Multi-view Hybrid Graph Convolutional Network for Volume-to-mesh Reconstruction in Cardiovascular MRI

Nicolás Gaggion[a], Benjamin A. Matheson[b], Yan Xia[b], Rodrigo Bonazzola[b], Nishant Ravikumar[b], Zeike A. Taylor[b], Diego H. Milone[a], Alejandro F. Frangi[c,d,e], Enzo Ferrante[a,*]

[a]*Institute for Signals, Systems, and Computational Intelligence, sinc(i) CONICET-UNL, Santa Fe, Argentina*
[b]*Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), School of Computing, University of Leeds, Leeds, UK*
[c]*Christabel Pankhurst Institute, The University of Manchester, Manchester, UK*
[d]*Centre for Computational Imaging and Modelling in Medicine (CIMIM), Department of Computer Science, School of Engineering, and Division of Informatics Imaging and Data Science, School of Health Sciences, The University of Manchester, Manchester, UK*
[e]*Medical Imaging Research Centre (MIRC), Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium*

Abstract

Cardiovascular magnetic resonance imaging is emerging as a crucial tool to examine cardiac morphology and function. Essential to this endeavour are anatomical 3D surface and volumetric meshes derived from CMR images, which facilitate computational anatomy studies, biomarker discovery, and in-silico simulations. Traditional approaches typically follow complex multi-step pipelines, first segmenting images and then reconstructing meshes, making them time-consuming and prone to error propagation. In response, we introduce HybridVNet, a novel architecture for direct image-to-mesh extraction seamlessly integrating standard convolutional neural networks with graph convolutions, which we prove can efficiently handle surface and volumetric meshes by encoding them as graph structures. To further enhance accuracy, we propose a multi-view HybridVNet architecture which processes both long axis and short axis CMR, showing that it can increase the performance of cardiac MR mesh generation. Our model combines traditional convolutional networks with variational graph generative models, deep supervision and mesh-specific regularisation. Experiments on a comprehensive dataset from the UK Biobank confirm the potential of HybridVNet to significantly advance cardiac imaging and computational cardiology by efficiently generating high-fidelity meshes from CMR images. Multi-view HybridVNet outperforms the state-of-the-art, achieving improvements of up to ∼27% reduction in Mean Contour Distance (from 1.86 mm to 1.35 mm for the LV Myocardium), up to ∼18% improvement in Hausdorff distance (from 4.74 mm to 3.89mm, for the LV Endocardium), and up to ∼8% in Dice Coefficient (from 0.78 to 0.84, for the LV Myocardium), highlighting its superior accuracy.

Keywords: Cardiac Imaging, Geometric Deep Learning, Hybrid Graph Convolutional Neural Network, Volume-to-Mesh

1. Introduction

Cardiovascular magnetic resonance (CMR) imaging has become an indispensable tool in the diagnosis, treatment planning, and management of cardiovascular diseases. A critical component of advanced cardiac imaging is the extraction of accurate 3D meshes from CMR images. These meshes serve as the foundation for various applications, including computational simulations [1], biomarker discovery [2], and analysis of heart deformation and dynamics [3].

Despite its importance, cardiac mesh extraction remains a challenging task. Traditional methods, such as active shape models [9] and multi-atlas segmentation [10], often require extensive computational resources and can be time-consuming. The inherent variability in heart shapes, sizes, and pathologies further complicates the extraction process, necessitating robust and adaptable methods.

Traditional mesh generation pipelines are complex, involving multiple steps and often requiring manual interventions. Figure 1 and Table 1 illustrate this complexity, comparing different mesh generation approaches and highlighting the numerous steps, algorithms, and manual interventions typically required in common pipelines.

A particular challenge lies in transitioning from 2D image slices to a cohesive 3D representation, especially when modeling tetrahedral meshes. Current methodologies often require intricate post-processing steps to refine the meshes and make them suitable for simulations [1, 11]. These additional steps can introduce errors and prolong the overall processing time.

Existing approaches to cardiac mesh generation can be broadly categorized into two main strategies. The first strategy follows a multi-stage pipeline that begins with voxel-level segmentation using techniques like U-Net [12] or V-Net [13], followed by surface mesh extraction and volumetric mesh generation [1, 11, 4, 5, 6]. However, this approach often introduces errors at each stage –segmentation models can produce unrealistic masks with holes or artifacts [14], and the subsequent mesh generation steps can compound these errors. Recently, Chen

*Corresponding author: eferrante@sinc.unl.edu.ar

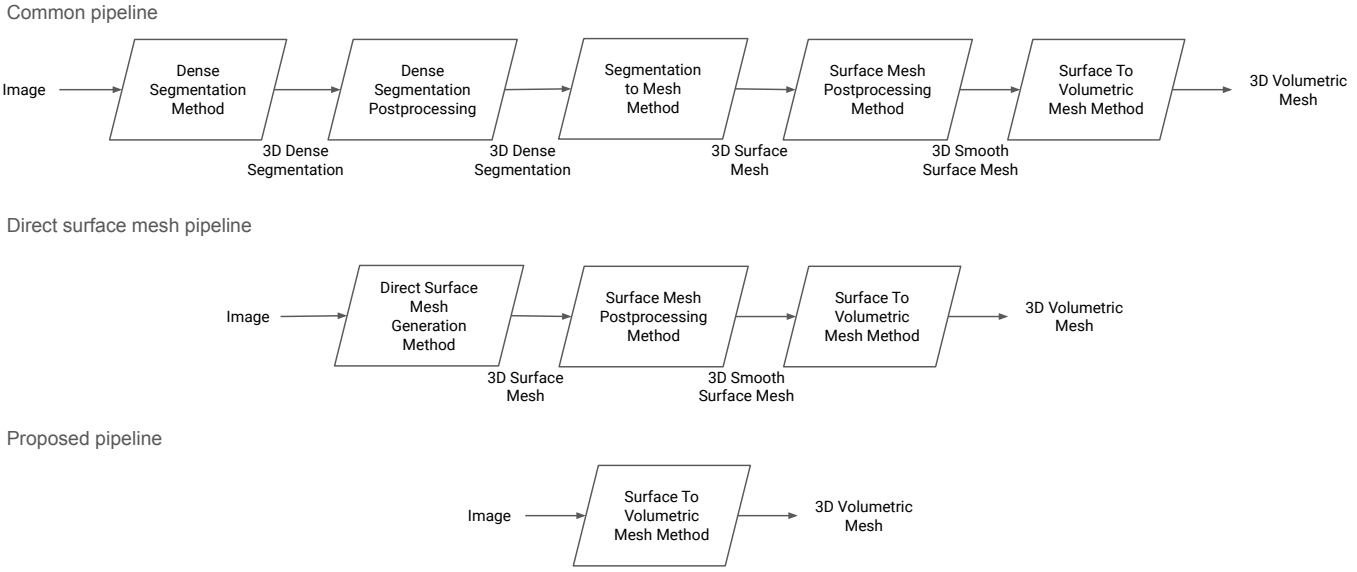


Figure 1: Mesh generation pipelines

et al. [7] proposed MR-Net, which improves this pipeline by using a deep learning-based dense segmentation-to-point cloud registration approach. While MR-Net achieves faster inference times compared to traditional registration methods, it still relies on an initial segmentation step and is therefore limited by the quality of these segmentations. Other methods attempt to improve this pipeline by estimating mesh node displacements [15, 16] or by deforming a simulation-ready template [17], but their accuracy remains constrained by the quality of the estimated deformations.

The second strategy aims to bypass these intermediate steps by generating meshes directly from images. Recent work has explored end-to-end neural networks that use convolutional architectures to estimate parameterized shapes [18, 19]. While these methods are promising, they typically rely on Principal Component Analysis (PCA) shape models, which are inherently limited by their linear nature and struggle to capture the full complexity of cardiac structures. More recent advances have focused on developing alternative approaches that learn to deform template meshes directly from medical images through various techniques such as differentiable mesh voxelization, graph convolutional networks, and mesh-based motion tracking [20, 8, 21]. Although these methods successfully eliminate complex multi-step processing pipelines, they remain constrained by their reliance on template geometries and deformation field estimations, potentially limiting their ability to capture patient-specific anatomical variations.

We propose HybridVNet, a novel architecture that advances the direct image-to-mesh approach by combining the strengths of volumetric image processing and geometric deep learning. We work under the hypothesis that direct generation can improve the accuracy of the resulting meshes while being computationally efficient. Motivated by this hypothesis, our method produces high-quality surface and volumetric meshes directly from CMR images through an end-to-end learning approach. Unlike previous methods, HybridVNet uses a hybrid architecture that combines standard 3D convolutions for volumetric image encoding with a spectral graph convolutional decoder for mesh generation. This combination allows us to better capture both global anatomical context and local geometric details, producing meshes that are immediately suitable for computational models without requiring additional processing steps. Notably, while our primary contribution is in direct mesh generation, our experiments also demonstrate that HybridVNet significantly outperforms existing segmentation-to-mesh pipelines, including MR-Net, achieving substantially lower reconstruction errors and better mesh quality.

Contributions: Our primary contributions encompass the development of HybridVNet, a multi-view volumetric hybrid graph convolutional model capable of seamlessly integrating multiple CMR views within a jointly learned latent space, directly producing meshes from images. Our model exhibits versatility in creating both cardiac surface and tetrahedral meshes which could potentially be employed for finite element simulations, both from images or segmentations as input. We explore classic regularisation techniques for surface meshes and introduce a novel differentiable regularisation term specifically tailored for tetrahedral meshes, markedly enhancing element quality. Notably, while previous works often relied on cropped regions of volumetric images, our model demonstrates exceptional performance in both cropped areas and complete images, showcasing its robustness and adaptability. The performance of HybridVNet is evaluated using the UK Biobank CMR dataset [22], providing a comprehensive assessment in the context of cardiac imaging.

2. Datasets and Reference Meshes

2.1. UK Biobank CMR Dataset

Data for this study were collected from the UK Biobank (UKB) under access applications 2,964 and 11,350. The

Mesh generation	Process Step	Algorithms involved	Hyper-parameters	Comments
Common pipeline [4, 5, 6, 7]	Segmentation from an 3D image set with dense coverage	Convolutional neural networks, multi-atlas segmentation	CNN architecture selection, neural network training hyperparameters, number of atlases and atlas selection	Explicitly limited by voxel-size and by the rectangular shape of each voxel
	Postprocessing of densely-covering segmentation masks	Artifact removal, smoothing, resampling, hole filling	Size of the artifacts to remove, kernel sizes for smoothing, interpolation method, manual hole filling	Semi automatic procedure, open problem of generating anatomically plausible segmentations
	Generation of surface mesh from segmentation masks	Marching cubes, marching tetrahedra	Grid resolution, isovalue, interpolation method	Grid dependency, staircase effect, limited handling of noise, triangle quality
	Postprocessing surface meshes	Laplacian smoothing, mesh decimation, hole filling, normal smoothing, topological cleaning, edge smoothing	Smoothing factor and iterations, target vertex count, maximum hole size, hole filling method, normal computation method, cleaning methods	Manual intervention required
	Volumetric mesh generation from surface meshes	Delaunay tetrahedralization, quality control, mesh optimization	Tetrahedralization algorithm parameters, surface mesh constraints, boundary preservation thresholds, quality metric selection and threshold, smoothing parameters and optimization objectives	Slow and with manual intervention required
Surface mesh pipeline [8]	Surface mesh extraction directly from 3D images	Convolutional neural networks mixed with point-distribution models	CNN architecture, point distribution model selection, neural network training hyperparameters	Can incorporate anatomical information in the PDM model and generate meshes with no topological artifacts
	Mesh postprocessing	Laplacian smoothing, normal smoothing, edge smoothing	Smoothing factor and iterations, normal computation method	Manual intervention may be required
	Generation of volumetric meshes from surface meshes	Delaunay tetrahedralization, quality control, mesh optimization	Tetrahedralization algorithm parameters, surface mesh constraints, boundary preservation thresholds, quality metric selection and threshold, smoothing parameters and optimization objectives	Slow and with manual intervention required
Proposed volumetric mesh pipeline	Extraction of surface or volumetric meshes directly from raw images	**Hybrid graph convolutional neural networks (proposed)**	CNN encoder and decoder architectures, neural network training hyperparameters	Incorporates topological information via the adjacency matrix, learns anatomical embeddings in the hybrid network bottleneck, and direct specification of mesh quality via regularization terms at training time.

Table 1: Comparison of Mesh Generation Pipelines

study adhered to the guidelines outlined in the Declaration of Helsinki and received ethical approval from the National Research Ethics Service of the National Health Service on 17 June 2011 (Ref 11/NW/0382) and extended on 10 May 2016 (Ref 16/NW/0274). Informed consent was obtained from all participants.

The rationale behind the UKB imaging study is explained in [23]. The UKB resource contains cine-CMR sequences acquired using a balanced steady-state free precession (bSSFP) pulse sequence. The imaging protocol includes a short-axis (SAX) stack covering the full heart with spatial resolution of $1.8 \times 1.8 \times 10$ mm, and three long-axis (LAX) views: two-chamber, three-chamber, and four-chamber views. All sequences were acquired with a temporal resolution of 50 frames per cardiac cycle, as detailed in [22]. For this study, we focused on both end-diastolic (ED) and end-systolic (ES) cardiac phases, representing the points of maximum and minimum ventricular volume in the cardiac cycle, respectively.

2.2. Reference Surface Meshes

The foundation of our study is a reference cohort of 3D surface meshes introduced by Xia et al. (2022) [19]. These meshes were created through a process of registering a high-resolution atlas of the human heart [24] to manually delineated 2D contours at ED and ES. The atlas used in this process comprises a mesh that includes six distinct cardiac structures: the left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), pulmonary artery and the ascending aorta; however, it must be noted that the two latter structures were not present in the manual contours and were inferred from the rest of the structures during registration. The selection criteria for subjects chosen for manual segmentation and the methodology followed are detailed in [25]. Their quality control process included both quantitative and qualitative steps. They first computed the point-to-point distance of the generated shape to the stack of manual contours annotated by medical experts, and if the average error was less than half of the in-plane pixel spacing, then they kept the mesh, otherwise it was discarded. After that, they visually checked all the shapes overlaid on the stack of manual contours to discard any sub-optimal shapes from the dataset. This resulted in the 4525 subjects ultimately included in our study.

A key characteristic of this cohort is that each final ground-truth mesh maintains the same number of nodes and set of faces, resulting in an identical adjacency matrix across all meshes. This consistency is a direct result of the atlas registration pro-

cess and is particularly advantageous for our graph-based approach, as it allows for uniform processing across all samples.

To ensure fair comparison with baseline methods, we followed the same data splits used in previous works. For image-to-mesh experiments, we used the splits from [19], with 3925 subjects for training and 600 for testing (considering ED and ES phases, giving a total of 1200 meshes). Similarly, for segmentation-to-mesh experiments, we adopted the splits from [7], using 957 subjects as the test set (considering ED and ES phases, giving a total of 1914 meshes).

2.3. Volumetric Mesh Generation

We derived volumetric mesh ground-truth annotations from the surface meshes through a systematic process that preserves anatomical connectivity. First, we constructed a volumetric atlas mesh using Simpleware software (Version Medical T-2022.03, Synopsys Inc., Mountain View, USA) [26]. We imported heart structures from the human heart atlas [24] as individual closed surface meshes of triangular elements. The hollow surface meshes were then populated with tetrahedral elements, ensuring that elements at the interfaces between different cardiac structures share nodes to maintain anatomical connectivity. This resulted in a reference volumetric atlas containing 408,764 elements.

To generate subject-specific volumetric meshes, we leveraged the one-to-one correspondence between the surface nodes to register the volumetric atlas to the surface mesh of each subject. This correspondence, inherited from the surface mesh generation process, provides the key landmarks needed for TPS warping [27], which was implemented through the Vedo library [28]. TPS warping provides a smooth interpolation between corresponding points while minimizing the bending energy of the transformation, making it particularly suitable for preserving anatomical relationships.

3. HybridVNet: Volume-to-Mesh extraction in cardiovascular MR

In this section we introduce the proposed HybridVNet architecture for volume-to-mesh direction extraction. As shown in Figure 2, our HybridVNet model receives multiple CMR views as input: the short-axis view (SAX), which is a 3D cross-sectional view of the heart acquired perpendicular to the long axis, and three different 2D long-axis views (LAX), for two, three and four chambers of the heart (LAX 2CH, LAX 3CH and LAX 4CH, respectively), providing 2D cross-sectional views acquired parallel to the long axis. Given these four images (one volumetric and three 2D), we aim to generate a (surface or tetrahedral) mesh representing the structures of interest.

Consider a dataset $\mathcal{D} = \{(\mathbf{I}, \mathbf{G})_n\}_{0<n\leq N}$, composed of N samples of multi-view CMR images $\mathbf{I} = (\mathbf{I}^{\text{LAX 2CH}}, \mathbf{I}^{\text{LAX 3CH}}, \mathbf{I}^{\text{LAX 4CH}}, \mathbf{I}^{\text{SAX}})$, and their associated cardiac meshes as graphs $\mathbf{G} = \langle V, \mathbf{A}, \mathbf{X}\rangle$, where V is the set of M nodes or vertices ($|V| = M$), $\mathbf{A} \in \{0,1\}^{M\times M}$ is the adjacency matrix indicating the connectivity between pairs of nodes ($a_{ij} = 1$ indicates an edge connecting vertices i and j, and $a_{ij} = 0$ otherwise), and $\mathbf{X} \in \mathbb{R}^{M\times s}$ is a function (represented as a matrix) assigning a feature vector to every node. It assigns a 3-dimensional spatial coordinate (the mesh vertex position, $s = 3$). Since our dataset includes meshes with the same number of nodes and the same connectivity by construction, we can use spectral graph convolutions to decode meshes from a latent space [29, 2].

The proposed model consists of a hybrid variational encoder-decoder architecture with multiple inputs. An image convolutional encoder, learns a latent representation of the input images, and a spectral graph convolutional decodergenerates a graph representation of the organ. Since our input consists of four images with varying shapes and views, we use a multi-view encoder to handle it. To this end, independent encoder branches are defined for each image view, and a joint latent space is constructed by concatenating their outputs. For all types of LAX images, we use 2D convolutional encoders, $f_e^{\text{LAX 2CH}}$, $f_e^{\text{LAX 3CH}}$ and $f_e^{\text{LAX 4CH}}$, with residual convolutions [30]. For the 3D SAX image, we use a 3D convolutional encoder, f_e^{SAX}, consisting of 3D residual blocks interleaved by max-pooling operations.

Consequently, our model uses a variational encoder-decoder architecture to generate a graph representation of a desired organ from multi-view input images. The encoder maps the input to a lower-dimensional embedding which represents the parameters of a latent distribution, $\mathbf{z} = f_e^I(\mathbf{I}^{\text{LAX 2CH}}, \mathbf{I}^{\text{LAX 3CH}}, \mathbf{I}^{\text{LAX 4CH}}, \mathbf{I}^{\text{SAX}})$. This latent distribution is sampled in training by using the reparametrisation trick [31] to ensure a smooth latent space, while at inference time we directly take the mean of the distribution. The sampled latent vector is then passed through a fully connected layer, and reshaped to obtain initial node features for the graph convolutional decoder, f_d^G. Following the variational autoencoder formulation, the latent code is assumed to be sampled from a multivariate Gaussian, $Q(\mathbf{z}|\mathbf{I}) = \mathcal{N}(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}))$. The distribution is parameterised by the concatenation of outputs from the joint multi-view encoder, $(\boldsymbol{\mu}, \boldsymbol{\sigma}) = f_e^I(\mathbf{I})$. Given a sample of the latent code, $\mathbf{z}$, the graph representation of the organ can be obtained through the decoder $f_d^G(\mathbf{z})$.

The model is trained by minimising a loss function defined as

$$\mathcal{L} = \mathcal{L}_r(f_d(f_e(\mathbf{I})), \mathbf{G}) + \lambda_{KL}\, \mathcal{L}_{KL}\left(Q(\mathbf{z}|\mathbf{I}) \| \mathcal{N}(0,1)\right), \quad (1)$$

where the first term is the reconstruction loss based on the mean squared error (MSE) of the vertex positions, the second term imposes a unit Gaussian prior $\mathcal{N}(0,1)$ for the latent posteriors via the KL divergence loss ($\mathcal{L}_{KL}$) and λ_{KL} is a weighting factor.

3.1. Deeply-supervised spectral graph decoder

To generate the graph representation of the target organ, we employed a decoder constructed using spectral graph convolutional neural networks (GCNN). Spectral convolutions are based on the eigendecomposition of the graph Laplacian matrix. In this context, we adopt the spectral convolutions introduced by Defferrard et al. (2016) [32], which constrain the filters to polynomial filters. This constraint arises from the observation that polynomial filters exhibit strict localisation in the vertex domain, consequently reducing the computational complexity

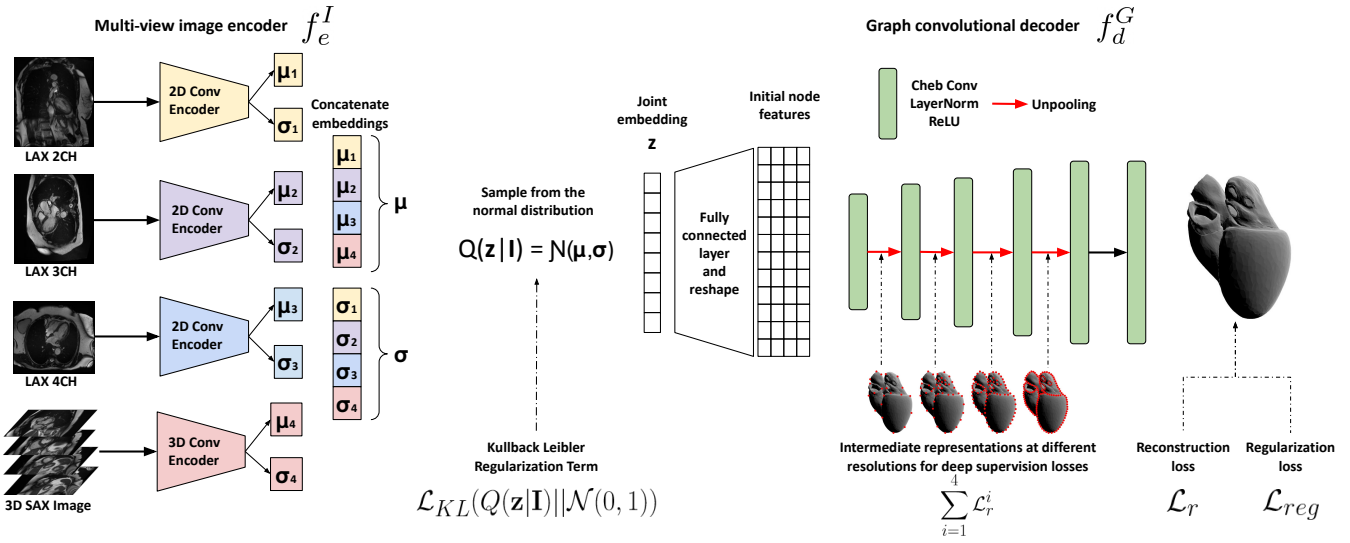


Figure 2: **Multi-view HybridVNet model architecture:** The proposed model uses a variational encoder-decoder architecture to generate a graph representation of a desired organ from multi-view input images. The encoder consists of independent branches for each input view, concatenated to obtain a joint latent space. The latent code is then passed through a fully connected layer and reshaped to obtain the initial node features for the graph convolutional decoder. This decoder uses the initial node features to generate a final graph representation of the organ.

of the convolutional operation. For an in-depth understanding of spectral convolutions, please refer to [32].

A spectral convolutional layer operates as standard convolutions applied to images and feature maps. It takes an input feature matrix $\mathbf{X}^\ell$ and produces filtered versions $\mathbf{X}^{\ell+1}$ as output. Our spectral decoder architecture comprises five graph-convolutional layers, each complemented by ReLU nonlinearities with previous Layer Normalisation [33]. These layers are strategically interleaved with four fixed graph unpooling layers, allowing the network to learn representations at multiple resolutions.

We implement the technique outlined by Ranjan et al. (2018) [29] to obtain these multiple resolutions to construct pairs of pooling and unpooling layers. The process begins by estimating the pooling matrix, achieved through an iterative contraction of vertex pairs while maintaining precise surface error approximations using quadric matrices into the atlas surface mesh. Simultaneously, the unpooling matrix is derived to enable the reversal of the pooling transformation. This process is repeated four times (in the previously pooled version of the atlas), resulting in four sets of pooling and unpooling layers, each reducing and increasing the number of nodes by a factor of two, respectively. Importantly, these pooling and unpooling matrices remain fixed during training, as they are estimated only once for the atlas surface mesh.

To increase our model's performance, we apply the concept of deep supervision [34], which involves supervising the network at various resolution levels. During training, we utilise the estimated pooling operation to obtain down-sampled versions of the ground-truth meshes, enabling us to minimise the reconstruction error at each resolution level. Ultimately, we employ a final graph-convolutional layer, without bias and identity activation function, to predict the final vertex positions.

The incorporation of deep supervision terms leads to the following loss function:

$$\mathcal{L} = \mathcal{L}_r + \lambda_{KL}\mathcal{L}_{KL} + \lambda_{DS}\sum_{i=1}^{4}\mathcal{L}_r^i, \tag{2}$$

where $\mathcal{L}_{KL}$ is the previously defined KL term, λ_{DS} is a weighting factor, and the index i indicates the resolution level of the graph. Best results were obtained with λ_{DS} set to 1, giving all different reconstructions the same weight in the final loss.

3.2. Mesh regularisation loss functions

To ensure smooth meshes, state-of-the-art approaches to surface mesh generation often use regularisers such as normal regularisation, edge length regularisation, and Laplacian smoothing ($\mathcal{L}_{lap}$), as introduced in [35], which we also incorporate. However, these existing metrics were initially designed for triangular surface meshes and, therefore, do not consider the structure of tetrahedral elements in a volumetric mesh [16]. We propose a new regularisation loss function designed to generate tetrahedral volumetric meshes to address this limitation directly. We introduce our new *tetrahedral element regularisation* loss,

$$\mathcal{L}_{ter} = \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{1}{6}\sum_{j=1}^{6}\left(\|\mathbf{e}_j^i\|_2 - \frac{1}{6}(\sum_{k=1}^{6}\|\mathbf{e}_k^i\|_2)\right)^2, \tag{3}$$

where N_t is the number of tetrahedra, i represents the i^{th} tetrahedron and $\mathbf{e}_j^i$ and $\mathbf{e}_k^i$ represent the edges of that tetrahedron. This regularisation term encourages the formation of well-shaped tetrahedral elements by penalizing large variations in edge lengths within each tetrahedron, promoting more uniform and stable volumetric meshes. Such meshes are crucial for accurate finite element simulations in computational cardiology. The final loss function used to train the model is:

$$\mathcal{L} = \mathcal{L}_r + \lambda_{KL}\mathcal{L}_{KL} + \lambda_{DS}\sum_{i=1}^{4}\mathcal{L}_r^i + \lambda_{reg}\mathcal{L}_{reg}, \tag{4}$$

where $\mathcal{L}_{reg}$ can be any of the regularisation losses mentioned above: $\mathcal{L}_{lap}$ for the surface case or $\mathcal{L}_{ter}$ for the volumetric case, and λ_{reg} is the corresponding weighting factor. We explored combining both regularisation terms for volumetric meshes by applying the Laplacian regularisation solely to surface faces, but this combination showed no improvements when used alongside the tetrahedral regularisation term. The choice between surface or volumetric mesh generation is determined by the adjacency matrix used in the model.

4. Experimental Setup

4.1. Image and mesh pre-processing

CMR images were pre-processed by normalising intensities to the range [0, 1]. SAX images had dimensions ranging from (100, 100, 6) to (200, 200, 16) and a voxel spacing of [1.82, 1.82, 10] mm, while LAX images had varying dimensions depending on the associated SAX image. To handle different sizes of SAX images between subjects, we evaluated our model in two settings: (1) *Full image input*, where we padded all SAX images to (210, 210, 16), and (2) *Cut input*, where we followed previous work [36, 19] and cropped SAX images to (100, 100, 16), padding slices as needed. In all cases, the LAX images were zero-padded to have a square shape of size (224, 224).

Inspired by classic object detection approaches, we align the vertex positions of the mesh with their relative position inside the SAX image, which is effective when using graph generative models for landmark detection [37]. We first remove the origin of the SAX image and divide each direction by the corresponding voxel spacing to obtain the positions in the voxel space. For the full-image pipeline, we add the padding applied to the positions and divide by the image size. For the cropped-image pipeline, we subtract the origin of the bounding box and divide it by the image size. With this, we obtain a *relative positional space* for training the models, with a value of (0.5, 0.5, 0.5) indicating a node in the centre of the SAX image. To evaluate the results, we reversed this operation and recovered the original positions in millimetres. No additional pre-processing was performed on the SAX images.

4.2. Data augmentation

All models were trained using online data enhancement, including intensity enhancement, random rotations of the SAX images (between -10 and 10 degrees), and arbitrary scaling on the x and y axes. The LAX images were scaled to match the scaling performed in the associated SAX image using each LAX image's respective direction vector. We added a step to randomly choose the cropping centre for the cropped model, ensuring that the entire heart is always inside the region. This helps the model avoid dependence on a perfectly centred crop and is an extra data augmentation step.

4.3. Model implementation and training details

All models were implemented in Python using the PyTorch framework [38]. The PyTorch Geometric library [39] was used for the spectral graph convolutional neural network (GCNN) layers. Hyperparameters were selected through grid search, with the k hop neighbourhood parameter [32] set to 6. We conducted training for 600 epochs using the Adam optimiser with a learning rate of $1E-4$. The batch size was set to 4, and the weight decay was applied at $1E-5$. A KL divergence weight factor of $\lambda_{KL} = 1E-5$ was introduced, and a learning rate decline with a factor of 0.99 occurred after each epoch. The 2D and 3D Convolutional Neural Network (CNN) encoders consisted of six residual blocks [40]. In 2D encoders, the maxpooling layers were interleaved with these blocks. In 3D encoders, max-pooling was applied on the X and Y axes between each residual block, with Z-axis max-pooling at the third layer. After a grid search hyperparameter selection, the latent representations were obtained using fully connected layers in the encoders, with a dimension of 32 for the 3D encoder and 8 for all 2D encoders. GCNN decoders, in both 2D and 3D models, comprised six layers of Chebyshev convolutions with Layer Normalisation [33] and ReLU nonlinearities. Classic surface regularisation losses from the PyTorch3D library [41] were used. These losses included edge length, normal vector, and Laplacian regularisation terms.

Source code is available at `https://github.com/ngaggion/HybridVNet`.

4.4. Model comparison

We implemented different single and multi-view variants of the HybridVNet architecture and compared our approach against both direct mesh generation and traditional segmentation-to-mesh pipelines.

For direct mesh generation, we compared with the Multi-Cue Shape Inference Network (MCSI-Net) [19], which constitutes the state-of-the-art point distribution model for this dataset. MCSI-Net combines two different networks: a position-inference network that predicts the central coordinates of the mesh and a rotation vector, and a shape-inference network that uses CNN layers to infer the parameters of a point distribution model (PDM) based on PCA. This model uses the same SAX and multiple LAX views as ours, but also incorporates patient metadata information into the PDM learning process, such as demographic data (age, weight, height, and body mass) and cardiovascular-related parameters, including lifestyle factors, blood pressure, and laboratory-derived biological markers (comprising a total of 29 variables in addition to the imaging data). On the contrary, our model does not require patient metadata. By comparing our HybridVNet with MCSI-Net, we aim to demonstrate the effectiveness of our approach in generating high-quality cardiac meshes without relying on patient metadata.

A fundamental question in cardiac mesh generation is whether direct mesh estimation from images is more efficient than the traditional pipeline of image segmentation followed by mesh reconstruction. To address this, we also compared HybridVNet against baseline approaches that follow the traditional pipeline: first performing automated image-based segmentation, then reconstructing surface/volumetric meshes from the segmentations via sparse point clouds. For this comparison, we evaluated several state-of-the-art methods representing

	Metrics	MCSI-Net SAX	HybridVNet		MCSI-Net SAX-LAX	MV-HybridVNet	
		Cropped	Full Image	Cropped	Cropped	Full Image	Cropped
LV Endo	DC ↑	0.87 (0.05)	0.89 (0.05)	**0.90 (0.04)**	0.88 (0.05)	0.90 (0.04)	**0.91 (0.04)**
	HD (mm) ↓	5.13 (1.97)	4.48 (1.32)	**4.08 (1.22)**	4.74 (1.75)	4.22 (1.22)	**3.89 (1.18)**
	MCD (mm) ↓	1.93 (0.83)	1.67 (0.55)	**1.49 (0.49)**	1.86 (0.79)	1.55 (0.51)	**1.39 (0.46)**
LV Myo	DC ↑	0.76 (0.09)	0.80 (0.06)	**0.83 (0.05)**	0.78 (0.08)	0.81 (0.05)	**0.84 (0.04)**
	HD (mm) ↓	5.31 (1.98)	4.71 (1.36)	**4.23 (1.27)**	4.75 (1.76)	4.40 (1.26)	**3.96 (1.23)**
	MCD (mm) ↓	1.97 (0.95)	1.71 (0.56)	**1.49 (0.51)**	1.86 (0.82)	1.57 (0.52)	**1.35 (0.46)**
RV Endo	DC ↑	0.85 (0.06)	0.85 (0.05)	**0.86 (0.05)**	0.85 (0.06)	0.86 (0.05)	**0.87 (0.05)**
	HD (mm) ↓	7.11 (2.78)	6.97 (2.31)	**6.44 (2.19)**	7.06 (2.64)	6.79 (2.23)	**6.13 (2.23)**
	MCD (mm) ↓	2.34 (0.98)	2.10 (0.64)	**1.90 (0.57)**	2.27 (0.95)	1.99 (0.59)	**1.76 (0.59)**

Table 2: Quantitative evaluation of surface mesh segmentation ($n = 1,200$). Arrows indicate whether metrics improve with higher (↑) or lower (↓) values. Bold values indicate statistically significant improvements ($p < 0.05$, independent t-test) compared to baseline models.

different approaches to mesh reconstruction from segmentation masks, including PointNet++ [42], PU-Net [43], Pixel2mesh [35], Coherent Point Drift (CPD) [44], Gaussian mixture models (GMMREG) [45], and MR-Net [7], the current state-of-the-art in segmentation-to-mesh reconstruction for this dataset. These methods were applied to point clouds generated from automated segmentations.

5. Results and Discussion

We conducted a comprehensive series of experiments to evaluate the performance of the proposed HybridVNet model alongside the baseline models and their various configurations. These experiments covered surface and tetrahedral volumetric mesh scenarios, including a sensitivity analysis of the proposed regularisation losses. All evaluations were carried out on the same test dataset comprising 600 subjects, as presented in [19], for the ground-truth meshes associated with this dataset.

5.1. Surface mesh extraction

To evaluate the quality of cardiac meshes, we used mesh metrics (Table 3) and mask-based metrics (Table 2). As shown in Figure 3, our model generates high-quality surface meshes and accurate segmentations across different test subjects. First, to enable a direct comparison with MCSI-Net, which was evaluated directly on the segmentation masks generated by the model in the SAX image space, we derived dense segmentation masks from the surface meshes. Then, we evaluated classic segmentation metrics such as Dice coefficient (DC), Hausdorff distance (HD), and the average distance between the reference and predicted contours in each slice (MCD).

In our initial comparison, we evaluated our HybridVNet against the SAX-only MCSI-Net with full images and cropped versions centred on the structure of interest (Table 2). Remarkably, HybridVNet outperforms SAX MCSI-Net for all metrics and structures. Next, we compare our MV-HybridVNet with the standard MCSI-Net, which also incorporates multiple views and is the current state of the art for this data set. The results demonstrate the superiority of our MV-HybridVNet, as it outperforms the standard MCSI-Net across all segmentation metrics for both the left and right ventricle segmentation tasks.

Our *full image* variant of the model achieves better results compared to the baselines, all while eliminating the need for an additional step to detect the region of interest during the segmentation process. Furthermore, the MV-HybridVNet model on *cropped images* beats the results with significant differences relative to the full image.

To account for structures that may not be visible in SAX images and to provide more insight into how the incorporation of long axis views in our model helps the model learn more details about the complete heart structure, we conducted a thorough evaluation of our proposed models directly on various subparts of the output mesh. Standard mesh evaluation metrics, including vertex mean squared error (MSE) and mean average error (MAE) between reference and predicted surface meshes, were calculated in millimetres. Table 3 summarises the results in our models, comparing HybridVNet with its multi-view version for *cropped images* and *full images* versions independently. Evaluation was performed at the nodes of the left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA) and aorta.

Comparing the performance of the HybridVNet with and without the inclusion of LAX images, we observed a significant improvement in accuracy for all parts of the mesh. This improvement is particularly pronounced for the left and right atria (LA and RA) and the aorta, which are not fully visible in SAX images. The base HybridVNet model demonstrates the ability to approximate the positions of these structures, with further refinement achieved through the integration of LAX images. Interestingly, note that our method does not require any kind of pre-alignment between LAX and SAX images. Even though we could potentially improve performance even further by doing this pre-alignment, this would imply an additional step, making model use more complicated and prone to errors that may be introduced during the registration process.

Finally, we evaluated the reconstruction performance across different regions of the LV and at the two cardiac phases, ED and ES, using the 17-segment model defined by the American Heart Association (AHA). As shown in Fig. 4, reconstruction errors were consistently higher at ES compared to ED, while regional differences were less pronounced overall, with slightly elevated errors observed around the anterolateral region.

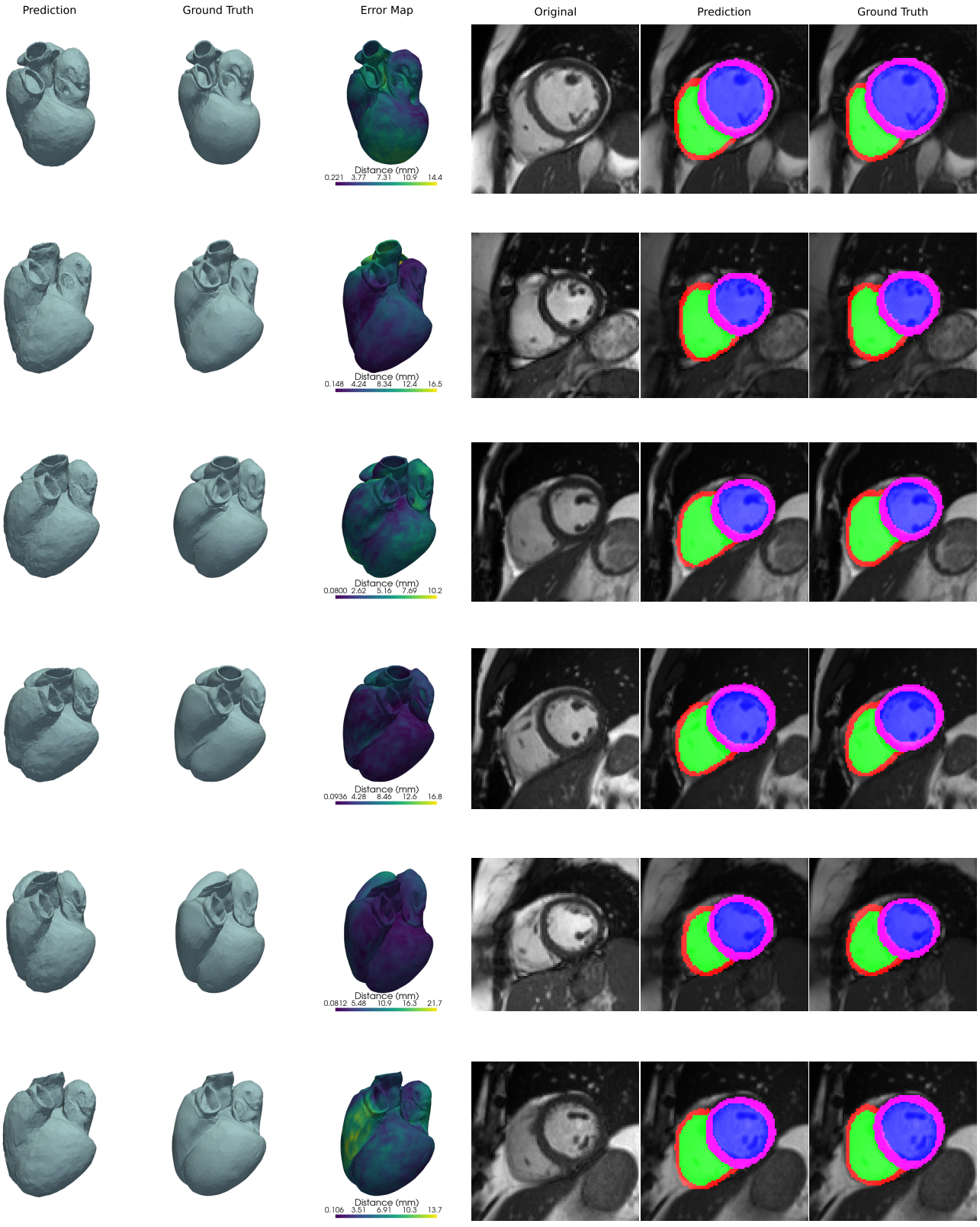
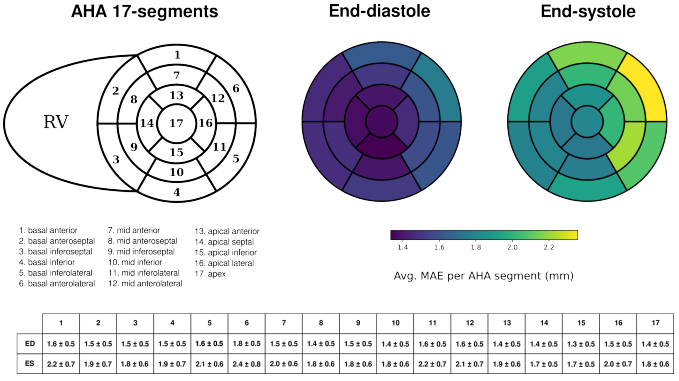


Figure 3: Qualitative performance evaluation of MV-HybridVNet (with $\lambda_{lap} = 0.01$) on cardiac MRI segmentation across six test subjects, showing the predictions against ground truth. For each subject, we present mesh-based visualizations (left three columns) showing the predicted surface, ground-truth surface, and their error map comparison, alongside 2D visualizations (right three columns) at mid-ventricular level displaying the original middle slice MRI, predicted segmentation overlay, and ground-truth segmentation overlay. The top four rows demonstrate the performance on healthy subjects, while the bottom two rows showcase segmentation results from subjects with myocardial infarction.

Subpart	Metric	Full SAX Image		Cropped SAX Image	
		HybridVNet	MV-HybridVNet	HybridVNet	MV-HybridVNet
Full Mesh	MAE (mm) ↓	2.56 (0.62)	**2.26 (0.55)**	2.43 (0.59)	**2.18 (0.54)**
	MSE (mm^2) ↓	12.20 (7.11)	**9.29 (5.48)**	11.27 (6.69)	**8.80 (5.31)**
LV	MAE (mm) ↓	1.90 (0.57)	**1.79 (0.55)**	1.75 (0.54)	**1.70 (0.54)**
	MSE (mm^2) ↓	6.23 (4.28)	**5.60 (4.03)**	5.35 (3.83)	5.11 (3.67)
RV	MAE (mm) ↓	2.18 (0.64)	**2.08 (0.60)**	2.00 (0.58)	1.97 (0.59)
	MSE (mm^2) ↓	8.39 (5.64)	**7.69 (4.93)**	7.12 (4.84)	7.04 (4.72)
LA	MAE (mm) ↓	2.90 (1.00)	**2.37 (0.78)**	2.84 (0.99)	**2.30 (0.77)**
	MSE (mm^2) ↓	15.40 (13.73)	**10.07 (9.74)**	14.88 (13.29)	**9.58 (9.24)**
RA	MAE (mm) ↓	3.07 (0.96)	**2.57 (0.76)**	2.98 (0.93)	**2.51 (0.80)**
	MSE (mm^2) ↓	17.46 (13.65)	**12.00 (9.42)**	16.67 (13.13)	**11.75 (10.16)**
AORTA	MAE (mm) ↓	2.66 (0.93)	**2.37 (0.84)**	2.56 (0.89)	**2.34 (0.83)**
	MSE (mm^2) ↓	13.17 (11.05)	**10.24 (8.71)**	12.38 (10.52)	**10.04 (8.43)**

Table 3: Comparison of single-view (HybridVNet) versus multi-view (MV-HybridVNet) approaches on full and cropped SAX images ($n = 1,200$). Arrows (↑,↓) indicate the desired direction for each metric. Bold values indicate statistically significant improvements ($p < 0.05$, independent t-test) between models within each image type.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ED	1.6 ± 0.5	1.5 ± 0.5	1.5 ± 0.5	1.5 ± 0.5	1.6 ± 0.5	1.8 ± 0.5	1.5 ± 0.5	1.4 ± 0.5	1.5 ± 0.5	1.4 ± 0.5	1.6 ± 0.5	1.6 ± 0.5	1.4 ± 0.5	1.4 ± 0.5	1.3 ± 0.5	1.5 ± 0.5	1.4 ± 0.5
ES	2.2 ± 0.7	1.9 ± 0.7	1.8 ± 0.6	1.9 ± 0.7	2.1 ± 0.6	2.4 ± 0.8	2.0 ± 0.6	1.8 ± 0.6	1.8 ± 0.6	1.8 ± 0.6	2.2 ± 0.7	2.1 ± 0.7	1.9 ± 0.6	1.7 ± 0.5	1.7 ± 0.5	2.0 ± 0.7	1.8 ± 0.6

Figure 4: Average segmentation error (measured as MAE) across the 17 AHA left-ventricular segments on the test set of 600 subjects, evaluated at end-diastole (ED) and end-systole (ES). A consistently higher error is observed for ES segmentation, with slightly increased errors around the antero-lateral region in both phases. The table shows the values per segment as average ± standard deviation.

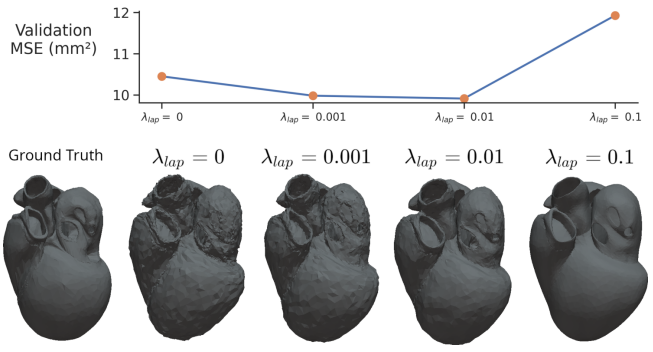


Figure 5: Qualitative analysis of the impact of Laplacian regularisation term on surface mesh smoothness. It demonstrates the influence of adjusting the regularisation parameter on mesh quality. The best quantitative results regarding MSE for the validation split were achieved when $\lambda_{lap} = 0.01$.

Surface mesh regularisation effect

In the context of the surface mesh experiment, we performed a comprehensive evaluation of various surface regularisation loss functions to enhance the performance of our HybridVNet model. Specifically, we investigated the efficacy of three distinct regularisation approaches: normal regularisation, edge-length regularisation, and Laplacian smoothing. For more information on these regularisers, see [35].

Notably, while commonly employed in mesh regularisation tasks, normal regularisation, and edge length regularisation did not yield significant improvements in our model's performance. This observation aligns with the intuitive understanding that these metrics are better suited for meshes with varying node counts and highly irregular target shapes. This is not the case in our dataset. In contrast, the incorporation of Laplacian smoothing produced notably smoother surface meshes. This can be visually appreciated in Figure 5, which presents a qualitative analysis of the meshes obtained as the regularisation parameter for the Laplacian regularisation loss was increased. Figures clearly illustrate the enhanced smoothness and quality of the meshes as the regularisation strength is adjusted.

To assess the impact of different loss terms during the training process, we refer to Figure 6. This figure provides a comparison of the MSE values throughout both the training and validation phases. Notably, due to the resource-intensive nature of the validation process, we adjusted the intervals when recording loss values, with smaller intervals as more training time elapsed.

Significantly, the red curve in Figure 6 illustrates that the best performance is achieved when combining both deep supervision and Laplacian regularisation losses. This combination eases the training process and leads to improved model performance. The optimal regularisation strength for Laplacian smoothing, resulting in the best MSE for the entire image and cropped models, was determined to be $\lambda_{lap} = 0.01$. This finding was consistent with both qualitative and quantitative evaluations, as over-smoothed meshes appeared when using high values of the regularisation term.

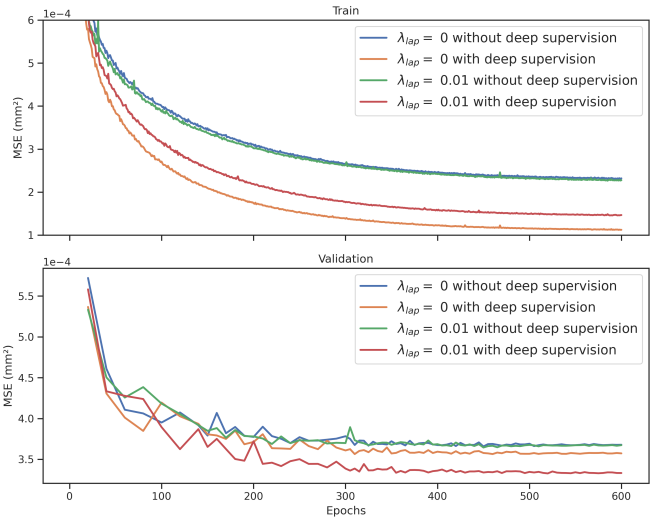


Figure 6: MSE values throughout training and validation for different configurations of hyperparameters, measured in the *relative positional space*. The red curve highlights the significant impact of combining deep supervision and Laplacian regularisation losses on model performance. Smaller intervals were used for loss recording as training progressed.

5.2. Clinical Validation

We performed a clinical validation of our automated cardiac analysis method against manual expert measurements across 600 test subjects, following the procedure in [19]. To derive the clinical cardiac functional indices, we first extract contours corresponding to the intersection between the 3D triangular meshes obtained by our best MV-HybridVNet model and the CMR image slices. Following standard clinical practice, ventricular volumes are calculated using the method of disks, where the total cardiac volume is approximated by summing the areas within 2D segmentation contours and multiplying by the inter-slice spacing.

Figure 7 presents the correlation and Bland-Altman analyses for key ventricular parameters. Our method demonstrated good agreement with manual measurements for ventricular volumes, achieving high correlation coefficients for both left and right ventricles (LVEDV: r=0.957, LVESV: r=0.905, RVEDV: r=0.949, RVESV: r=0.886). The Bland-Altman analysis revealed minimal systematic bias in volume measurements (LVEDV: -2.62±9.94 mL, LVESV: -2.28±7.83 mL, RVEDV: -1.31±12.24 mL, RVESV: 0.02±10.07 mL). For ejection fractions, while the correlations were moderate (LVEF: r=0.572, RVEF: r=0.565), the mean differences were minimal (LVEF: 0.84±4.82%, RVEF: -0.47±5.50%), indicating good clinical agreement.

5.3. Comparison with Segmentation-to-Mesh Pipeline

As shown in Table 4, HybridVNet significantly outperforms the existing segmentation-to-mesh pipelines across all metrics. We considered baseline methods reported in the literature which achieve Chamfer distances (CD) ranging from 12.10 to 20.90 mm and Hausdorff distances (HD) from 13.05 to 18.57 mm. MR-Net, the current state-of-the-art method, achieved a CD of 4.39 ± 1.48 mm and HD of 6.89 ± 1.88 mm. In contrast, HybridVNet improves upon these results with a surface mesh CD of 4.13 ± 1.16 mm and HD of 5.17 ± 1.02 mm, and an even more impressive volumetric mesh CD of 2.80 ± 2.40 mm and HD of 6.09 ± 2.09 mm.

These results demonstrate that direct mesh estimation from dense segmentation masks using HybridVNet is not only more efficient in terms of computational pipeline but also significantly more accurate than the traditional segmentation-to-mesh approach. This superior performance can be attributed to the ability of HybridVNet to learn end-to-end shape features directly from the image data, avoiding error accumulation that occurs in multi-stage pipelines (in this case, the step of obtaining point-clouds from dense segmentation masks). By eliminating intermediate processing steps, HybridVNet reduces computational overhead and minimizes potential sources of error, resulting in more precise mesh reconstructions.

5.4. Generation and quality improvement of tetrahedral meshes

Our last experiment focused on the creation of tetrahedral meshes, which could potentially be used for simulations, specifically examining the trade-off between mesh quality and anatomical accuracy. We evaluated various weighting factors (λ_{ter}) for the regularisation term defined in (3), to understand its influence on both mesh quality and segmentation accuracy. Table 5 presents results for different λ_{ter} values.

To assess mesh quality comprehensively, we employed several standard metrics. The primary metric used was the scaled Jacobian, widely adopted in the field. The Jacobian of a tetrahedron is a matrix that describes how the tetrahedron's shape changes under deformation. The scaled Jacobian provides a quantitative measure of regularity and symmetry, falling within the range [-1, 1] and not affected by scale or units. A high-scaled Jacobian value implies high regularity, low distortion, and therefore high quality [46]. As a complement, we also evaluated additional tetrahedral quality metrics, which capture different aspects of element quality: aspect ratio indicates elongation, mean ratio assesses deviation from equilateral shape, skewness measures angular deformation, and shape quality evaluates overall element regularity. As shown in Table 5, all quality metrics improve as the regularisation strength increases, with λ_{ter} = 1E-2 achieving the best absolute scores.

Our exploration reveals that λ_{ter} directly mediates the balance between segmentation accuracy and mesh quality. Lower values (λ_{ter} = 1E-4) optimize segmentation performance, closely matching the non-regularized model. In contrast, λ_{ter} = 1E-3 provides the best compromise between maintaining anatomical accuracy while achieving acceptable mesh quality for most applications. Higher values (λ_{ter} = 1E-2) further improve element quality but at the cost of reduced accuracy in vertex positions.

A closer examination of the training dynamics, as illustrated in Figure 8, reinforces the benefits of using small values of λ_{ter}. These values result in improved validation performance in terms of vertex MSE without substantial fluctuations in the training curves. On the contrary, the highest regularisation strength (λ_{ter} = 1E-2) leads to decreased performance in both training and validation.

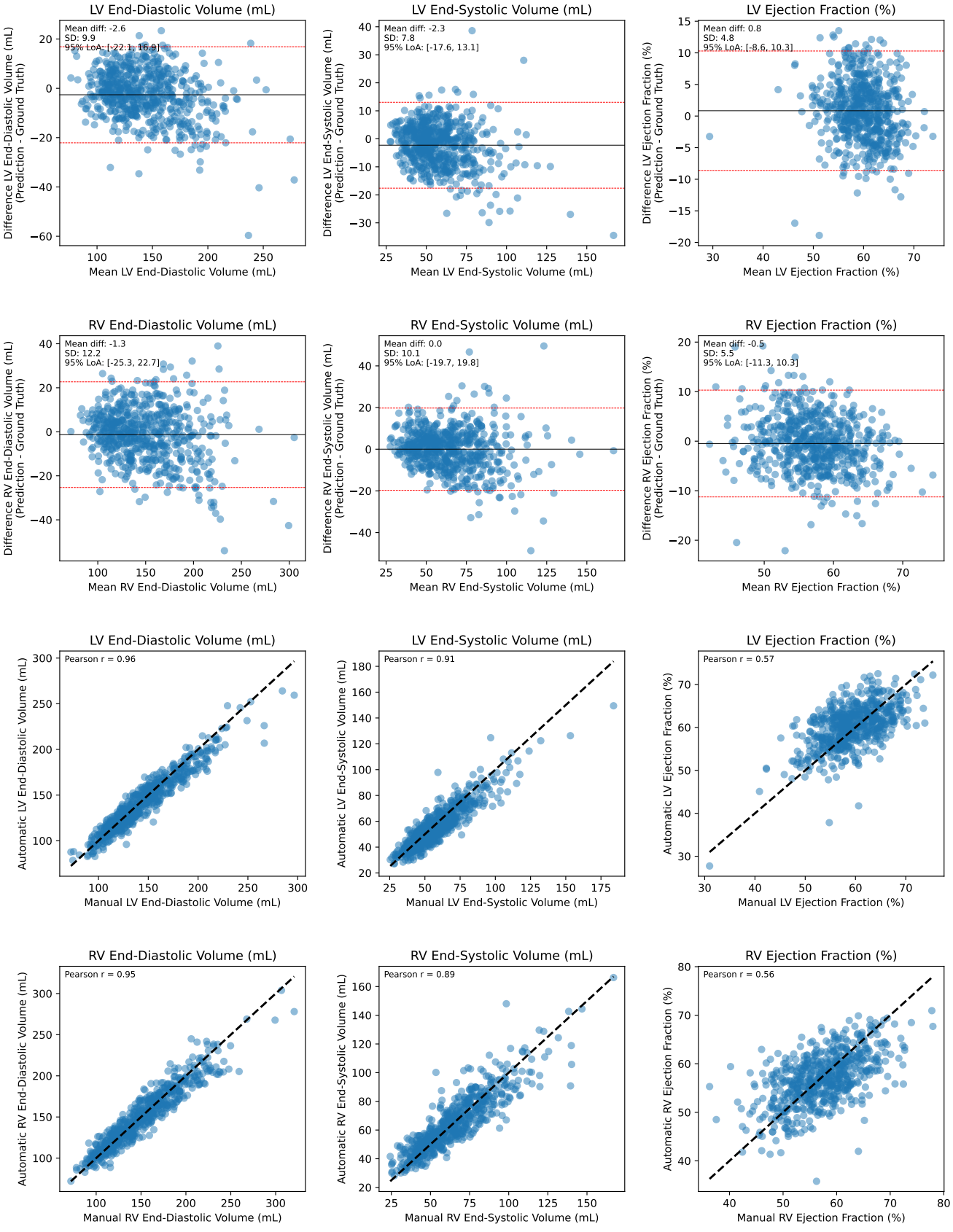


Figure 7: Bland-Altman and Correlation analyses comparing automated measurements derived from meshes generated with HybridVNet and manual measurements of cardiac parameters across 600 test subjects. Top two rows: Bland-Altman plots showing the agreement between automated and manual measurements, with mean difference (solid line) and 95% limits of agreement (red dashed lines). Bottom two rows: Correlation plots between automated and manual measurements.

Methods	**CD (mm)**	**HD (mm)**	**Inference Time (ms)**
PointNet+	13.03 (2.96)	17.04 (3.57)	< 0.1
PU-Net	12.15 (2.88)	15.74 (3.37)	< 0.1
Pixel2mesh	19.38 (5.54)	16.20 (3.30)	< 0.1
CPD	12.10 (6.63)	13.05 (7.04)	37.45
GMMREG	20.90 (7.18)	18.57 (3.04)	60.90
MR-Net	4.39 (1.48)	6.89 (1.88)	< 0.1
HybridVNet (Surface)	**4.13 (1.16)**	**5.17 (1.02)**	< 0.1
HybridVNet (Volumetric)	**2.80 (2.40)**	**6.09 (2.09)**	< 0.1

Table 4: Comparison of segmentation-to-mesh methods on bi-ventricular mesh generation ($n = 1,914$). Bold values indicate statistically significant improvements ($p < 0.05$, independent t-test) over the best baseline (MR-Net). Baseline results were considered as reported in [7].

Metrics		**MV-HybridVNet**			
		λ_{ter} = 0	λ_{ter} = 1E-4	λ_{ter} = 1E-3	λ_{ter} = 1E-2
Mesh	MAE ↓	2.08 (0.63)	2.07 (0.64)	2.04 (0.61)*	2.11 (0.61)*
	MSE ↓	8.25 (6.14)	8.22 (6.12)	7.93 (5.63)	8.39 (6.00)
LV Endo	DC ↑	**0.90 (0.04)**	**0.90 (0.04)**	0.90 (0.05)	0.88 (0.05)
	HD (mm) ↓	**4.36 (1.22)**	**4.32 (1.24)**	**4.41 (1.35)**	5.21 (1.42)
	MCD (mm) ↓	**1.52 (0.46)**	**1.51 (0.49)**	1.58 (0.54)	1.89 (0.62)
LV Myo	DC ↑	**0.78 (0.04)**	**0.78 (0.04)**	0.76 (0.05)	0.74 (0.06)
	HD (mm) ↓	5.27 (1.47)	**4.98 (1.40)**	5.17 (1.50)	5.30 (1.57)
	MCD (mm) ↓	1.86 (0.61)	**1.81 (0.64)**	1.95 (0.72)	1.96 (0.77)
RV Endo	DC ↑	**0.85 (0.06)**	**0.86 (0.05)**	**0.85 (0.05)**	0.85 (0.06)
	HD (mm) ↓	7.22 (2.76)	**6.97 (2.54)**	7.38 (2.67)	7.55 (2.80)
	MCD (mm) ↓	**2.05 (0.64)**	**2.02 (0.63)**	2.09 (0.64)	2.13 (0.69)
Mesh Quality	Scaled Jacobian ↑	0.22 (0.23)	0.23 (0.23)	0.43 (0.21)	**0.50 (0.31)**
	Aspect Ratio ↓	52.46 (16302.31)	24.49 (2212.50)	8.16 (861.52)	**4.57 (702.27)**
	Mean Ratio ↑	0.38 (0.35)	0.38 (0.35)	0.66 (0.25)	**0.69 (0.40)**
	Skewness ↓	0.69 (0.15)	0.68 (0.15)	0.54 (0.16)	**0.43 (0.15)**
	Shape Quality ↑	0.43 (0.26)	0.44 (0.26)	0.67 (0.21)	**0.74 (0.24)**

Table 5: Evaluation of volumetric mesh segmentation and quality metrics ($n = 1,200$). Arrows (↑,↓) next to each metric indicate whether higher or lower values are better. Bold values indicate the best performing methods exhibiting significant differences with respect to the non-bold values ($p < 0.05$, independent t-test), but no significant differences among themselves. Asterisk (*) marks significant differences between columns.

For an in-depth analysis of mesh quality, Table 6 provides a comprehensive overview of scaled Jacobian values under different conditions, comparing our approaches with volumetric atlases, ground-truth meshes, and a subset of surface meshes converted to volumetric meshes using Simpleware's ScanIP [26]. The analysis reveals that ground-truth meshes used for training, obtained through atlas registration, exhibit poor quality characteristics with a mean Jacobian of 0.355, and contain degenerated tetrahedra as evidenced by the minimum Jacobian value of -0.207. Our regularized approach demonstrates significant quality improvements, with λ_{ter} = 1E-2 achieving a mean Jacobian of 0.501, surpassing the average element quality of both ground truth and atlas meshes, and approaching the quality of Simpleware-generated meshes (0.524), despite being trained on imperfect data. Our regularised models surpass ground-truth elements in terms of quality, beginning from the 25% quartile and onwards, for λ_{ter} = 1E-3 and higher, underscoring that the regularisation loss significantly enhances mesh quality. Figure 9 visually summarises this improvement, positioning our method competitively with Simpleware meshes, except for a small number of elements, potentially due to the original low quality of the ground truth.

Overall, our model demonstrates competitive results compared to the conventional approach of directly converting surface to volumetric meshes. Moreover, it addresses a challenge posed by direct conversion, where degenerate triangles can obstruct the creation of volumetric meshes, affecting approximately 10% of cases in our experiments. When comparing the time required for generating a volumetric mesh, Simpleware's ScanIP procedure consumes approximately 6 minutes on average for each mesh, employing the same configuration as used in the atlas generation procedure. In contrast, our approach requires less time for generating the vertex set of volumetric meshes. When executed on an NVIDIA A100-SXM4 GPU, it accomplishes this task in just 0.04 seconds for each set of CMR images during the forward pass, resulting in a substantial speed improvement. Even in cases where GPU computing is unavailable, when running on an Intel(R) Core(TM) i7-7700 CPU operating at 3.60GHz, the forward pass requires only 5 seconds on average, providing a significant acceleration.

5.5. Limitations

Although our method exhibits promising results when compared with state-of-the-art methods, it is not free from limita-

		Mean	Std	Min	Max	1%	5%	25%	50%	75%
Reference Meshes	Atlas	0.491	0.174	0.092	0.984	0.115	0.194	0.367	0.494	0.617
	Ground Truth	0.355	0.156	-0.207	0.838	0.04	0.103	0.238	0.353	0.47
	Simpleware	0.524	0.185	0.064	0.992	0.128	0.202	0.387	0.535	0.667
MV-HybridVNet	λ_{ter} = 0	0.222	0.225	-0.759	0.876	-0.327	-0.144	0.065	0.219	0.384
	λ_{ter} = 1E-4	0.229	0.23	-0.771	0.871	-0.337	-0.151	0.068	0.231	0.397
	λ_{ter} = 1E-3	0.433	0.206	-0.719	0.904	-0.138	0.059	0.307	0.457	0.585
	λ_{ter} = 1E-2	0.501	0.309	-0.931	0.943	-0.681	-0.298	0.434	0.577	0.688

Table 6: Quality assessment of volumetric mesh elements. Values represent scaled Jacobian, with higher values indicating better quality.

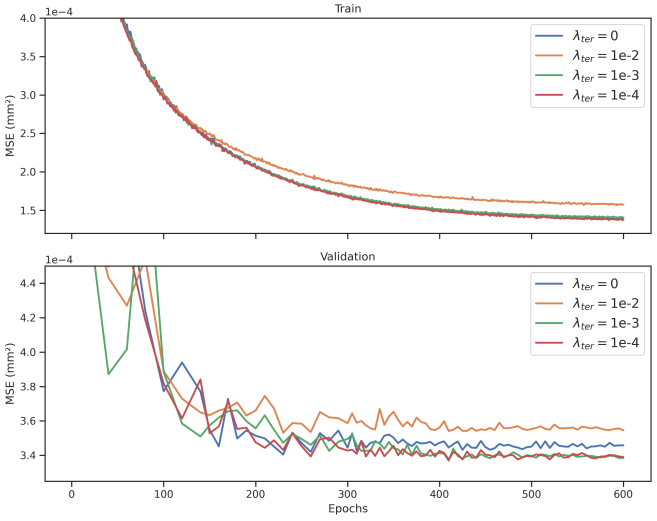


Figure 8: MSE values throughout training and validation for volumetric meshes, exploring different configurations of λ_{ter}, with values measured in the *relative positional space*. Noticeably, λ_{ter} = 1E-2 (Orange) shows a high-performance decay for both train and validation curves. Smaller intervals were used for loss recording as training progressed.

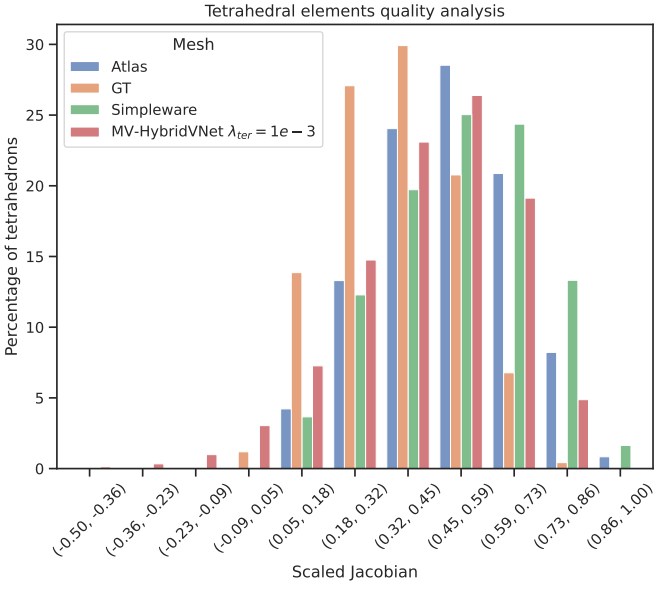


Figure 9: Histogram of tetrahedral mesh quality using scaled Jacobian values. The x-axis represents the scaled Jacobian values, and the y-axis shows the percentage of tetrahedral elements within each range.

tions. One of its main limitations is the requirement in training time of ground-truth meshes with fixed topology, i.e. all subject meshes should have the same number of vertices and faces. While we acknowledge this limitation, it is worth noting that this is well aligned with computational anatomy applications, where correspondences between subjects are essential for statistical analysis. This is particularly true in cardiac imaging, where most existing meshes are created using atlas-based or point-based approaches, which naturally generate fixed topology meshes and correspondences between subjects. Nevertheless, in future research we plan to extend our framework to handle varying mesh topologies, enabling its use in new application scenarios where such meshes are not currently available.

Another limitation is related to the lack of fine-grained evaluation of our model in highly anomalous or pathological cases. Here we considered the UK Biobank CMR Dataset [22], because this has become a standard benchmark for evaluating cardiac mesh extraction methods (see for example [8, 7]). Even though most of the patients correspond to healthy controls, this dataset also includes a smaller proportion of samples with myocardial infarction (about 150 patients) that were considered during the evaluation. Illustrative examples of the resulting meshes are included in Figure 3. Nonetheless, in future work we will conduct a comprehensive performance analysis of HybridVNet focusing on highly anomalous cases, to gain deeper insight into its robustness when handling outliers and complex pathological anatomies.

Additionally, it is important to note that even though the cardiac atlas mesh used in the study (available from [24]) includes the base of the aorta and pulmonary artery, and all cardiac valve planes in addition to all four cardiac chambers, the quality of the meshes inferred for aorta and pulmonary artery is not as reliable as the rest of the structures. This is due to the fact that pulmonary and aortic arteries were not included in the original manual contours; instead, these structures were inferred from the rest of the delineated anatomy during registration.

Finally, since our meshes are based on a generative decoder, we cannot guarantee by construction that the different cardiac structures will not intersect. However, several factors make this highly unlikely. First, our method preserves the topological connectivity in the atlas of Rodero et al. [24], ensuring shared nodes at interfaces, which inherently mitigates intersection risks. Second, the use of regularisation losses, both for surface and volumetric meshes, prevents extreme deformations that could lead to self-intersections. Additionally, the ground-truth meshes used for training were generated via a mesh-to-

contour registration process that incorporated both 3D coordinates and normal vectors, further ensuring accurate correspondence and preventing overlap [19]. As a result, our trained model inherits the non-intersecting property of the ground-truth meshes, while our combined approach of anatomical connectivity and geometric regularisation offers strong safeguards against mesh irregularities.

6. Conclusions

This study introduces HybridVNet, a novel method for directly generating surface and tetrahedral meshes from images. Our comprehensive experiments and evaluations reveal that HybridVNet significantly enhances mesh accuracy and versatility compared to state-of-the-art point distribution models that depend on linear PCA component decoding. In particular, integrating short- and long-axis views improves the accuracy of the generated meshes. HybridVNet stands out for its efficiency and speed, substantially reducing vertex set generation time compared to conventional approaches, a precious trait for large-scale processing such as in studies on the UK Biobank.

The generic nature of HybridVNet opens doors to broader applications in medical image analysis, with potential extensions to tasks such as cortical surface reconstruction from brain magnetic resonance images. Future work will direct efforts toward improving the quality of the tetrahedral ground truth used for model training, as the current training dataset contains suboptimal elements due to the atlas registration process. This could potentially lead to even better mesh quality in the predicted results.

Acknowledgments

AFF acknowledges support from the Royal Academy of Engineering under the RAEng Chair in Emerging Technologies (INSILEX CiET1919/19) and the ERC Advanced Grant – UKRI Frontier Research Guarantee (INSILICO EP/Y030494/1). EF acknowledges support from Nvidia for the donation of GPU computing, the Argentinian Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación (PICT PRH 2019-0009), and Universidad Nacional del Litoral (CAID project).

References

[1] M. Fedele and A. Quarteroni, "Polygonal surface processing and mesh generation tools for the numerical simulation of the cardiac function," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 37, no. 4, p. e3435, 2021.

[2] R. Bonazzola, N. Ravikumar, R. Attar, E. Ferrante, T. Syeda-Mahmood, and A. F. Frangi, "Image-derived phenotype extraction for genetic discovery via unsupervised deep learning in CMR images," in *MICCAI*. Springer, 2021, pp. 699–708.

[3] M. Beetz, J. C. Acero, A. Banerjee, I. Eitel, E. Zacur, T. Lange, T. Stiermaier, R. Evertz, S. J. Backhaus, H. Thiele, A. Bueno-Orovio, P. Lamata, A. Schuster, and V. Grau, "Mesh u-nets for 3d cardiac deformation modeling," in *Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers*, O. Camara, E. Puyol-Antón, C. Qin, M. Sermesant, A. Suinesiaputra, S. Wang, and A. Young, Eds. Cham: Springer Nature Switzerland, 2022, pp. 245–257.

[4] J. J. Kim, J. Nam, and I. G. Jang, "Fully automated segmentation of a hip joint using the patient-specific optimal thresholding and watershed algorithm," *Computer Methods and Programs in Biomedicine*, vol. 154, pp. 161–171, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016926071730826X

[5] S. P. Väänänen, L. Grassi, M. S. Venäläinen, H. Matikka, Y. Zheng, J. S. Jurvelin, and H. Isaksson, "Automated segmentation of cortical and trabecular bone to generate finite element models for femoral bone mechanics," *Medical Engineering & Physics*, vol. 70, pp. 19–28, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350453319301092

[6] D. H. Pak, M. Liu, T. Kim, C. Ozturk, R. McKay, E. T. Roche, R. Gleason, and J. S. Duncan, "Robust automated calcification meshing for biomechanical cardiac digital twins," 2024. [Online]. Available: https://arxiv.org/abs/2403.04998

[7] X. Chen, N. Ravikumar, Y. Xia, R. Attar, A. Diaz-Pinto, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Shape registration with learned deformations for 3d shape reconstruction from sparse and incomplete point clouds," *Medical image analysis*, vol. 74, p. 102228, 2021.

[8] F. Kong and S. C. Shadden, "Learning whole heart mesh generation from patient images for computational simulations," *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 533–545, 2023.

[9] S. Ordas, E. Oubel, R. Leta, F. Carreras, and A. F. Frangi, "A statistical shape model of the heart and its application to model-based segmentation," in *Medical Imaging 2007: Physiology, Function, and Structure from Medical Images*, vol. 6511. SPIE, 2007, pp. 490–500.

[10] W. Bai, W. Shi, C. Ledig, and D. Rueckert, "Multi-atlas segmentation with augmented features for cardiac mr images," *Medical image analysis*, vol. 19, no. 1, pp. 98–109, 2015.

[11] A. Neic, M. A. Gsell, E. Karabelas, A. J. Prassl, and G. Plank, "Automating image-based mesh generation and manipulation tasks in cardiac modeling workflows using meshtool," *SoftwareX*, vol. 11, p. 100454, 2020.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[13] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[14] A. J. Larrazabal, C. Martínez, B. Glocker, and E. Ferrante, "Post-DAE: anatomically plausible segmentation via post-processing with denoising autoencoders," *IEEE transactions on medical imaging*, vol. 39, no. 12, pp. 3813–3820, 2020.

[15] E. Puyol-Anton *et al.*, "A multimodal spatiotemporal cardiac motion atlas from mr and ultrasound data," *Medical image analysis*, vol. 40, pp. 96–110, 2017.

[16] D. H. Pak, M. Liu, T. Kim, L. Liang, R. McKay, W. Sun, and J. S. Duncan, "Distortion energy for deep learning-based volumetric finite element mesh generation for aortic valves," in *MICCAI*. Springer International Publishing, 2021, pp. 485–494.

[17] F. Kong and S. C. Shadden, "Whole heart mesh generation for image-based computational simulations by learning free-from deformations," in *MICCAI*. Springer, 2021, pp. 550–559.

[18] K. Tóthová, S. Parisot, M. Lee, E. Puyol-Antón, A. King, M. Pollefeys, and E. Konukoglu, "Probabilistic 3D surface reconstruction from sparse MRI information," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 813–823.

[19] Y. Xia *et al.*, "Automatic 3D+t four-chamber CMR quantification of the UK biobank: integrating imaging and non-imaging data priors at scale," *Medical Image Analysis*, vol. 80, p. 102498, 2022.

[20] T. Joyce, S. Buoso, C. T. Stoeck, and S. Kozerke, "Rapid inference of personalised left-ventricular meshes by deformation-based differentiable mesh voxelization," *Medical Image Analysis*, vol. 79, p. 102445, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841522000901

[21] Q. Meng, W. Bai, D. P. O'Regan, and D. Rueckert, "Deepmesh: Mesh-based cardiac motion tracking using deep learning," *IEEE transactions on medical imaging*, 2023.

[22] S. E. Petersen *et al.*, "UK biobank's cardiovascular magnetic resonance protocol," *Journal of cardiovascular magnetic resonance*, vol. 18, no. 1, pp. 1–7, 2015.

[23] ——, "Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, no. 1, p. 46, Dec. 2013.

[24] C. Rodero *et al.*, "Linking statistical shape models and simulated function in the healthy adult human heart," *PLOS Computational Biology*, vol. 17, no. 4, pp. 1–28, 04 2021.

[25] S. E. Petersen, N. Aung, M. M. Sanghvi, F. Zemrak, K. Fung, J. M. Paiva, J. M. Francis, M. Y. Khanji, E. Lukaschuk, A. M. Lee *et al.*, "Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (cmr) in caucasians from the uk biobank population cohort," *Journal of cardiovascular magnetic resonance*, vol. 19, no. 1, pp. 1–19, 2017.

[26] Synopsys, "Simpleware," 2021, [Online]. Available: https://www.synopsys.com/simpleware.html. [Accessed: Nov. 8, 2023].

[27] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[28] M. M. et al., "vedo, a python module for scientific analysis and visualization of 3d objects and point clouds," Oct. 2022, version 2022.4.1, Zenodo. Online]. Available: https://github.com/marcomusy/vedo/. [Accessed: Jan. 12, 2025].

[29] A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, "Generating 3D faces using convolutional mesh autoencoders," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 704–720.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[31] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[32] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *arXiv preprint arXiv:1606.09375*, 2016.

[33] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[34] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Artificial intelligence and statistics*, 2015, pp. 562–570.

[35] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2mesh: Generating 3D mesh models from single rgb images," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 52–67.

[36] R. Attar *et al.*, "Quantitative CMR population imaging on 20,000 subjects of the UK biobank imaging study: Lv/rv quantification pipeline and its evaluation," *Medical image analysis*, vol. 56, pp. 26–42, 2019.

[37] N. Gaggion, L. Mansilla, C. Mosquera, D. H. Milone, and E. Ferrante, "Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: applications to chest x-ray analysis," *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 546–556, 2022.

[38] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019.

[39] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[41] N. Ravi *et al.*, "Accelerating 3D deep learning with pytorch3d," *arXiv:2007.08501*, 2020.

[42] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[43] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "Pu-net: Point cloud upsampling network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2790–2799.

[44] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[45] B. Jian and B. C. Vemuri, "Robust point set registration using gaussian mixture models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1633–1645, 2010.

[46] A. Johnen, C. Geuzaine, T. Toulorge, and J.-F. Remacle, "Efficient computation of the minimum of shape quality measures on curvilinear finite elements," *Computer-Aided Design*, vol. 103, pp. 24–33, 2018.