Single-View 3D Reconstruction of the Left Ventricle from Echocardiographic Silhouette Using Shape Priors and Graph-Constrained Deformation

Dowoo Kim
Dept. of Mathematics and Statistics
McGill University
Montréal, Canada
dowoo.kim@mail.mcgill.ca

Melisa Mateu
LIVE Lab & Dept. of Software and IT Engineering
ÉTS & CHU Sainte-Justine Research Center
Montréal, Canada
mateu.melisa@gmail.com

Faten M'hiri
Department of Computer Science
McGill University
Montréal, Canada
faten.mhiri@mcgill.ca

Luc Duong
LIVE Lab & Dept. of Software and IT Engineering
ÉTS & CHU Sainte-Justine Research Center
Montréal, Canada
luc.duong@etsmtl.ca

***Abstract*—Robust 3D left-ventricular (LV) geometry is essential for cardiac function assessment and therapy planning, yet full 3D imaging (CT/MRI) is costly or impractical in time-sensitive settings, and exposes patients to radiation in the case of CT. Echocardiography is safer and ubiquitous, but often provides only a limited 2D view. We present a learning-based pipeline that infers a patient-specific 3D LV mesh from a single 2D silhouette. From a cohort of segmented 3D meshes we build a statistical shape model (SSM) after Procrustes alignment, extracting principal deformation modes via PCA to encode plausible anatomical variability. A CNN encodes the input silhouette into a global descriptor that conditions a topology-aware graph-convolutional deformation module to predict per-vertex displacements from a mean template. Training combines a Chamfer surface loss with a latent-space consistency term that encourages reconstructions to lie within the SSM span. In experiments on a dataset of 557 LV meshes (subject-wise splits), the approach yields accurate and anatomically plausible reconstructions. Distances are reported in both normalized units and millimetres using a cohort scale of 1.0 unit corresponding to $82\,\mathrm{mm}$; we obtain low mean surface error and high silhouette overlap, demonstrating the feasibility of single-view LV 3D recovery for downstream volumetry and guidance.**

***Index Terms*—Echocardiography, 3D reconstruction, Statistical Shape Model, Graph Neural Network, Chamfer distance**

I. Introduction

Patient-specific three-dimensional models of the left ventricle (LV) are central to many tasks in cardiovascular care, including the estimation of end-diastolic and end-systolic volumes (EDV/ESV), computation of left-ventricular ejection fraction (LVEF), and pre-procedural planning such as valve or device sizing and placement. While cardiac MRI and CT can deliver high-fidelity 3D anatomy, they require dedicated scanner time and staff, incur non-negligible cost, and CT entails exposure to ionizing radiation. These modalities are therefore not ideal for bedside or intra-procedural use.

By contrast, transthoracic echocardiography is portable, inexpensive, and widely available. In many clinical workflows, however, decisions are still based on a small set of 2D views (e.g., apical four-chamber, two-chamber), and volumetric quantities are derived using simplified geometric assumptions (e.g., biplane Simpson's rule). This limited 2D information makes it difficult to reason about detailed 3D morphology, regional wall motion, or device–tissue interactions, especially when anatomy deviates from idealized shapes.

Reconstructing a full 3D LV surface from a single 2D echocardiographic silhouette is a highly ill-posed inverse problem: many distinct 3D shapes can project to nearly identical outlines. To obtain clinically meaningful reconstructions, one must inject prior knowledge about plausible cardiac anatomy. Statistical shape models (SSMs) provide a classical, well-studied way to encode such anatomical variability in a low-dimensional space of deformations around a mean anatomy [1], [2]. In parallel, deep learning methods have begun to incorporate anatomical constraints explicitly, for instance through anatomically constrained neural networks (ACNNs) for cardiac image enhancement and segmentation [13].

Single-view 3D reconstruction has also been explored in computer vision, where learned point-set and mesh decoders can infer 3D shape from a single RGB image or silhouette by relying on strong learned priors [4], [9], [14]. Mesh-based and graph-convolutional architectures such as Mesh R-CNN [8] and graph-based approaches to vascular reconstruction from angiography [12] highlight the benefits of operating directly on surface graphs. Together, these lines of work suggest that combining an explicit statistical prior with a topology-aware neural decoder is a promising strategy for recovering 3D cardiac geometry from limited-view data.

Motivated by these developments, we hypothesize that even with a single 2D LV outline, strong shape priors coupled

with graph-based mesh deformation can reconstruct 3D LV geometry within clinically acceptable error bounds. We focus on a minimal-input setting where the only test-time input to the reconstruction module is a binary silhouette of the LV. In a clinical deployment scenario, this silhouette would be produced by an upstream LV segmentation network applied to an echocardiographic frame, for example the SegFormer-based method of Mateu *et al.* [11]. In the present work we isolate and study the reconstruction module itself by training on synthetic silhouettes rendered from 3D meshes; this clarifies the relationship between echocardiography, segmentation output, and 3D recovery.

In this paper we propose a single-view LV reconstruction framework that couples a PCA-based SSM with a graph-conditioned deformation network. A CNN encoder maps the input silhouette to a global feature vector, which conditions a graph convolutional network (GCN) defined on a template LV mesh. The network predicts per-vertex displacements that deform the template, while a latent-space penalty ensures that predicted shapes remain close to the learned SSM manifold.

Our evaluation targets both geometric fidelity (mean and worst-case surface deviation) and agreement with echo-like 2D projections, using a subject-held-out test split. The design emphasizes minimal dependencies at inference time (a single binary mask), aligning with point-of-care scenarios where multi-view acquisition, calibration, or cine volumes are unavailable. Beyond volumetry, the reconstructed meshes could support visualization, procedural planning, and simulation-based analyses.

Contributions. The main contributions of this work are:

- A single-view LV reconstruction pipeline that combines a PCA-based SSM with a graph-conditioned deformable decoder operating on a mesh template.
- An SSM-coefficient consistency loss that encourages reconstructed meshes to remain close to a low-dimensional, anatomically plausible shape manifold.
- A subject-wise evaluation on 557 LV meshes with distances reported in both normalized units and millimetres (1.0 unit corresponds to 82 mm), including ablations on the number of SSM modes and the impact of graph conditioning.

II. Methodology

Overview. Our pipeline takes a single 2D LV silhouette (binary mask) and produces a 3D, patient-specific LV mesh. The core idea is to constrain the solution to a population-derived statistical shape space while allowing a topology-aware decoder to make view-conditioned, local adjustments, as summarized in Fig. 1. Concretely, we (i) build an SSM from Generalized Procrustes–aligned meshes and retain the leading PCA modes; (ii) extract a global feature from the input silhouette with a CNN encoder; and (iii) feed this feature to a graph-constrained mesh decoder that predicts per-vertex displacements from a mean template.

SSMs encode LV variability as deformations around a mean anatomy in a low-dimensional linear subspace [1], [2]. Deep neural networks complement this by mapping image evidence to either shape parameters or mesh displacements, and can be regularized using anatomical constraints as in ACNN [13]. Convolutional encoders distill global cues—silhouette curvature, aspect ratio, contour context—into feature vectors correlated with underlying 3D structure [4], [9], [14]. Operating directly on mesh graphs respects surface topology: graph convolutions propagate information over the surface, enabling fine geometric adjustments while preserving connectivity [3], [8], [12]. Our design combines these elements by using an SSM both as an explicit prior and as a latent space in which we penalize deviations, while a GCN-based decoder performs template deformation.

A. Dataset

We used a total of 557 patient-derived LV *surface meshes* (STL) from the `heart_ventricle_left` category of MedShapeNet [10]. We relied on the public STL assets provided with MedShapeNet rather than raw images. An example mesh is shown in Fig. 2.

Public MedShapeNet LV meshes do not include demographics (sex/age) or clinical labels, so our study focuses on geometry-only evaluation. During preprocessing, we excluded atypical or corrupted meshes (e.g., truncated surfaces, non-manifold components, extreme self-intersections) detected by automated mesh checks (valid manifoldness, watertightness) and manual inspection. We split the cohort *by subject* into train/validation/test = 446/56/55 ($\approx$80/10/10) using a fixed random seed to prevent view leakage and ensure reproducibility.

B. Silhouette source: training vs. inference

For supervised training and validation, we render orthographic binary silhouettes from each ground-truth mesh using fixed canonical cameras, forming paired samples of the form (silhouette, 3D mesh / SSM coefficients). This pairing allows us to supervise both vertex-level geometry and the corresponding low-dimensional shape codes.

At inference time (deployment scenario), the 2D silhouette is not rendered from a mesh but is instead provided by an upstream LV segmentation model operating on an echocardiographic frame (e.g., a SegFormer-based approach as in Mateu *et al.* [11]). Our reconstruction module then consumes only this binary mask to predict a 3D mesh. Separating training (rendered silhouettes) from deployment (segmentation-derived silhouettes) makes explicit the gap between "echocardiography" and "mask input" and clarifies that this work focuses on the reconstruction stage; in future, segmentation noise will be incorporated into training.

C. Rigid Normalization and GPA

All meshes are centered, PCA-oriented, and uniformly scaled so that the *apex–base* (major principal axis) length of the mean template equals 1.0 unit. We then apply similarity Procrustes alignment to a latent mean by alternating closed-form per-mesh updates (scale s_i, rotation R_i, translation t_i) with mean recomputation until convergence [5], [6]. This

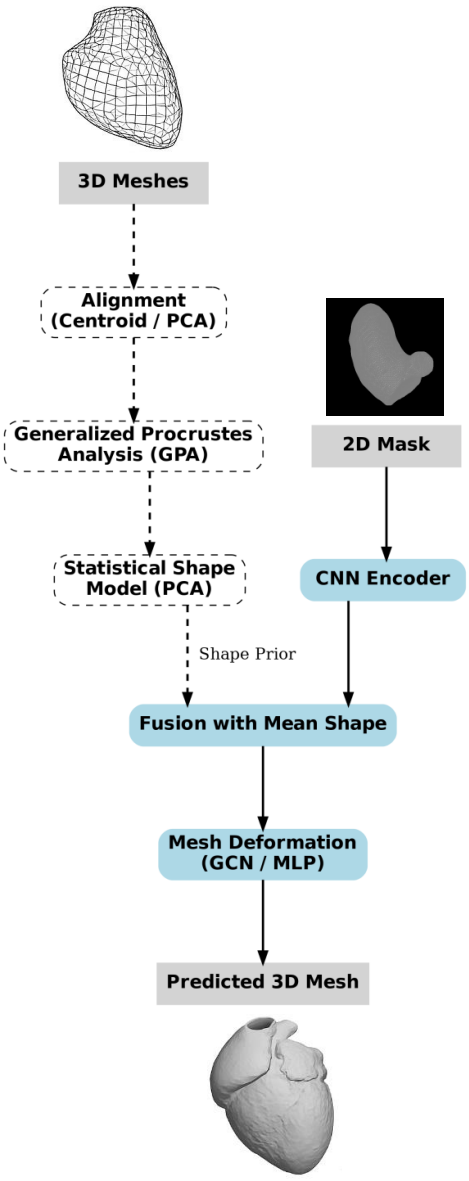

Fig. 1. End-to-end pipeline. Left: build an LV statistical shape space from Generalized Procrustes–aligned meshes (compute mean template and PCA modes). Right: extract a global feature from the input silhouette with a CNN, then use a graph-conditioned decoder to predict per-vertex displacements that deform the template into the final reconstruction. At inference, the input silhouette comes from an echocardiographic LV segmentation (e.g., Mateu *et al.* [11]).

yields a mean template $\bar{\mathbf{S}}$ and a set of aligned subjects $\{\tilde{\mathbf{S}}_i\}$. For physical interpretability we report distances in both normalized units and millimetres using the cohort scale 1.0 unit $= 82\,\mathrm{mm}$.

D. PCA Statistical Shape Model (SSM)

Vectorizing each aligned mesh as $\mathbf{x}_i \in \mathbb{R}^{3K}$ (K=7,268 template vertices), we estimate the mean vector $\boldsymbol{\mu}$ and covariance matrix, and perform PCA to obtain eigenpairs $\{(\boldsymbol{\phi}_j, \lambda_j)\}$. We

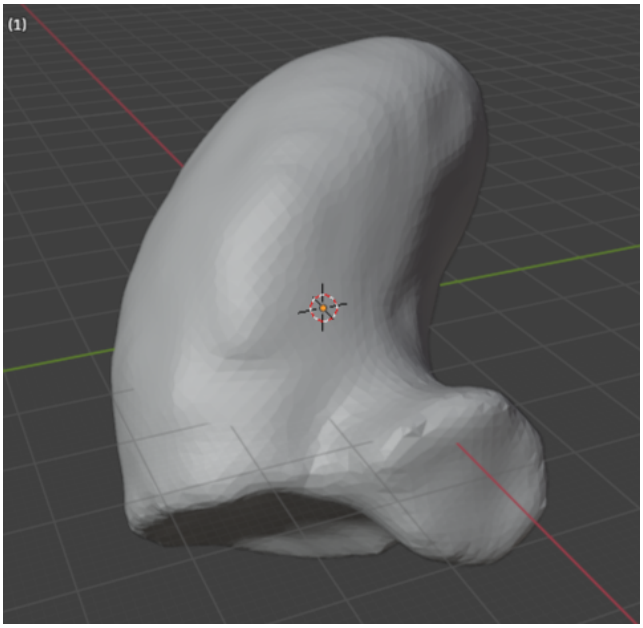

Fig. 2. Example LV mesh (STL) from the MedShapeNet `heart_ventricle_left` class. This cleaned, watertight surface is used as part of the SSM construction. During preprocessing, atypical or corrupted meshes (e.g., truncated surfaces, severe self-intersections, non-manifold components) were excluded.

retain M=25 modes ($\approx 92\%$ variance; an ablation at M=35 yields $\approx 95\%$), representing meshes as

$$\mathbf{x} \approx \boldsymbol{\mu} + \sum_{j=1}^{M} \alpha_j \boldsymbol{\phi}_j, \tag{1}$$

which constrains deformations to plausible LV anatomy [1], [2].

E. Latent Coefficients and Supervision

Each subject is projected to the SSM to obtain coefficients $\mathbf{c}_i = \boldsymbol{\Phi}_M^{\mathsf{T}}(\mathbf{x}_i - \boldsymbol{\mu})$, and decoded displacements $\Delta\mathbf{V}_i = (\boldsymbol{\Phi}_M \mathbf{c}_i).\mathrm{reshape}(K, 3)$. When necessary, we resample meshes to the template vertex set using nearest-neighbour correspondence before projection (empirically negligible error). We persist $\mathbf{c}_i$ for I/O and use $\Delta\mathbf{V}_i$ during training. Figure 3 illustrates this coefficient pipeline.

F. Paired Feature–Code Dataset and Encoder

For each subject i and canonical view v, we pair a global silhouette feature $\mathbf{f}_i^{(v)}$ with the view-invariant code $\mathbf{c}_i$. The feature is extracted by a ResNet-50 encoder (to global average pooling) applied to 256×256 RGB masks with light data augmentation (small rotations, flips) [7]–[9]. Splits are subject-wise to avoid leakage between training, validation, and test sets.

G. Graph-Conditioned Mesh Deformation Network

We treat the template $\bar{\mathbf{S}}$ as a graph $\mathcal{G}$=$(\mathcal{V}, \mathcal{E})$ with edges from the mesh connectivity and use the symmetric normalized Laplacian $\widetilde{\mathbf{L}}$ for graph convolutions. The CNN feature $\mathbf{f}$ is

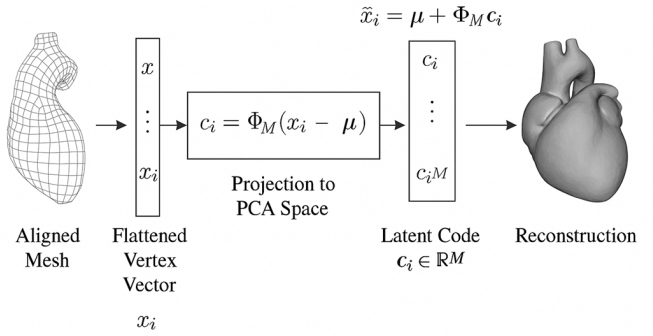

Fig. 3. Coefficient pipeline. An aligned mesh is vectorized, projected onto the SSM to obtain its low-dimensional coefficient vector $\mathbf{c}_i$, and then decoded back into a dense vertex displacement field $\Delta\mathbf{V}_i$, which adds to the mean template to recover the subject-specific shape.

mapped to a latent vector $\mathbf{h}$ via two fully connected layers with ReLU activations, broadcast to all nodes, and concatenated with node coordinates to form initial node features. A stack of GCN layers propagates $\mathbf{h}$ over $\mathcal{G}$ and outputs per-vertex displacements $\widehat{\Delta\mathbf{V}}$, forming $\widehat{\mathbf{S}} = \bar{\mathbf{S}} + \widehat{\Delta\mathbf{V}}$ [3]. Compared with a parameter-matched MLP, the GCN respects mesh topology, improving surface smoothness and sample efficiency while remaining anchored by the SSM prior [1]–[3], [8], [9].

H. Losses and Optimization

We minimize the symmetric (bidirectional) Chamfer distance between predicted and target vertex sets for geometric fidelity [4], together with a latent penalty

$$\mathcal{L}_{\text{coeff}} = \left\|\widehat{\mathbf{c}} - \mathbf{c}\right\|_2^2, \quad \widehat{\mathbf{c}} = \mathbf{\Phi}_M^{\mathsf{T}}\big(\text{vec}(\widehat{\mathbf{S}}) - \boldsymbol{\mu}\big), \tag{2}$$

which encourages predictions to stay close to the learned shape manifold. The total loss is

$$\mathcal{L} = \mathcal{L}_{\text{Cham}} + \lambda\, \mathcal{L}_{\text{coeff}}, \tag{3}$$

with λ selected on a subject-held-out validation set; early stopping is applied based on validation Chamfer/IoU.

I. Evaluation setup and metrics

We evaluate on a *subject-held-out* test split (10%; 55 subjects) constructed from the paired feature–coefficient dataset (Sec. II-F). To quantify geometric fidelity in 3D we report: (i) the symmetric (bidirectional) Chamfer distance, which averages nearest-neighbour point-to-surface distances in both directions between predicted and ground-truth vertex sets and is sensitive to overall surface mismatch [4]; and (ii) the symmetric Hausdorff distance, which reports the largest of the two directed maximum nearest-neighbour distances and therefore highlights worst-case local deviations.

To assess 2D agreement with echo-like views, we render orthographic projections (canonical cameras, 256×256) of the predicted/ground-truth meshes and compute: (iii) Intersection-over-Union (IoU), the ratio of overlapping foreground pixels to the union; and (iv) Dice coefficient, defined as

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}, \tag{4}$$

which emphasizes overlap while being less punitive on boundary jitter than IoU. All distances are given in normalized units and in millimetres using the cohort scale 1.0 unit $= 82\,\text{mm}$.

III. Results

A. Quantitative analysis

Table I summarizes performance on the subject-held-out test set; Fig. 4 provides qualitative examples.

Surface accuracy. The mean symmetric Chamfer distance was 1.28×10^{-3} units (≈ 0.105 mm), with a range $[2.59\times10^{-4}, 2.32\times10^{-2}]$ units ($\approx [0.021, 1.90]$ mm). The symmetric Hausdorff distance averaged 7.43×10^{-2} units (≈ 6.09 mm), with minimum and maximum values of 4.80×10^{-2} and 1.20×10^{-1} units ($\approx 3.94/9.84$ mm). Together, these indicate low average surface discrepancy with worst-case errors localized to challenging regions.

Silhouette fidelity. Projection agreement remained high (mean IoU 0.936; mean Dice 0.967). Even in the worst case, IoU remained above 0.71, suggesting strong consistency with 2D contours representative of echocardiographic silhouettes.

Ablations. Replacing the graph-conditioned decoder with a parameter-matched MLP degraded validation Chamfer by $+13.6\%$ and IoU by -2.1%, supporting the benefit of leveraging mesh connectivity. Varying the number of retained SSM modes showed performance saturation around $M = 25$ (about 92% variance explained); increasing to $M = 35$ (about 95%) yielded marginal gains (< 0.05 mm), so we fixed $M = 25$ for all main results.

Clinical interpretation and units. Expressing distances in millimetres (with 1.0 unit corresponding to 82 mm) facilitates clinical interpretation. A mean surface error of roughly 0.1 mm is below typical echo in-plane resolution, whereas worst-case Hausdorff deviations on the order of 10 mm highlight localized discrepancies that may affect volumetry. Consequently, downstream validation should compare EDV/ESV/LVEF derived from reconstructed meshes against volumetric ground truth (cine MRI or 3D echo) and assess agreement in clinically relevant corridors.

TABLE I
METRICS ON THE SUBJECT-HELD-OUT TEST SET (55 SUBJECTS). DISTANCES ARE IN NORMALIZED UNITS; CONVERT TO MM WITH 1.0 UNIT = 82 MM.

Metric	Mean	Min value	Max value
Chamfer	1.28×10^{-3}	2.59×10^{-4}	2.32×10^{-2}
Hausdorff	7.43×10^{-2}	4.80×10^{-2}	1.20×10^{-1}
IoU	0.936	0.715	0.974
Dice	0.967	0.834	0.987

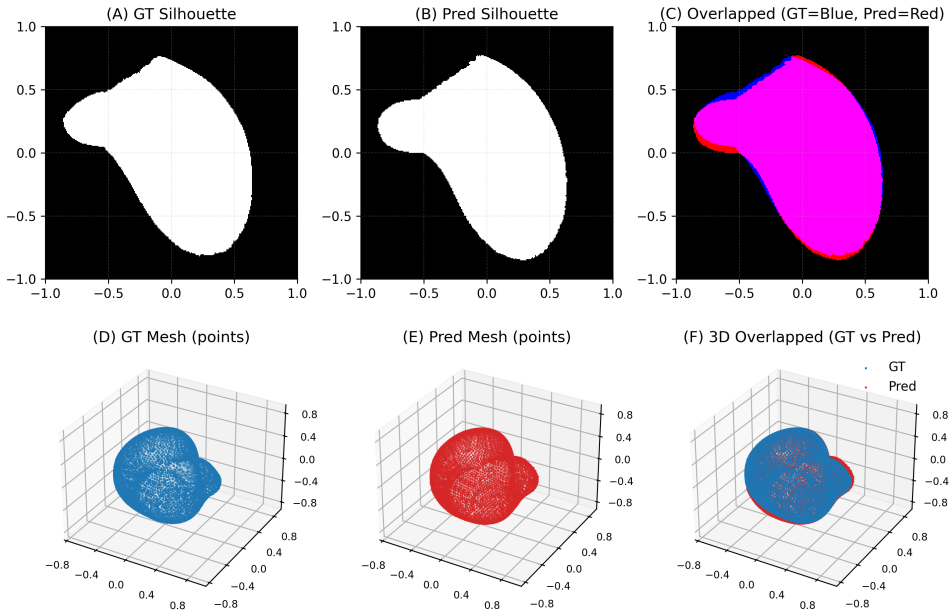

Fig. 4. **Qualitative results on held-out subjects.** (A) Ground-truth silhouette, (B) predicted silhouette, (C) GT–prediction overlap *(GT = blue, prediction = red; overlapping pixels appear pink)*; (D) ground-truth point cloud, (E) predicted point cloud, (F) 3D overlap (GT = blue, prediction = red; normalized pose).

B. Qualitative analysis and error patterns

Figure 4 shows 2D mask overlays (blue = ground truth, red = prediction, pink = overlap) and 3D point-cloud overlays for held-out subjects. Reconstructions match the global LV shape and preserve local curvature, with residuals concentrated near the basal rim and apical tip where single-view ambiguity is greatest.

The largest Hausdorff outliers are typically observed near the basal plane and apex. These regions are weakly constrained by a single outline and sensitive to small pose variations, so multiple 3D configurations can produce similar silhouettes. The SSM prior reduces but does not completely eliminate this ambiguity. Additional information—such as multi-view consistency, temporal coherence across cardiac frames, or explicitly learned boundary smoothness priors—would be expected to reduce these localized errors.

C. Relation to input source (echo segmentation)

As detailed in Sec. II, the inference-time input mask is supplied by an upstream LV segmentation model operating on an echocardiographic frame (e.g., a SegFormer-based architecture [11]). Our experiments use rendered silhouettes from ground-truth meshes to isolate reconstruction quality from segmentation noise. In deployment, errors from the segmentation stage will compound with reconstruction error. A natural next step is therefore an end-to-end study that injects realistic segmentation masks at training time and quantifies how segmentation quality affects 3D reconstruction performance.

IV. Conclusion

We presented a single-view LV reconstruction framework that transforms a 2D echocardiographic silhouette—obtained from an upstream segmentation stage—into a 3D patient-specific mesh by combining GPA/PCA-based shape priors with a graph-conditioned deformation network. On a subject-held-out cohort, the method achieved low mean surface error and high silhouette agreement, indicating that strong anatomical priors plus topology-aware decoding can mitigate the limitations of single-view input.

This work fits within the broader literature on single-view 3D reconstruction and mesh-based prediction [4], [8], [9], [14], and is conceptually related to graph-convolutional reconstructions of vascular structures from angiography [12]. Our contribution is to adapt these ideas to LV anatomy and to explicitly embed a statistical shape model within the deformation pipeline, improving plausibility and interpretability while retaining the flexibility of graph-based local adjustments.

As future work, we plan to (i) perform end-to-end evaluations using realistic echocardiographic segmentations (e.g., Mateu *et al.* [11]) to quantify the interaction between segmentation and reconstruction errors; (ii) incorporate temporal coherence across cardiac frames and weak multi-view cues when available; (iii) extend the prior from linear PCA to learned nonlinear shape manifolds while preserving interpretability; and (iv) validate clinical readouts (EDV/ESV/LVEF) against volumetric ground truth to assess potential impact on decision-making in real-world workflows.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) – Undergraduate Student Research Award (USRA). We also acknowledge support from the Fonds de recherche du Québec (FRQ). FRQ grant DOI: https://doi.org/10.69777/377062. All data were de-identified and used in accordance with institutional guidelines.

REFERENCES

[1] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Medical Image Analysis*, vol. 13, no. 4, pp. 543–563, Aug. 2009.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[3] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017.

[4] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *Proc. CVPR*, 2017, pp. 605–613.

[5] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 53, no. 2, pp. 285–339, 1991.

[6] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. Chichester, U.K.: Wiley, 1998.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016, pp. 770–778.

[8] G. Gkioxari, J. Malik, and J. Johnson, "Mesh R-CNN," in *Proc. ICCV*, 2019, pp. 9785–9794.

[9] N. Wang *et al.*, "Pixel2Mesh: Generating 3D mesh models from single RGB images," in *Proc. ECCV*, 2018, pp. 52–67.

[10] J. Li, Z. Zhou, J. Yang, A. Pepe, C. Gsaxner, G. Luijten, *et al.*, "MedShapeNet: a large-scale dataset of 3D medical shapes for computer vision," *arXiv preprint*, arXiv:2308.16139, 2023.

[11] M. Mateu, J. Olveres-Montiel, B. Escalante-Ramírez, M.-J. Raboisson, J. Miró, and L. Duong, "Automatic segmentation of the left ventricle through the cardiac cycle in pediatric echocardiography videos using SegFormer architecture," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2025.

[12] K. M. Bransby, V. Tufaro, M. Cap, G. Slabaugh, C. Bourantas, and Q. Zhang, "3D coronary vessel reconstruction from bi-plane angiography using graph convolutional networks," *arXiv preprint*, arXiv:2302.14795, 2023.

[13] O. Oktay *et al.*, "Anatomically constrained neural networks (ACNN): Application to cardiac image enhancement and segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, 2018.

[14] N. Wang *et al.*, "Deep single-view 3D object reconstruction with visual hull embedding," in *Proc. CVPR*, 2018.