VesselTok: Tokenizing Vessel-like 3D Biomedical Graph Representations for Reconstruction and Generation

Chinmay Prabhakar[1,2], Bastian Wittmann[1,2], Tamaz Amiranashvili[1,3], Paul Büschl[1], Ezequiel de la Rosa[1], Julian McGinnis[3,4], Benedikt Wiestler[3,4], Bjoern Menze*[1,2], and Suprosanna Shit*[1,2]

[1] University of Zurich, Switzerland
[2] ETH AI Center, Zurich, Switzerland
[3] Technical University of Munich, Germany
[4] Munich Center for Machine Learning (MCML), Germany
chinmay.prabhakar@uzh.ch

Abstract. Spatial graphs provide a lightweight and elegant representation of curvilinear anatomical structures such as blood vessels, lung airways, and neuronal networks. Accurately modeling these graphs is crucial in clinical and (bio-)medical research. However, the high spatial resolution of large networks drastically increases their complexity, resulting in significant computational challenges. In this work, we aim to tackle these challenges by proposing VesselTok, a framework that approaches spatially dense graphs from a parametric shape perspective to learn latent representations (tokens). VesselTok leverages centerline points with a pseudo radius to effectively encode tubular geometry. Specifically, we learn a novel latent representation conditioned on centerline points to encode neural implicit representations of vessel-like, tubular structures. We demonstrate VesselTok's performance across diverse anatomies, including lung airways, lung vessels, and brain vessels, highlighting its ability to robustly encode complex topologies. To prove the effectiveness of VesselTok's learned latent representations, we show that they (i) generalize to unseen anatomies, (ii) support generative modeling of plausible anatomical graphs, and (iii) transfer effectively to downstream inverse problems, such as link prediction.

Keywords: Reconstruction · Generative Modeling · Tokenizer · Biomedical Graph Representations · Blood Vessels · Tubular Structures

1 Introduction

Vessel-like, tubular networks are ubiquitous in physiological systems (*e.g.*, blood vessel network, neuronal network, lung airways, or lymphatic network) and serve

* Equal senior contribution

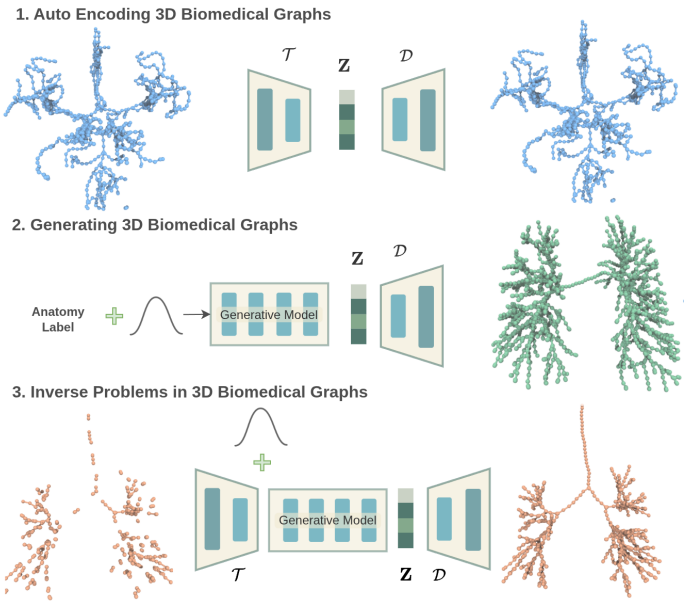

Fig. 1: VesselTok provides expressive latent representations **Z** which can be effectively leveraged to several downstream tasks, including generalization to unseen anatomies, generative modeling (*i.e.*, sample additional graphs), and inverse problems (*e.g.*, repair of incomplete structures).

the purpose of connecting different anatomical regions, facilitating the transportation of signals and substrates. Studying their normal and abnormal characteristics is of immense importance in clinical and biomedical research, particularly in disease diagnosis, including cerebrovascular diseases [13, 23, 42], pulmonary disorders [3,31], and diabetic peripheral neuropathy [12]. In this context, a 3D spatial graph representation comprised of the centerlines of these tubular structures efficiently encodes the necessary geometric and functional properties (lengths, radii, branch topology, and inputs for, *e.g.*, computational flow modeling) [5, 7]. However, at high spatial resolution, these graphs typically contain a large number of nodes and edges, rendering them computationally challenging to process using off-the-shelf algorithms. For this reason, previous graph modeling methods have only considered simplified structures [33], small components or subgraphs [14], and tree-like graphs [2]. The objective of this work is to find *high-fidelity* and *computationally tractable* latent embeddings (tokens) of *topologically complex, large 3D spatial graphs*, a shortcoming of existing techniques.

Although vessel-like structures exhibit substantial radius variation, we argue that explicitly modeling radius is not the main bottleneck for learning compact representations of complex tubular networks. In many anatomical systems (*e.g.*, airways and cerebral vasculature), radii are anatomically constrained and vary smoothly along the centerline [18,38], making them amenable to reliably regress

from centerline coordinates (see supplementary Sec. B). We therefore assign a fixed pseudo-radius to centerline points and treat the true radius as an inferable attribute based on the structure. This choice is not meant to trivialize the task. Rather, it concentrates model capacity on encoding large-scale 3D geometry and topology while decoupling structural learning from the severe scale imbalance introduced by wide-radius distributions.

Recognizing the increased computational complexity from a pure graph perspective, we take an alternative approach and model the 3D spatial graph as a tubular surface (with a fixed pseudo-radius) represented as a continuous 3D occupancy field. However, vascular networks differ from generic 3D shapes. They are sparse, thin, and highly branched, resulting in a high surface-to-volume ratio. Consequently, surface representations can contain far more samples than the underlying centerline. For example, a representative ATM case contains roughly 64,000 surface points but only 3,000 centerline points. Generic surface-point tokenizers [50, 55] may therefore spend much of their fixed query budget on redundant tubular surface samples, while undersampling small branches, endpoints, and bifurcations. Therefore, we propose to operate on centerline points, allowing the model to focus its representational capacity on the intrinsic structure of high surface-to-volume ratio tubular networks rather than redundant surface geometry, resulting in more expressive latent representations.

Building on this intuition, we introduce VesselTok, a tokenizer that generates high-fidelity, yet computationally tractable, latent representations of vascular graphs. Our key insight is to represent a 3D graph as an implicit occupancy field and to tokenize its intrinsic centerline structure through an implicit decoder. This implicit representation enables a compact, expressive latent space that faithfully encodes the geometry and topology of large, complex vascular networks, making it suitable as a shared token space for reconstruction, generation, and infilling. VesselTok is trained on diverse datasets, resulting in a generalizable model that robustly encodes complex topologies. We further demonstrate the applicability of our generated 3D graph tokens in generative modeling and inverse problems, such as link prediction (see Fig. 1).
In summary, our contributions are as follows:

1. We propose VesselTok, a state-of-the-art tokenizer for 3D spatial graphs that leverages the implicit representation of tubular shapes and is capable of generalizing to diverse curvilinear anatomical structures.
2. We demonstrate downstream use of VesselTok's embeddings for unconditional and conditional generative modeling, highlighting its representational strength to capture complex 3D topology.
3. Further, we show VesselTok's generative prior in the token space can be used to solve inverse problems, such as link prediction.

2 Related Literature

Graph Tokenizer: Tokenizing graphs to learn semantically rich latent representations has been extensively explored. Pioneering works, such as VQGraph [49],

learn a VQ–VAE–style tokenizer that assigns each node's local substructure to a discrete code, forming a codebook that captures diverse local patterns. Similarly, Liu et al. [25] systematically study different tokenizers (node, edge, motif, and GNN-based) on molecules and show that subgraph-level tokenization, combined with an expressive decoder, substantially improves self-supervised representation learning. More recently, Wang et al. [41] propose a Graph Quantized Tokenizer trained via multi-task self-supervised objectives and decoupled from the downstream Transformer. OpenGraph [45] similarly introduces a unified tokenizer within a graph foundation model: graphs from diverse domains are mapped into a shared token space, which allows a single model to generalize to unseen graph properties and datasets. However, all above described methods operate at native graph resolution, *i.e.*, the latent retains the original node count. Hence, such methods are effective for downstream analysis but ill-suited to the computational demands of generating large graphs. In contrast, we explicitly prioritize compression, learning a compact latent that enables high-fidelity synthesis at scales far beyond prior approaches [33, 39].

Shape Tokenizer: Recent efforts increasingly compress 3D shapes into compact tokens that serve as an interface to generative models. Early text–to–3D and image–to–3D pipelines, such as Shap-E [16] and Meta 3D Gen [4], demonstrated the effectiveness of latents for shape synthesis. Among other notable works, 3DShape2VecSet [50] encodes surface point clouds into a set of latent vectors for diffusion-based shape modeling. Hunyuan3D 2.0 [55] proposes an improved architecture for shape encoding with important query point sampling. 3D Shape Tokenization [6] leverages continuous shape tokens learned via flow matching for image-to-3D and neural rendering. Dora [8] addresses the sampling bias in Shape-VAEs through sharp-edge-aware sampling. Michelangelo [56] aligns shape–image–text latents for conditional 3D generation. Other relevant work includes VAT [51], a variational tokenizer that compresses unordered 3D features into hierarchical latent tokens for autoregressive generation, Structured 3D Latents [46], which uses an augmented sparse 3D grid to enable fast, high-quality text/image-conditioned 3D generation, and G3PT [52], which uses transformer tokenizers for point clouds with occupancy decoders. Although these methods aim for compact, faithful, cross-modal tokens, they are designed for generic shapes using surface-centric cues and omit vessel-specific, graph-oriented design. In contrast, VesselTok exploits the high surface-to-volume ratio of tubular networks by operating on centerline points.

Vessel Tokenizer: While early works explored generating vascular trees directly in uncompressed centerline space [20, 44], more recent approaches for tubular and tree-like anatomies have increasingly shifted toward learned latent representations and tokenized encodings, enabling more compact and scalable modeling. VesselVAE [14] formulates a recursive variational autoencoder that maps a vessel tree to a compact latent vector. VesselGPT [15] builds a discrete codebook for vessels with a VQ-VAE applied to represent vascular structure as a short token sequence. Kuipers et al. [19] adapt the 3DShape2VecSet framework [50]

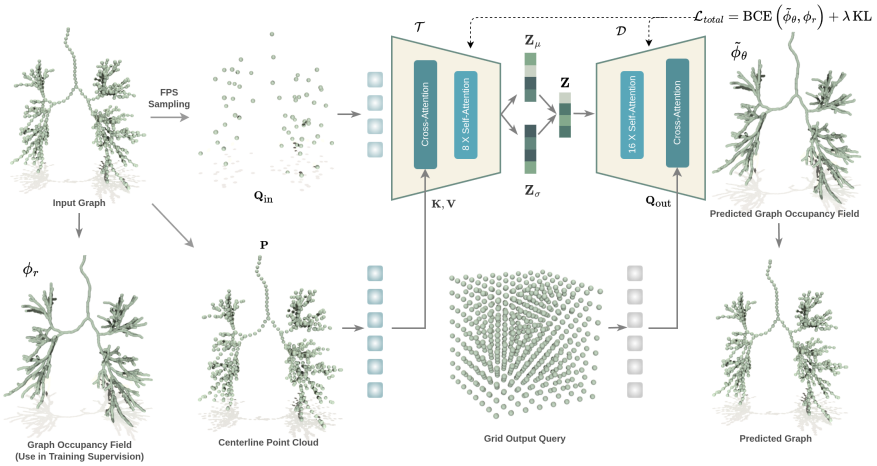

Fig. 2: Architectural overview of VesselTok. VesselTok consists of an encoder $\mathcal{T}$, which extracts features from a pre-processed centerline point cloud $\mathbf{P}$ to generate a continuous, expressive, and compressed latent token $\mathbf{Z}$. A decoder $\mathcal{D}$ subsequently reconstructs the graph occupancy field $\tilde{\phi}_\theta$ from $\mathbf{Z}$ via cross-attention to output query points $\mathbf{Q}_{\text{out}}$. The final graph is subsequently reconstructed from $\tilde{\phi}_\theta$.

to vascular data, but evaluate it only on the relatively small Circle of Willis anatomy. Batten et al. [2] propose to learn vector representations of vessel trees at two levels: a segment autoencoder that encodes branch centerlines, and a tree autoencoder that aggregates segment codes into a single vector. Similarly, Chen et al. [9] propose a hierarchical part-based model for generation. It first samples a global binary tree topology, then generates segment-level geometry, and finally assembles the full vessel tree by composing the synthesized segments according to the sampled topology. Beyond blood vessels, Zhang et al. [54] utilize geometric correspondence in the segmentation space to encode airway trees using neural implicit representation. While these works take early steps toward vessel-specific latent modeling, they are largely limited to small geometries, tree-only structures, and simple vessel segments. In contrast, VesselTok scales to large geometries and supports complex topologies, including non-tree connectivity.

3 Methodology

3.1 Definitions

Problem Statement: Let $G = (V, E, \mathbf{P})$ be a 3D spatial graph with vertices $V = \{1, \ldots, n\}$, edges $E \subseteq V \times V$, and associated coordinates $\mathbf{P} \in \mathbb{R}^{n \times 3}$. We aim to learn a graph tokenizer $\mathcal{T}$ that maps G to a l-length token sequence with channel size c (continuous tokens $\mathbf{Z} \in \mathbb{R}^{l \times c}$ in case of VAE latents) together with a decoder $\mathcal{D}$ such that the decoded graph $\tilde{G} : \mathcal{D}(\mathcal{T}(G))$ has the same topology and structure of the input graph G.

3D Graph Occupancy Field: Since the position of the centerline points and the total number of points can vary without altering the underlying vessel structure (non-injective mapping), we aim to construct a robust latent representation of the graph structure. To this end, we adopt a representation invariant to such choices by treating each edge as a straight segment in $\mathbb{R}^3$ and thickening it based on a pseudo radius $r > 0$. We then define a continuous graph occupancy field ϕ_r, with a pseudo radius r assigned to each of its edges. To obtain the graph occupancy field at a point $p \in \mathbb{R}^3$, we compute the distance $d_G(p, G)$ to the closest edge (and, if present, isolated nodes) as follows:

$$d_G(p,G) = \min\left\{ \min_{(i,j)\in E} \big\| \, p - \big(p_i + \alpha_{ij}(p)\,(p_j - p_i)\big) \, \big\|, \; \min_{i\in V} \|p - p_i\| \right\},$$

where p_i, p_j are node coordinates associated with edge $(i, j) \in E$, and

$$\alpha_{ij}(p) = \mathrm{clamp}\left(\frac{(p - p_i)\cdot(p_j - p_i)}{\|p_j - p_i\|^2}, \, 0, \, 1 \right), \quad \mathrm{clamp}(t, 0, 1) = \min\{1, \max\{0, t\}\}.$$

The clamping ensures the closest point lies on the *segment* and $\min_{i\in V} \|p - p_i\|$ is used to handle boundary nodes. From this, we obtain the graph occupancy field ϕ_r as follows:

$$\phi_r(p, G) = \begin{cases} 1, & \text{if } d_G(p, G) \le r \\ 0, & \text{otherwise} \end{cases}. \tag{1}$$

Note that the choice of the pseudo-radius r is critical as it determines how faithfully the occupancy field represents the graph topology. Too small r could hinder the learning capabilities of the model, whereas larger r could miss topologically important structures and thus, introduce artifacts. We provide a detailed ablation analysis of the pseudo radius choice in our experiments (see Section 4.6).

3.2 VesselTok

We present VesselTok, a method to generate latent representations arising from graph occupancy fields ϕ_r conditioned on centerline graphs G. Prior methods developed for generic 3D shapes [50, 55] primarily rely on surface point samples, which we argue are suboptimal for high surface-to-volume structures such as vascular graphs under realistic computational constraints. In contrast, VesselTok operates on centerline points, a substantially more computationally tractable representation than dense surface meshes, and leverages them as structural primitives to model the graph occupancy field directly. This design allows VesselTok to jointly process all centerline points without subsampling, efficiently producing expressive latent representations even for large and complex structures. We adopt a VAE-style encoder-decoder architecture to obtain a continuous token $\mathbf{Z}$ for the occupancy field ϕ_r, where the encoder $\mathcal{T}$ and decoder $\mathcal{D}$ consist of transformer blocks following recent works [55, 56]. An architectural overview of VesselTok is presented in Fig. 2.

Encoder: The encoder $\mathcal{T}$ aims to extract features from the centerline point cloud $\mathbf{P}$ and subsequently convert them into continuous tokens $\mathbf{Z}$. For this purpose, we initialize L query points $\mathbf{Q}_{\text{in}}$ by using farthest point sampling (FPS) from $\mathbf{P}$. This results in an initial representation of the underlying structure, which is refined by the encoder. We use Fourier positional encoding for both the input point cloud $\mathbf{P}$ and the query points $\mathbf{Q}_{\text{in}}$, followed by a linear embedding layer to expand the feature dimension to the hidden dimension d. In the first layer of the encoder, we contextualize the input point cloud via the query points by applying a cross-attention layer between the embeddings of $\mathbf{P}$ and $\mathbf{Q}_{\text{in}}$. Next, we employ a series of self-attention layers to obtain the encoder hidden representation $\mathbf{H} \in \mathbb{R}^{l \times d}$. We obtain the mean and variance denoted by $\mathbf{Z}_\mu, \mathbf{Z}_\sigma \in \mathbb{R}^{l \times c}$, respectively, via two final linear layers.

Decoder: The decoder $\mathcal{D}$ reconstructs the graph occupancy field from the encoded latent. The first decoder layer maps the latent to the hidden dimension. Subsequently, the embedding goes through a series of self-attention layers. Next, we sample query points $\mathbf{Q}_{\text{out}}$ from the 3D grid, and embed them using Fourier positional encoding followed by a linear projection. Finally, the embeddings of the $\mathbf{Q}_{\text{out}}$ are forwarded to a cross-attention layer to predict the graph occupancy field $\tilde{\phi}_\theta$ parameterized by the model parameters θ.

Training: We apply a reconstruction loss on the decoder output and the KL-divergence at the latent to train our model. For the reconstruction loss, we compute the Binary Cross Entropy (BCE) between the reconstructed graph occupancy and the reference on the query points. The total training loss is formalized as follows:

$$\mathcal{L}_{\text{total}} = \mathbb{E}_{p \in \mathbb{R}^3}[\text{BCE}(\tilde{\phi}_\theta(p), \phi(p, G))] + \lambda \cdot \text{KL}(\mathcal{N}(\mathbf{Z}_\mu, \text{diag}(\mathbf{Z}_\sigma)), \mathcal{N}(\mathbf{0}, \mathbf{I})). \quad (2)$$

Since the graph occupancy field is quite sparse throughout the entire domain, we employ an imbalance-aware query selection strategy similar to [50], emphasizing points located on the boundary of the object.

Inference: For graph reconstruction, we decode the latent representation into a continuous occupancy field and then recover a discrete centerline graph. We evaluate the predicted field on a regular grid $\mathcal{X} \subset \mathbb{R}^3$, apply a threshold $\tau \in (0, 1)$, and obtain a discretized occupancy field:

$$\Omega_\tau \;=\; \{\, p \in \mathcal{X} \;:\; \tilde{\phi}_\theta(p) \geq \tau \,\}.$$

Next, we extract centerline points from the discretized occupancy field using a skeletonization algorithm [22], yielding a set of predicted skeleton points $\hat{V}$. We convert $\hat{V}$ to a graph by a deterministic neighborhood-based connectivity prediction ($\hat{E}$).

This procedure yields a discrete, topology-preserving graph $\hat{G}$ consistent with the continuous graph occupancy predicted by the decoder, closing the loop from latent decoding back to a usable centerline representation. We apply the same

graph-extraction procedure to both our method and all baselines for a fair comparison. Importantly, this step is modular, and more robust graph extraction methods (e.g., [28]) can be substituted without changing the rest of the pipeline.

4 Experiments

Datasets: Similar to recent works [43], we curate an expressive publicly available dataset spanning diverse anatomies (see supplementary Sec. A). To this end, we extract graphs from airway datasets (ATM [53], AIIB [30], AeroPath [37]), cerebral vasculature (COSTA [29]), and pulmonary vessels (HiPas [11], PARSE [26], Pulmonary-AV [10]). For out-of-distribution validation, we evaluate our method on TopCoW [48] and renal vasculature data [21,40,47]. Graphs are derived from segmentation masks using the Voreen graph extraction tool [28], yielding centerline graphs with per-edge geometry. The resulting dataset spans a wide range of sizes and topologies (*e.g.*, node counts, branching factors, and loop prevalence). Importantly, all graphs are derived from real biomedical images, and no additional post-processing or refinement steps that could potentially perturb the graph structure were applied.

Metrics: To assess reconstruction quality, we compare the topology of reconstructed graphs to the ground truth using Betti numbers. Specifically, for topology, we report absolute Betti differences, $|\Delta\beta_0|$ (connected components) and $|\Delta\beta_1|$ (loops), where lower values indicate better preservation of graph structure. However, Betti-based errors alone are insufficient since, although they capture homology equivalence, they do not measure spatial coverage or geometric overlap between structures. Therefore, we complement them with clDice [35] and Chamfer Distance (CD). clDice emphasizes centerline fidelity and connectivity by quantifying overlap between the reconstructed and reference centerlines, whereas CD measures geometric discrepancy between the reconstructed and reference shapes. For all evaluations, we render graphs into a continuous occupancy field by dilating edges with a pseudo-radius of r (see Section 4.6) on a 512^3 voxel grid, and compute clDice and CD between the reconstruction and the corresponding ground-truth occupancy fields.

For generative evaluation, we report both point-based and graph-based metrics. We follow Zhang et al. [50] and report Fréchet Inception Distance (FID), Maximum Mean Discrepancy based on Chamfer Distance (MMD-CD), and Earth Mover Distance (MMD-EMD) along with Coverage (COV-CD, COV-EMD). To compute FID, we train a PointNet++ encoder [34] to predict the anatomical labels of the centerlines contained in the training set. The FID is computed between the embeddings of the training set samples and 1,000 generated samples. We also compute MMD and Coverage on Betti summaries (MMD–Betti, COV–Betti) of graphs to assess distributional alignment in terms of connected components and loops. For further details, please see supplementary Sec. C. The source code is publicly available at https://github.com/chinmay5/vessel_tok

Table 1: Quantitative results achieved on our testing anatomies. We report metrics for each dataset.

Data	Method	clDice ↑	CD ↓	$\lvert\Delta\beta_0\rvert$ ↓	$\lvert\Delta\beta_1\rvert$ ↓
ATM	Hunyuan3D 2.0 [55]	86.75	4.08	7.28	6.40
	3DShape2VecSet [50]	90.08	2.14	7.46	6.53
	VesselTok (ours)	**96.33**	**0.96**	**4.91**	**6.19**
AIIB	Hunyuan3D 2.0 [55]	85.13	2.90	8.40	11.00
	3DShape2VecSet [50]	87.22	2.74	8.6	**10.85**
	VesselTok (ours)	**94.85**	**0.66**	**4.47**	11.00
COSTA	Hunyuan3D 2.0 [55]	56.53	4.28	79.15	56.19
	3DShape2VecSet [50]	59.65	2.29	70.85	54.35
	VesselTok (ours)	**77.26**	**1.64**	**26.81**	**26.83**
HiPas	Hunyuan3D 2.0 [55]	52.58	3.95	137.31	**23.26**
	3DShape2VecSet [50]	55.58	2.92	132.56	28.65
	VesselTok (ours)	**67.59**	**1.82**	**115.88**	38.42
PARSE	Hunyuan3D 2.0 [55]	42.64	3.42	159.06	37.06
	3DShape2VecSet [50]	42.44	3.19	155.06	**32.31**
	VesselTok (ours)	**57.03**	**0.69**	**128.06**	34.06
Pulm.	Hunyuan3D 2.0 [55]	79.46	6.99	31.26	5.61
	3DShape2VecSet [50]	86.40	1.59	25.52	**5.04**
	VesselTok (ours)	**94.30**	**0.74**	**9.35**	5.52

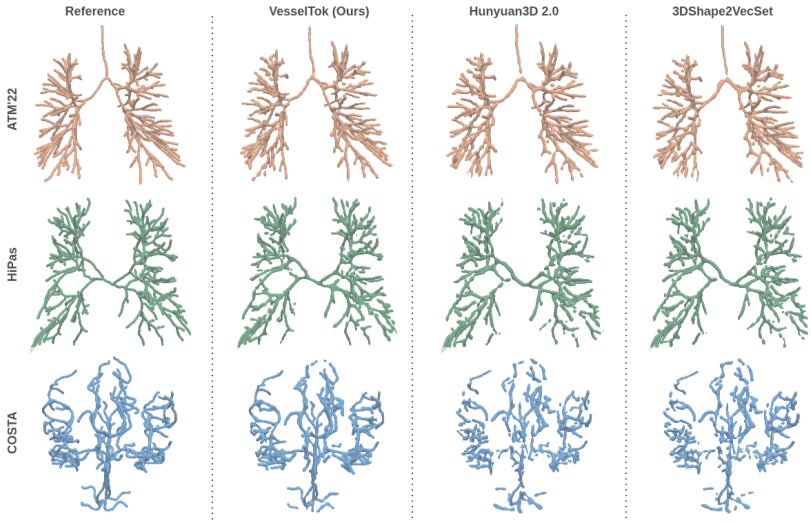

Fig. 3: Qualitative results for the graph reconstruction task. We find that VesselTok demonstrates superior reconstruction capabilities.

4.1 Reconstruction

VesselTok delivers consistent gains over the strong baselines 3DShape2VecSet [50] and Hunyuan3D 2.0 [55] across most datasets and metrics. The per-dataset train-test splits and implementation details are provided in the supplementary Sec. A and D, respectively. Tab. 1 shows that VesselTok attains higher clDice, lower

Table 2: Comparison against vessel-specific and other baselines on the ATM dataset

Method	clDice ↑	CD ↓	$\lvert\Delta\beta_0\rvert$ ↓	$\lvert\Delta\beta_1\rvert$ ↓
VesselGPT [15]	77.09	0.008	**0.05**	10.28
Hunyuan3D 2.0 [55]	87.31	0.007	0.11	9.31
3DShape2VecSet [50]	90.11	0.007	0.09	9.48
VesselTok (ours)	**96.61**	**0.005**	0.07	**8.97**

Chamfer Distance, and lower $|\Delta\beta_0|$ error consistently and competitive $|\Delta\beta_1|$ error. A per-dataset breakdown on airways (ATM and AIIB), cerebral vasculature (COSTA), and pulmonary vessels (HiPas, PARSE, and Pulmonary-AV) confirms that these improvements are consistent across anatomies. Notably, VesselTok better preserves connectivity and loop structure on average while reducing geometric discrepancy, indicating closer adherence to the underlying graph topology. Qualitatively, Fig. 3 shows superior reconstruction capabilities of our proposed method over the previous shape-encoding methods. Supplementary Sec. E shows additional qualitative results.

4.2 Vessel Specific Baselines

The state-of-the-art vessel-generation approach [33] can handle branches and loops but operates in the uncompressed point cloud space, which makes it computationally intractable for the number of nodes in our dataset. Alternatively, we benchmark the reconstruction capabilities of VesselTok against established autoencoder VesselGPT [15]. However, VesselGPT is a vessel-tree-only method.

Hence, we compare VesselTok and other baselines only on the airway tree (ATM) dataset, training them from scratch. Please note that airway trees are expected to be loop-free but often contain small spurious cycles due to minute inaccuracies in segmentation. Since VesselGPT assumes a strict tree structure, we needed to remove these loops during preprocessing. In contrast, VesselTok requires no such cleanup. We also experimented with VesselVAE [14], but it fails to converge. Tab. 2 shows that VesselTok achieves the best clDice, CD, and $|\Delta\beta_1|$, reflecting higher geometric fidelity and better recovery of cyclic structure, while VesselGPT performs best on $|\Delta\beta_0|$, likely benefiting from the simplification step in its preprocessing pipeline.

4.3 Generalization to Unseen Anatomies

To assess VesselTok's generalization, we evaluate it on anatomies and graph scales outside the training distribution. VesselTok is trained on airways, whole-brain vessels, and pulmonary trees, with graphs ranging from $\approx$2500 to $>$ 10000 nodes. To demonstrate its generalizability, we applied VesselTok on three datasets: (i) large-scale renal vasculature (RV), (ii) compact human Circle of Willis (CoW), and (iii) synthetic data created by partitioning airway trees and brain vessels along the sagittal plane to induce anatomically implausible edits (probing topological breaks and long-range consistency). CoW graphs are small ($<$ 1,000 nodes

Table 3: Quantitative results achieved on unseen anatomies and scales.

Data	Method	clDice ↑	CD ↓	$\lvert\Delta\beta_0\rvert$ ↓	$\lvert\Delta\beta_1\rvert$ ↓
ATM*	Hunyuan3D 2.0 [55]	89.81	0.125	2.66	5.34
	3DShape2VecSet [50]	96.26	0.097	**0.049**	5.34
	VesselTok (ours)	**99.16**	**0.001**	**0.049**	**5.29**
COSTA*	Hunyuan3D 2.0 [55]	81.16	**0.121**	0.85	5.20
	3DShape2VecSet [50]	82.65	0.123	0.59	4.36
	VesselTok (ours)	**95.44**	0.125	**0.24**	**3.49**
TopCoW	Hunyuan3D 2.0 [55]	97.82	0.003	0.09	0.21
	3DShape2VecSet [50]	97.96	**0.001**	**0.06**	0.14
	VesselTok (ours)	**99.42**	0.002	0.07	**0.04**
RV	Hunyuan3D 2.0 [55]	73.81	0.086	15.29	6.68
	3DShape2VecSet [50]	79.66	0.019	**11.91**	6.45
	VesselTok (ours)	**88.86**	**0.017**	13.09	**6.25**

* Synthetic set created by partitioning vessels along the sagittal plane.

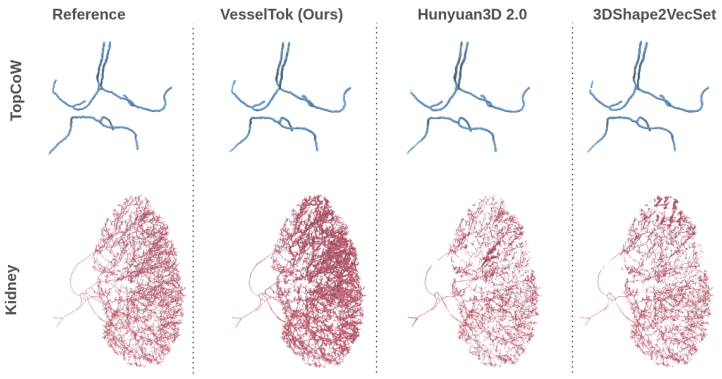

Fig. 4: Qualitative results for the graph reconstruction task of previously unseen domains. This demonstrates VesselTok's strong prior, resulting in robust reconstructions.

on average), while renal graphs exceed 100,000 nodes (over an order of magnitude larger than typical training examples). Although not limited by computational scalability, to obtain higher fidelity, we ran inference via spatial chunking with 150^3 grid size, yielding heterogeneous chunks ranging from sparse ($<$ 200 nodes) to dense ($>$ 8,000 nodes). For all graphs, coordinates are normalized to [-1,1] and centered by subtracting the center of mass.

Tab. 3 and Fig. 4 show that VesselTok robustly encodes heterogeneous out-of-distribution structures, achieving a clDice of 99.42 on CoW and 88.86 on RV, while also attaining lower β_1 error and competitive β_0 error, indicating more faithful topology preservation than the baselines. On the synthetic sagittal-cut ATM and COSTA samples, VesselTok maintains clDice scores of 99.16 and 95.44, respectively, demonstrating resilience to anatomically implausible topological edits. In contrast, baseline methods consistently exhibit lower clDice and higher β_1 error (i.e., more spurious or missed loops), while 3DShape2VecSet achieves competitive β_0 error. For additional details, see supplementary Sec F.

Table 4: FID computed in the PointNet++ latent space, Maximum Mean Discrepancy on Chamfer and Earth Mover distance, and Betti error for the generated samples using EDM [17] on the latent tokens. We present conditional and unconditional results. MMD-CD and MMD-EMD values are reported in 10^{-2}.

	Method	FID ↓	MMD ↓				COV ↑			
			CD	EMD	β_0	β_1	CD	EMD	β_0	β_1
cond.	Hunyuan3D 2.0 [55]	74.61	1.89	0.17	28.06	5.19	0.43	0.46	0.84	0.68
	3DShape2VecSet [50]	73.58	1.92	0.18	24.07	6.82	0.44	0.50	0.89	0.61
	VesselTok (ours)	**43.13**	**0.25**	**0.14**	**7.25**	**1.01**	**0.48**	**0.53**	**0.95**	**0.73**
uncon.	Hunyuan3D 2.0 [55]	107.57	1.85	3.51	125.68	3.74	0.40	0.28	0.22	0.12
	3DShape2VecSet [50]	146.57	1.88	3.45	127.60	3.22	0.39	0.26	0.21	0.08
	VesselTok (ours)	**96.87**	**0.30**	**2.35**	**78.36**	**2.64**	**0.42**	**0.32**	**0.32**	**0.18**

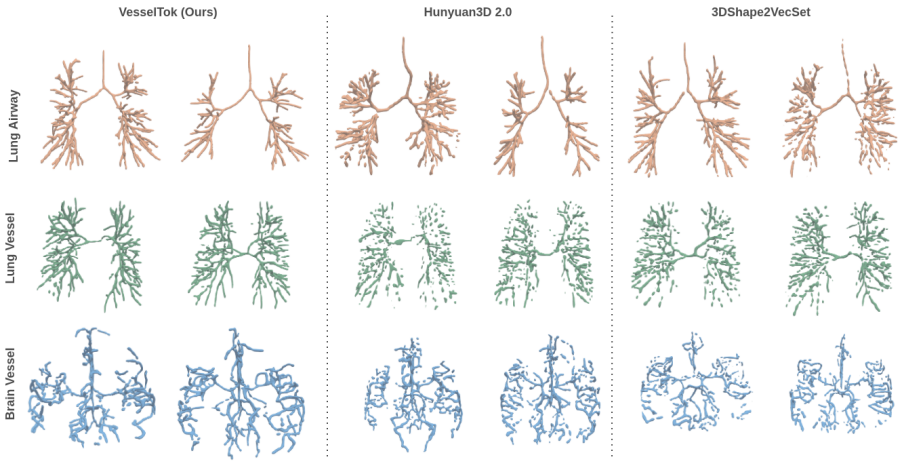

Fig. 5: Qualitative results for conditional generation. VesselTok consistently generates more realistic vessels than previous methods.

4.4 Generative Modeling

We generate anatomical graphs by training a diffusion model in the token space. Using a fixed-length sequence of 512 tokens per graph keeps training and sampling tractable compared to native graph generation [33]. We train both unconditional and class-conditional models (conditioned on anatomical categories), using the EDM backbone [17] and hyperparameters matching 3DShape2VecSet [50]. We compare our method against the strong baselines Hunyuan3D [55] and 3DShape2VecSet, while keeping the architecture and training budget fixed to ensure fairness. Please see Supplementary Sec. G for details.

As reported in Table 4 and Fig. 5, our latent yields lower FID in both unconditional and conditional regimes and also improves complementary metrics (MMD-CD, MMD-EMD, COV-CD, and COV-EMD). Further, the lower Betti errors indicate that graphs generated by our method align with the ground truth

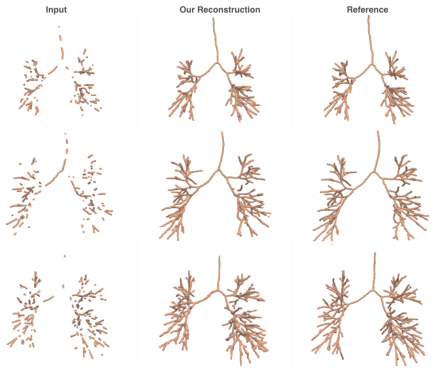

Fig. 6: Qualitative examples for link prediction on the airway tree (ATM) dataset. Given an incomplete airway graph with missing connections, our model infers and inserts plausible edges to repair the centerline while preserving anatomical plausibility.

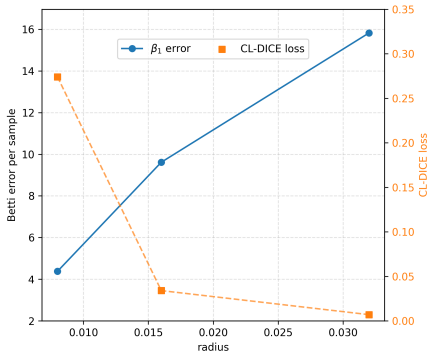

Fig. 7: Comparison of reconstruction fidelity vs. topological characteristics retained as we change the pseudo radius r. Lower values of r improve topological characteristics ($|\Delta\beta_1|$) but degrade reconstruction performance (clDice loss). We find $r = 0.016$ strikes a good balance.

data distribution better. This demonstrates that a semantically structured token space materially benefits generative modeling of biomedical, spatial graphs. Additional training and ablation details are mentioned in supplementary Sec. D.

4.5 Downstream Tasks

Beyond unconditional generation, the learned token representation supports clinically relevant downstream tasks such as link prediction for repairing incomplete vasculature. Specifically, given a vessel graph with missing connections, the goal is to infer plausible edges that are consistent with both local geometry and global topology. For this, we train a Diffusion Transformer [32] with a conditional flow-matching objective [24]. The incomplete graph serves as a conditioning signal, and the model learns to denoise toward the distribution of the complete (clean) graph. Training pairs are generated by the description provided in supplementary Sec. H, resulting in a graph with ≈40% of missing links. The model learns to recover the missing connectivity. In inference, the model samples plausible

Table 5: VesselTok's performance on inverse problems. On the ATM infill task, VesselTok outperforms the baseline [1], producing more coherent vascular structures.

Method	clDice ↑	CD↓	$\lvert\Delta\beta_0\rvert$ ↓	$\lvert\Delta\beta_1\rvert$ ↓
Autodecoder [1]	83.49	0.066	4.32	8.68
VesselTok (ours)	**88.13**	**0.043**	**3.19**	**8.58**

completions that "infill" the missing edges. Tab. 5 shows that our approach performs competitively with a strong autodecoder baseline [1]. Fig. 6 shows some qualitative examples.

4.6 Ablation Studies

We ablate two core design choices of VesselTok: (i) the pseudo-radius r used to form the occupancy field, and (ii) the number K and channel dimension C of the tokens in the VAE latent space.

Pseudo Radius: We model each biomedical graph as a 3D centerline. Conventional neural fields struggle with extremely thin structures [36], while vector-field–based approaches for curve-like geometry have shown limited fidelity [27]. In contrast, our objective prioritizes topological connectivity over recovering physical thickness. Hence, we adopt a simple design choice: convert the graph to a continuous occupancy field by dilating each edge with a fixed pseudo-radius r. Please note that using this continuous representation decouples synthesis from an a priori node count, an assumption required by several existing methods [33, 39].

To study the trade-off between reconstruction fidelity and topology preservation as a function of the dilation radius r, we train three encoders on ATM with $r \in 0.008, 0.016, 0.032$. Fig. 7 shows that larger r simplifies learning and improves reconstruction (clDice loss ≈ 0.007) but degrades fine topology (β_1 error ≈ 16 per sample), whereas smaller r better preserves topology (β_1 error ≈ 4) but reduces reconstruction quality (clDice loss ≈ 0.28). Hence, we select $r = 0.016$ as a balanced operating point, which maintains salient topological features while achieving competitive reconstruction quality. Importantly, this pseudo-radius is used as a single global hyperparameter across experiments, rather than being tuned separately for each dataset.

VAE Latent Space: The number of tokens per graph, K, and the latent channel dimension, C, govern a tradeoff between compression and reconstruction fidelity. Since biomedical graphs exhibit substantial variation across individual

Table 6: Effect of varying the number of keypoints K on reconstruction quality. Performance drops sharply at 256 or fewer keypoints, while 512 provides a sweet spot between reconstruction quality and compression ratio. ($C = 4$ for all models).

K	clDice ↑	CD ↓	$\lvert\Delta\beta_0\rvert$ ↓	$\lvert\Delta\beta_1\rvert$ ↓	κ
768	**97.13**	**0.005**	**0.07**	**8.93**	4.68
512	96.61	**0.005**	**0.07**	8.97	7.03
256	80.35	0.006	25.61	10.02	14.06
128	12.29	0.114	1.90	10.39	28.12
64	8.42	0.116	2.52	10.27	56.24

Table 7: Effect of channel dimension C on reconstruction quality. Increasing the channel dimension improves the reconstruction, but at the expense of the compression ratio κ. We fix K to 512 for this ablation.

C	clDice ↑	CD ↓	$\lvert\Delta\beta_0\rvert$ ↓	$\lvert\Delta\beta_1\rvert$ ↓	κ
16	96.32	**0.004**	**0.06**	**8.52**	1.76
8	95.95	**0.004**	**0.06**	8.80	3.51
4	**96.61**	0.005	0.07	8.97	7.03
2	7.08	0.058	1.68	10.36	14.06

anatomies, we use the average compression ratio, κ, to quantify this tradeoff. For a dataset consisting of M samples, with each sample consisting of N_i nodes, κ is defined as: $\kappa = \frac{1}{M}\sum_{i=1}^{M}\frac{N_i \cdot 3}{K \cdot C}$ Our goal is to learn a *compact* latent space that can faithfully reconstruct the graphs. We ablate both factors K and C by training VAEs from scratch on the representative ATM dataset. Reconstruction quality is measured with clDice scores, CD, and Betti error.

Tab. 6 varies K and demonstrates that increasing K improves reconstruction quality (clDice 97.13 and CD 0.005), but reduces κ as a tradeoff. Small K sharply degrades reconstruction quality (*e.g.*, clDice 8.42 and CD 0.116 at $K = 64$). Tab. 7 varies C and exhibits the same trend: higher C yields better fidelity (*e.g.*, $C = 16$ reaches clDice 96.32 and CD 0.004) at the expense of compression. We find that overall, $K = 512$ and $C = 4$ offer a balanced operating point, achieving clDice of 96.61, CD of 0.005, and low Betti error. We adopt this configuration for our experiments.

Limitation A natural limitation of fixed-capacity latent representations is that performance can degrade as input complexity increases. However, our analysis suggests that graph size alone is not the primary bottleneck. We stratify samples into medium ($\leq$5.5K nodes) and large (>5.5K nodes) groups, using node count as a coarse proxy for complexity. Under this stratification, clDice drops only modestly on ATM (98.90 $\rightarrow$ 91.43), but decreases sharply on the more topologically complex COSTA graphs (89.19 $\rightarrow$ 75.20). This suggests that reconstruction performance is influenced more strongly by topological complexity than by graph size alone.

5 Conclusion

Existing methods struggle to generate large graphs due to computational bottlenecks, rendering them unsuitable for analyzing network-like structures in biomedical systems (*e.g.*, vascular networks or airways). To address this issue, we present VesselTok, a novel graph tokenization method designed to learn a semantically rich, compressed latent representation of large networks. Unlike shape tokenizers that operate on mesh surfaces, VesselTok uses a centerline-based representation (occupancy around dilated centerlines), which is especially well-suited to high surface-to-volume geometries, such as dense vasculature. We demonstrate that VesselTok generalizes to anatomies not encountered during training and can learn compact latent representations of diverse anatomical networks. These latents can be efficiently leveraged for both generative modeling (unconditional and conditional synthesis) and inverse problem tasks (*e.g.*, link prediction), where VesselTok consistently outperforms the state of the art, paving the way for large-scale, targeted analysis of biomedical networks.

Acknowledgments

This work has been supported by the Helmut Horten Foundation. S. Shit is supported by the UZH Postdoc Grant (K-74851-03-01).

References

1. Amiranashvili, T., Lüdke, D., Li, H.B., Zachow, S., Menze, B.H.: Learning continuous shape priors from sparse data with neural implicit functions. Medical Image Analysis **94**, 103099 (2024)
2. Batten, J., Schaap, M., Sinclair, M., Bai, Y., Glocker, B.: Vector representations of vessel trees. arXiv preprint arXiv:2506.11163 (2025)
3. Belchi, F., Pirashvili, M., Conway, J., Bennett, M., Djukanovic, R., Brodzki, J.: Lung topology characteristics in patients with chronic obstructive pulmonary disease. Scientific reports **8**(1), 5341 (2018)
4. Bensadoun, R., Monnier, T., Kleiman, Y., Kokkinos, F., Siddiqui, Y., Kariya, M., Harosh, O., Shapovalov, R., Graham, B., Garreau, E., et al.: Meta 3d gen. arXiv preprint arXiv:2407.02599 (2024)
5. Bumgarner, J.R., Nelson, R.J.: Open-source analysis and visualization of segmented vasculature datasets with VesselVio. Cell reports methods **2**(4) (2022)
6. Chang, J.H.R., Wang, Y., Martin, M.A.B., Gu, J., Zhao, X., Susskind, J., Tuzel, O.: 3d shape tokenization via latent flow matching. arXiv preprint arXiv:2412.15618 (2024)
7. Chapman, B.E., Berty, H.P., Schulthies, S.L.: Automated generation of directed graphs from vascular segmentations. Journal of biomedical informatics **56**, 395–405 (2015)
8. Chen, R., Zhang, J., Liang, Y., Luo, G., Li, W., Liu, J., Li, X., Long, X., Feng, J., Tan, P.: Dora: Sampling and benchmarking for 3d shape variational auto-encoders. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 16251–16261 (2025)
9. Chen, S., Zhang, G., Lai, J., Shen, B., Zhang, S., Dong, C., Chen, X., Li, Y.: Hierarchical part-based generative model for realistic 3d blood vessel. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 257–267. Springer (2025)
10. Cheng, H., Zheng, L., Yan, Z., Zhang, H., Meng, B., Xu, X.: Fusion of machine learning and deep neural networks for pulmonary arteries and veins segmentation in lung cancer surgery planning. In: International Conference on Pattern Recognition. pp. 422–438. Springer (2024)
11. Chu, Y., Luo, G., Zhou, L., Cao, S., Ma, G., Meng, X., Zhou, J., Yang, C., Xie, D., Mu, D., et al.: Deep learning-driven pulmonary artery and vein segmentation reveals demography-associated vasculature anatomical differences. Nature Communications **16**(1), 2262 (2025)
12. Dabbah, M.A., Graham, J., Petropoulos, I.N., Tavakoli, M., Malik, R.A.: Automatic analysis of diabetic peripheral neuropathy using multi-scale quantitative morphology of nerve fibres in corneal confocal microscopy imaging. Medical image analysis **15**(5), 738–747 (2011)
13. Epp, R., Glück, C., Binder, N.F., El Amki, M., Weber, B., Wegener, S., Jenny, P., Schmid, F.: The role of leptomeningeal collaterals in redistributing blood flow during stroke. PLoS computational biology **19**(10), e1011496 (2023)
14. Feldman, P., Fainstein, M., Siless, V., Delrieux, C., Iarussi, E.: Vesselvae: Recursive variational autoencoders for 3d blood vessel synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 67–76. Springer (2023)
15. Feldman, P., Sinnona, M., Delrieux, C., Siless, V., Iarussi, E.: Vesselgpt: Autoregressive modeling of vascular geometry. In: International Conference on Medi-

cal Image Computing and Computer-Assisted Intervention. pp. 662–672. Springer (2025)
16. Jun, H., Nichol, A.: Shap-e: Generating conditional 3d implicit functions. arXiv preprint arXiv:2305.02463 (2023)
17. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. Advances in neural information processing systems **35**, 26565–26577 (2022)
18. Kirst, C., Skriabine, S., Vieites-Prado, A., Topilko, T., Bertin, P., Gerschenfeld, G., Verny, F., Topilko, P., Michalski, N., Tessier-Lavigne, M., et al.: Mapping the fine-scale organization and plasticity of the brain vasculature. Cell **180**(4), 780–795 (2020)
19. Kuipers, T.P., Konduri, P.R., Bekkers, E.J., Marquering, H.: Self-supervised synthetic cerebral vessel tree generation using semantic signed distance fields. In: Medical Imaging with Deep Learning (2025)
20. Kuipers, T.P., Konduri, P.R., Marquering, H., Bekkers, E.J.: Generating cerebral vessel trees of acute ischemic stroke patients using conditional set-diffusion. In: Medical Imaging with Deep Learning (2024)
21. Kuo, W., Rossinelli, D., Schulz, G., Wenger, R.H., Hieber, S., Müller, B., Kurtcuoglu, V.: Terabyte-scale supervised 3d training and benchmarking dataset of the mouse kidney. Scientific data **10**(1), 510 (2023)
22. Lee, T.C., Kashyap, R.L., Chu, C.N.: Building skeleton models via 3-d medial surface axis thinning algorithms. CVGIP: graphical models and image processing **56**(6), 462–478 (1994)
23. Lin, E., Kamel, H., Gupta, A., RoyChoudhury, A., Girgis, P., Glodzik, L.: Incomplete circle of Willis variants and stroke outcome. European Journal of Radiology **153**, 110383 (2022)
24. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
25. Liu, Z., Shi, Y., Zhang, A., Zhang, E., Kawaguchi, K., Wang, X., Chua, T.S.: Rethinking tokenizer and decoder in masked graph modeling for molecules. Advances in Neural Information Processing Systems **36**, 25854–25875 (2023)
26. Luo, G., Wang, K., Liu, J., Li, S., Liang, X., Li, X., Gan, S., Wang, W., Dong, S., Wang, W., et al.: Efficient automatic segmentation for multi-level pulmonary arteries: The parse challenge. arXiv preprint arXiv:2304.03708 (2023)
27. Mello Rella, E., Chhatkuli, A., Konukoglu, E., Van Gool, L.: Neural vector fields for implicit surface representation and inference. International Journal of Computer Vision **133**(4), 1855–1878 (2025)
28. Meyer-Spradow, J., Ropinski, T., Mensmann, J., Hinrichs, K.: Voreen: A rapid-prototyping environment for ray-casting-based volume visualizations. IEEE Computer Graphics and Applications **29**(6), 6–13 (2009)
29. Mou, L., Lin, J., Zhao, Y., Liu, Y., Ma, S., Zhang, J., Lv, W., Zhou, T., Liu, J., Frangi, A.F., et al.: Costa: A multi-center tof-mra dataset and a style self-consistency network for cerebrovascular segmentation. IEEE transactions on medical imaging **43**(12), 4442–4456 (2024)
30. Nan, Y., Xing, X., Wang, S., Tang, Z., Felder, F.N., Zhang, S., Ledda, R.E., Ding, X., Yu, R., Liu, W., et al.: Hunting imaging biomarkers in pulmonary fibrosis: benchmarks of the aiib23 challenge. Medical Image Analysis **97**, 103253 (2024)
31. Ortiz-Puerta, D., Diaz, O., Retamal, J., Hurtado, D.E.: Morphometric analysis of airways in pre-copd and mild COPD lungs using continuous surface representations of the bronchial lumen. Frontiers in Bioengineering and Biotechnology **11**, 1271760 (2023)

32. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023)
33. Prabhakar, C., Shit, S., Musio, F., Yang, K., Amiranashvili, T., Paetzold, J.C., Li, H.B., Menze, B.: 3d vessel graph generation using denoising diffusion. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–13. Springer (2024)
34. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
35. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16560–16569 (2021)
36. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
37. Støverud, K.H., Bouget, D., Pedersen, A., Leira, H.O., Langø, T., Hofstad, E.F.: Aeropath: An airway segmentation benchmark dataset with challenging pathology. arXiv preprint arXiv:2311.01138 (2023)
38. Todorov, M.I., Paetzold, J.C., Schoppe, O., Tetteh, G., Shit, S., Efremov, V., Todorov-Völgyi, K., Düring, M., Dichgans, M., Piraud, M., et al.: Machine learning analysis of whole mouse brain vasculature. Nature methods **17**(4), 442–449 (2020)
39. Vignac, C., Osman, N., Toni, L., Frossard, P.: Midi: Mixed graph and 3d denoising diffusion for molecule generation. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 560–576. Springer (2023)
40. Walsh, C.L., Tafforeau, P., Wagner, W., Jafree, D., Bellier, A., Werlein, C., Kühnel, M., Boller, E., Walker-Samuel, S., Robertus, J., et al.: Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. Nature methods **18**(12), 1532–1541 (2021)
41. Wang, L., Hassani, K., Zhang, S., Fu, D., Yuan, B., Cong, W., Hua, Z., Wu, H., Yao, N., Long, B.: Learning graph quantized tokenizers. arXiv preprint arXiv:2410.13798 (2024)
42. Wang, X., Liu, E., Wu, Z., Zhai, F., Zhu, Y.C., Shui, W., Zhou, M.: Skeleton-based cerebrovascular quantitative analysis. BMC medical imaging **16**(1), 68 (2016)
43. Wittmann, B., Wattenberg, Y., Amiranashvili, T., Shit, S., Menze, B.: vesselFM: A Foundation Model for Universal 3D Blood Vessel Segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20874–20884 (2025)
44. Wolterink, J.M., Leiner, T., Isgum, I.: Blood vessel geometry synthesis using generative adversarial networks. arXiv preprint arXiv:1804.04381 (2018)
45. Xia, L., Kao, B., Huang, C.: Opengraph: Towards open graph foundation models. arXiv preprint arXiv:2403.01121 (2024)
46. Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., Chen, D., Tong, X., Yang, J.: Structured 3d latents for scalable and versatile 3d generation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 21469–21480 (2025)
47. Yagis, E., Aslani, S., Jain, Y., Zhou, Y., Rahmani, S., Brunet, J., Bellier, A., Werlein, C., Ackermann, M., Jonigk, D., et al.: Deep learning for vascular segmentation and applications in phase contrast tomography imaging. arXiv preprint arXiv:2311.13319 (2023)
48. Yang, K., Musio, F., Ma, Y., Juchler, N., Paetzold, J.C., Al-Maskari, R., Höher, L., Li, H.B., Hamamci, I.E., Sekuboyina, A., et al.: Benchmarking the cow with the

topcow challenge: Topology-aware anatomical segmentation of the circle of willis for CTA and MRA. arXiv preprint arXiv:2312.17670 (2025)
49. Yang, L., Tian, Y., Xu, M., Liu, Z., Hong, S., Qu, W., Zhang, W., Cui, B., Zhang, M., Leskovec, J.: Vqgraph: Rethinking graph representation space for bridging gnns and mlps. arXiv preprint arXiv:2308.02117 (2023)
50. Zhang, B., Tang, J., Niessner, M., Wonka, P.: 3DShape2VecSet: A 3D Shape Representation for Neural Fields and Generative Diffusion Models. ACM Transactions On Graphics (TOG) **42**(4), 1–16 (2023)
51. Zhang, J., Xiong, F., Xu, M.: 3d representation in 512-byte: Variational tokenizer is the key for autoregressive 3d generation. arXiv preprint arXiv:2412.02202 (2024)
52. Zhang, J., Xiong, F., Xu, M.: G3pt: Unleash the power of autoregressive modeling in 3d generation via cross-scale querying transformer. arXiv preprint arXiv:2409.06322 (2024)
53. Zhang, M., Wu, Y., Zhang, H., Qin, Y., Zheng, H., Tang, W., Arnold, C., Pei, C., Yu, P., Nan, Y., et al.: Multi-site, multi-domain airway tree modeling. Medical image analysis **90**, 102957 (2023)
54. Zhang, M., Zhang, H., You, X., Yang, G.Z., Gu, Y.: Implicit representation embraces challenging attributes of pulmonary airway tree structures. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 546–556. Springer (2024)
55. Zhao, Z., Lai, Z., Lin, Q., Zhao, Y., Liu, H., Yang, S., Feng, Y., Yang, M., Zhang, S., Yang, X., et al.: Hunyuan3d 2.0: Scaling diffusion models for high resolution textured 3d assets generation. arXiv preprint arXiv:2501.12202 (2025)
56. Zhao, Z., Liu, W., Chen, X., Zeng, X., Wang, R., Cheng, P., Fu, B., Chen, T., Yu, G., Gao, S.: Michelangelo: Conditional 3d shape generation based on shape-image-text aligned latent representation. Advances in neural information processing systems **36**, 73969–73982 (2023)