Anatomica: Localized Control over Geometric and Topological Properties for Anatomical Diffusion Models

Karim Kadry[1*] Abdalla Abdelwahed[2*] Shoaib Goraya[3] Ajay Manicka[1]
Naravich Chutisilp[4] Farhad R. Nezami[3] Elazer R. Edelman[1]

[1]MIT, Cambridge, MA, USA [2]American University in Cairo, New Cairo, Egypt
[3]Brigham and Women's Hospital, Boston, MA, USA [4]EPFL, Lausanne, Switzerland

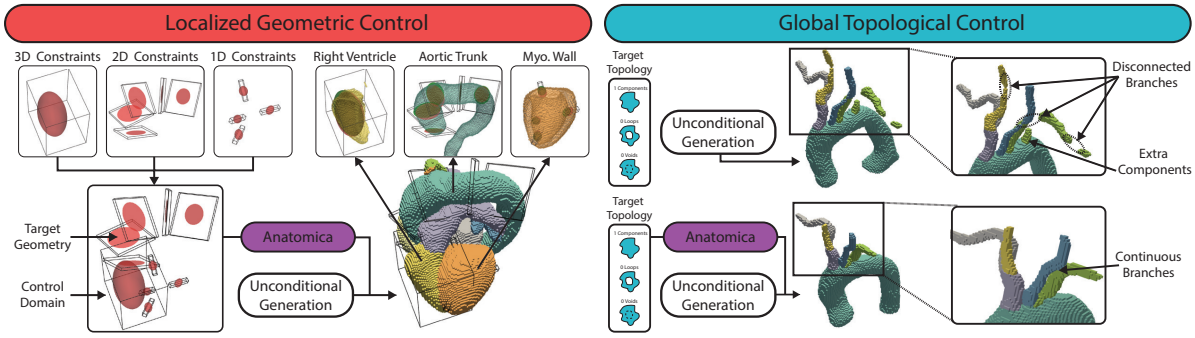

Figure 1. **Anatomica is a compositional diffusion-guidance framework for generating segmentations based on anatomical features that are localized within cuboidal control domains. Left**: We generate voxel maps according to localized target geometry (size, shape, and position) visualized as red ellipsoids. **Right**: We generate voxel maps according to target topology (components, loops, and voids).

Abstract

We present Anatomica: an inference-time framework for generating multi-class anatomical voxel maps with localized geo-topological control. During generation, we use cuboidal control domains of varying dimensionality, location, and shape, to slice out relevant substructures. These local substructures are used to compute differentiable penalty functions that steer the sample towards target constraints. We control geometric features such as size, shape, and position through voxel-wise moments, while topological features such as connected components, loops, and voids are enforced through persistent homology. Lastly, we implement Anatomica for latent diffusion models, where neural field decoders partially extract substructures, enabling the efficient control of anatomical properties. Anatomica applies flexibly across diverse anatomical systems, composing constraints to control complex structures over arbitrary dimensions and coordinate systems, thereby enabling the rational design of synthetic datasets for virtual trials or machine learning workflows.

[*]Equal contribution.

1. Introduction

Anatomical form plays a vital role in dictating the function and dysfunction of physiological systems. By virtually modelling patient-specific organ systems as 3D voxelized segmentations, we can leverage numerical simulators to reveal structure-function relationships that inform clinical research and medical device design. Such use cases include the simulation of clinical trials to evaluate medical devices [1, 42, 44], or simulating image-formation to create robust datasets for machine learning workflows [4, 11, 16, 19, 20].

Due to the sparsity and imbalances inherent to real-world datasets, there has been growing interest in augmenting anatomical datasets with synthetic data. A key advantage of using generative models over patient datasets lies in their controllability. Conditional generation of medical images based on anatomical or demographic information has been shown to improve the performance of machine learning classifiers and segmentation networks [16, 35, 38]. However, conditional generation of 3D multi-class segmentations based on anatomical features remains difficult. These features encompass both *geometry* (shape and size) and *topology* (connected components, loops, or voids). Moreover, such features are defined compositionally over mul-

tiple substructures within the segmentation, with varying dimensionality (e.g., 3D vs 2D), and across varying coordinate systems (e.g., Cartesian vs curvilinear). The ideal generative model must not only control such features in a *precise* and *compositional* manner, but also offer control mechanisms that are *intuitive* to use.

We introduce Anatomica: an inference-time framework for controlling anatomical latent diffusion models based on arbitrarily localized properties related to geometry and topology. We formulate guidance through two key stages for each sampling step. First, we differentiably parse voxel-space segmentations to extract anatomical substructures with varying dimensionality over arbitrary coordinate systems. Second, we measure geometric and topological properties in a differentiable manner and apply potential functions to guide the reverse sampling process. Lastly, we adapt this guidance framework for latent diffusion models through neural field decoders which map arbitrary query points in latent space to voxel space, enabling the efficient measurement of anatomical properties from latent space. We advance the state-of-the-art in the following ways:

- **Differentiable and Localized Substructure Extraction**: We introduce a modular method to differentiably parse *localized* and *anatomically relevant* substructures from voxel-space segmentations (V-parsing). We base our method on cuboidal control domains with varying scales, positions, and orientations. By arranging multiple control domains of varying dimensionality across relevant coordinate systems, we enable the characterization of a wide array of anatomical systems and structures.
- **Unified Geo-Topological Measurement and Guidance**: We demonstrate that applying differentiable measurement and potential functions over anatomical substructures allows us to constrain localized properties through diffusion guidance. This includes geometric properties such as size, shape, position, and orientation, as well as topological properties such as the number of components, loops, or voids. We show that, by combining different control domains and potential, we unlock a rich design space for *compositional* anatomical control, within which a wide variety of structures can be controllably generated.
- **Latent Diffusion Guidance with Neural fields**: We show that neural field decoders enable the efficient measurement of voxel-space properties within control domains *directly from latent space* during sampling (L-parsing). By exploiting the ability of neural fields to decode arbitrarily discretized point grids, we avoid the computational overhead of full-volume decoding. We introduce two partial decoding strategies: *coarse L-parsing* decodes globally at reduced spatial resolution, while *localized L-parsing* decodes local regions at high resolution.

2. Related Work

Geometric Control for Generative Models of Anatomy Geometric features such as size and shape play a crucial role in biophysical dynamics [14, 26]. Modelling anatomy with simple shapes such as cylinders [3, 37] provides control over form but not realism. Statistical shape models [12, 41, 47] represent realistic variation via global shape vectors [22, 47] but are not as interpretable or editable. To bridge this gap, recent studies conditionally train generative models based on size-based measures [9, 28]. Recently, Kadry et al. [29] proposed inference-time geometric guidance via differentiable geometry, expanding control to size, position, and shape, in a compositional manner over multi-class anatomy. However, this method was limited to globally defined geometric properties in 3D. In this work, we extend geometric guidance to arbitrarily localized attributes based on cuboidal control domains of varying scale, position, orientation, and dimensionality. By arranging control domains over non-Cartesian coordinate systems, we enable significantly more complex compositional control of physiologically relevant geometric features.

Topological Deep Learning Topological properties such as the number of components, loops, or voids also play a crucial role in modulating biophysical dynamics [33]. To regularize machine learning workflows in a differentiable manner, persistent homology (PH) can be used [5] to measure the continuous-valued persistence of topological features. PH-based topological losses have been used for the training [7] and test-time adaptation [6] of segmentation networks. Similarly, PH has been used to conditionally train diffusion models of 2D binary label maps [21] and 3D surfaces [24]. In contrast to using topological losses to update network weights, we use PH for inference-time control generative models that sample multi-class 3D anatomical segmentations without conditional training. This enables us to flexibly constrain topological features in a plug-and-play manner without retraining.

Spatial Conditioning for Generative Models Spatial control of generative models relies on two main strategies. The first conditions models on mid-level representations (e.g., bounding boxes, ellipsoid parameters) [2, 15, 23, 34, 39]. The second involves guidance methods, such as self-guidance [13], which employs attention-based losses for basic geometric control (size, position) in text-to-image models, but it is not suited for multi-label segmentations, nor is it adapted for complex constraints needed to describe anatomical shape. In our work, we extend energy-based guidance to localized control over geometry and topology by introducing differentiable potentials for 3D multi-component anatomical voxel maps based on substructure-specific properties. We show that this enables a rich design space for anatomical control, within which a wide variety of organs can be controllably generated.

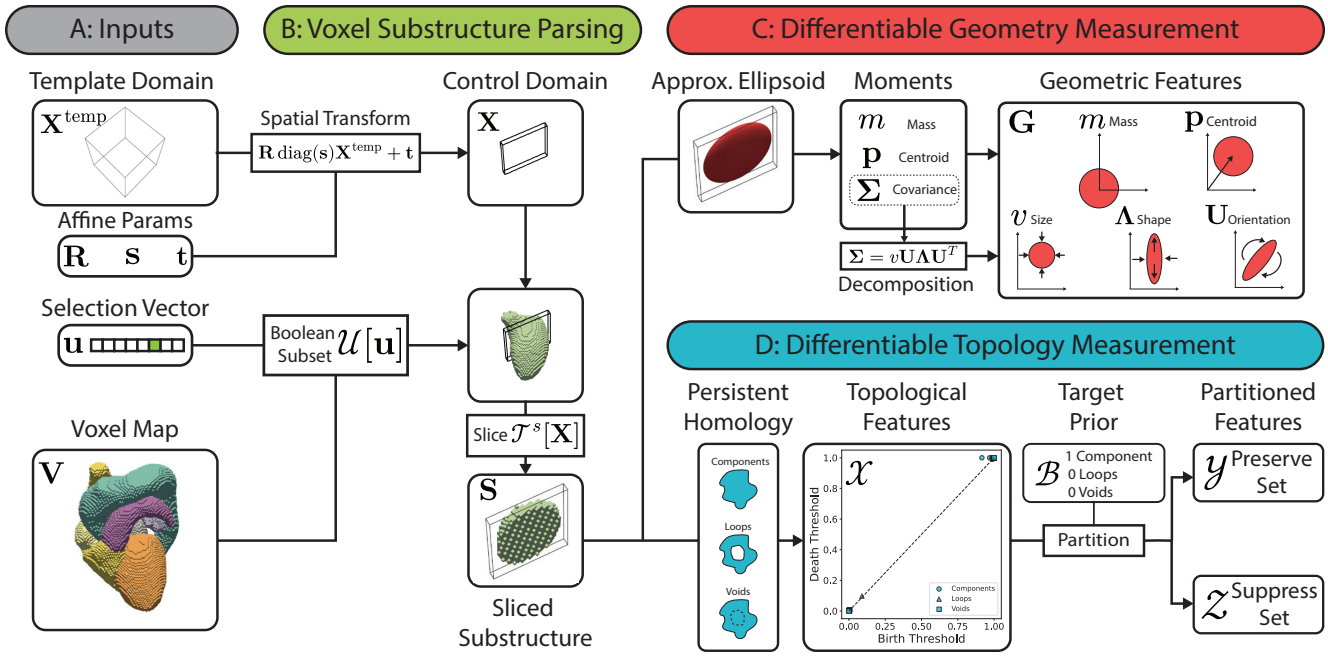

Figure 2. **Differentiable measurement of anatomical properties from multi-class voxel maps.** **A**: We differentiably parse relevant substructures from anatomical voxel maps for localized measurement. **B**: We spatially transform cuboidal primitives (template domains) into control domains that slice into anatomical structures (V-parsing). **C**: The substructure is then differentiably measured in terms of geometric properties; as well as **D**: persistent homology-based topological properties.

3. Methodology

3.1. Anatomical Latent Diffusion Models

Autoencoder with Neural Field Decoder We develop our variational autoencoder based on hybrid implicit-explicit representations [40]. Our dataset consists of 3D segmentation volumes $\mathbf{V} \in \mathbb{R}^{C\times H\times W\times D}$ with C tissue channels and (H, W, D) spatial dimensions. During training, a convolutional encoder $\mathcal{E}$ first encodes the voxelized segmentation map $\mathbf{V}$ into a voxelized latent grid representation $\mathbf{z} = \mathcal{E}(\mathbf{V})$, where $\mathbf{z} \in \mathbb{R}^{c\times h\times w\times d}$ comprises c channels and spatial dimensions $(h, w, d) = (H/f,\ W/f,\ D/f)$ for an integer downsampling factor f. To decode back into voxel space, a 3D query point grid $\mathbf{X}^q \in \mathbb{R}^{H\times W\times D\times 3}$ is used to compute a latent point grid through the latent slice operator $\mathcal{T}^l[\mathbf{X}^q] : \mathbb{R}^3 \to \mathbb{R}^c$ which applies trilinear interpolation for each query point as in Jaderberg et al. [25]. The resulting latent grid is then pointwise decoded with a neural field decoder $\mathcal{F} : \mathbb{R}^c \to \mathbb{R}^C$ into the predicted voxel map $\bar{\mathbf{V}} \in \mathbb{R}^{C\times H\times W\times D}$:

$$\bar{\mathbf{V}} = \underbrace{\mathcal{F}[\mathbf{X}^q]}_{\text{Decode Latent}} \circ \underbrace{\mathcal{T}^l[\mathbf{X}^q](\mathbf{z})}_{\text{Slice Latent}}. \tag{1}$$

Here, $\mathcal{F}$ is parametrized as a multi-layer perceptron that takes in the interpolated latent point grid and positionally encoded query points, and $\circ$ denotes function composition.

Unconditional Diffusion Model We use an unconditional latent diffusion model (LDM) as a prior over 3D anatomical segmentations. In the forward process, data samples are progressively corrupted by adding Gaussian noise $\mathbf{n}$ through the relation $\mathbf{z}_\sigma = \mathbf{z} + \mathbf{n}$ where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Similar to Karras et al. [32], we aim to learn the score function $\nabla_{\mathbf{z}_\sigma} \log p(\mathbf{z}_\sigma; \sigma)$ that defines the reverse diffusion process:

$$d\mathbf{z}_\sigma = -2\sigma\nabla_{\mathbf{z}_\sigma} \log p(\mathbf{z}_\sigma; \sigma)\, dt + \sqrt{2\sigma}\, d\mathbf{w} \tag{2}$$

where $d\mathbf{w}$ is the Wiener process. This score function $\nabla_{\mathbf{z}_\sigma} \log p(\mathbf{z}_\sigma; \sigma) = (D_\theta(\mathbf{z}_\sigma; \sigma) - \mathbf{z}_\sigma)/\sigma^2$ can be expressed via a denoising function D_θ parametrized by a 3D U-Net. The neural network is trained by minimizing the clean data prediction objective $L = \mathbb{E}_{\sigma,\mathbf{z},\mathbf{n}}\left[\omega(\sigma)\|D_\theta(\mathbf{z}_\sigma; \sigma) - \mathbf{z}\|_2^2\right]$, with $\omega(\sigma)$ balancing loss contributions across noise levels.

3.2. Guidance via Anatomical Potential Functions

We assume the voxel map $\mathbf{V}$ contains K substructures $\mathbf{S}_k \in [0, 1]^{\alpha\times\beta\times\gamma}$ of interest, where α, β, γ are grid size dimensions. Substructures represent anatomical regions of interest that may be comprised of single tissues (e.g., cross-sectional slices of an aorta) or combinations thereof (e.g., left and right atria). As shown in Fig. 2, each substructure can be measured in terms of geometric properties (mass, position, size, shape, orientation) and topological properties (connectivity, presence of loops or voids).

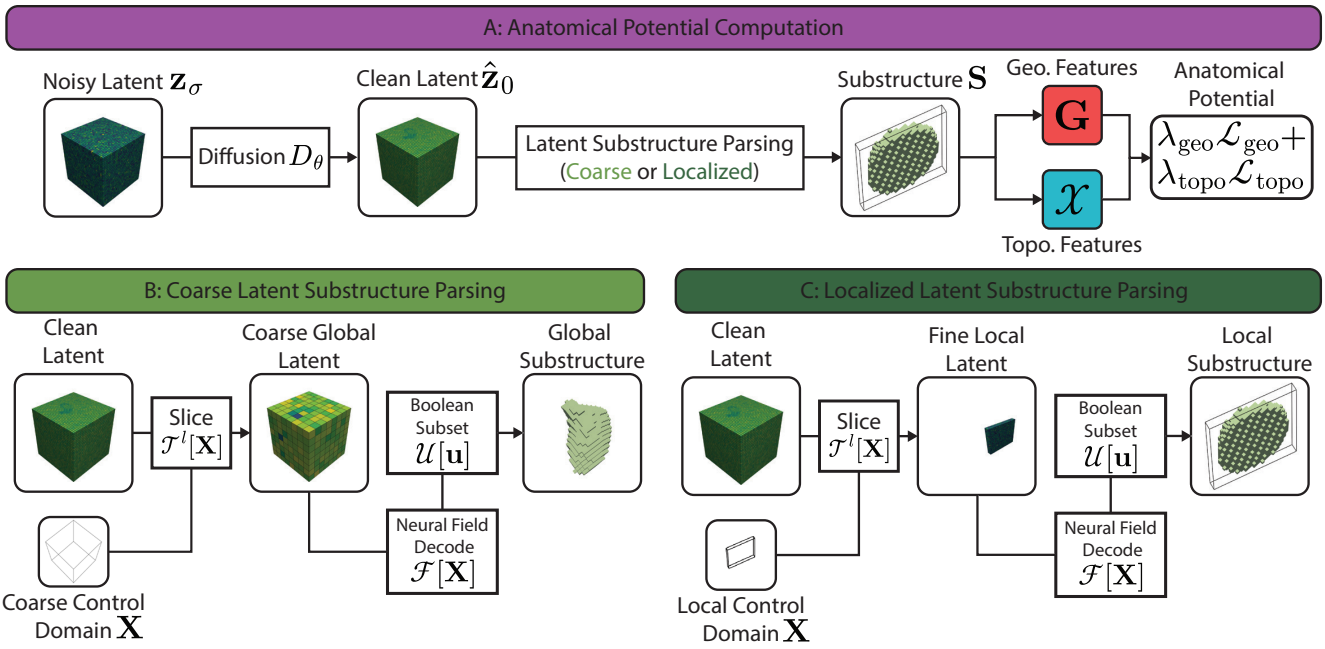

Figure 3. **Efficient parsing of anatomical substructures during diffusion guidance. A**: During guidance, we parse relevant substructures directly from the clean latent prediction with a neural field decoder (L-parsing). **B**: In coarse L-parsing, we use a coarse grid to decode globally defined substructures at low spatial resolution. **C**: In localized L-parsing, we use a similar grid size but spatially transform the template point grid to decode localized substructures at high spatial resolution.

Differentiable Geometric Measurement Given a single parsed substructure $\mathbf{S}_k$, we aim to differentiably extract the geometric moments by numerically integrating the zeroth, first, and second-order moments of the substructure. Here, the zeroth moment represents the mass $m_k \in \mathbb{R}$, the first moment represents the centroid $\mathbf{p}_k \in \mathbb{R}^3$, and the second moment represents the covariance matrix $\mathbf{\Sigma}_k \in \mathbb{R}^{3\times3}$. To do this, we follow Kadry et al. [29], but instead compute the moments for substructures inside arbitrarily sized cuboidal control domains, rather than globally. We first define $\mathbf{\Omega}_k \in \mathbb{R}^{(\alpha\beta\gamma)\times1}$ as the flattened substructure voxel grid $\mathbf{S}_k$ and $\mathbf{r}_k \in \mathbb{R}^{(\alpha\beta\gamma)\times3}$ as the normalized voxel coordinates between 0 and 1. The moments are then computed as:

$$m_k = \mathbf{1}^T \cdot \mathbf{\Omega}_k \quad \text{and} \quad \mathbf{p}_k = \frac{\mathbf{\Omega}_k^T \mathbf{r}_k}{m_k},$$
$$\mathbf{\Sigma}_k = \frac{1}{m_k}\mathbf{r}_k^T \,\text{diag}(\mathbf{\Omega}_k)\,\mathbf{r}_k - \mathbf{p}_k\mathbf{p}_k^T. \tag{3}$$

Here, $\mathbf{1}^T$ is the all-ones vector, and $\text{diag}(\cdot)$ refers to diagonal matrix embedding.

Geometric Decomposition and Size Normalization As the covariance matrix implicitly contains information on mass, we aim to obtain a scale-normalized covariance matrix that relates only to orientation and relative aspect ratios. We decompose the covariance as $\mathbf{\Sigma}_k = \mathbf{U}_k\tilde{\mathbf{\Lambda}}_k\mathbf{U}_k^T = v_k\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^T$, where size $v_k \in \mathbb{R} = \text{tr}(\mathbf{\Sigma}_k)$ is the trace of the covariance matrix, shape $\mathbf{\Lambda}_k \in \mathbb{R}^{3\times3} = \tilde{\mathbf{\Lambda}}_k/v_k$ is the eigenvalue matrix $\mathbf{\Lambda}_k$ normalized by the trace, and orientation $\mathbf{U}_k \in \mathbb{R}^{3\times3}$ is an orthonormal matrix. We thus define the scale-normalized covariance matrix as $\mathbf{\Sigma}_k^n = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^T$.

Geometric Potential Functions After computing the geometric features, we aim to penalize the deviations from the target geometric features $\bar{\mathbf{G}}_k$ through a geometric potential function $\mathcal{L}_k^{\text{geo}}$. We formulate this potential function as a weighted combination of mean squared error losses $\mathcal{L}_{\text{MSE}}$:

$$\mathcal{L}_k^{\text{geo}} = \lambda_0\mathcal{L}_{\text{MSE}}(m_k, \bar{m}_k) + \lambda_1\mathcal{L}_{\text{MSE}}(\mathbf{p}_k, \bar{\mathbf{p}}_k) + \lambda_2\mathcal{L}_{\text{MSE}}(\mathbf{\Sigma}_k^n, \bar{\mathbf{\Sigma}}_k^n). \tag{4}$$

Here, $[\lambda_0, \lambda_1, \lambda_2]$ are the weighting factors for the mass, position, and normalized covariance losses, respectively.

Adaptive Mass Weighting Given that geometric features are now defined locally, rather than globally, our potential functions must account for numerical instability in the centroid and covariance-based losses resulting from empty voxels. We address this by adaptively setting $\lambda_1 = \lambda_2 = 0$ when the mass is below a specified threshold.

Differentiable Topological Measurement We use persistent homology (PH) to measure the presence of topological features within the substructure $\mathbf{S}_k$ considered as a cubical complex, similar to Gupta et al. [21]. We consider super-level sets of $\mathbf{S}_k$, which are the set of voxels above a threshold value τ. By decreasing τ, we obtain a filtration

of nested sets, all of which are used to extract topological features such as connected components (0D features), loops (1D features), and voids (2D features). The output of this process is a set of persistence points $p \in \mathcal{X}_k$ which include the birth b and death d thresholds for all topological features. Intuitively, persistent features have a large interval between their birth and death thresholds. We use the Cubical Ripser library [31] to compute the PH of $\mathbf{S}_k$.

Softmax Temperature Tuning As our substructure is derived from softmaxed multi-class probability maps, we found that the gradient of the topological potential was too low in regions where the probability was close to 0 or 1. To address this, when decoding the substructure from the latent space, we apply softmax with an increased temperature for topological guidance, enabling the gradient to pass through during backpropagation.

Topological Potential Functions To enforce topological structures within $\mathbf{S}_k$ during the reverse diffusion process, we partition the persistence set into disjoint sets $\mathcal{X}_k = \mathcal{Y}_k \cup \mathcal{Z}_k$ consisting of points that should be preserved $\mathcal{Y}_k$ or suppressed $\mathcal{Z}_k$ based on a topological prior $\mathcal{B}_k \in \mathbb{R}^3$, which specifies the desired features for the components, loops, and voids, respectively. To differentiably compute the topological potential, we sample the voxel intensities from $\mathbf{S}_k$ at the birth r_b^p and death r_d^p coordinates for each persistence point p. We then maximize or minimize the persistence of each feature through the following potential functions:

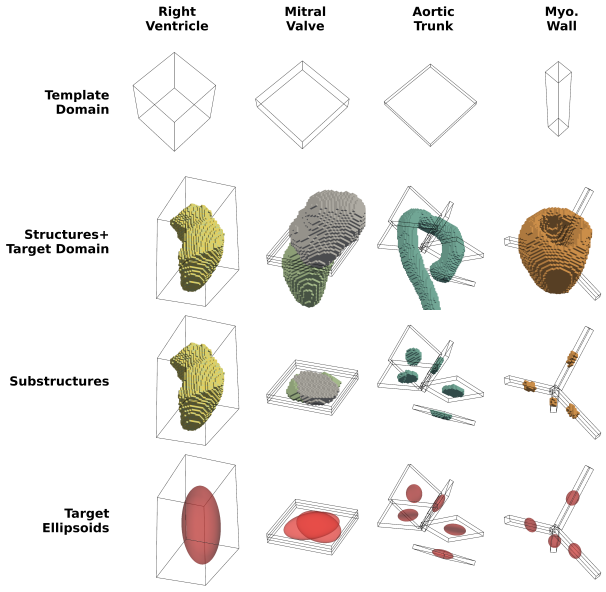

Figure 4. **Geometric Control Tasks**. We define a variety of relevant tasks by varying the selected tissues, template domain grid size, and control domain-specific spatial transforms.

$$\mathcal{L}_k^{\text{topo}} = -\sum_{p\in\mathcal{Y}_k} \underbrace{|\mathbf{S}_k(r_b^p) - \mathbf{S}_k(r_d^p)|^2}_{\text{Preserve Loss}} + \sum_{p\in\mathcal{Z}_k} \underbrace{|\mathbf{S}_k(r_b^p) - \mathbf{S}_k(r_d^p)|^2}_{\text{Suppress Loss}}. \quad (5)$$

Gradient-Based Guidance Following Kadry et al. [29], we guide the diffusion process using the gradient derived from anatomical potential functions. At each sampling step, we first denoise the intermediately noised latent $\mathbf{z}_\sigma$ to obtain a clean latent prediction $\hat{\mathbf{z}}_0 = D_\theta(\mathbf{z}_\sigma; \sigma)$, which is then *parsed* into K substructures $\mathbf{S}_k$ as described in Sec. 3.3. We compute a composite anatomical potential $\mathcal{L} = \frac{1}{K}\sum_k(\lambda_{\text{geo}}\mathcal{L}_k^{\text{geo}} + \lambda_{\text{topo}}\mathcal{L}_k^{\text{topo}})$ where λ_{geo} and λ_{topo} weight the geometric and topological potentials (Fig. 3, top row). The guided denoising step is then applied as follows:

$$\underbrace{D_\theta^w(\mathbf{z}_\sigma;\sigma)}_{\text{Guided Denoising}} = \underbrace{D_\theta(\mathbf{z}_\sigma;\sigma)}_{\text{Uncond. Denoising}} - \underbrace{\sigma^2 \cdot \nabla_{\mathbf{z}_\sigma}\mathcal{L}}_{\text{Anatomical Guidance}}. \quad (6)$$

3.3. Substructure Parsing for Diffusion Guidance

Voxel-Space Parsing Given a predicted voxel-space segmentation map $\hat{\mathbf{V}}$, we can parse K substructures $\mathbf{S}_k$ with voxel-space substructure parsing (Fig. 2), which we will refer to as *V-parsing*. V-parsing extracts a substructure by first using the boolean subset operator $\mathcal{U}[\mathbf{u}] : \mathbb{R}^C \to \mathbb{R}$; parametrized by a selection vector $\mathbf{u} \in \{0,1\}^C$ for elementwise extraction and recombination of tissues from $\hat{\mathbf{V}}$ into a structure voxel map $\hat{\mathbf{S}}_k$. By taking different tissue combinations, we enable the control of various structures within the segmentation. The structure slicing operator $\mathcal{T}^s[\mathbf{X}_k] : \mathbb{R} \to \mathbb{R}$ then samples voxelwise intensity values from $\hat{\mathbf{S}}_k$ at locations specified by a cuboidal control domain $\mathbf{X}_k \in \mathbb{R}^{\alpha\times\beta\times\gamma\times 3}$ discretized into a lattice-like point grid. The final relation is given by:

$$\mathbf{S}_k = \underbrace{\mathcal{T}^s[\mathbf{X}_k]}_{\text{Slice Structure}} \circ \underbrace{\mathcal{U}[\mathbf{u}](\hat{\mathbf{V}})}_{\text{Subset Voxel}}. \quad (7)$$

Spatial Transformation of Template Domains Control domains $\mathbf{X}_k$ are obtained by spatially transforming a template point grid $\mathbf{X}_k^{\text{temp}} \in \mathbb{R}^{\alpha\times\beta\times\gamma\times 3}$ using affine transformation parameters $\mathbf{A}_k = [\mathbf{R}_k, \mathbf{s}_k, \mathbf{t}_k]$, where $\mathbf{R}_k \in \mathbb{R}^{3\times 3}$ is a rotation matrix, $\mathbf{s}_k \in \mathbb{R}^3$ is a scaling vector, and $\mathbf{t}_k \in \mathbb{R}^3$ is a translation vector. The spatial transformation is defined as:

$$\mathbf{X}_k = \mathbf{R}_k\,\text{diag}(\mathbf{s}_k)\mathbf{X}_k^{\text{temp}} + \mathbf{t}_k. \quad (8)$$

This formulation enables flexible substructure parsing across multiple design axes (Fig. 4). For example, the template grid size controls the discretization (coarse to fine) and dimensionality (3D to 1D) of the parsed substructure. The affine parameters $\mathbf{A}_k$ can be defined based on custom coordinate systems and enable the localization of substructures under varying extents, orientations, and positions in 3D space.

Latent-Space Parsing Rather than decoding the entire voxel grid and subsequently slicing out substructures, we propose to use neural field decoders to directly parse substructures from latent space (*L-parsing*). This is accomplished by applying the latent slicing operator $\mathcal{T}^l[\mathbf{X}_k]$ on

the clean predicted latents $\hat{\mathbf{z}}_0$ followed by the neural field decoder $\mathcal{F}[\mathbf{X}_k]$ and the boolean subset operator $\mathcal{U}[\mathbf{u}]$:

$$\mathbf{S}_k = \underbrace{\mathcal{U}[\mathbf{u}]}_{\text{Subset Voxel}} \circ \underbrace{\mathcal{F}[\mathbf{X}_k]}_{\text{Decode Latent}} \circ \underbrace{\mathcal{T}^l[\mathbf{X}_k](\hat{\mathbf{z}}_0)}_{\text{Slice Latent}}. \quad (9)$$

Partial Decoding Strategies Neural field representations offer a key advantage in that they enable decoding from arbitrary point sets. We introduce two L-parsing strategies that apply *partial decoding* by using point grids with small grid sizes, enabling fast guidance at inference time. To efficiently measure global-level anatomical properties that are invariant to spatial resolution, we apply **coarse L-parsing** by using a template grid $\mathbf{X}_k^{\text{temp}}$ with a low grid discretization $[\alpha < H, \beta < W, \gamma < D]$ and setting the identity-like affine transformation parameters $\mathbf{A}^{\text{coarse}} = [\mathbf{I}, \mathbf{1}, \mathbf{0}]$ (Fig. 3 bottom left). To efficiently measure anatomical properties that are defined locally, we apply **localized L-parsing** by using a similarly low template grid discretization, but spatially transform the template point grid $\mathbf{X}_k^{\text{temp}}$ into a localized region, effectively achieving a high spatial resolution (Fig. 3 bottom right).

4. Experiments

4.1. Anatomical Datasets

For the **cardiac dataset**, we extract heart-related labels from the TotalSegmentator dataset [46], resulting in 596 11-channel segmentations. For the **aortic dataset**, we extract aorta-related labels from the TotalSegmentator dataset, resulting in 450 7-channel segmentations. For the **spinal dataset**, we extract spinal vertebrae-related labels from the CTSpine1k dataset [10], resulting in 784 25-channel segmentations. For the **coronary dataset**, we extract coronary artery-related labels from the DISRUPT-CAD dataset [45], resulting in approximately 360 unique 4-channel segmentations. All segmentations were resampled to a uniform resolution of 128^3, and were partitioned into training and validation sets with an 80/20 split.

4.2. Control Tasks

Geometric Control To evaluate geometric control, we use the cardiac validation set to determine appropriate target control domains and target geometric features (see Fig. 4). The first task is **Right Ventricle**, which constrains the right ventricle within a volumetric domain. The second task is **Mitral Valve**, which constrains two volumetric domains at the intersection of the left ventricle and atrium. The third task is **Aortic Trunk**, which constrains 5 planar domains along the aortic trunk centerline. The fourth task is **Myocardium Wall**, which constrains 4 linear domains emanating from the centroid of the myocardium. All quantitative experiments were conducted by sampling 128 synthetic samples.

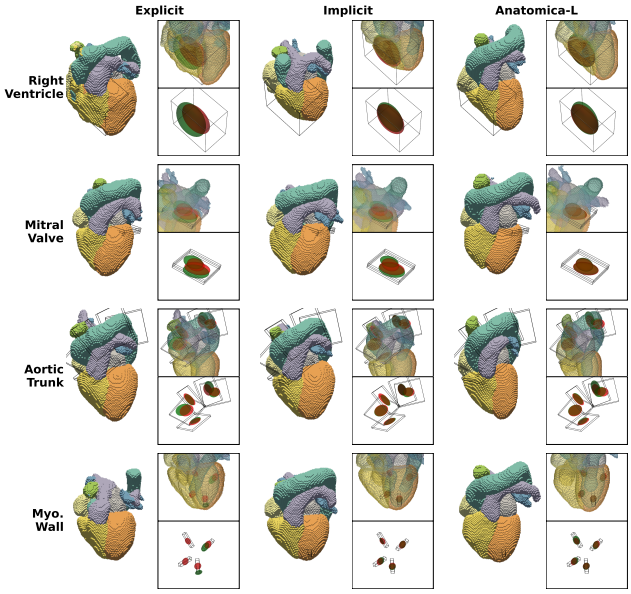

Figure 5. **Qualitative evaluation of geometric control experiments.** We generate anatomical segmentations based on target domains (black frames) and geometric features (red ellipsoids) for four cardiac tasks. Sample geometry shown as green ellipsoids.

Topological Control To evaluate topological control, we define a task for each anatomical dataset. We set a global target control domain as well as appropriate topological priors for each task: **Atria Separation** enforces 2 connected components for the left and right atria; **Branch Connectivity** enforces 1 connected component for the ascending aorta and all branches; **Vertebrae Connectivity** enforces 1 connected component and 9 loops for five thoracic spinal vertebrae (T6-T10); and **Calcium Count** enforces 2 calcium components in the coronary artery wall.

Table 1. **Comparison of geometric control task approaches**.

Approach	Decoder	Parsing	Control Method
Explicit	Neural Field	N/A	Conditioning
Implicit	Neural Field	N/A	Conditioning
Anatomica-V	Convolutional	V-parsing	Guidance
Anatomica-L	Neural Field	L-parsing	Guidance

4.3. Implementation Details

As Anatomica guides the unconditional sampling process for each task in a training-free manner, we only train a separate unconditional diffusion model for each anatomical dataset. We present two variants of Anatomica that utilize different decoding and parsing strategies for guidance (Tab. 1). **Anatomica-V** uses a convolutional decoder to first produce a global voxel grid from the predicted clean la-

Table 2. **Quantitative results for geometric control tasks**. We report geometric fidelity and generation quality for each task-approach combination. Fidelity values for mass, centroid, and covariance are multiplied by 1e5, 1e4, 1e5 respectively.

		Geometric Fidelity (↓)			Generation Quality (↓)	
Task	**Approach**	**Mass**	**Cent.**	**Cov.**	**FMD**	**1-NNA**
Right Vent.	Explicit	154.5	227.1	101.4	164.7	0.761
	Implicit	60.6	51.0	30.6	156.3	0.593
	Anatomica-L	17.5	48.6	22.1	93.7	**0.566**
	Anatomica-V	**12.3**	**30.2**	**21.6**	**84.9**	0.590
Mitral Valve	Explicit	29.0	246.5	37.4	97.1	0.577
	Implicit	8.91	87.0	17.3	314.8	0.661
	Anatomica-L	**3.22**	**11.4**	**7.89**	**88.8**	**0.577**
	Anatomica-V	3.81	41.4	7.99	93.8	0.586
Aortic Trunk	Explicit	2.53	272.6	13.9	114.3	0.587
	Implicit	**0.81**	**35.2**	**5.30**	104.8	0.565
	Anatomica-L	2.38	86.0	16.2	89.8	0.580
	Anatomica-V	2.40	84.9	14.4	**82.0**	**0.561**
Myo. Wall	Explicit	1.01	123.4	3.39	130.7	0.574
	Implicit	0.48	**22.3**	**1.67**	111.0	0.558
	Anatomica-L	**0.29**	34.6	1.87	**86.4**	0.609
	Anatomica-V	0.36	42.3	1.93	97.3	**0.554**

tents, upon which we extract localized substructures during guidance with V-parsing. In contrast, **Anatomica-L** uses a neural field decoder to directly parse substructures with L-parsing. For the geometric control tasks, we use local L-parsing, while the topological tasks use coarse L-parsing. Unless specified otherwise, we use 100 diffusion sampling steps.

4.4. Baselines

We also compare our general approach of guiding unconditional diffusion models against approaches that require conditional training for each task (Tab. 1). For the geometric control tasks, we implement conditional baselines to control target geometric features representing the size m, centroid $\mathbf{p}$, and covariance $\mathbf{\Sigma}$ of each substructure within the task. The first baseline is **Explicit Conditioning**, where we directly encode geometric attributes as scalar values in the conditioning signal [28, 30]. We flatten and stack all geometric moments (mass, centroid, covariance) into a 13-dimensional vector for all K substructures. We then expand this vector into a voxel grid $\mathbf{G}_{\text{exp}} \in \mathbb{R}^{13\times K\times h\times w\times d}$ which is concatenated to the latent grid $\mathbf{z}$ along the channel dimension. The second is **Implicit Conditioning**, where we indirectly encode geometric attributes in the concatenated conditioning signal through 3D heatmaps [30]. Here, we embed the geometric moments (centroid, covariance) as 3D Gaussians in voxel space. For each substructure, we create a voxel map $\mathbf{G}_{\text{imp}} \in \mathbb{R}^{K\times h\times w\times d}$ where the voxel values encode the Mahalanobis distance.

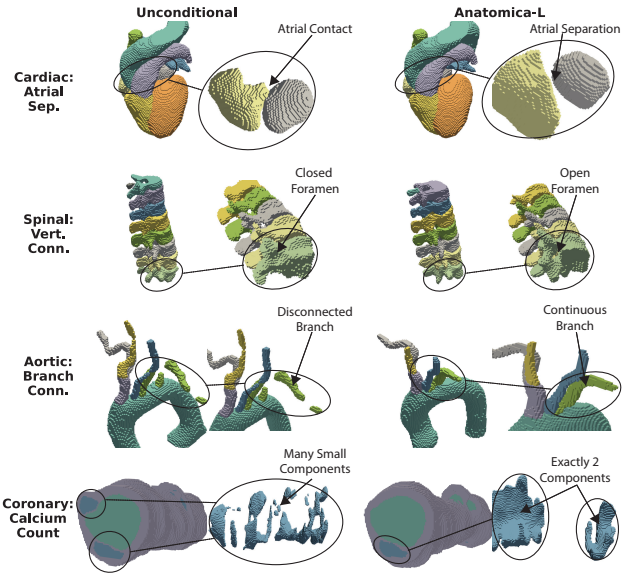

Figure 6. **Qualitative evaluation of topological control experiments.** We generate anatomical segmentations based on target topological priors for four anatomical datasets.

Table 3. **Quantitative evaluation for topological control tasks**. We report Betti precision for number of connected components B_0, loops B_1, and voids B_2, and generation quality (1-NNA) for each approach.

		Topo. Precision (%)			Gen. Qual.
Task	**Approach**	**B0** (↑)	**B1** (↑)	**B2** (↑)	**1-NNA** (↓)
Atrial Sep.	Uncond.	7.81	5.47	56.2	**0.578**
	Anatomica-L	**78.9**	**89.1**	**97.7**	0.606
Branch Conn.	Uncond.	55.5	12.5	63.3	0.559
	Anatomica-L	**77.3**	**17.2**	**64.1**	**0.532**
Vert. Conn.	Uncond.	28.9	8.59	**12.5**	**0.518**
	Anatomica-L	**74.2**	**26.6**	7.03	0.537
Calcium Count	Uncond.	0.00	2.34	95.3	0.653
	Anatomica-L	**60.9**	**79.7**	**98.4**	**0.618**

4.5. Evaluation Metrics

We measure morphological quality metrics by measuring the Fréchet morphological distance (FMD) [27] between real and synthetic distributions in morphological space. To compute such features, we consider all tissues as substructures and concatenate all masses, centroids, and normalized eigenvalues. We average the per-tissue 1-nearest neighbor accuracy (1-NNA) to compare point cloud distributions using the Earth Mover's Distance (EMD) [48]. We evaluate geometric control fidelity by taking the L_1-norm between the target and measured moments. Lastly, to measure topological control fidelity, we compute the topological precision for components B_0, loops B_1, and voids B_2 as the

fraction of samples with the correct Betti number.

4.6. Results

Precise Geometric Control We quantitatively evaluate geometric control on the cardiac dataset with four different tasks. We see in Tab. 2 and Fig. 5 that our inference-time approach (Anatomica) is competitive in controlling the size, shape, position, and orientation of various substructures in the cardiac dataset. The closest baseline is the specialized implicit concatenation method, which requires conditional retraining for every task.

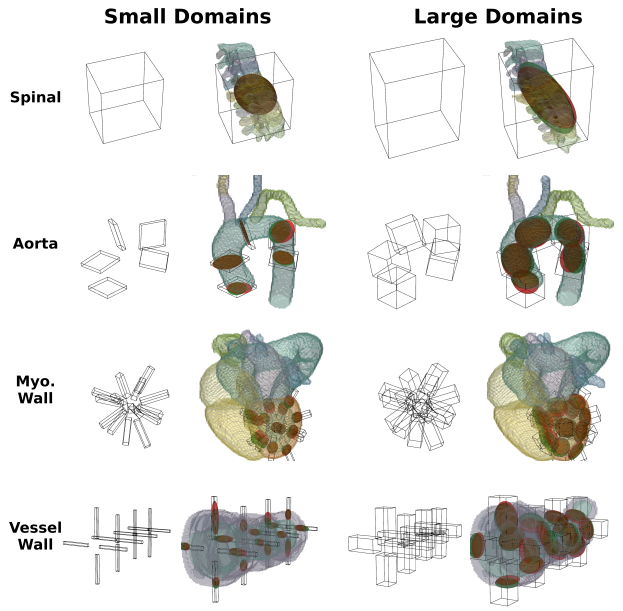

Figure 7. **Multi-scale geometric control of various anatomical substructures over different coordinate systems.** We generate anatomical segmentations based on domain size and anatomically relevant coordinate systems (Cartesian, curvilinear, cylindrical, and spherical).

Topological Control We evaluate topological control on four anatomical datasets. We see in Tab. 3 and Fig. 6 that our framework is able to control the number of connected components and loops for various types of anatomical structures. Adherence to the number of voids is not improved for the aortic and vertebrae datasets, possibly due to the existence of single-voxel voids that are not easily detected at coarser measurement resolutions.

Multi-scale Geometric Control We demonstrate Anatomica's flexibility for geometric control across multiple anatomical types including the cardiac, aortic, spinal, and coronary datasets. As shown in Fig. 7, our framework handles varying spatial scales (large to small control domains) and operates across different coordinate system types (Cartesian, cylindrical, and spherical).

Partial Decoding Ablation We evaluate the speed-fidelity trade-off for partial decoding guidance under various decoding resolutions and strategies. We use the geometric control task involving the right ventricle for evaluation. We see in Tab. 4 that low-resolution partial decoding with Anatomica-L maintains geometric fidelity while achieving substantial speedups against high-resolution decoding. We also find that given the same decoding resolution, using a neural field decoder (Anatomica-L) increases sampling speed at the cost of slightly reduced geometric fidelity, as compared to convolutional decoders (Anatomica-V).

Table 4. **Quantitative ablation study for partial decoding strategies**. We evaluate the geometric fidelity and generation quality, and sampling speed for different decoding strategies and resolutions. Speed is measured in terms of sampled label maps per second using the maximum allowable batch size on a single GPU, normalized to the slowest method. Fidelity values for mass, centroid, and covariance are multiplied by 1e5, 1e4, 1e5 respectively.

Methodology			Geometric Fidelity (↓)			Gen. Qual. (↓)		Speed (↑)
Approach	**Domain**	**Res.**	**Mass**	**Cent.**	**Cov.**	**FMD**	**1-NNA**	**Speed**
Anatomica-L	Local	High	17.02	**48.14**	**21.85**	**91.16**	0.57	2.48
		Med.	16.64	48.41	22.03	93.89	0.58	7.43
		Low	**16.43**	48.75	22.07	105.57	**0.55**	**10.40**
	Coarse	High	**17.50**	48.58	**22.77**	**114.62**	0.55	2.08
		Med.	17.75	48.78	23.03	119.93	**0.53**	7.43
		Low	20.30	**46.96**	25.60	123.31	0.54	**10.40**
Anatomica-V	Global	High	11.95	30.66	21.92	84.70	0.58	1.00

5. Discussion

Limitations The main limitation of our guidance approach is that loss weightings should be tuned to balance the contributions of each constraint. However, we found that our choice of weights transfers readily between the four anatomical datasets we consider in this study. Moreover, we found that computing persistent homology at high resolution is computationally demanding and limits the coarse decoding resolution for topological guidance.

Conclusion We propose an inference-time framework for controlling generative models of anatomical voxel maps based on localized geo-topological attributes. Our design space centers on cuboidal control domains that compositionally slice out substructures across varying dimensions and coordinate systems. Over these domains, we propose the use of geo-topological potential functions for diffusion guidance, as well as neural field decoders for efficient partial decoding from latent space. We demonstrate state-of-the-art performance for geometric and topological control across a variety of anatomical systems and structures. We believe this work opens new avenues for controllable anatomical generation, with applications to virtual clinical trials and synthetic data augmentation for machine learning.

References

[1] Ehsan Abadi, William P Segars, Benjamin MW Tsui, Paul E Kinahan, Nick Bottenus, Alejandro F Frangi, Andrew Maidment, Joseph Lo, and Ehsan Samei. Virtual clinical trials in medical imaging: a review. *Journal of Medical Imaging*, 7 (4):042805–042805, 2020. 1

[2] Anonymous. Protcomposer: Compositional protein structure generation with 3d ellipsoids. In *Submitted to The Thirteenth International Conference on Learning Representations*, 2024. under review. 2

[3] Reidmen Aróstica, David Nolte, Aaron Brown, Amadeus Gebauer, Elias Karabelas, Javiera Jilberto, Matteo Salvador, Michele Bucelli, Roberto Piersanti, Kasra Osouli, et al. A software benchmark for cardiac elastodynamics. *Computer Methods in Applied Mechanics and Engineering*, 435: 117485, 2025. 2

[4] Benjamin Billot, Douglas N Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V Dalca, Juan Eugenio Iglesias, et al. Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical image analysis*, 86:102789, 2023. 1

[5] Rickard Brüel-Gabrielsson, Bradley J Nelson, Anjan Dwaraknath, Primoz Skraba, Leonidas J Guibas, and Gunnar Carlsson. A topology layer for machine learning. *arXiv preprint arXiv:1905.12200*, 2019. 2

[6] Nick Byrne, James R Clough, Israel Valverde, Giovanni Montana, and Andrew P King. A persistent homology-based topological loss for cnn-based multiclass segmentation of cmr. *IEEE transactions on medical imaging*, 42(1):3–14, 2022. 2

[7] James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):8766–8778, 2020. 2

[8] Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000. 6

[9] Bram de Wilde, Max T Rietberg, Guillaume Lajoinie, and Jelmer M Wolterink. Steerable anatomical shape synthesis with implicit neural representations. *arXiv preprint arXiv:2504.03313*, 2025. 2

[10] Yang Deng, Ce Wang, Yuan Hui, Qian Li, Jun Li, Shiwei Luo, Mengke Sun, Quan Quan, Shuxin Yang, You Hao, et al. Ctspine1k: A large-scale dataset for spinal vertebrae segmentation in computed tomography. *arXiv preprint arXiv:2105.14711*, 2021. 6

[11] Neel Dey, Benjamin Billot, Hallee E. Wong, Clinton Wang, Mengwei Ren, Ellen Grant, Adrian V Dalca, and Polina Golland. Learning general-purpose biomedical volume representations using randomized synthesis. In *The Thirteenth International Conference on Learning Representations*, 2025. 1

[12] Haoran Dou, Seppo Virtanen, Nishant Ravikumar, and Alejandro F Frangi. A generative shape compositional framework: Towards representative populations of virtual heart chimaeras. *arXiv preprint arXiv:2210.01607*, 2022. 2

[13] Dave Epstein, Allan Jabri, Ben Poole, Alexei Efros, and Aleksander Holynski. Diffusion self-guidance for controllable image generation. *Advances in Neural Information Processing Systems*, 36:16222–16239, 2023. 2

[14] Enrico Fabris, Balasz Berta, Tomasz Roleder, Renicus S Hermanides, Alexander JJ IJsselmuiden, Floris Kauer, Fernando Alfonso, Clemens Von Birgelen, Javier Escaned, Cyril Camaro, et al. Thin-cap fibroatheroma rather than any lipid plaques increases the risk of cardiovascular events in diabetic patients: Insights from the combine oct–ffr trial. *Circulation: Cardiovascular Interventions*, 15(5):e011728, 2022. 2

[15] Weixi Feng, Chao Liu, Sifei Liu, William Yang Wang, Arash Vahdat, and Weili Nie. Blobgen-vid: Compositional text-to-video generation with blob video representations. *arXiv preprint arXiv:2501.07647*, 2025. 2

[16] Virginia Fernandez, Walter Hugo Lopez Pinaya, Pedro Borges, Mark S Graham, Petru-Daniel Tudosiu, Tom Vercauteren, and M Jorge Cardoso. Generating multi-pathological and multi-modal images and labels for brain mri. *Medical Image Analysis*, 97:103278, 2024. 1

[17] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouve, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690, 2019. 7

[18] Álvaro González. Measurement of areas on a sphere using fibonacci and latitude–longitude lattices. *Mathematical geosciences*, 42(1):49–64, 2010. 4

[19] Vivek Gopalakrishnan and Polina Golland. Fast auto-differentiable digitally reconstructed radiographs for solving inverse problems in intraoperative imaging. In *Workshop on Clinical Image-Based Procedures*, pages 1–11. Springer, 2022. 1

[20] Vivek Gopalakrishnan, Neel Dey, and Polina Golland. Intraoperative 2d/3d image registration via differentiable x-ray rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11662–11672, 2024. 1

[21] Saumya Gupta, Dimitris Samaras, and Chao Chen. Topodiffusionnet: A topology-aware diffusion model. In *The Thirteenth International Conference on Learning Representations*, 2025. 2, 4

[22] Uxio Hermida, Milou PM van Poppel, Malak Sabry, Hamed Keramati, Johannes K Steinweg, John M Simpson, Trisha V Vigneswaran, Reza Razavi, Kuberan Pushparajah, David FA Lloyd, et al. The onset of coarctation of the aorta before birth: Mechanistic insights from fetal arch anatomy and haemodynamics. *Computers in Biology and Medicine*, 182: 109077, 2024. 2

[23] Amir Hertz, Or Perel, Raja Giryes, Olga Sorkine-Hornung, and Daniel Cohen-Or. Spaghetti: Editing implicit shapes through part aware generation. *ACM Transactions on Graphics (TOG)*, 41(4):1–20, 2022. 2

[24] Jiangbei Hu, Ben Fei, Baixin Xu, Fei Hou, Weidong Yang, Shengfa Wang, Na Lei, Chen Qian, and Ying He. Topology-

aware latent diffusion for 3d shape generation. *arXiv preprint arXiv:2401.17603*, 2024. 2

[25] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015. 3

[26] Karim Kadry, Max L Olender, David Marlevi, Elazer R Edelman, and Farhad R Nezami. A platform for high-fidelity patient-specific structural modelling of atherosclerotic arteries: from intravascular imaging to three-dimensional stress distributions. *Journal of the Royal Society Interface*, 18 (182):20210436, 2021. 2

[27] Karim Kadry, Shreya Gupta, Farhad R Nezami, and Elazer R Edelman. Probing the limits and capabilities of diffusion models for the anatomic editing of digital twins. *npj Digital Medicine*, 7(1):1–12, 2024. 7

[28] Karim Kadry, Max L Olender, Andreas Schuh, Abhishek Karmakar, Kersten Petersen, Michiel Schaap, David Marlevi, Adam UpdePac, Takuya Mizukami, Charles Taylor, et al. Morphology-based non-rigid registration of coronary computed tomography and intravascular images through virtual catheter path optimization. *IEEE Transactions on Medical Imaging*, 2024. 2, 7, 1, 3

[29] Karim Kadry, Shoaib Goraya, Ajay Manicka, Abdalla Abdelwahed, Farhad Nezami, and Elazer Edelman. Cardiocomposer: Flexible and compositional anatomical structure generation with disentangled geometric guidance. *arXiv preprint arXiv:2509.08015*, 2025. 2, 4, 5, 1

[30] Karim Kadry, Shreya Gupta, Jonas Sogbadji, Michiel Schaap, Kersten Petersen, Takuya Mizukami, Carlos Collet, Farhad R Nezami, and Elazer R Edelman. A diffusion model for simulation ready coronary anatomy with morpho-skeletal control. In *European Conference on Computer Vision*, pages 396–412. Springer, 2025. 7

[31] Shizuo Kaji, Takeki Sudo, and Kazushi Ahara. Cubical ripser: Software for computing persistent homology of image and volume data. *arXiv preprint arXiv:2005.12692*, 2020. 5

[32] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *arXiv preprint arXiv:2206.00364*, 2022. 3, 2

[33] Fanwei Kong, Sascha Stocker, Perry S Choi, Michael Ma, Daniel B Ennis, and Alison L Marsden. Sdf4chd: Generative modeling of cardiac anatomies with congenital heart defects. *Medical Image Analysis*, 97:103293, 2024. 2

[34] Juil Koo, Seungwoo Yoo, Minh Hieu Nguyen, and Minhyuk Sung. Salad: Part-level latent diffusion for 3d shape generation and manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14441–14451, 2023. 2

[35] Ira Ktena, Olivia Wiles, Isabela Albuquerque, Sylvestre-Alvise Rebuffi, Ryutaro Tanno, Abhijit Guha Roy, Shekoofeh Azizi, Danielle Belgrave, Pushmeet Kohli, Taylan Cemgil, et al. Generative models improve fairness of medical classifiers under distribution shifts. *Nature Medicine*, 30(4):1166–1173, 2024. 1

[36] Joshua M. Long. Random fourier features pytorch. *GitHub. Note: https://github.com/jmclong/random-fourier-features-pytorch*, 2021. 1

[37] Ali Madani, Ahmed Bakhaty, Jiwon Kim, Yara Mubarak, and Mohammad RK Mofrad. Bridging finite element and machine learning modeling: stress prediction of arterial walls in atherosclerosis. *Journal of biomechanical engineering*, 141(8):084502, 2019. 2

[38] Stefania L Moroianu, Christian Bluethgen, Pierre Chambon, Mehdi Cherti, Jean-Benoit Delbrouck, Magdalini Paschali, Brandon Price, Judy Gichoya, Jenia Jitsev, Curtis P Langlotz, et al. Improving performance, robustness, and fairness of radiographic ai models with finely-controllable synthetic data. *arXiv preprint arXiv:2508.16783*, 2025. 1

[39] Weili Nie, Sifei Liu, Morteza Mardani, Chao Liu, Benjamin Eckart, and Arash Vahdat. Compositional text-to-image generation with dense blob representations. *arXiv preprint arXiv:2405.08246*, 2024. 2

[40] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *European Conference on Computer Vision*, pages 523–540. Springer, 2020. 3

[41] Mengyun Qiao, Kathryn A McGurk, Shuo Wang, Paul M Matthews, Declan P O'Regan, and Wenjia Bai. A personalized time-resolved 3d mesh generative model for unveiling normal heart dynamics. *Nature Machine Intelligence*, pages 1–12, 2025. 2

[42] Ali Sarrami-Foroushani, Toni Lassila, Michael MacRaild, Joshua Asquith, Kit CB Roes, James V Byrne, and Alejandro F Frangi. In-silico trial of intracranial flow diverters replicates and expands insights from conventional clinical trials. *Nature communications*, 12(1):3861, 2021. 1

[43] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020. 1

[44] Marco Viceconti, Luca Emili, Payman Afshari, Eulalie Courcelles, Cristina Curreli, Nele Famaey, Liesbet Geris, Marc Horner, Maria Cristina Jori, Alexander Kulesza, et al. Possible contexts of use for in silico trials methodologies: a consensus-based review. *IEEE Journal of Biomedical and Health Informatics*, 25(10):3977–3982, 2021. 1

[45] Zachary M Visinoni, Daniel L Jurewitz, Dean J Kereiakes, Richard Shlofmitz, Evan Shlofmitz, Ziad Ali, Jonathan Hill, and Michael S Lee. Coronary intravascular lithotripsy for severe coronary artery calcification: The disrupt cad i-iv trials. *Cardiovascular Revascularization Medicine*, 65:81–87, 2024. 6

[46] Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5): e230024, 2023. 6

[47] Jessica G Williams, David Marlevi, Jan L Bruse, Farhad R Nezami, Hamed Moradi, Ronald N Fortunato, Spandan Maiti, Marie Billaud, Elazer R Edelman, and Thomas G Gleason. Aortic dissection is determined by specific shape

and hemodynamic interactions. *Annals of Biomedical Engineering*, 50(12):1771–1786, 2022. 2

[48] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4541–4550, 2019. 7

Anatomica: Localized Control over Geometric and Topological Properties for Anatomical Diffusion Models

Supplementary Material

6. Overview

Methods In Sec. 7, we detail the methods relating to diffusion model training, substructure parsing, and geo-topological measurement.
Experimental Details In Sec. 8, we provide additional details on dataset creation, task setup, and evaluation metrics.
Ablations In Sec. 9, we study the influence of various hyperparamters such as individual loss weightings, decoding resolutions, and softmax temperature for geometric and topological guidance.

7. Methodological Details

7.1. Variational Autoencoder

For this study, we adapt the voxel map VAE architecture specified by Kadry et al. [29], which consists of a convolutional encoder and decoder. All architectural and training hyperparameters can be found in tables 5 and 6.
Decoder Architecture We introduce two variants of Anatomica for latent diffusion guidance, with the primary difference being the decoder architecture that converts latent grid representation $\mathbf{z} \in \mathbb{R}^{c\times h\times w\times d}$ into a voxel grid representation $\hat{\mathbf{V}} \in \mathbb{R}^{C\times H\times W\times D}$ that can be anatomically characterized. **Anatomica-V** uses a convolutional decoder that mirrors the encoder, where the latent grid can only be decoded to full voxel resolution. On the other hand, **Anatomica-L** uses a neural field-based decoder that takes as input an arbitrary point grid $\mathbf{X}^q \in \mathbb{R}^{H\times W\times D\times 3}$ and returns for each point, the probability vector denoting the most likely anatomical class.
Neural Field Decoder Our decoder $\mathcal{F}$ decodes voxel maps with neural fields by first applying a bottleneck convolution to the latent grid representation in order to aggregate features within a local neighborhood. We then use a set of query points to interpolate into the latent grid representation using the slice operator $\mathcal{T}^l[\mathbf{X}^q]$ to create a set of latent points which are then point-wise concatenated with random Fourier features [36, 43]. These features are then fed into a multi-layer perceptron (MLP) which consists of several hidden layers and finally outputs a logit vector for each query point. The logit vectors are then softmaxed to produce a segmentation probability vector.
Training We train all autoencoders with a combination of Dice-Cross Entropy reconstruction loss and KL divergence loss [28]. For neural field decoder training, we decode back to the full resolution global voxel grid with a query point grid $\mathbf{X}^q \in \mathbb{R}^{H\times W\times D\times 3}$.

Table 5. Autoencoder architecture hyperparameters

Conv. Encoder (shared)	**Value**
Num. Channels	[64, 128, 256]
Num. Res. Blocks	2
Final Downscaling Factor	4
Bottleneck Dim	3
Conv. Decoder (Anatomica-V)	**Value**
Num. Channels	[64, 128, 256]
Num. Res. Blocks	2
Final Upscaling Factor	4
Neural Field Decoder (Anatomica-L)	**Value**
Bottleneck Conv. Channels	64
Positional Encoding Dim	10
Positional Encoding Bandwidth	1
MLP Hidden Dim	128
MLP Num Layers	3
Normalization	LayerNorm
Activation	ReLU

Table 6. Autoencoder training hyperparameters

Hyperparameter	**Value**
Learning Rate	1×10^{-5}
Epochs	40
Batch Size	1
Dice-CE Loss Weight	1
KL Loss Weight	1×10^{-6}

7.2. Latent Diffusion Model

Architecture & Training For latent diffusion model architecture and training, we follow the formulation and architecture specified by Kadry et al. [29]. All architectural and training hyperparameters can be found in table 7. Our denoising model D_θ is parametrized in a skip-connection manner with a U-Net $\mathcal{K}_\theta$ with a convolutional encoder and decoder through the following relation:

$$D_\theta(\mathbf{z}_\sigma;\sigma) = c_{\text{skip}}(\sigma)\,\mathbf{z}_\sigma + c_{\text{out}}(\sigma)\,\mathcal{K}_\theta\big(c_{\text{in}}(\sigma)\,\mathbf{z}_\sigma;\, c_{\text{noise}}(\sigma)\big) \tag{10}$$

Where (c_{skip},c_{out},c_{in},c_{noise}) are noise-level-dependent scaling coefficients [32], σ is the noise level. We use the same hyperparameters for the scaling coefficients as in Karras et al. [32], but sample our noise level $p(\sigma)$ from a lognormal distribution with different parameters (see table 7).

Sampling Once the denoiser has been sufficiently trained, we define a specific noise level schedule governing the reverse process, in which the initial noise level, σ, starts at σ_{max} and decreases to σ_{min}:

$$\sigma_i = \left(\sigma_{\text{max}}^{\frac{1}{\rho}} + \frac{i}{N-1}(\sigma_{\text{min}}^{\frac{1}{\rho}} - \sigma_{\text{max}}^{\frac{1}{\rho}})\right)^{\rho} \tag{11}$$

where ρ, σ_{min} and σ_{max} are hyperparameters defined in table 7. We specifically use the stochastic sampling method proposed in Karras et al. [32] (see Tab. 7 for hyperparameters).

Table 7. Diffusion model hyperparameters

Training	**Value**
lr	2.5×10^{-5}
Epochs	50
Batch Size	1
Num. Channels	[64, 128, 196]
Num. Res. Blocks	2
Num. Attn. Heads	1
Attn. Res.	8, 4, 2
σ_{data}	1
$p(\sigma)$ mean	1
$p(\sigma)$ std	1.2
Sampling	**Value**
σ_{min}	1×10^{-2}
σ_{max}	80
ρ	1

7.3. Substructure Parsing

Anatomica revolves around parsing substructures through the use of selection vectors and control domains, enabling the measurement of anatomical properties for specified tissues within localized regions of interest. Selection vectors are binary vectors $\mathbf{u} \in \{0,1\}^C$ which select a subset of tissues from a voxel grid $\mathbf{V}$ through the Boolean subset operator $\mathcal{U}[\mathbf{u}]$ (see Algorithm 1). By varying the Boolean selection vector, we enable the measurement of anatomic structures that are composed of multiple tissue types. Control domains are instantiated as point grids $\mathbf{X} \in \mathbb{R}^{\alpha\times\beta\times\gamma\times 3}$ that are used to parse substructures from grid-like representations such as voxel grids $\mathbf{V}$ using the voxel slicing operator $\mathcal{T}^s[\mathbf{X}]$ with **V-parsing** (see Algorithm 2). Alternatively, we can parse substructures directly from the latent representation $\mathbf{z}$ using the latent slicing operator $\mathcal{T}^l[\mathbf{X}]$ with **L-parsing** (see Algorithm 3). To obtain control domains, we first define a template domain $\mathbf{X}^{\text{temp}} \in \mathbb{R}^{\alpha\times\beta\times\gamma\times 3}$ as a point grid centered at $\mathbf{0}$, with a grid size $\mathbf{g} = [\alpha, \beta, \gamma]$. We then apply a spatial transformation defined by affine transformation parameters $\mathbf{A} = [\mathbf{R}, \mathbf{s}, \mathbf{t}]$ to obtain anatomically relevant control domains.

Algorithm 1 Boolean Subset Operator

Require: $\mathbf{V} \in \mathbb{R}^{C\times H\times W\times D}$ ▷ Voxel map
Require: $\mathbf{u} \in \{0,1\}^C$ ▷ Boolean selection vector
$\hat{\mathbf{S}} \leftarrow \mathbf{0}$ ▷ Initialize
for tissue i where $\mathbf{u}_i = 1$ **do**
  $\hat{\mathbf{S}} \leftarrow \max(\hat{\mathbf{S}}, \mathbf{V}_i)$ ▷ Union via maximum
end for
return $\hat{\mathbf{S}} \in \mathbb{R}^{H\times W\times D}$

Algorithm 2 Voxel Substructure Parsing (V-parsing)

Require: $\mathbf{V} \in \mathbb{R}^{C\times H\times W\times D}$ ▷ Voxel map
Require: $\mathbf{u} \in \{0,1\}^C$ ▷ Selection vector
Require: $\{\mathbf{X}_k\}_{k=1}^K$ where $\mathbf{X}_k \in \mathbb{R}^{\alpha\times\beta\times\gamma\times 3}$ ▷ Control domains
Subset Tissues
$\hat{\mathbf{S}} \leftarrow \mathcal{U}[\mathbf{u}](\mathbf{V}) \in \mathbb{R}^{H\times W\times D}$ ▷ Boolean subset
Parse Substructures
for $k = 1, \ldots, K$ **do**
  $\mathbf{S}_k \leftarrow \mathcal{T}^s[\mathbf{X}_k](\hat{\mathbf{S}}) \in \mathbb{R}^{\alpha\times\beta\times\gamma}$ ▷ Voxel slice
end for
return $\{\mathbf{S}_k\}_{k=1}^K$

Algorithm 3 Latent Substructure Parsing (L-parsing)

Require: $\mathbf{z} \in \mathbb{R}^{c\times h\times w\times d}$ ▷ Latent representation
Require: $\mathbf{u} \in \{0,1\}^C$ ▷ Selection vector
Require: $\{\mathbf{X}_k\}_{k=1}^K$ where $\mathbf{X}_k \in \mathbb{R}^{\alpha\times\beta\times\gamma\times 3}$ ▷ Control domains
Parse Substructures
for $k = 1, \ldots, K$ **do**
  $\mathbf{z}_k \leftarrow \mathcal{T}^l[\mathbf{X}_k](\mathbf{z}) \in \mathbb{R}^{c\times\alpha\times\beta\times\gamma}$ ▷ Latent slice
  $\mathbf{S}_k \leftarrow \mathcal{U}[\mathbf{u}] \circ \mathcal{F}[\mathbf{X}_k](\mathbf{z}_k) \in \mathbb{R}^{\alpha\times\beta\times\gamma}$ ▷ Decode & subset
end for
return $\{\mathbf{S}_k\}_{k=1}^K$

7.4. Control Domains

Anatomica supports several methods for defining control domains $\mathbf{X}_k$, each useful for probing different anatomical or geometric properties. In this study, we primarily compute control domain parameters from real anatomical voxel maps and measure geometric properties within such domains to define targets for diffusion guidance. This approach is not limited to guidance use-cases, and can potentially be used

for other machine-learning tasks that use differentiable loss functions. We now detail the algorithmic procedures for computing control domain parameters across different coordinate systems from anatomical voxel maps.

Global Domain Computation The global control domain can be used to measure properties over the entire voxel grid without geometric feature extraction at a variable spatial resolution. We compute global domains through Algorithm 4.

Algorithm 4 Global Domain Computation

Require: (α, β, γ) with $\alpha \approx \beta \approx \gamma$ ▷ Volumetric grid size

Set Affine Parameters
$\mathbf{R} \leftarrow \mathbf{I} \in \mathbb{R}^{3\times3}$ ▷ Rotation
$\mathbf{t} \leftarrow \mathbf{0} \in \mathbb{R}^3$ ▷ Translation
$\mathbf{s} \leftarrow \mathbf{1} \in \mathbb{R}^3$ ▷ Scale
return $\mathbf{A} = [\mathbf{R}, \mathbf{s}, \mathbf{t}]$

Cartesian Domain Computation Cartesian domains enable the measurement of anatomical properties within localized bounding boxes that contain structures of interest. We compute Cartesian domains through Algorithm 5.

Algorithm 5 Cartesian Domain Computation

Require: $\mathbf{u} \in \{0,1\}^C$ ▷ Tissue selection vector
Require: (α, β, γ) with $\alpha \approx \beta \approx \gamma$ ▷ Volumetric grid size

Extract Bounding Box
$\hat{\mathbf{S}} \leftarrow \mathcal{U}[\mathbf{u}](\mathbf{V}), \tilde{\mathbf{S}} \leftarrow \mathbb{I}[\hat{\mathbf{S}} > 0.9]$ ▷ Subset & binarize
$\mathbf{r}^{\text{upper}}, \mathbf{r}^{\text{lower}} \leftarrow \text{ExtractLimits}(\tilde{\mathbf{S}})$ where $\mathbf{r}^{\text{upper}}, \mathbf{r}^{\text{lower}} \in \mathbb{R}^3$
Set Affine Parameters
$\mathbf{R} \leftarrow \mathbf{I} \in \mathbb{R}^{3\times3}$ ▷ Rotation
$\mathbf{t} \leftarrow (\mathbf{r}^{\text{upper}} + \mathbf{r}^{\text{lower}})/2 \in \mathbb{R}^3$ ▷ Translation
$\mathbf{s} \leftarrow (\mathbf{r}^{\text{upper}} - \mathbf{r}^{\text{lower}}) \oslash [\alpha, \beta, \gamma]^T \in \mathbb{R}^3$ ▷ Scale
return $\mathbf{A} = [\mathbf{R}, \mathbf{s}, \mathbf{t}]$

Interface Domain Computation Interface domains enable the measurement of local anatomical properties at the interfacial region between two or more structures, such as valve annuli or branch points. We compute interface domains through Algorithm 6.

Algorithm 6 Interface Domain Computation

Require: $\mathbf{u}^A, \mathbf{u}^B \in \{0,1\}^C$ ▷ Tissue selection vectors
Require: (α, β, γ) with $\alpha \ll \beta \approx \gamma$ ▷ Planar grid size
Require: $k_{\text{dil}}, \mathbf{R}^r \in \mathbb{R}^{3\times1}$ ▷ Kernel size, ref vector

Extract Interface Regions
$\hat{\mathbf{S}}^A \leftarrow \mathcal{U}[\mathbf{u}^A](\mathbf{V}), \hat{\mathbf{S}}^B \leftarrow \mathcal{U}[\mathbf{u}^B](\mathbf{V})$ ▷ Subset
$\hat{\mathbf{S}}^A_{\text{dil}} \leftarrow \text{maxpool}_{k_{\text{dil}}}(\hat{\mathbf{S}}^A), \hat{\mathbf{S}}^B_{\text{dil}} \leftarrow \text{maxpool}_{k_{\text{dil}}}(\hat{\mathbf{S}}^B)$ ▷ Dilate
$\mathbf{M} \leftarrow \min(\hat{\mathbf{S}}^A_{\text{dil}}, \hat{\mathbf{S}}^B_{\text{dil}})$ ▷ Combine
$\hat{\mathbf{S}}^A_{\text{int}} \leftarrow \hat{\mathbf{S}}^A \odot \mathbf{M}, \hat{\mathbf{S}}^B_{\text{int}} \leftarrow \hat{\mathbf{S}}^B \odot \mathbf{M}$ ▷ Mask interface
Compute Interface Frame Orientations
$\mathbf{p}^A, \mathbf{p}^B \leftarrow \text{Centroid}(\hat{\mathbf{S}}^A_{\text{int}}), \text{Centroid}(\hat{\mathbf{S}}^B_{\text{int}})$ ▷ (Alg. 12)
$\mathbf{R}^\alpha \leftarrow (\mathbf{p}^B - \mathbf{p}^A)/\|\mathbf{p}^B - \mathbf{p}^A\| \in \mathbb{R}^{3\times1}$ ▷ Interface vector
$\mathbf{R}^\beta, \mathbf{R}^\gamma \leftarrow \text{Orthonorm.}(\mathbf{R}^\alpha, \mathbf{R}^r) \in \mathbb{R}^{3\times1}$ ▷ (Alg. 11)
Set Affine Parameters
$\mathbf{R} \leftarrow [\mathbf{R}^\alpha, \mathbf{R}^\beta, \mathbf{R}^\gamma] \in \mathbb{R}^{3\times3}$ ▷ Rotation
$\mathbf{s} \leftarrow [\alpha/H, \beta/W, \gamma/D]^T \in \mathbb{R}^3$ ▷ Scale
$\mathbf{t}^A \leftarrow \mathbf{p}^A \in \mathbb{R}^3, \mathbf{t}^B \leftarrow \mathbf{p}^B \in \mathbb{R}^{3\times1}$ ▷ Translation
return $\mathbf{A}^A = [\mathbf{R}, \mathbf{s}, \mathbf{t}^A], \mathbf{A}^B = [\mathbf{R}, \mathbf{s}, \mathbf{t}^B]$

Curvilinear Domain Computation Curvilinear domains enable the measurement of cross-sectional anatomical properties along tubular structures such as blood vessels. We compute curvilinear domains through Algorithm 7. For skeletonization, we follow the methods and hyperparameters detailed in Kadry et al. [28] for non-differentiable hard skeletonization.

Algorithm 7 Curvilinear Domain Computation

Require: $\mathbf{u} \in \{0,1\}^C$ ▷ Tissue selection vector
Require: (α, β, γ) with $\alpha \ll \beta \approx \gamma$ ▷ Planar grid size
Require: $\mathbf{i}_{\text{sub}}, \mathbf{R}^r \in \mathbb{R}^{3\times1}$ ▷ Subsampling Indices, Ref vector

Extract Centerline
$\hat{\mathbf{S}} \leftarrow \mathcal{U}[\mathbf{u}](\mathbf{V}), \tilde{\mathbf{S}} \leftarrow \mathbb{I}[\hat{\mathbf{S}} > 0.9]$ ▷ Subset & binarize
$\mathbf{C} \leftarrow \text{Skeletonize}(\tilde{\mathbf{S}})$ where $\mathbf{C} \in \mathbb{R}^{N_{\text{center}}\times3}$
Compute Curvilinear Frames
$\mathbf{F}^\alpha \leftarrow \text{FiniteDifference}(\mathbf{C}) \in \mathbb{R}^{N_{\text{center}}\times3}$ ▷ Tangent vectors
$\mathbf{F}^\beta_0, \mathbf{F}^\gamma_0 \leftarrow \text{Orthonorm.}(\mathbf{F}^\alpha_0, \mathbf{R}^r)$ ▷ (Alg. 11)
$\mathbf{F}^\beta, \mathbf{F}^\gamma \leftarrow \text{ParallelTransport}(\mathbf{F}^\alpha, \mathbf{F}^\beta_0, \mathbf{F}^\gamma_0)$ ▷ (Alg. 10)
$\mathbf{C}^{\text{sub}} \leftarrow \text{Subsample}(\mathbf{C}, \mathbf{i}_{\text{sub}})$ where $\mathbf{C}^{\text{sub}} \in \mathbb{R}^{N_{\text{planes}}\times3}$
$\mathbf{R}^\alpha, \mathbf{R}^\beta, \mathbf{R}^\gamma \leftarrow \text{Subsample}(\mathbf{F}^\alpha, \mathbf{F}^\beta, \mathbf{F}^\gamma, \mathbf{i}_{\text{sub}}) \in \mathbb{R}^{N_{\text{planes}}\times3}$
Set Affine Parameters
for domain $k = 1, \ldots, N_{\text{planes}}$ **do**
  $\mathbf{R}_k \leftarrow [\mathbf{R}^\alpha_k, \mathbf{R}^\beta_k, \mathbf{R}^\gamma_k] \in \mathbb{R}^{3\times3}$ ▷ Rotation
  $\mathbf{s}_k \leftarrow [\alpha/H, \beta/W, \gamma/D]^T \in \mathbb{R}^3$ ▷ Scale
  $\mathbf{t}_k \leftarrow \mathbf{C}^{\text{sub}}_k \in \mathbb{R}^3$ ▷ Translation
end for
return $\{\mathbf{A}_k\}_{k=1}^{N_{\text{planes}}}$

Spherical Domain Computation Spherical domains enable the measurement of radial anatomical properties of shell-like structures such as myocardial walls. We compute spherical domains through Algorithm 8. Instead of sam-

pling equidistant points in polar and azimuthal space, we compute equally distributed points on the sphere surface using the Fibonacci lattice algorithm [18].

Algorithm 8 Spherical Domain Computation

Require: $\mathbf{u} \in \{0,1\}^C$ ▷ Tissue selection vector
Require: (α, β, γ) with $\alpha \approx \beta \ll \gamma$ ▷ Ray-like grid size
Require: N_{rays}, N_q, $\mathbf{R}^r \in \mathbb{R}^{3\times 1}$ ▷ Number of rays, query ray resolution, ref vector

Generate Radial Directions
$\hat{\mathbf{S}} \leftarrow \mathcal{U}[\mathbf{u}](\mathbf{V})$ ▷ Subset tissues
$\mathbf{p} \leftarrow \text{Centroid}(\hat{\mathbf{S}}) \in \mathbb{R}^3$ ▷ (Alg. 12)
$\mathbf{R}^\gamma \leftarrow \text{FibonacciLattice}(N_{\text{rays}}, \mathbf{p})$ where $\mathbf{R}^\gamma \in \mathbb{R}^{N_{\text{rays}}\times 3}$
$\mathbf{R}^\beta, \mathbf{R}^\alpha \leftarrow \text{Orthonorm.}(\mathbf{R}^\gamma, \mathbf{R}^r) \in \mathbb{R}^{N_{\text{rays}}\times 3}$ ▷ (Alg. 11)
Find Wall Centroids and Set Affine Parameters
for domain $k = 1, \ldots, N_{\text{rays}}$ **do**
  $\mathbf{X}_k^{\text{ray}} \leftarrow \text{MakeQueryRay}(\mathbf{p}, \mathbf{R}_k^\gamma, N_q)$ where $\mathbf{X}_k^{\text{ray}} \in \mathbb{R}^{1\times 1\times N_q\times 3}$
  $\mathbf{S}_k^{\text{ray}} \leftarrow \mathcal{T}^s[\mathbf{X}_k^{\text{ray}}](\hat{\mathbf{S}})$ ▷ Slice along ray
  $\mathbf{p}_{\text{wall},k} \leftarrow \text{Centroid}(\mathbf{S}_k^{\text{ray}}) \in \mathbb{R}^3$ ▷ (Alg. 12)
  $\mathbf{R}_k \leftarrow [\mathbf{R}_k^\alpha, \mathbf{R}_k^\beta, \mathbf{R}_k^\gamma] \in \mathbb{R}^{3\times 3}$ ▷ Rotation
  $\mathbf{s}_k \leftarrow [\alpha/H, \beta/W, \gamma/D]^T \in \mathbb{R}^3$ ▷ Scale
  $\mathbf{t}_k \leftarrow \mathbf{p}_{\text{wall},k} \in \mathbb{R}^3$ ▷ Translation
end for
return $\{\mathbf{A}_k\}_{k=1}^{N_{\text{rays}}}$

Cylindrical Domain Computation Cylindrical domains enable the measurement of radial anatomical properties of walled tubular structures such as coronary arteries. We compute cylindrical domains through Algorithm 9. We compute cylindrical domains by defining equidistant ray centers along the z-axis and equally sampling the polar directions according to predefined sampling resolutions.

Algorithm 9 Cylindrical Domain Computation

Require: $\mathbf{u} \in \{0,1\}^C$ ▷ Tissue selection vector
Require: (α, β, γ) with $\alpha \approx \beta \ll \gamma$ ▷ Ray-like grid size
Require: N_z, N_θ, N_q, $\mathbf{R}^r \in \mathbb{R}^{3\times 1}$ ▷ Z-levels, angles, query ray resolution, ref vector

Generate Cylindrical Directions
$\hat{\mathbf{S}} \leftarrow \mathcal{U}[\mathbf{u}](\mathbf{V})$ ▷ Subset tissues
$\mathbf{R}^\gamma \leftarrow \text{CylindricalLattice}(N_z, N_\theta)$ where $\mathbf{R}^\gamma \in \mathbb{R}^{N_{\text{rays}}\times 3}$, $N_{\text{rays}} = N_z \times N_\theta$
$\mathbf{R}^\beta, \mathbf{R}^\alpha \leftarrow \text{Orthonorm.}(\mathbf{R}^\gamma, \mathbf{R}^r) \in \mathbb{R}^{N_{\text{rays}}\times 3}$ ▷ (Alg. 11)
Find Wall Centroids and Set Affine Parameters
for domain $k = 1, \ldots, N_{\text{rays}}$ **do**
  $\mathbf{X}_k^{\text{ray}} \leftarrow \text{MakeQueryRay}(\mathbf{R}_k^\gamma, N_q)$ where $\mathbf{X}_k^{\text{ray}} \in \mathbb{R}^{1\times 1\times N_q\times 3}$
  $\mathbf{S}_k^{\text{ray}} \leftarrow \mathcal{T}^s[\mathbf{X}_k^{\text{ray}}](\hat{\mathbf{S}})$ ▷ Slice along ray
  $\mathbf{p}_{\text{wall},k} \leftarrow \text{Centroid}(\mathbf{S}_k^{\text{ray}}) \in \mathbb{R}^3$ ▷ (Alg. 12)
  $\mathbf{R}_k \leftarrow [\mathbf{R}_k^\alpha, \mathbf{R}_k^\beta, \mathbf{R}_k^\gamma] \in \mathbb{R}^{3\times 3}$ ▷ Rotation
  $\mathbf{s}_k \leftarrow [\alpha/H, \beta/W, \gamma/D]^T \in \mathbb{R}^3$ ▷ Scale
  $\mathbf{t}_k \leftarrow \mathbf{p}_{\text{wall},k} \in \mathbb{R}^3$ ▷ Translation
end for
return $\{\mathbf{A}_k\}_{k=1}^{N_{\text{rays}}}$

Parallel Transport Procedure For curvilinear coordinate systems, we aim to maintain consistent frame orientations as we move along the centerline. To do this, we apply parallel transport by propagating an initial orthonormal frame along a centerline using the Rodrigues rotation formula.

Algorithm 10 ParallelTransport

Require: $\mathbf{F}^1 \in \mathbb{R}^{N_{\text{center}}\times 3}$ ▷ Normalized tangent vectors
Require: $\mathbf{F}_0^2, \mathbf{F}_0^3 \in \mathbb{R}^3$ ▷ Initial normalized frame vectors
for $i = 1, \ldots, N_{\text{center}} - 1$ **do**
  $\mathbf{a}_i \leftarrow (\mathbf{F}_{i-1}^1 \times \mathbf{F}_i^1)/\|\mathbf{F}_{i-1}^1 \times \mathbf{F}_i^1\|$ ▷ Rotation axis
  $\theta_i \leftarrow \cos^{-1}(\mathbf{F}_{i-1}^1 \cdot \mathbf{F}_i^1)$ ▷ Rotation angle
  $\mathbf{F}_i^2, \mathbf{F}_i^3 \leftarrow \text{Rodrigues}(\mathbf{F}_{i-1}^2, \mathbf{F}_{i-1}^3, \mathbf{a}_i, \theta_i)$
end for
return $\mathbf{F}^2, \mathbf{F}^3 \in \mathbb{R}^{N_{\text{center}}\times 3}$

Orthonormalization Procedure For interface, curvilinear, spherical, and cylindrical coordinate systems, we wish to compute a set of orthonormal frame vectors from an initial vector. To do this, we define an arbitrary reference vector $\mathbf{R}^r$ and compute orthonormal frame vectors from a primary direction vector by taking successive cross products. For numerical stability, we use an alternate reference vector if the reference and initial vectors are perfectly aligned.

Algorithm 11 Orthonormalization

Require: $\mathbf{U}^0 \in \mathbb{R}^{3\times 1}$ ▷ Primary direction vector
Require: $\mathbf{R}^r \in \mathbb{R}^{3\times 1}$ ▷ Reference vector
$\mathbf{U}^1 \leftarrow (\mathbf{U}^0 \times \mathbf{R}^r)/\|\mathbf{U}^0 \times \mathbf{R}^r\| \in \mathbb{R}^{3\times 1}$ ▷ Second frame vector
$\mathbf{U}^2 \leftarrow \mathbf{U}^0 \times \mathbf{U}^1 \in \mathbb{R}^{3\times 1}$ ▷ Third frame vector (auto-normalized)
return $\mathbf{U}^1, \mathbf{U}^2$

7.5. Geometric Measurement & Guidance

Scale Standardization of Mass We normalize the measured mass m_k by the total number of voxels in the control domain $\alpha\beta\gamma$ in order to remain invariant to control domain discretization, allowing us to maintain similar geometric loss weightings across different discretization levels.

Local to Global Transformation of Moments Our geometric moment formulation can be sensitive to control domain discretization and coordinate system choice. For example, our substructure can inhabit 80% of the control domain, but the control domain may be a small region within the global domain, resulting in a large measured mass m_k. Another example would be the case of a localized control domain with a measured centroid $\mathbf{p}_k$ that is measured to be in the center of the control domain, but is at the periphery of the global domain. We therefore aim to express our geometric measurements in a manner that is invariant to control domain choice. This is important when applying MSE-based geometric loss functions across different tasks due to varying scales. Therefore, we express all geometric moments in the global coordinate system using the inverse of the control domain transformation parameters $\mathbf{A}_k = [\mathbf{R}_k, \mathbf{s}_k, \mathbf{t}_k]$.

Stabilizing Covariance Normalization As we normalize the covariance matrix by the trace, we stabilize the gradient in the case of empty substructures by adding a small epsilon (1e-9) to the diagonal of the covariance matrix.

Adaptive Mass Weighting To avoid centroid and covariance gradient explosion in the case of near-empty segmentations, we adaptively weight the centroid and covariance losses by the mass of the substructure. Below a predefined threshold, we set the centroid and covariance weightings $\lambda_1 = \lambda_2 = 0$. This mass threshold is determined on a task-by-task basis, where we multiply the average measured mass for the task by a factor of 0.1.

Algorithm 12 Geometric Measurement

Require: $\mathbf{S}_k \in \mathbb{R}^{\alpha\times\beta\times\gamma}$ ▷ Substructure
Require: $\mathbf{R}_k \in \mathbb{R}^{3\times 3}$ ▷ Rotation matrix
Require: $\mathbf{s}_k \in \mathbb{R}^3$ ▷ Scale vector
Require: $\mathbf{t}_k \in \mathbb{R}^3$ ▷ Translation vector
Compute Local Moments
$m_k^{\text{raw}} \leftarrow \text{ComputeMass}(\mathbf{S}_k)$ ▷ (Eq. 3)
$\mathbf{p}_k^{\text{local}} \leftarrow \text{ComputeCentroid}(\mathbf{S}_k, m_k^{\text{raw}})$ ▷ (Eq. 3)
$\mathbf{\Sigma}_k^{\text{local}} \leftarrow \text{ComputeCovariance}(\mathbf{S}_k, \mathbf{p}_k^{\text{local}}, m_k^{\text{raw}})$ ▷ (Eq. 3)
$m_k^{\text{local}} \leftarrow m_k^{\text{raw}}/(\alpha\beta\gamma)$ ▷ Normalize by voxel count
Local to Global Transformation
$\mathbf{J}_k \leftarrow \mathbf{R}_k \,\text{diag}(\mathbf{s}_k)$ ▷ Rotation-scale matrix
$m_k^{\text{global}} \leftarrow m_k^{\text{local}} \cdot |\det(\mathbf{J}_k)|$ ▷ Transform mass
$\mathbf{d}_k^{\text{local}} \leftarrow \mathbf{p}_k^{\text{local}} - \frac{1}{2}\mathbf{1}$ ▷ Local displacement from center
$\mathbf{d}_k^{\text{global}} \leftarrow \mathbf{J}_k \mathbf{d}_k^{\text{local}}$ ▷ Transform displacement
$\mathbf{p}_k^{\text{global}} \leftarrow \mathbf{t}_k + \mathbf{d}_k^{\text{global}}$ ▷ Transform centroid
$\mathbf{\Sigma}_k^{\text{global}} \leftarrow \mathbf{J}_k \mathbf{\Sigma}_k^{\text{local}} \mathbf{J}_k^T$ ▷ Transform covariance
return $(m_k^{\text{global}}, \mathbf{p}_k^{\text{global}}, \mathbf{\Sigma}_k^{\text{global}})$

7.6. Topological Measurement & Guidance

We partition the persistence set into disjoint sets $\mathcal{Y}_k$ and $\mathcal{Z}_k$ consisting of points that should be preserved or suppressed based on a topological prior $\mathcal{B}_k \in \mathbb{R}^3 = [\mathbf{B}0, \mathbf{B}1, \mathbf{B}2]$, which specifies the desired features for the components, loops, and voids, respectively. For each dimension, we sort the points by persistence and select the top $\mathcal{B}_i$ points for each dimension i specified by the prior.

7.7. Parallelization

For our geometric measurement operations, we take advantage of parallel GPU computation. We parallelize across different batch indices, constraints, and substructures. When computing control domains, some domain types allow for invalid domains, such as in the case of spherical ray domains, where the ray may not intersect with the substructure. We handle these invalid domains by masking out the computed loss. The only exception is our skeletonization step for curvilinear control domains, as our implementation is computed on a CPU. For topological measurement, we do not parallelize the persistent homology computation as no GPU-compatible implementation is publicly available, and CPU-parallelization over several cores did not provide significant speedups.

8. Experimental Details

8.1. Baselines

Explicit Conditioning To ensure that the elements of $\mathcal{G}\text{exp}$ are roughly between 0 and 1, we min-max normalize the masses m_k, centroids $\mathbf{p}_k$, and normalized covariances $\mathbf{\Sigma}_k^n$ with values calculated from the real dataset (Tab. 8). The LDM input channel count is increased to accommodate the

Table 8. Normalizing constants for geometric moments during explicit conditioning across different tasks.

	Geometric Control Task			
Parameter	**RV**	**Mitral**	**Aortic**	**Myo**
Mass Min m_k	3.19×10^{-3}	3.67×10^{-4}	0	0
Mass Max m_k	1.3×10^{-2}	1.36×10^{-3}	8.59×10^{-4}	1.95×10^{-5}
Centroid Min $\mathbf{p}_k$	0	0	-7.81×10^{-3}	0
Centroid Max $\mathbf{p}_k$	1	1	0.91	0.64
Covariance Min $\mathbf{\Sigma}_k$	-1×10^{-4}	-8.66×10^{-4}	-5.59×10^{-4}	2.88×10^{-4}
Covariance Max $\mathbf{\Sigma}_k$	1×10^{-2}	2.34×10^{-3}	1.56×10^{-3}	8.03×10^{-4}

concatenated input. This method does not readily permit the use of dropout to train a diffusion model in an unconditional manner because the null condition is defined as zero, which is equivalent to the minimum moment values.

Implicit Conditioning To compute the ellipsoidal distance map, we use the centroids $\mathbf{p}_k$ and non-normalized covariances $\mathbf{\Sigma}_k$ for each component to compute the Mahalanobis distance [8] for each voxel position. We then apply a shifted sigmoid transform to constrain the outputs between 0 and 1, and subsequently concatenate the resulting grid to the latents. To enable unconditional generation, we randomly drop out each substructure channel of $\mathcal{G}$imp with a probability of 0.1.

8.2. Datasets

Cardiac Dataset For our study, we utilize TotalSegmentator v2 [46] to create the cardiac segmentations, with 596 3D segmentations manually selected based on segmentation quality assessment. Cardiac structures include the myocardium (Myo), left and right atria (LA & RA), left and right ventricles (LV & RV), aorta (Ao), and pulmonary artery (PA), were segmented using a specialized TotalSegmentator model trained on sub-millimeter resolution data. For the inferior vena cava (IVC), superior vena cava (SVC), and pulmonary veins (PV), we retain the labels from the original dataset. This results in 11 channels per segmentation. To ensure anatomical validity, we perform topological filtration on all structures except the pulmonary veins, where we extract only the largest connected component. The resulting segmentations are standardized by resampling to a uniform voxel resolution of 2mm and subsequently cropped to a fixed range. The crop center is determined from the union of all four chamber segmentations, and the crop length is set to 128 voxels for each side.

Aortic Dataset For the aorta dataset, we extract labels directly from the original TotalSegmentator v2 [46] segmentations, without applying a specialized model, resulting in 450 3D segmentations manually selected based on segmentation quality assessment. The labels include the main aortic trunk and the ascending branches, which comprise the brachiocephalic trunk (BCT), left common carotid artery (LCCA), right common carotid artery (RCCA), left subclavian artery (LSCA), and right subclavian artery (RSCA), for a total of 7 channels per segmentation. All segmentations are resampled to an isotropic voxel size of 2 mm and cropped to a spatial size of 128^3 using a crop center determined from the center of all combined tissues.

Table 9. Task-specific hyperparameters and configurations for geometric control tasks.

	Geometric Control Task			
Parameter	**RV**	**Mitral**	**Aortic**	**Myo**
Domain	Cartesian	Interface	Curvilinear	Spherical
Selection Vector	[RV]	[LV], [LA]	[Ao]	[Myo]
Num. Substructures	1	2	5	4
Grid Resolution	[64,64,64]	[4,32,32]	[1,32,32]	[4,4,16]
Mass Threshold	10^{-5}	10^{-4}	10^{-6}	10^{-6}
λ_{geo}	1	1	1	1
λ_0 (Mass)	10^7	10^9	10^9	10^9
λ_1 (Centroid)	10^5	10^6	10^5	10^5
λ_2 (Covariance)	10^4	10^4	10^3	10^4

Spinal Dataset For the spinal dataset, we utilize the CT-Spine1K dataset [10] and extract all vertebral body segmentations, resulting in 784 3D segmentations. The segmentations include 7 cervical vertebrae (C1–C7), 12 thoracic vertebrae (T1–T12), and 5 lumbar vertebrae (L1–L5), for a total of 25 channels per segmentation. To ensure spatial consistency and anatomical completeness, all segmentations are first resampled to an isotropic voxel spacing of $1\,\text{mm}$. The center of the crop box is determined from the union (voxelwise sum) of all vertebral structures in each scan, and a fixed crop of 128^3 voxels is applied for each case.

Coronary Dataset For the coronary dataset, we extract coronary artery-related labels from the DISRUPT-CAD dataset [45], consisting of 120 patients with approximately 375 OCT frames in the longitudinal (z) direction. The segmentations include lumen (Lu), calcium (Ca), and vessel wall (Ve), for a total of 4 channels per segmentation. Training samples are generated by resampling the x and y directions to 128×128 pixels while preserving the original z resolution, then randomly cropping 128 consecutive frames along the z-axis from each patient scan. This yields approximately 360 unique 3D segmentations of size 128^3 with an isotropic in-plane voxel spacing of approximately $0.1\,\text{mm}$.

8.3. Tasks

Geometric Control Tasks We detail the task-specific hyperparameters and configurations for the geometric control tasks in Tab. 9.

Topological Control Tasks We detail the task-specific hyperparameters and configurations for the topological control

Table 10. Task-specific hyperparameters and configurations for topological control tasks.

Parameter	Topological Control Task			
	Atrial Separation	Branch Connectivity	Vert. Connectivity	Calcium Count
Domain	Global	Global	Global	Global
Selection Vector	[LA, RA]	All Tissues	[T6–T10]	[Ca]
Num. Substructures	1	1	1	1
Grid Resolution	[64,64,64]	[64,64,64]	[64,64,64]	[64,64,64]
Softmax Value	4	4	4	4
λ_{topo}	5	1	5	50
Prior B0	2	1	1	2
Prior B1	0	0	9	0
Prior B2	0	0	0	0

Table 11. Task-specific hyperparameters and configurations for multiscale control tasks. Hyperparameters marked with a slash / indicate smaller and larger domain configurations, respectively.

Parameter	Spinal	Aorta	Myo Wall	Vessel Wall
Domain	Cartesian	Curvilinear	Spherical	Cylindrical
Selection Vector	[T5–T10]/[T6–T8]	[Ao]	[Myo]	[Ca, Ve]
Num. Substructures	1	5	16	16
Grid Resolution	[64,64,64]	[1,32,32]	[1,32,32]	[1,32,32]
Mass Threshold	10^{-4}	10^{-6}	10^{-6}	10^{-6}
λ_{geo}	1	1	1	1
λ_0 (Mass)	10^7	10^9	10^9	10^9
λ_1 (Centroid)	10^5	10^5	10^5	10^5
λ_2 (Covariance)	10^4	10^3	10^4	10^4
Domain Grid	[64,64,64]	[1,16,16]/[16,16,16]	[4,4,16]/[8,8,16]	[4,4,32]/[16,16,32]

tasks in Tab. 10.

Multiscale Control We detail the task-specific hyperparameters and configurations for the multiscale control tasks in Tab. 11. For the spinal task, we achieve multiscale control by changing the selection vector to include fewer or more vertebral bodies. For all other tasks, we change the control domain grid resolution along specified axes.

Partial Decoding For the partial decoding experiments, we used a Cartesian domain with different resolutions. For Anatomica-L, both coarse and local L-parsing used grid resolutions of $[32, 32, 32]$, $[64, 64, 64]$, and $[128, 128, 128]$ for low, medium, and high resolutions respectively. For Anatomica-V, we used global decoding with a fixed resolution of $[128, 128, 128]$. We measured speed in terms of the maximum number of label maps sampled per second using the maximum allowable batch size on a single GPU. We used an A100 with 40 GB of memory for benchmarking. For geometric guidance, the wall clock time was approximately 50 seconds per sample for the highest decoding resolution with a convolutional decoder.

8.4. Evaluation

Frechet Morphological Distance To compute the morphological features, the features are normalized by the mean and standard deviation of the real data.

Pointcloud evaluation metrics: To compute the point cloud metrics, we calculate NNA for every tissue label using 256 points sampled using farthest point sampling. The metric is then averaged over the number of components. To compute the pointcloud distances, we approximate Earth Mover's Distance (EMD) through the Sinkhorn divergence [17].

Topological Precision To compute the Betti numbers, we take the argmax of the predicted segmentation and compute persistent homology. For a binary segmentation, the barcodes are 1 or 0 depending on the existence of the structure. We then take the sum of barcodes per dimension as the Betti number. The topological precision is then the fraction of samples with the correct Betti number per dimension.

9. Ablation Studies

9.1. Geometric Guidance Ablations

We aim to study the influence of individual geometric loss weightings on the geometric fidelity and generation quality. We specifically examine the influence of *disentangled* geometric guidance, where, for example, we only constrain the centroid but let size and shape free to vary. To do this, we sweep over the composite geometric loss weighting λ_{geo} for all tasks, and apply different combinations of loss weightings $[\lambda_0, \lambda_1, \lambda_2]$ to activate or deactivate different geometric loss terms (see Tab. 12). We sample 128 samples for each experiment, with 100 sampling steps.

Effect of Guidance Weight In Fig. 8, we see that increasing geometric guidance weight when all loss weightings are activated (Full) improves geometric fidelity up to a certain weight, after which sample quality degrades, decreasing geometric fidelity. This is especially pronounced in the case of centroid-only guidance for the mitral valve and myocardium wall tasks. For generation quality, we see similar trends where increasing guidance weights can reduce FMD up to a certain guidance weight.

Effect of Disentangled Guidance In Fig. 8, we demonstrate that our framework supports disentangled geometric guidance across all tasks. For instance, centroid-only guidance achieves centroid fidelity comparable to full guidance, without significantly affecting mass fidelity, shape fidelity, or generation quality as measured by FMD.

9.2. Topological Guidance Ablations

We aim to study the influence of topological loss weightings, softmax temperature, and partial decoding strategy on topological fidelity. We first sample 64 segmentations for several combinations of guidance weight and softmax temperature and evaluate topological fidelity for every combination (Fig. 9). We then sample 128 samples for various coarse decoding resolutions and guidance weights while

Table 12. Loss weight configurations for geometric guidance ablation study.

Guidance Loss	λ_0	λ_1	λ_2
Full	✓	✓	✓
Mass Only	✓	×	×
Centroid Only	×	✓	×
Covariance Only	×	×	✓

evaluating topological fidelity (Fig. 10) and sampling speed (Tab. 13).

Effect of Guidance Weight We see in Fig. 9 that increasing guidance weights broadly improves topological fidelity but can decrease fidelity with extreme guidance weights.

Effect of Softmax Temperature Similarly, in Fig. 9, we see that increasing softmax temperature can improve topological fidelity for the same guidance weight, but also improves robustness against the negative effects of exceedingly high guidance weights. The atrial separation task is an exception to this, where topological precision for loops and voids is maximized by using a softmax temperature of 1.

Effect of Partial Decoding Strategy We see in Fig. 10 that applying partial decoding with increased resolution can significantly improve topological fidelity at an increased computational cost. We find that the benefits of increased decoding resolution vary based on the topological feature and task. For example, the number of extra loops in the atrial separation task is minimized at a decoding resolution of 128, while the number of extra components for the aortic branch task is invariant after a decoding resolution of 32. We also see from Tab. 13 that a decoding resolution of 64 represents a good trade-off between computational cost and topological fidelity, providing a speedup of 11x over the next highest resolution. For topological guidance, the wall clock time was approximately 420 seconds per sample for the highest decoding resolution with a convolutional decoder.

Table 13. **Topological sampling speed comparison for partial decoding strategies**. Speed is measured in terms of sampled label maps per second using the maximum allowable batch size on a single GPU, normalized to the slowest method.

Methodology			
Approach	**Domain**	**Res.**	**Speed (↑)**
Anatomica-L	Coarse	16	**32.00**
		32	26.25
		64	11.00
		128	1.14
Anatomica-V	Global	128	1.00

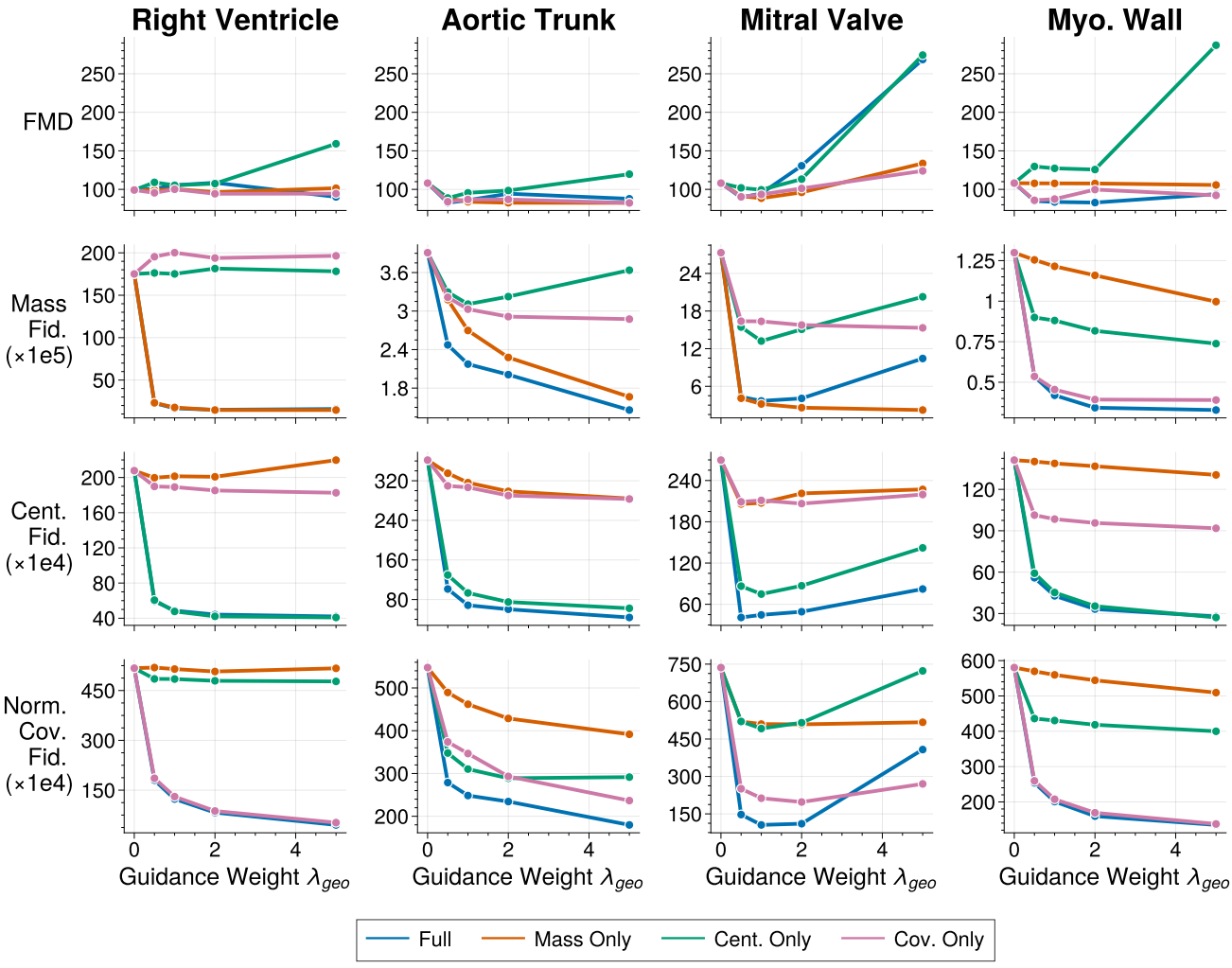

Figure 8. **Geometric guidance and disentangled guidance ablation study.**

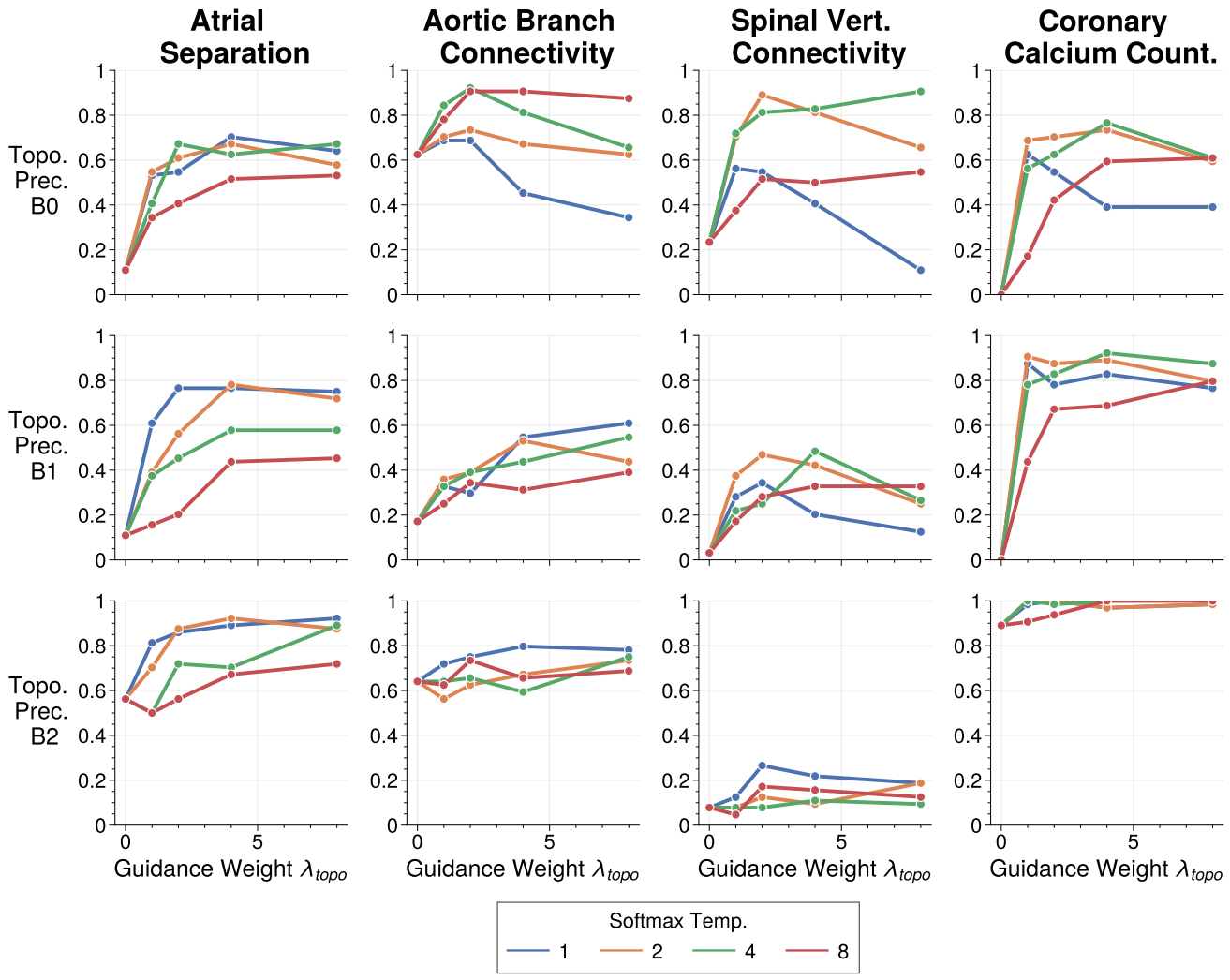

Figure 9. **Topological guidance and softmax temperature ablation study.**

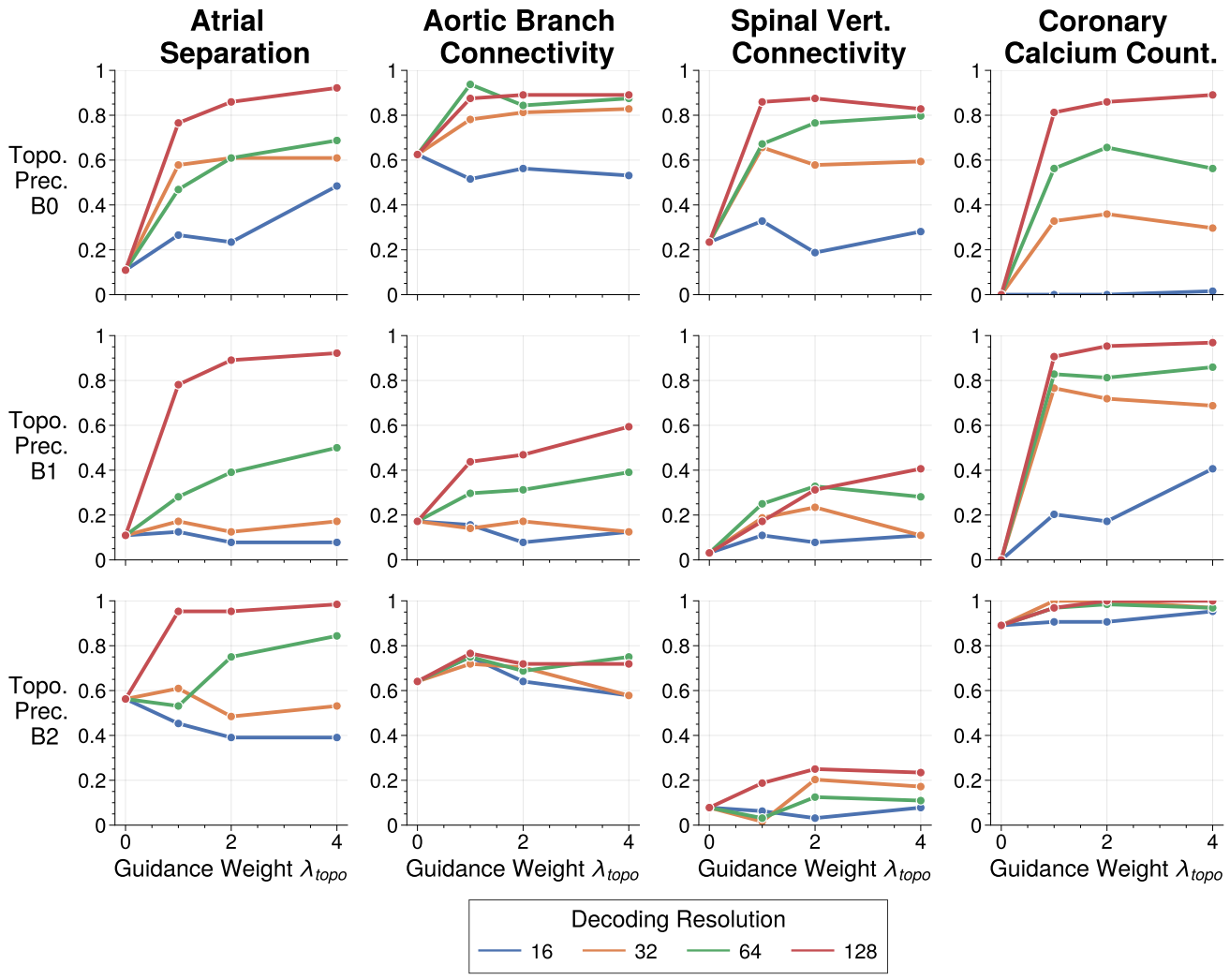

Figure 10. **Topological guidance and partial decoding resolution ablation study.**