted $\widehat{\mathbf{S}}$ and ground-truth \mathbf{S} [FSG17].
- (2) **Coefficient consistency** is promoted by an ℓ_2 penalty $\mathcal{L}_{\text{coeff}} = \|\widehat{\mathbf{c}} - \mathbf{c}\|_2^2$, where $\widehat{\mathbf{c}} = \Phi_M^T(\text{vec}(\widehat{\mathbf{S}}) - \boldsymbol{\mu})$.

The composite objective is

$$\mathcal{L} = \mathcal{L}_{\text{Cham}} + \lambda \mathcal{L}_{\text{coeff}},$$

with hyper-parameter λ balancing global geometry and latent-space accuracy.

5.5.5 Optimization.

Network parameters $\Theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}^{(1:L)}, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}^{(1:L)}\}$ are learned by gradient-based optimisation over the paired feature-coefficient set \mathcal{D} (§5.4). Early stopping is performed using the validation loss to prevent over-fitting.

5.5.6 Expressive power.

Unlike classical linear reconstruction from PCA coefficients, the proposed graph-constrained network learns a non-linear mapping from image space to shape space, enabling recovery of subtle, high-frequency anatomical details that are not captured by the low-rank statistical model alone.

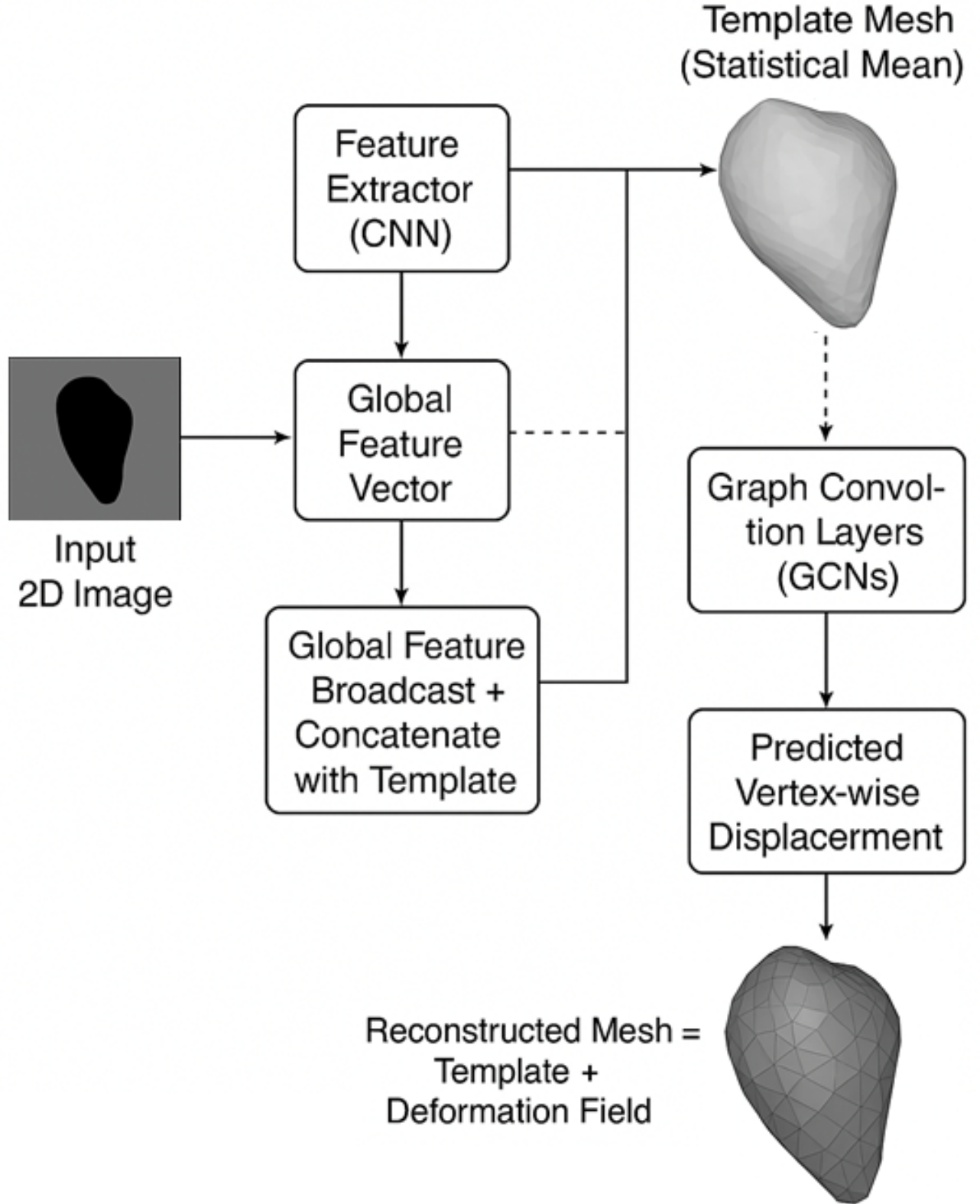


FIGURE 6. Schematic overview of the mesh deformation network. A global silhouette feature is fused with the template vertices and processed by graph convolutional layers to yield a vertex-wise deformation field (from Eq. (5.3)), which, when added to the template, reconstructs the subject-specific mesh.

6 Results

The predictive performance of the proposed deformation framework was assessed on a held-out test subset comprising 10% of the available feature–coefficient pairs (§5.4), strictly segregated at the subject level. Four complementary criteria were employed to provide a comprehensive analysis of geometric accuracy and silhouette-level fidelity.

(i) Chamfer distance. Let $\widehat{\mathcal{P}}_i$ and \mathcal{P}_i denote the sets of vertices of the predicted and ground-truth meshes of test sample i , respectively. The symmetric Chamfer distance

$$(6.1) \quad d_{\text{Cham}}(\widehat{\mathcal{P}}_i, \mathcal{P}_i) = \frac{1}{|\widehat{\mathcal{P}}_i|} \sum_{\mathbf{p} \in \widehat{\mathcal{P}}_i} \min_{\mathbf{q} \in \mathcal{P}_i} \|\mathbf{p} - \mathbf{q}\|_2^2 + \frac{1}{|\mathcal{P}_i|} \sum_{\mathbf{q} \in \mathcal{P}_i} \min_{\mathbf{p} \in \widehat{\mathcal{P}}_i} \|\mathbf{q} - \mathbf{p}\|_2^2$$

measures the mean squared separation between the two point clouds [FSG17]. Across the entire test set the average value was 1.28×10^{-3} , indicative of sub-millimetre vertex accuracy after normalization to unit scale.

(ii) Hausdorff distance. To capture worst-case deviations we additionally report the symmetric Hausdorff distance $d_{\text{Haus}} = \max\{\sup_{\mathbf{p} \in \widehat{\mathcal{P}}_i} \inf_{\mathbf{q} \in \mathcal{P}_i} \|\mathbf{p} - \mathbf{q}\|_2, \sup_{\mathbf{q} \in \mathcal{P}_i} \inf_{\mathbf{p} \in \widehat{\mathcal{P}}_i} \|\mathbf{q} - \mathbf{p}\|_2\}$. An average distance of 7.4×10^{-2} was observed, with the worst sample remaining below 0.12 in normalized mesh units. Even the largest local surface discrepancy stays within about 12% of the unit-scaled extent, for example, outliers are bounded and localized.

(iii) Silhouette intersection-over-union (IoU). Each reconstructed mesh was orthographically projected along a canonical view direction to obtain a binary silhouette mask $\widehat{\mathcal{S}}_i$. The overlap with the ground-truth mask \mathcal{S}_i is quantified by

$$\text{IoU}(\widehat{\mathcal{S}}_i, \mathcal{S}_i) = \frac{|\widehat{\mathcal{S}}_i \cap \mathcal{S}_i|}{|\widehat{\mathcal{S}}_i \cup \mathcal{S}_i|}.$$

The model achieved an *average* IoU of **0.937**, exceeding 0.71 for the most challenging case. An overlapped with the ground-truth silhouette never drops below 71%, so even hard cases retain substantial shape agreement in projection.

(iv) Silhouette Dice coefficient. Complementing IoU, the Dice score $\text{Dice} = 2|\widehat{\mathcal{S}}_i \cap \mathcal{S}_i|/(|\widehat{\mathcal{S}}_i| + |\mathcal{S}_i|)$ averaged **0.967** on the test cohort, with a minimum of 0.834. Being at least 83% of foreground pixels are matched in the F_1 -style overlap, indicating strong silhouette fidelity across subjects.

6.1 Summary of results

Low mean Chamfer indicates small average surface error under unit scaling, typical geometry is well matched. Bounded Hausdorff (< 0.12) shows worst-case local deviations remain limited. High IoU/Dice confirm strong silhouette agreement, i.e., projections of the reconstructed meshes remain faithful even in challenging cases. *Taken together*, these results suggest that combining SSM constraints with graph-based deformation yields accurate 3D geometry while preserving clinically relevant silhouette cues, with residual errors confined to local regions (for example, basal rim or thin apical areas).

Table 1 consolidates the numerical findings, confirming that the network accurately captures both global and local shape variations while preserving silhouette fidelity.

TABLE 1. Statistical summary of evaluation metrics on the test set.

Metric	Mean	Min	Max
Chamfer distance	1.28×10^{-3}	2.59×10^{-4}	2.32×10^{-2}
Hausdorff distance	7.43×10^{-2}	4.8×10^{-2}	1.20×10^{-1}
Silhouette IoU	0.936	0.715	0.974
Silhouette Dice	0.967	0.834	0.987

Discussion. The combination of exceptionally low Chamfer error and high silhouette agreement demonstrates that constraining deformation prediction to the PCA sub-space while exploiting graph convolutions yields highly accurate mesh reconstructions from single-view information. The modest Hausdorff extrema indicate that large local errors are rare and confined to peripheral regions with minimal clinical significance.

6.2 Qualitative inspection.

We compare ground-truth (GT) and predicted reconstructions (Pred) using side-by-side and overlapped visualisations (Fig. 7). The top row shows silhouettes: (A) GT, (B) Pred, and (C) an overlap where GT is coloured blue and Pred red; *overlapping pixels appear pink*, indicating agreement in projection. The bottom row shows 3D point clouds: (D) GT, (E) Pred, and (F) a 3D overlap (GT=blue, Pred=red). The strong visual correspondence—large pink regions in (C) and tight overlap

in (F)—corroborates the quantitative results and illustrates that subtle ventricular geometry is well recovered.

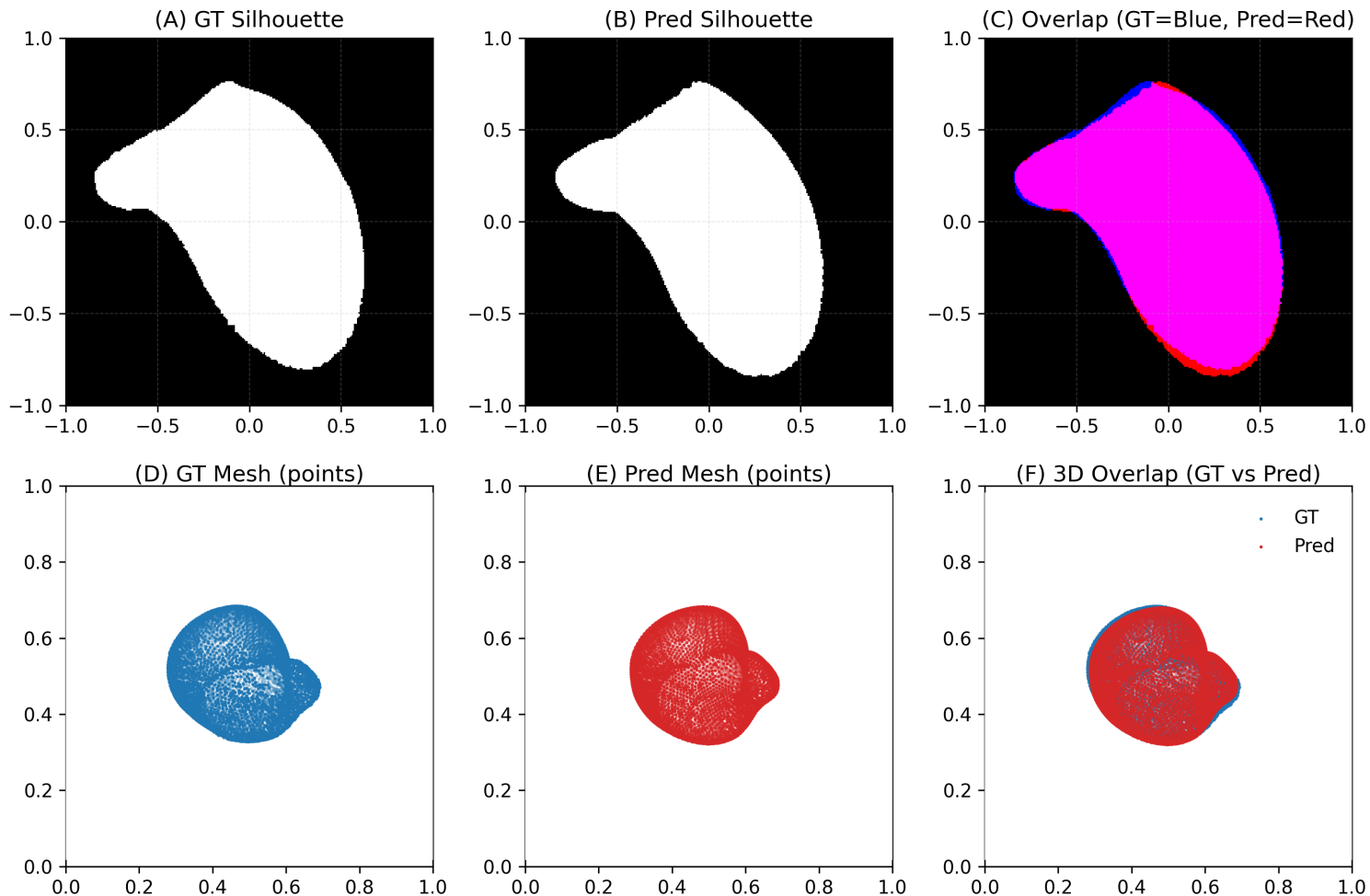


FIGURE 7. **Qualitative results on held-out subjects.** (A) GT silhouette, (B) Predicted silhouette, (C) GT–Pred overlap (*GT=blue, Pred=red; overlapping pixels appear pink*); (D) GT point cloud, (E) Predicted point cloud, (F) 3D overlap (GT=blue, Pred=red; normalized pose).

7 Conclusion

In this work, we presented a comprehensive pipeline for reconstructing 3D anatomical shapes of the left ventricle (LV) from a single 2D silhouette. By integrating statistical shape modeling with a graph-constrained deformation network, the method leverages anatomical priors to produce geometrically consistent, patient-specific meshes. The preprocessing stage ensures pose/scale normalization across meshes, while a PCA-based statistical shape model provides a compact and interpretable latent space for learning.

On a held-out cohort, the model achieves high reconstruction fidelity—low Chamfer (small mean surface error), bounded Hausdorff (worst-case deviations remain limited), and strong silhouette agreement (high IoU/Dice). These outcomes indicate that constraining predictions with shape priors while propagating global-to-local context over mesh topology is effective for single-view 3D inference. The modular design also makes the pipeline adaptable to clinical contexts with limited views (for example, echography), and amenable to real-time deployment.

The contributions of this work were: Building upon the deformation-based framework, we streamlined coefficient estimation and mesh deformation, added Hausdorff distance to characterise worst-case error, and emphasised silhouette consistency via IoU/Dice—guided by differentiable rendering losses. The resulting pipeline requires only a single binary mask at test time, improving practicality for intraoperative scenarios without multi-view calibration or explicit depth.

7.1 Future work

The proposed single-view, prior-guided pipeline recovers clinically plausible left ventricle geometry with strong quantitative and qualitative agreement. With uncertainty estimation, temporal cues, and domain adaptation, the approach can mature into a practical tool for functional assessment, planning, and guidance in real-world interventional workflows.

- **Uncertainty quantification & calibration:** predict per-vertex (or mode-wise) aleatoric/epistemic uncertainty; report calibrated credibility intervals on derived clinical indices (EDV/ESV/SV/LVEF).
- **Temporal modelling from echographic sequences:** incorporate short cine windows (EKG-gated) to regularise single-view ambiguity with motion consistency; explore sequence-to-shape or shape-flow priors.

- **Physics/regularisers:** add Laplacian/ARAP smoothness and thickness constraints; try differentiable silhouette normals/contours and perspective-correct rendering.
- **Domain adaptation:** unsupervised/adversarial adaptation from CT/MRI-derived SSM to echography; stain/style and device/institution shift robustness.
- **Richer priors:** extend beyond linear PCA (for example, VAEs, diffusion in shape space) while retaining clinical interpretability via mode attribution.
- **External validation & prospection:** multi-centre evaluation, robustness across pathologies (for example, DCM/HCM), and prospective cath-lab feasibility with human-in-the-loop segmentation.

7.2 Limitations and remedies

- **Single-view ambiguity:** apical/basal regions and out-of-plane concavities remain underconstrained. Remedy: multi-frame temporal cues, weak multi-view constraints when available, and contour/edge-aware silhouette losses.
- **Linear shape subspace:** PCA may underfit complex pathology. Remedy: hybrid PCA+ nonlinear latent models with mode attribution for interpretability.
- **Segmentation noise & calibration:** binary masks carry boundary noise and scale ambiguity. Remedy: uncertainty-aware training, perspective-correct rendering, and metric calibration using catheter or anatomical landmarks.
- **Generalisation/bias:** domain shifts across vendors/institutions/pathologies. Remedy: domain adaptation, augmentation, and subgroup analyses with fairness reporting.
- **Runtime integration:** ensure stable <100 ms inference, light-weight models for intraoperative use, and UI hooks for clinician feedback.

7.3 Potential applications

- **Functional assessment from limited views:** fast estimation of chamber volumes (EDV/ESV), stroke volume (SV), and LVEF when 3D imaging is unavailable.
- **Planning & guidance:** device sizing and trajectory planning (for example, LV lead placement, mitral interventions), pre/post comparison for therapy response.

- **Education & communication:** 3D visualisation/printing from a single echographic mask for patient briefing and trainee education.
- **Downstream modelling:** CFD or electromechanical simulations initialised from the reconstructed left ventricle geometry.

References

- [Ber+18] Olivier Bernard et al. “Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?” In: *IEEE Transactions on Medical Imaging* 37.11 (2018), pp. 2514–2525.
- [Coo+95] Timothy F Cootes et al. “Active shape models—their training and application”. In: *Computer vision and image understanding* 61.1 (1995), pp. 38–59.
- [DM98] Ian L. Dryden and Kanti V. Mardia. *Statistical Shape Analysis*. Wiley, 1998.
- [FSG17] Haoqiang Fan, Hao Su, and Leonidas Guibas. “A Point Set Generation Network for 3D Object Reconstruction from a Single Image”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 605–613.
- [GMJ19] Georgia Gkioxari, Jitendra Malik, and Justin Johnson. “Mesh R-CNN”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019.
- [Goo91] Colin Goodall. “Procrustes methods in the statistical analysis of shape”. In: *Journal of the Royal Statistical Society: Series B* 53.2 (1991), pp. 285–339.
- [He+16] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.
- [HM09] Tobias Heimann and Hans-Peter Meinzer. “Statistical shape models for 3D medical image segmentation: a review”. In: *Medical Image Analysis* 13.4 (2009), pp. 543–563.
- [KW17] Thomas Kipf and Max Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In: *International Conference on Learning Representations*. 2017.
- [Okt+17] Ozan Oktay et al. “Anatomically constrained neural networks (ACNN): Application to cardiac image enhancement and segmentation”. In: *IEEE Transactions on Medical Imaging* 37.2 (2017), pp. 384–395.
- [Wan+18] Nanyang Wang et al. “Deep Single-View 3D Object Reconstruction with Visual Hull Embedding”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.