MANIFOLD-AWARE PERSISTENT HOMOLOGY REPRESENTATION FOR POINT CLOUD REGISTRATION WITH TOPOLOGY PRESERVATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Point clouds sampled from real-world objects often exhibit complex non-convex geometry, in which global structure such as connectivity, holes and part adjacency is essential for shape understanding, yet topological analysis based on Euclidean distances measures proximity in ambient space rather than intrinsic connectivity and can create shortcuts across narrow gaps or concave boundaries. We introduce a manifold-aware topological representation that computes persistent homology from geodesic distances to provide faithful global shape cues, and condition a registration network on it: TopoReg routes the case-level descriptor into a standard coarse-to-fine backbone at a single deep level, so that the gain comes from the representation rather than from added capacity. We further prove that any stable descriptor built on the Euclidean filtration changes by at most an amount proportional to the gap width when a narrow gap is bridged, whereas the graph geodesic metric separates the two sides by an unbounded amount. Experiments on the large-scale lung CT dataset Lung250M-4B and the KITTI autonomous driving dataset show state-of-the-art performance and robust generalization.

1 INTRODUCTION

Point cloud registration is a cornerstone of medical image analysis, enabling intraoperative navigation, longitudinal disease monitoring and radiation therapy planning. Given two point clouds representing the same anatomical structure at different time points, the goal is to recover a dense deformation field that accounts for soft-tissue motion. Despite decades of research, robust registration of organs with complex topology, such as the lung's branching vascular trees or the liver's vasculature, remains an open challenge Zhang et al. (2026a).

Classical methods Besl & McKay (1992); Myronenko & Song (2010b) and modern deep networks Liu et al. (2019); Wu et al. (2020), together with recent variants that adopt more powerful deformation models Lu et al. (2025); Jiang et al. (2025); Liu et al. (2025a); Wang et al. (2025); Muniya et al. (2026); Yuan et al. (2025); Fang & Chen (2025), all construct local neighborhoods from Euclidean distances. On non-convex manifolds with closely apposed but physically separate surfaces—pulmonary fissures separating lung lobes, or vessel walls enclosing blood flow—Euclidean k-nearest-neighbor graphs therefore generate spurious edges that cross anatomical boundaries, and these shortcuts corrupt the correspondence field by forcing structures that should slide independently to fuse, yielding deformation fields that are anatomically implausible.

Persistent homology characterises multi-scale topological structure by tracking the birth and death of connected components, loops and voids across a filtration, and the resulting diagrams are stable, provably robust descriptors of shape topology Cohen-Steiner et al. (2007); Adams et al. (2017). It has since been extended to feature matching Kerber & Xu (2025) and topology-aware transformers Pan et al. (2026), and surveys have summarised its role in optimization Carriere et al. (2026a). Yet despite its growing use in shape analysis and segmentation Dey & Wang (2022); Carlsson (2009); Oudot (2015) and the emergence of topology-preserving registration methods **?**Xu et al. (2025); Chen et al. (2025b); Wang et al. (2026); Bai et al. (2025); Serrano de Haro Iváñez et al. (2025); Miura et al. (2025); Zhang et al. (2026b); Xia et al. (2026); Carriere et al. (2026b), persistent homology remains largely unexplored for non-rigid registration, the barrier being that standard Vietoris–Rips

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

filtrations (Edelsbrunner & Harer, 2010) are built on Euclidean distances and therefore inherit the very shortcut problem they are meant to solve. A transition to manifold-aware distance metrics is thus necessary but has not been systematically investigated.

We introduce TopoReg, a manifold-aware representation of point-cloud structure together with a mechanism that conditions a registration network on it; Figure 1 gives an overview. Our contributions are as follows:

- A manifold-aware topological representation is proposed to compute persistent homology from geodesic distances, capturing the intrinsic connectivity and holes of non-convex shapes rather than ambient proximity, and thereby provides faithful global shape cues for registration.
- A conditioning scheme is designed to let the case-level descriptor modulate local features at a single deep level of a standard backbone, so that the gain comes from the representation rather than from added capacity.
- A proof that any stable descriptor built on the Euclidean filtration changes by at most an amount proportional to the gap width when a narrow gap is bridged, whereas the graph geodesic metric separates the two sides by an unbounded amount.
- The lowest mean TRE on the large-scale lung CT dataset Lung250M-4B and robust generalisation to the KITTI autonomous driving dataset, indicating broad clinical and robotic applicability.

2 Related Work

2.1 Topological Methods in Point Cloud Registration

Persistent homology captures multi-scale structure such as connected components, loops and voids, and yields stable, interpretable descriptors for 3D data Adams et al. (2017); Carlsson (2009). Topology-aware registration frameworks preserve topological invariants during deformation or use skeleton-based features to guide correspondence **?**Xu et al. (2025); Chen et al. (2025a), and topology-informed medical networks enforce anatomical consistency Serrano de Haro Iváñez et al. (2025); Miura et al. (2025). Most such approaches nevertheless build their filtrations from Euclidean distances, which introduce shortcuts across non-convex boundaries; our geodesic-aware persistent homology module addresses this limitation.

2.2 Geometric Deep Learning for Point Cloud Registration

Point cloud registration has been studied extensively in computer vision and medical image analysis. Classical methods such as Coherent Point Drift Myronenko & Song (2010a) and iterative closest point variants Besl & McKay (1992) have been largely superseded by deep networks that learn feature representations and flow estimation from large-scale data, among which PointPWC Wu et al. (2020) introduces a coarse-to-fine cost-volume architecture; subsequent work has extended this paradigm with transformers Vaswani et al. (2017), state space models Liang et al. (2024a) and clustering-based correspondence Zhao et al. (2024b), while VoxelMorph++ Heinrich & Hansen (2022) achieves strong accuracy for intra-patient lung CT alignment. Most of these methods still construct neighbourhoods and smoothness terms from Euclidean distances, which fails near anatomical fissures where sliding motion occurs; we address this by supplying the network with a manifold-aware representation of global structure rather than by enlarging the backbone.

3 Method

3.1 Motivation

Deep registration networks represent a point cloud through features built from local Euclidean neighbourhoods, whose receptive fields remain local and whose values vary from point to point. Such features lack a case-level description of global structure (connectivity, holes and the adjacency of

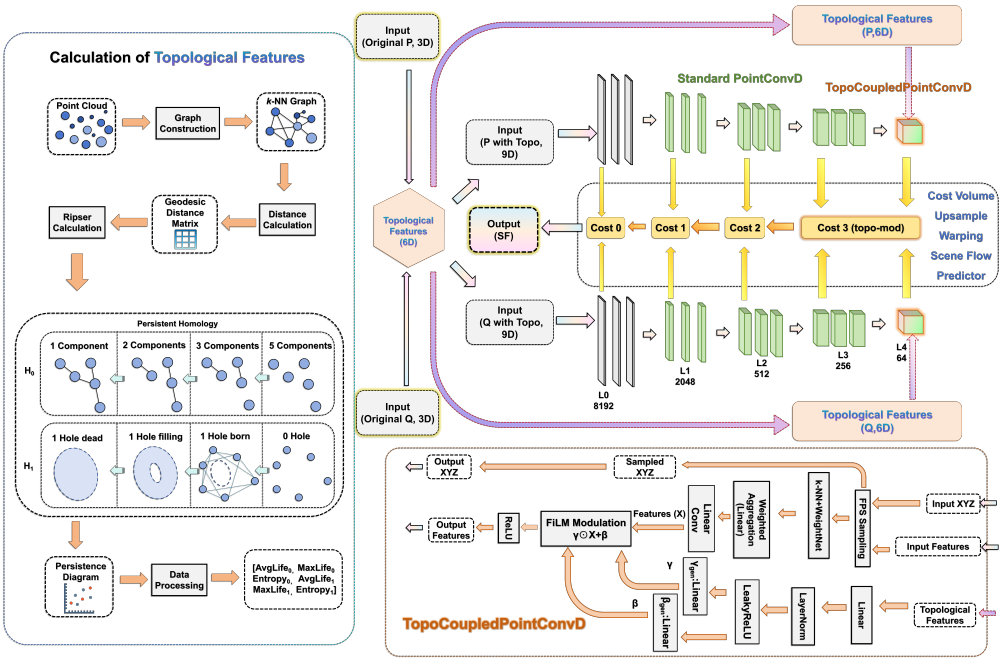


Figure 1: Overview of TopoReg.

parts), which distinguishes anatomically complex shapes from simple ones. Because Euclidean distance measures proximity in ambient space rather than intrinsic connectivity, a global descriptor derived from it inherits the shortcuts that cross narrow gaps or concave boundaries, and summarises the accidental geometry of the sampling rather than the organisation of the shape. We therefore design the descriptor on an intrinsic, manifold-aware metric.

3.2 Manifold Topology Representation

We represent each point cloud as a discrete sample of an underlying three-dimensional manifold. Let $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^{n}$ denote an input point cloud, where $\mathbf{p}_i \in \mathbb{R}^3$. To keep the representation compact and the cost of topological analysis bounded, at most 500 points are drawn at random from $\mathcal{P}$, yielding a subsampled cloud $\widetilde{\mathcal{P}}$.

A k-NN graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is then constructed on $\widetilde{\mathcal{P}}$, where $k = 20$, each edge being weighted by the Euclidean distance between its incident points. The metric under which a representation is built decides which parts of a shape count as close, and we therefore replace these direct by shortest-path distances. Euclidean distances admit shortcuts across non-convex boundaries, which approximate the geodesic distances of the underlying manifold. The resulting matrix preserves the local connectivity of the sampled manifold and suppresses the non-manifold shortcuts of ambient Euclidean space.

Given the geodesic distance matrix, we construct a Vietoris–Rips filtration: at scale ϵ the complex contains every simplex whose pairwise geodesic distances are at most ϵ, and increasing ϵ yields a nested sequence of simplicial complexes. The filtration is what makes the representation multi-scale, a feature being retained only if it survives over a range of scales. For homology dimension d, each such feature is a persistence interval with birth and death scales $b_j^{(d)}$ and $\delta_j^{(d)}$, whose persistence lifetime is $\ell_j^{(d)} = \delta_j^{(d)} - b_j^{(d)}$.

A large lifetime indicates that the corresponding structure persists over a wide range of scales and is therefore more significant topologically. We use the 0- and 1-dimensional homology groups, the former describing connected components and the latter loop-like structures and holes.

To obtain a fixed-dimensional representation, three statistics are extracted from the persistence lifetimes of each homology dimension; write $\mathcal{I}_d$ for the index set of its finite lifetimes. Since a diagram is a set, every statistic of its lifetimes is invariant to the ordering of the input points, and the average and maximum persistence are

$$\mathrm{AvgLife}_d = \frac{1}{|\mathcal{I}_d|} \sum\nolimits_{j \in \mathcal{I}_d} \ell_j^{(d)}, \qquad \mathrm{MaxLife}_d = \max_{j \in \mathcal{I}_d} \ell_j^{(d)}. \tag{1}$$

To characterize the distribution of persistence across topological features, we normalize each lifetime and take the corresponding entropy,

$$q_j^{(d)} = \frac{\ell_j^{(d)}}{\sum_{r \in \mathcal{I}_d} \ell_r^{(d)}}, \qquad \mathrm{Ent}_d = -\sum\nolimits_{j \in \mathcal{I}_d} q_j^{(d)} \log\left(q_j^{(d)} + \varepsilon\right), \tag{2}$$

where ε is a small positive constant for numerical stability. Average persistence measures overall significance, maximum persistence identifies the most persistent structure, and entropy describes the diversity of persistence across structures.

The final manifold topology descriptor is obtained by concatenating the statistics from the 0- and 1-dimensional persistence diagrams:

$$\phi(P) = \big(\mathrm{AvgLife}_0,\, \mathrm{MaxLife}_0,\, \mathrm{Ent}_0,\, \mathrm{AvgLife}_1,\, \mathrm{MaxLife}_1,\, \mathrm{Ent}_1\big)^\top \in \mathbb{R}^6. \tag{3}$$

The descriptor is compact, order-invariant and, as Section A shows, stable with respect to the bottleneck distance, which is what allows it to serve as a conditioning signal rather than merely as a shape statistic. It is broadcast to all points and concatenated with the coordinates to form the per-point input

$$C = \big[\, P,\ \mathbf{1}\,\phi(P)^\top \big] \in \mathbb{R}^{n \times 9}, \tag{4}$$

so that the three coordinate channels are augmented by the six-dimensional descriptor and the global shape code is visible to the network from its first layer.

For a pair of point clouds $\mathcal{P}_1$ and $\mathcal{P}_2$, the representations $\phi(\mathcal{P}_1)$ and $\phi(\mathcal{P}_2)$ are computed independently through the same subsampling, graph construction, geodesic computation, persistent homology and aggregation pipeline. Unlike the per-point features produced by the convolution operators of Section 3.3, whose receptive fields are local, ϕ is a case-level variable summarising the connectivity and hole structure that persists across geometric scales; the two therefore describe a shape at different levels of granularity and are complementary rather than redundant.

3.3 Manifold Topology-based Registration

The descriptor ϕ of Section 3.2 is a case-level variable, whereas the features a registration backbone computes are local and vary with position. Using the representation therefore calls for a mechanism that lets a global variable reweight local features rather than a new spatial operator. We realise it with Feature-wise Linear Modulation (FiLM) at a single deep level of a standard coarse-to-fine backbone Wu et al. (2020), leaving all other levels and the cost-volume/flow modules unchanged.

Feature Pyramid. Given normalised inputs $X_s, X_t \in \mathbb{R}^{n \times 9}$, we embed them via two shared $1{\times}1$ convolutions and build a five-level pyramid via farthest point sampling (FPS) and PointConvD aggregation. The point counts are $8192, 2048, 512, 256, 64$ from finest to coarsest, with progressively increasing channel widths.

Topology-FiLM Coupling. At the deepest pyramid level (64 points), we replace PointConvD with TopoCoupledPointConvD, which performs standard PointConv aggregation:

$$h_i^l = \sum\nolimits_{j \in \mathcal{N}_K(i)} \omega_\theta(\Delta p_{ij})\, \psi_\theta\big([h_j^{l-1}, \Delta p_{ij}]\big), \tag{5}$$

where $\Delta p_{ij} = p_i - p_j$ is the relative coordinate offset in the K-nearest neighbourhood. A FiLM generator $\mathcal{F}_\theta$ maps the 6-dimensional topology vector ϕ to per-channel scale and shift parameters $\gamma, \beta \in \mathbb{R}^C$ through a hidden layer of width 64, normalised before the nonlinearity,

$$t = \sigma\big(\mathrm{LN}(W_1\phi + b_1)\big), \qquad \gamma = W_\gamma t + b_\gamma, \qquad \beta = W_\beta t + b_\beta, \tag{6}$$

with σ the leaky ReLU and LN layer normalisation, followed by the modulation $\hat{h}_i^l = \gamma \odot h_i^l + \beta$. The modulation is multiplicative because a per-channel scale acts as a soft gate selecting which channels carry global shape information, whereas a shift could only offset them; and since γ and β depend on a case-level variable, the same weights realise a different transformation for each input, so shape-dependent behaviour needs no additional spatial operator. Initialising the generator with $W_\gamma = \mathbf{0}$, $b_\gamma = \mathbf{1}$, $W_\beta = \mathbf{0}$ and $b_\beta = \mathbf{0}$ gives $\gamma = \mathbf{1}$, $\beta = \mathbf{0}$ initially, so topology enters gradually rather than through a randomly initialised modulation. BatchNorm is disabled within the module, as it conflicts with topology-conditioned feature scaling, and FiLM is applied only at this deepest level (Table 3).

The descriptor reaches the network through three routes: equation (4) augments the input coordinates, so the global code is visible from the first layer; the FiLM block above modulates the deepest level, a route which survives the spatial downsampling applied between levels; and the topology consistency loss of Section 3.4 constrains the warped geometry during optimisation. The three act on input channels, intermediate activations and the training signal respectively, and are complementary rather than redundant.

Cost Volume and Flow Estimation. Cross-frame matching at each level is performed via a learned PointConvFlow cost volume, and scene flow is predicted coarse-to-fine from the deepest level and upsampled between levels by inverse-distance weighting over the three nearest sparse anchors,

$$\tilde{F}_i = \sum\nolimits_{j \in \mathcal{N}_3(i)} w_{ij} F_j, \qquad w_{ij} = \frac{1/d_{ij}}{\sum_{l \in \mathcal{N}_3(i)} 1/d_{il}}, \tag{7}$$

with $d_{ij} = \|p_i - p_j\|_2$. The target is warped at each stage to concentrate subsequent matching on residual motion, and the flows predicted at four pyramid levels (L3 to L0) are supervised jointly.

3.4 Training Strategy

Multi-scale Flow Loss. We supervise at all pyramid levels using the multi-scale ℓ_2 loss between predicted and ground-truth flow, the ground truth being downsampled with the same FPS indices used to construct the pyramid.

Topology Consistency Loss. A topology consistency loss is essential for training stability, its removal degrading TRE from 2.18 to 2.39 mm (Table 3, lower block). We sample anchor points on the source cloud by deterministic linear spacing and construct k-nearest-neighbour graphs in Euclidean space around each anchor, so that for every anchor a and neighbour j we may compare the edge length under the predicted flow with that under the ground-truth flow:

$$\mathcal{L}_{\text{topo}} = \frac{1}{|\mathcal{A}| \cdot k} \sum\nolimits_{a \in \mathcal{A}} \sum\nolimits_{j \in \mathcal{N}_k(a)} \ell\big(\|p_{s,a}^{\text{pred}} - p_{s,j}^{\text{pred}}\|_2,\ \|p_{s,a}^{\text{gt}} - p_{s,j}^{\text{gt}}\|_2\big), \tag{8}$$

where $p_{s,i}^{\text{pred}} = p_{s,i} + \hat{F}_i$, $p_{s,i}^{\text{gt}} = p_{s,i} + F_{\text{GT},i}$, $\mathcal{A}$ is the anchor set and ℓ the smooth ℓ_1 loss. Because both warped positions are built from the same source samples, every local edge retains a valid point-wise correspondence however the source and target clouds were independently sampled. The total objective is $\mathcal{L} = \mathcal{L}_{\text{flow}} + \lambda_{\text{topo}} \mathcal{L}_{\text{topo}}$, and optimisation uses Adam with a multi-step learning rate schedule and gradient clipping, the topology loss weight being warmed up linearly from zero over the initial epochs; full hyper-parameters are provided in the supplementary material.

Inference. At test time, the predicted flow $\hat{F}$ in normalised space is mapped back to original coordinates and composed with the pre-alignment displacement: $F = F_{\text{pre}} + S \cdot \hat{F}$, yielding the final warped source $P_s^{\text{warp}} = P_s + F$. For landmark-based evaluation, the predicted displacement field is interpolated to the landmark positions via thin-plate spline interpolation and the target registration error (TRE) is computed as the Euclidean distance between warped and target landmarks in physical millimetres.

4 Theoretical Analysis

Our theory establishes an asymmetry between the Euclidean filtration and the graph geodesic metric. For any admissible descriptor built on the Euclidean Vietoris–Rips filtration, bridging a narrow gap changes the descriptor by at most a quantity proportional to the gap width w, whereas the graph

geodesic metric separates the two sides by an amount that is unbounded relative to w; this motivates conditioning the FiLM module on ϕ_G rather than ϕ_E, the statements below proving only that no admissible Euclidean descriptor can register the bridging by more than $\frac{1}{2}C_0 w$.

Let $P \subset \mathbb{R}^3$ be a finite point cloud with $|P| = n$. Given P, build the Euclidean k-nearest-neighbour graph $G_k(P) = (P, E)$ and define the graph geodesic metric (a metric whenever $G_k(P)$ is connected)

$$d_G(x, y) \;=\; \min_{\pi:\, x \rightsquigarrow y} \sum\nolimits_{(u,v)\in\pi} \|u - v\|_2, \tag{9}$$

where the minimum is taken over paths in $G_k(P)$. Let d_E denote the ambient Euclidean metric. For a metric d on P, let $\mathcal{D}_j(P, d)$ be the j-th persistence diagram of the Vietoris–Rips filtration, and define

$$\phi(P, d) = \big(\mathrm{AvgLife}_0, \mathrm{MaxLife}_0, \mathrm{Ent}_0, \mathrm{AvgLife}_1, \mathrm{MaxLife}_1, \mathrm{Ent}_1\big)(P, d), \tag{10}$$

using only finite lifetimes. Write $\phi_E := \phi(\cdot, d_E)$ and $\phi_G := \phi(\cdot, d_G)$. In the implementation, the essential H_0 class is retained with a death time of $1.1\times$ the largest finite death. $\mathrm{MaxLife}_0$ as computed is therefore a strictly increasing affine function of $\mathrm{MaxLife}_0$ as defined above, so the separation argument below, which uses only that component, is unaffected.

Definition 4.1 (Admissible filtration descriptor). *Fix $C_0 > 0$ and a metric d on P. A descriptor ψ is C_0-admissible for d if it factors through the persistence diagram, $\psi(P, d) = F(\mathcal{D}_\bullet(P, d))$, and is C_0-Lipschitz with respect to bottleneck distance:*

$$\|\psi(P, d) - \psi(Q, d)\|_2 \;\le\; C_0 \cdot \max_j d_B\big(\mathcal{D}_j(P, d), \mathcal{D}_j(Q, d)\big). \tag{11}$$

Theorem 4.1 (Stability of lifetime statistics). *Let P, Q be point clouds with $|P| = |Q| = n$, and let $\delta := \max_j d_B(\mathcal{D}_j(P, d), \mathcal{D}_j(Q, d))$. Then*

$$|\mathrm{MaxLife}_j(P) - \mathrm{MaxLife}_j(Q)| \;\le\; 2\delta, \tag{12}$$

with constant independent of n, $L_{\min}$ and $L_{\max}$. If the two diagrams contain equally many finite lifetimes and all finite lifetimes lie in $[L_{\min}, L_{\max}]$ with $L_{\min} > 0$, then $\mathrm{AvgLife}_j$ is also 2-Lipschitz and Ent_j is Lipschitz with constant

$$C_{\mathrm{Ent}} = \frac{2\big(2\log n + \log(L_{\max}/L_{\min})\big)}{L_{\min}}. \tag{13}$$

Theorem 4.2 (Separation by the graph geodesic metric). *Let $A, B \subset P$ be disjoint, with gap width $w := \min_{a\in A, b\in B} \|a - b\|_2$ and detour length $L := \min_{a\in A, b\in B} d_G(a, b)$. Let r be the maximum edge length of $G_k(P)$. Assume* (H1) *$r < w$ and* (H2) *$G_k(P)$ is connected. Then no edge joins A to B, and*

$$\min_{a\in A, b\in B} d_E(a, b) = w \;<\; L = \min_{a\in A, b\in B} d_G(a, b). \tag{14}$$

Moreover, for every $C > 1$ there is a configuration satisfying (H1)–(H2) *with $L/w > C$.*

Theorem 4.3 (Limits of the Euclidean filtration). *Let P, A, B, w, L be as above, and let (a^*, b^*) realise the gap width. Define the bridged point cloud*

$$Q = P \cup \{z_1, \dots, z_m\}, \qquad z_i \in [a^*, b^*], \qquad \|z_i - z_{i+1}\|_2 < s < w, \tag{15}$$

the z_i being ordered along the segment with z_1 adjacent to a^* and z_m to b^*, and with Q again satisfying (H1)–(H2). *Then:*

- *(i) $d_{GH}\big((P, d_E), (Q, d_E)\big) \le w/2$, and consequently every C_0-admissible Euclidean descriptor satisfies*
$$\big\|\psi(P, d_E) - \psi(Q, d_E)\big\|_2 \;\le\; \tfrac{1}{2}C_0 w. \tag{16}$$
- *(ii) while $\min_{a\in A, b\in B} d_G(a, b) = L$ in P, the added points form an a^*–b^* path of length at most w, so $\min_{a\in A, b\in B} d_G(a, b) \le w$ in Q.*

The graph geodesic metric, by contrast, changes by at least $L - w$, which can be made arbitrarily large relative to w. Full proofs are in Appendix A–C.

Table 1: Performance comparison on the Lung250M-4B dataset.

Methods	Mean(mm)	25th Percentile(mm)	Median(mm)	75th Percentile(mm)	Min(mm)	Max(mm)
SVM	10.65	6.24	9.70	14.17	0.70	35.68
Random Forest	16.25	10.14	15.94	21.76	1.04	42.92
Transformer Vaswani et al. (2017) (NeurIPS'17)	5.34	2.55	4.32	7.17	0.17	35.09
PLADE Chen et al. (2020) (TGRS'20)	10.96	6.68	10.06	14.35	0.24	35.34
NIE Jiang et al. (2023) (CVPR'23)	7.12	2.83	5.16	9.51	0.17	41.22
KAN Liu et al. (2025b) (ICLR'25)	10.40	6.28	9.28	13.44	0.61	35.14
PointPWC Wu et al. (2020) (ECCV'20)	2.21	1.16	1.77	2.71	0.11	17.97
TopoReg (Ours)	**2.18**	**1.12**	**1.75**	**2.62**	**0.15**	**17.15**

Table 2: Performance comparison on the KITTI dataset.

Method	Mean (m)	25th pct. (m)	Median (m)	75th pct. (m)	Min (m)	Max (m)
SVM	1.9951	0.7703	1.2655	2.5220	0.1584	11.6227
Random Forest	1.8208	0.7071	1.1314	2.1932	0.1598	11.8106
NIE Jiang et al. (2023) (CVPR'23)	0.2433	0.2122	0.2420	0.2762	0.0913	0.4250
ClusterReg Zhao et al. (2024a) (CVPR'24)	0.3044	0.2689	0.3017	0.3403	0.1182	0.4521
PointMamba Liang et al. (2024b) (NeurIPS'24)	0.2206	0.1780	0.2095	0.2538	0.0954	0.5737
3DFMNet Zhang et al. (2024) (NeurIPS'24)	1.4694	0.3438	0.4818	1.0730	0.0949	45.3346
TurboReg Yan et al. (2025) (ICCV'25)	0.3444	0.2122	0.2784	0.3739	0.0931	11.0393
PointPWC Wu et al. (2020) (ECCV'20)	0.2394	0.1856	0.2157	0.2765	0.1035	0.5008
KAN Liu et al. (2025b) (ICLR'25)	1.9350	1.2858	1.9549	2.4984	0.4798	3.6113
TopoReg (Ours)	**0.1944**	**0.1356**	**0.1612**	**0.1884**	**0.0847**	**0.3533**

5 Experimental Results

5.1 Datasets and Settings

We evaluate on two datasets. Lung250M-4B Falta et al. (2023) contains 124 paired 3D CT-derived lung point clouds across respiratory phases (deformation $>$30 mm) with ground-truth landmarks for non-rigid registration; models are trained for 1,500 epochs on NVIDIA GeForce L40 GPUs, the 124 cases being split into 97 training cases together with validation and test sets, and each baseline is trained with its original hyperparameters for fair comparison. KITTI Geiger et al. (2012) provides dense LiDAR point clouds with 3D motion ground truth, for which we follow the standard split and preprocessing.

5.2 Overall Performance

On Lung250M-4B (Table 1), TopoReg achieves a mean TRE of 2.18 mm, outperforming PointPWC (2.21 mm) Wu et al. (2020) and all other methods, while the median falls from 1.77 to 1.75 mm and the 75th percentile from 2.71 to 2.62 mm. Deformation analysis confirms physical plausibility: displacements are moderate (median 11.91 mm) with a sharply peaked smoothness distribution (median 1.76), and the Z-component dominates (Fig. 4b,c), consistent with diaphragm-driven breathing.

On KITTI (Table 2), TopoReg attains a mean Chamfer distance of 0.1944 m, an 18.8% reduction over PointPWC (0.2394 m) Wu et al. (2020), with the median improving from 0.2157 to 0.1612 m and the maximum falling from 0.5008 to 0.3533 m, indicating consistent rather than outlier-driven gains.

5.3 Ablation Studies

We ablate four design choices on the Lung250M-4B test set, all results being reported after full 1,500-epoch convergence. Table 3 (upper block) examines the injection level. L4-only conditioning yields the best TRE (2.18 mm), whereas shallower or multi-level injection degrades performance, all-level conditioning reaching 2.49 mm. The lower block evaluates coupling mechanisms, in which pure multiplicative FiLM outperforms residual FiLM with BatchNorm ($+0.07$ mm) and low-rank adapters ($+0.06$–0.11 mm), while the largest degradation follows from removing $\mathcal{L}_{\text{topo}}$ ($+0.21$ mm).

Table 3: Architecture ablation under a fixed training seed (seed 123) on both datasets. The upper block varies the FiLM injection level; the lower block evaluates coupling mechanisms and design choices at L4. Lung is landmark TRE (mm); KITTI is Chamfer distance (m).

	Configuration	Level	Points	TRE (mm)	CD (m)
Injection level ablation (pure $\gamma \cdot x + \beta$ FiLM)	PointPWC (baseline)	—	—	2.21	**0.2343**
	Ours (L4-FiLM)	L4	64	**2.18**	**0.1944**
	L1-FiLM	L1	2048	2.29	0.2201
	L2-FiLM	L2	512	2.37	0.2377
	L3-FiLM	L3	256	2.38	0.1973
	L1+L4-FiLM	L1, L4	—	2.36	0.1991
	L3+L4-FiLM	L3, L4	—	2.37	0.2307
	All-level FiLM	L1–L4	—	2.49	0.2228
Coupling mechanism & design choice ablation (at L4)	**Ours (pure $\gamma \cdot x + \beta$, no BN)**	L4	64	**2.18**	**0.1944**
	Residual FiLM + BN	L4	64	2.25	0.2264
	Residual FiLM (hidden 32)	L4	64	2.22	0.2290
	Low-rank + BN ($s = 0.5$)	L4	64	2.29	0.2395
	Low-rank + BN (rank=1)	L4	64	2.29	0.2329
	Low-rank, no BN	L4	64	2.24	0.2393
	L3+L4 FiLM	L3, L4	—	2.37	0.2307
	All-level residual FiLM	L1–L4	—	2.35	0.2271
	w/o $\mathcal{L}_{\text{topo}}$	L4	64	2.39	0.2428
	w/ Euclidean filtration (vs. geodesic)	L4	64	2.22	0.2200

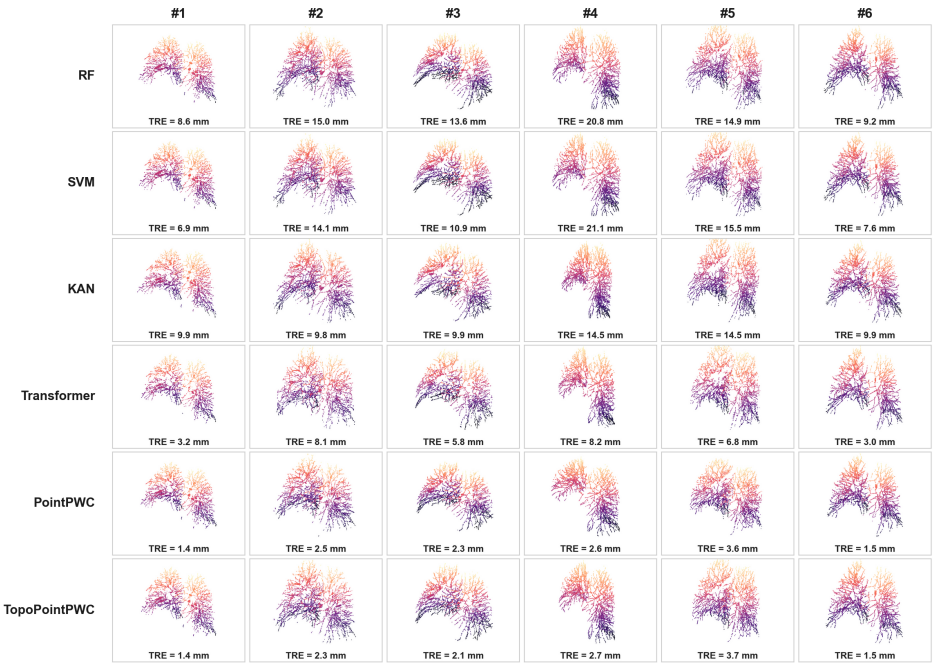


Figure 2: Visual comparison of performance with other models on the Lung250M-4B dataset.

5.4 Visualization Analysis

Figure 2 compares TopoReg with SVM, Random Forest, Transformer and PointPWC over six representative Lung250M-4B pairs, on which the Euclidean-based methods produce spurious cross-boundary artefacts near pulmonary fissures and vascular bifurcations, fusing lobes that should slide independently; TopoReg suppresses these shortcuts and preserves lobar boundaries, in agreement with the lowest mean TRE in Table 1.

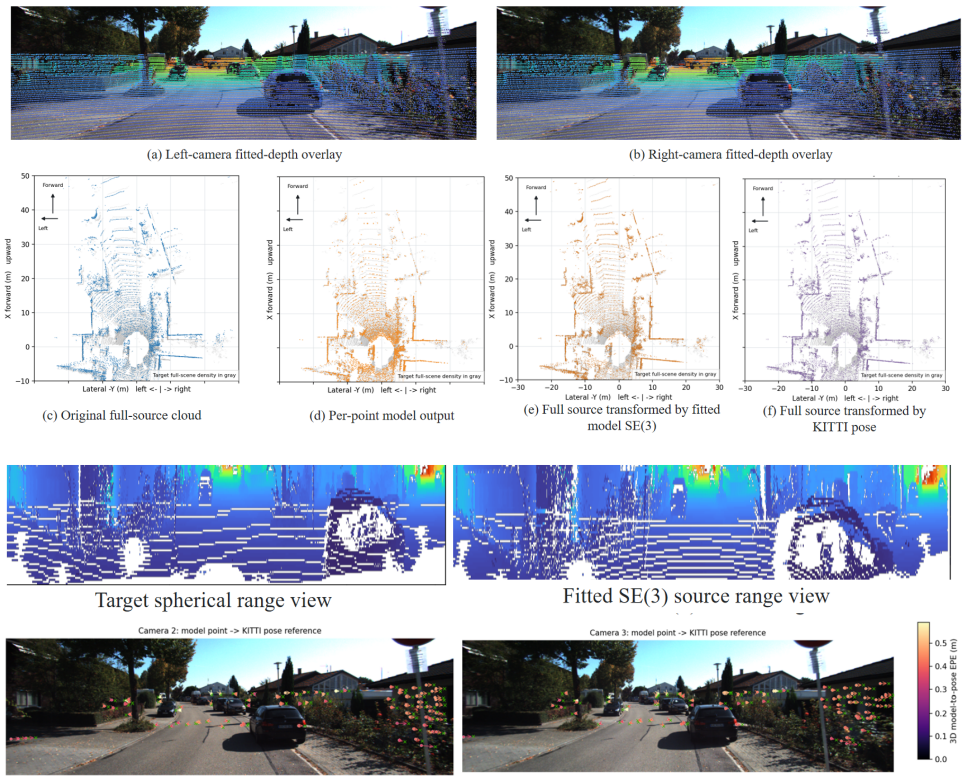


Figure 3: Qualitative results on the KITTI dataset. Top: visual comparison of performance with other methods. Bottom: registration of a KITTI Odometry frame pair, showing the target and fitted SE(3) spherical range views together with the projected point clouds.

Figure 3 extends the evaluation to KITTI. Its upper panel compares fitted-depth overlays and the sources transformed by the fitted SE(3) and by the ground-truth pose, where the prediction yields smoother transitions at dynamic object boundaries; its lower panel shows the corresponding spherical range views and projections from Cameras 2 and 3, both consistent with the reference pose.

5.5 Effectiveness of the Manifold Topology Representation

We evaluate the manifold topology representation independently of the registration backbone. Table 4 shows that geodesic k-NN graphs on Lung250M-4B have $11.9\times$–$20.6\times$ larger diameters than Euclidean ones, with $2.5\times$–$3.6\times$ higher long-edge ratios, indicating that shortcut edges across non-convex boundaries are suppressed (Figure 4a). Table 5 shows correspondingly larger H_1 and H_2 persistence gaps and higher topological SNR for all $k \in \{6, 20, 30\}$ ($p < 0.0001$), the H_2 gap increasing by 2.4–$2.8\times$; the advantage vanishes at $k = 30$, as larger neighbourhoods make geodesics approach Euclidean distances. The default $k = 20$ retains significant gains and is used for ϕ_G, whose conditioning yields consistent improvements in Table 3.

6 Conclusion

We presented TopoReg, which represents a point cloud by the persistent homology of a geodesic filtration and conditions a registration network on that representation. Geodesic distances replace Euclidean ones, removing shortcuts across non-convex boundaries and preserving structures such as pulmonary fissures, while our analysis characterises descriptor stability and the separation carried by the geodesic metric. Experiments on Lung250M-4B and KITTI show strong accuracy and generalisation, indicating that global topological priors can be incorporated into geometric deep learning for clinical and robotic settings where anatomical integrity is essential.

AI USE STATEMENT

Generative AI tools were used solely to assist with LaTeX typesetting, the formatting and organization of mathematical expressions, and limited language polishing. The authors reviewed these edits and take full responsibility for the scientific content and final manuscript.

REFERENCES

H. Adams, T. Emerson, M. Kirby, R. Neville, J. Singer, H. Carmesin, Y. Mileyko, and P. Bubenik. Persistence images: A stable vector representation of persistent homology. *J. Mach. Learn. Res.*, 18(8):1–35, 2017.

Y. Bai, J. Chang, and H. Zhou. Topology-preserving implicit template learning via bidirectional deformation for 3d shape reconstruction and registration. In *Proc. Int. Conf. Multimed. Expo (ICME)*, 2025.

P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992.

G. Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46(2):255–308, 2009.

J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *Proc. ACM SIGGRAPH*, pp. 67–76, 2001.

M. Carriere, Y. Ike, T. Lacombe, and N. Nishikawa. Persistence-based topological optimization: a survey. *arXiv preprint arXiv:2603.24613*, 2026a.

M. Carriere, Y. Ike, T. Lacombe, and N. Nishikawa. Persistence-based topological optimization: a survey. *arXiv preprint arXiv:2603.24613*, 2026b.

Z. Chen, K. Sun, F. Yang, and W. Tao. Plade: A plane-based descriptor for point cloud registration with small overlap. *IEEE Trans. Geosci. Remote Sens.*, 58(4):2530–2540, 2020.

Z. Chen, Y. Li, and X. Wang. Skeleton-based robust registration framework for corrupted 3d point clouds. *Appl. Soft Comput.*, 181:113461, 2025a.

Z. Chen, Y. Li, and X. Wang. Skeleton-based robust registration framework for corrupted 3d point clouds. *Appl. Soft Comput.*, 181:113461, 2025b.

D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.

T. K. Dey and Y. Wang. *Computational Topology for Data Analysis*. Cambridge University Press, 2022.

H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, Providence, RI, 2010.

F. Falta, C. Großbröhm, A. Hering, A. Bigalke, and M. Heinrich. Lung250m-4b: A combined 3d dataset for ct- and point cloud-based intra-patient lung registration. In *Adv. Neural Inf. Process. Syst. (NeurIPS)*, volume 36, pp. 54819–54832, 2023.

H. Fang and Y. Chen. Non-rigid registration based on rotation-invariant feature flow. *IEEE Access*, 13:12500–12512, 2025.

A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3354–3361, 2012.

Mattias P. Heinrich and Lasse Hansen. Voxelmorph++: Going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation. *Biomedical Image Registration: 10th International Workshop, WBIR 2022, Munich, Germany, July 10–12, 2022, Proceedings*, pp. 85–95, 2022.

H. Jiang, Z. Deng, J. Shi, Z. Zhao, and G. Chen. Robust point cloud registration framework based on deep graph matching. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 21835–21844, 2023.

Y. Jiang, Y. Zhao, X. Lu, and J. Wang. Nfr: Neural feature-guided non-rigid shape registration. *arXiv preprint arXiv:2505.22445*, 2025.

M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proc. Symp. Geometry Processing*, pp. 61–70, 2006.

M. Kerber and C. Xu. Persistent feature matching. *J. Mach. Learn. Res.*, 2025.

D. Liang, X. Zhou, W. Xu, X. Zhu, Z. Zou, X. Ye, X. Tan, and X. Bai. Pointmamba: A simple state space model for point cloud analysis. In *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2024a.

D. Liang, X. Zhou, W. Xu, X. Zhu, Z. Zou, X. Ye, X. Tan, and X. Bai. Pointmamba: A simple state space model for point cloud analysis. In *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2024b.

J. Liu, W. Zhang, and Q. Chen. Deftransnet: A transformer-based method for non-rigid point cloud registration in the simulation of soft tissue deformation. *arXiv preprint*, 2025a.

X. Liu, C. R. Qi, and L. J. Guibas. Flownet3d: Learning scene flow in 3d point clouds. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 529–537, 2019.

Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark. Kan: Kolmogorov–arnold networks. In *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2025b.

X. Lu, J. Wang, M. Bennamoun, and B. Sheng. Non-rigid point cloud registration via anisotropic hybrid field harmonization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025.

S. Miura, M. Nakayama, and B. Zhang. Point cloud registration algorithm using liver vascular skeleton feature with computed tomography and ultrasonography image fusion. *Int. J. Comput. Assist. Radiol. Surg.*, 2025.

E. J. Muniya, W. Zhang, T. K. Lee, and S. Park. Deep learning based non-rigid volume-to-surface registration for brain shift compensation using point cloud. *arXiv preprint arXiv:2604.17389*, 2026.

A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12):2262–2275, 2010a.

A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12):2262–2275, 2010b.

S. Oudot. *Persistence Theory: From Quiver Representations to Data Analysis*. American Mathematical Society, 2015.

W. Pan, Y. Zhang, and S. Liu. Topformer: Topology-aware transformer for point cloud registration. In *Int. Conf. Learn. Represent. (ICLR)*, 2026.

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Trans. Med. Imag.*, 18 (8):712–721, 1999.

S. Serrano de Haro Iváñez, M. E. L. van Ginneken, and J. Schnabel. Toproi: A topology-informed network approach for tissue partitioning. *arXiv preprint arXiv:2510.12772*, 2025.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Adv. Neural Inf. Process. Syst. (NeurIPS)*, volume 30, pp. 5998–6008, 2017.

J. Wang, H. Zhang, and L. Wang. Robust point cloud registration via multi-scale perception and geometric support weighted svd. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2026.

S. Wang, L. Chen, and H. Li. Ernet: Efficient non-rigid registration network for point sequences. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2025.

W. Wu, Z. Y. Wang, Z. Li, W. Liu, and L. Fuxin. Pointpwc-net: Cost volume on point clouds for (self-)supervised scene flow estimation. In *Proc. Eur. Conf. Comput. Vis. (ECCV)*, volume 12350, pp. 88–107, 2020.

Y. Xia, T. Ye, and J. Huang. Liver point cloud registration via multi-feature fusion and hyperbolic embedding for augmented reality surgical navigation. *Comput. Methods Programs Biomed.*, 2026.

M. Xu, H. Ye, and F. Cao. Neighborhood-attention and topology-aware graph convolution method for robust point cloud registration. *J. Comput. Appl.*, 45(9):2893–2901, 2025.

S. Yan, P. Shi, Z. Zhao, K. Wang, K. Cao, J. Wu, and J. Li. Turboreg: Turboclique for robust and efficient point cloud registration. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 26371–26381, 2025.

M. Yuan, X. Li, and H. Tan. An elastic fine-tuning dual recurrent framework for non-rigid point cloud registration. *arXiv preprint*, 2025.

L. Zhang, L. Hui, Q. Liu, B. Li, and Y. Dai. 3d focusing-and-matching network for multi-instance point cloud registration. In *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2024.

Y. Zhang, W. Xu, and H. Huang. Deep learning-based point cloud registration: A comprehensive survey and taxonomy. *Int. J. Comput. Vis.*, 134:168, 2026a.

Y. Zhang, Y. Zou, and P. X. Liu. Gia-net: Geometry-informed attention network for 3d point cloud registration in liver surgery. *Int. J. Comput. Assist. Radiol. Surg.*, 2026b.

M. Zhao, J. Jiang, L. Ma, S. Xin, G. Meng, and D.-M. Yan. Correspondence-free nonrigid point set registration using unsupervised clustering analysis. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024a.

M. Zhao, J. Jiang, L. Ma, S. Xin, G. Meng, and D.-M. Yan. Correspondence-free nonrigid point set registration using unsupervised clustering analysis. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024b.

A THEORETICAL ANALYSIS

Our theoretical contribution is a bound on the separation that a topologically stable descriptor can exhibit between two point clouds, and a proof that this bound is governed by the filtration metric rather than by the descriptor design. Lipschitz stability with respect to the bottleneck distance controls how far apart two point clouds can land under any such descriptor. We show that no descriptor built on the Euclidean filtration and satisfying this stability can register the bridging of a narrow gap by more than a quantity that vanishes with the gap, whereas the graph geodesic metric separates the two parts by an amount that can be made arbitrarily large. Section A.1 fixes the admissible class, Section A.2 gives the stability of the lifetime statistics that carry the argument, Section A.3 quantifies the separation of the graph geodesic metric, and Section A.4 shows why the Euclidean filtration cannot reproduce it. All statements are made for finite metric spaces and require no manifold, curvature, or sampling assumption.

A.1 ADMISSIBLE DESCRIPTORS

A point cloud is a finite set $P \subset \mathbb{R}^3$ with $|P| = n$. Given P, we build a k-nearest-neighbour graph $G_k(P) = (P, E)$ with Euclidean edge weights and define the graph geodesic metric

$$d_G(x, y) \;=\; \min_{\pi:\, x \rightsquigarrow y} \sum_{(u,v)\in\pi} \|u - v\|_2, \tag{17}$$

the minimum taken over paths in $G_k(P)$; d_E denotes the ambient Euclidean metric. Then (P, d_E) is a finite metric space, and so is (P, d_G) whenever $G_k(P)$ is connected. For a metric d on P, let $\mathcal{D}_k(P, d)$ be the k-th persistence diagram of the Vietoris–Rips filtration and

$$\phi(P, d) = \big(\mathrm{AvgLife}_0, \mathrm{MaxLife}_0, \mathrm{Ent}_0, \mathrm{AvgLife}_1, \mathrm{MaxLife}_1, \mathrm{Ent}_1\big)(P, d) \in \mathbb{R}^6 \tag{18}$$

the descriptor extracted from it, abbreviated $\phi_E := \phi(\cdot, d_E)$ and $\phi_G := \phi(\cdot, d_G)$. Here $\mathrm{MaxLife}_k$ and $\mathrm{AvgLife}_k$ are the maximum and mean of the finite lifetimes $\ell = d - b$ over $\mathcal{D}_k(P, d)$, and Ent_k is the entropy of the normalised finite lifetimes; classes with infinite death, when present, are excluded. In the implementation the essential H_0 class is retained with a death time of $1.1\times$ the largest finite death, so that $\mathrm{MaxLife}_0$ as computed is a strictly increasing affine function of $\mathrm{MaxLife}_0$ as defined here; all orderings and comparisons below are unaffected. Our network conditions the FiLM module on ϕ_G.

Definition A.1 (Admissible filtration descriptor)**.** *Fix $C_0 > 0$ and a metric d on a point cloud. A map ψ that assigns a vector in $\mathbb{R}^m$ to each finite point cloud is a C_0-admissible filtration descriptor for d if it factors through the persistence diagram of the Vietoris–Rips filtration of (P, d), $\psi(P, d) = F(\mathcal{D}_\bullet(P, d))$, with F being C_0-Lipschitz with respect to the bottleneck distance:*

$$\|\psi(P, d) - \psi(Q, d)\|_2 \;\le\; C_0 \cdot \max_k d_B\big(\mathcal{D}_k(P, d), \mathcal{D}_k(Q, d)\big), \tag{19}$$

where both diagrams are computed under the same metric d, so that the bottleneck distance is well defined.

This class includes persistence landscapes and all lifetime statistics used in practice, including the $\mathrm{MaxLife}$ components of our own ϕ (Theorem A.1). Admissibility is precisely the property that makes a topological descriptor usable as a learned feature, since a smaller C_0 means lower sensitivity to sampling noise. The price of that insensitivity is quantified in Section A.4.

A.2 STABILITY OF THE LIFETIME STATISTICS

Reliable conditioning requires the descriptor not to amplify the perturbations it meets in clinical data, such as noise, occlusion and inter-phase deformation.

Theorem A.1 (Stability of the lifetime statistics)**.** *Let P, Q be point clouds and let $\delta := \max_k d_B(\mathcal{D}_k(P, d), \mathcal{D}_k(Q, d))$. Then $\mathrm{MaxLife}_k$ is 2-Lipschitz,*

$$|\mathrm{MaxLife}_k(P) - \mathrm{MaxLife}_k(Q)| \;\le\; 2\delta, \tag{20}$$

with constant independent of n, $L_{\min}$ and $L_{\max}$. If in addition all finite lifetimes lie in $[L_{\min}, L_{\max}]$ with $L_{\min} > 0$, and the two diagrams contain equally many finite lifetimes m_k, then $\mathrm{AvgLife}_k$ is also 2-Lipschitz and Ent_k satisfies

$$|\mathrm{Ent}_k(P) - \mathrm{Ent}_k(Q)| \;\le\; C_{\mathrm{Ent}}\, \delta, \qquad C_{\mathrm{Ent}} = \frac{2\left(2\log n + \log \frac{L_{\max}}{L_{\min}}\right)}{L_{\min}}. \tag{21}$$

Proof sketch. An optimal bottleneck matching moves each birth and death by at most δ, hence each lifetime by at most 2δ; applying this at the lifetime maximiser gives equation 20, and averaging over a bijective matching of the finite lifetimes gives the same bound for $\mathrm{AvgLife}_k$. Entropy is nonlinear, so we bound its gradient instead: writing $p_i = \ell_i / \sum_j \ell_j$, the constraints $L_{\min} \le \ell_i \le L_{\max}$ and $m_k \le n$ give $\|\nabla H\|_1 \le L_{\min}^{-1}(2\log n + \log \frac{L_{\max}}{L_{\min}})$, and the mean value theorem completes the proof. Supplementary Note A gives the derivation in full. □

The bound on $\mathrm{MaxLife}_k$ is unconditional and is the one the argument below uses; the remaining four components require the two diagrams to contain equally many finite lifetimes. On Lung250M-4B that condition holds exactly for H_0, whose count is 499 for every source–target pair we measured, but not for H_1, where the counts differ by up to 30 on roughly 130 lifetimes. A lifetime threshold $\ell_{\min} = 10^{-6}$ applied before computing ϕ guarantees $L_{\min} \ge \ell_{\min} > 0$.

Geodesic graphs also yield significantly larger persistence gaps and topological SNR than Euclidean graphs across all $k \in \{6, 20, 30\}$ ($p < 0.0001$; Table 5).

A.3 Separation under the graph geodesic metric

Euclidean k-NN graphs generate spurious cross-boundary connections that geodesic distances suppress, as we confirm directly on pulmonary point clouds (Supplementary Fig. 4a). We now quantify how far the two metrics separate two parts of a point cloud.

Theorem A.2 (Separation of the graph geodesic metric)**.** *Let P be a point cloud and $A, B \subset P$ disjoint, with gap width $w := \min_{a\in A, b\in B} \|a - b\|_2$ and detour length $L := \min_{a\in A, b\in B} d_G(a, b)$. Let r be the maximum edge length of $G_k(P)$. Assume* (H1) *$r < w$ and* (H2) *$G_k(P)$ is connected. Then no edge of $G_k(P)$ joins A to B, and*

$$\min_{a\in A, b\in B} d_E(a, b) \;=\; w \;<\; L \;=\; \min_{a\in A, b\in B} d_G(a, b), \tag{22}$$

Proof. The first equality holds by the definition of w, and (H2) makes L finite. Every edge of $G_k(P)$ has length at least the Euclidean distance between its endpoints, so summing along a path gives $d_G \ge d_E$ pointwise and hence $L \ge w$. Suppose $L = w$. Then a shortest A–B path has length w, and since its endpoints $a \in A$ and $b \in B$ satisfy $\|a - b\|_2 \ge w$, the path consists of the single edge $\{a, b\}$ of length exactly w. This contradicts (H1), which bounds every edge length by $r < w$; hence $L > w$. Finally, (H1) also implies that no edge joins A to B, so every A–B path meets $P \setminus (A \cup B)$. Supplementary Note B gives the derivation in full. □

Both hypotheses are directly verifiable on a given point cloud: (H1) compares two measurable scalars, r read off the graph and w read off the data, and (H2) is a connectivity check. Under (H1) every A–B path must detour through $P \setminus (A \cup B)$, so the geodesic metric separates the two parts strictly more than the ambient metric does. The ratio L/w is moreover unbounded: it grows without limit as the gap narrows while the detour stays fixed, which is the configuration of a fissure between two lobes that remain joined around the hilum (Supplementary Note B). The consequence for the graphs themselves is visible in Table 4: geodesic k-NN graphs have diameters $11.9\times$–$20.6\times$ larger than their Euclidean counterparts (485–720 versus 24–61 hops), a direct result of excluding the shortcut edges that (H1) rules out.

A.4 Limits of the Euclidean filtration

The result above concerns the two metrics. We now ask what remains of that separation once the point cloud is passed through a filtration and summarised by a descriptor.

Theorem A.3 (Limits of the Euclidean filtration)**.** *Let P, A, B, w, L be as in Theorem A.2, and let (a^*, b^*) realise the gap width. Define the bridged point cloud*

$$Q \;=\; P \cup \{z_1, \dots, z_m\}, \qquad z_i \in [a^*, b^*], \quad \|z_i - z_{i+1}\|_2 < s, \tag{23}$$

with $s < w$, the z_i chosen so that Q again satisfies (H1)–(H2)*. Then* (i) *$d_{GH}\big((P, d_E), (Q, d_E)\big) \le w/2$, and consequently every C_0-admissible descriptor ψ satisfies*

$$\big\|\psi(P, d_E) - \psi(Q, d_E)\big\|_2 \;\le\; \tfrac{1}{2}\, C_0\, w; \tag{24}$$

Table 4: Anatomical fidelity of Euclidean versus geodesic graphs across neighbourhood sizes k.

	$k=6$		$k=20$		$k=30$	
Metric	Euc.	Manif.	Euc.	Manif.	Euc.	Manif.
Avg. edge length (mm)↑	8.70±0.18	**9.22±0.19**	12.61±0.25	**13.36±0.27**	14.48±0.28	**15.43±0.30**
Long edge ratio (%)↑	1.46±0.18	**3.70±0.24**	1.89±0.31	**6.78±0.33**	2.18±0.37	**7.59±0.27**
Graph diameter (hops)↑	60.70±8.97	**719.76±121.63**	29.81±5.78	**528.63±100.37**	23.57±3.94	**484.61±82.81**
Avg. path length↑	30.98±6.20	**292.92±58.58**	15.71±3.14	**222.77±44.55**	12.53±2.51	**206.58±41.32**

Table 5: Quantitative comparison between Euclidean and manifold topology under different k-NN settings. Higher values indicate better structural preservation.

Metric	Euclidean	Manifold $K=6$	Manifold $K=20$	Manifold $K=30$
H_1 persistence gap	14.04 ± 3.86	$\mathbf{46.59 \pm 12.84}$ (p<0.0001)	22.15 ± 7.99 (p<0.0001)	22.31 ± 18.08 (p<0.0001)
H_2 persistence gap	2.45 ± 0.95	$\mathbf{6.87 \pm 1.61}$ (p<0.0001)	6.28 ± 2.19 (p<0.0001)	5.81 ± 3.19 (p<0.0001)
Topological SNR	6.29 ± 1.29	7.47 ± 1.60 (p<0.0001)	9.01 ± 2.62 (p<0.0001)	$\mathbf{9.16 \pm 5.51}$ (p<0.0001)
Total structural persistence	309.34 ± 52.45	$\mathbf{854.91 \pm 147.40}$ (p<0.0001)	338.77 ± 53.60 (p<0.0001)	310.57 ± 64.22 (p=0.1197)

(ii) *whereas* $\min_{a\in A, b\in B} d_G(a,b) = L$ *in* P*, the added points form an* a^*–b^* *path of total length at most* w*, so* $\min_{a\in A, b\in B} d_G(a,b) \le w$ *in* Q*.*

Proof. (i) Since $P \subset Q$, one direction of the Hausdorff distance vanishes. Each z_i lies on the segment $[a^*, b^*]$ of length w, hence within $w/2$ of $\{a^*, b^*\} \subset P$; therefore $d_H(P,Q) \le w/2$ and, both clouds lying in the same metric space, $d_{GH} \le d_H \le w/2$. By the bottleneck stability theorem Cohen-Steiner et al. (2007), $\max_k d_B(\mathcal{D}_k(P,d_E), \mathcal{D}_k(Q,d_E)) \le d_{GH} \le w/2$, and equation 24 follows from Definition A.1. No property of ψ beyond admissibility is used.

(ii) The z_i lie on $[a^*, b^*]$ with consecutive spacings below $s < w$ and with z_1 adjacent to a^* and z_m to b^*, so they form an a^*–b^* path in $G_k(Q)$ whose length is at most $\|a^* - b^*\|_2 = w$; hence $\min_{a,b} d_G(a,b) \le w$ in Q, against $L > w$ in P. Supplementary Note C gives the derivation in full. □

Statement (i) bounds the change of an admissible descriptor and (ii) the change of the metric; they are not two bounds on the same quantity, and the theorem does not claim that they are. What it establishes is an asymmetry. Bridging a narrow gap moves the Euclidean Gromov–Hausdorff distance by at most $w/2$, so no admissible Euclidean descriptor can move by more than $\frac{1}{2}C_0 w$, which vanishes as the gap narrows. The graph geodesic metric, by contrast, moves by at least $L - w$, which Supplementary Note B shows can be made arbitrarily large relative to w. The mechanism is that stability, which makes a topological descriptor usable, also limits its sensitivity: a C_0-Lipschitz map cannot separate inputs that are $w/2$-close in the input metric. The graph geodesic metric is therefore not an incremental design choice but the object that carries the separation, and this is why the conditioning signal supplied to the FiLM block is ϕ_G and not ϕ_E. Full derivations for all three theorems are given in Appendix A–C, and additional geodesic-versus-Euclidean graph-structure experiments in Supplementary Note D.

B Bi-phase Pulmonary Surface Reconstruction

As a further auxiliary validation that geodesic topology translates into physically plausible downstream deformation, we construct a physically consistent 4D deformable surface model across the respiratory cycle in three stages.

Surface reconstruction. The reference surface $\mathcal{S}_0$ is reconstructed from $\mathcal{P}_{\text{exp}}$ via screened Poisson surface reconstruction Kazhdan et al. (2006):

$$\min_{\tilde{\chi}} \|\nabla\tilde{\chi} - \mathbf{V}\|_{L^2}^2 + \alpha \sum_{i=1}^{N} w_i \big(\tilde{\chi}(\mathbf{p}_i) - \tfrac{1}{2}\big)^2, \tag{25}$$

with octree depth $d = 9$ (standard) or $d = 11$ (high-fidelity), followed by density-based filtering ($\rho(\mathbf{v}) > \tau_{0.1}$).

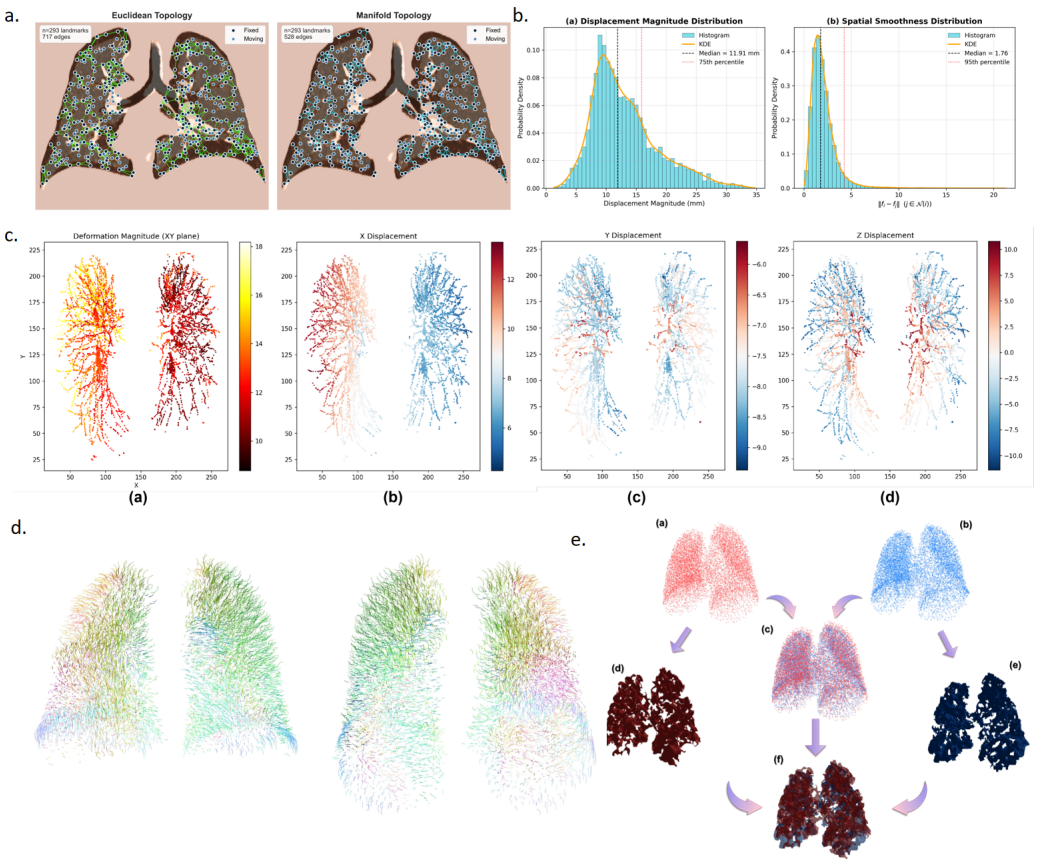


Figure 4: Topological and deformational analysis of pulmonary registration. a. Comparison of Euclidean k-NN (left) versus geodesic (right) graphs, showing geodesic constraints prevent cross-fissure edges. b. Dynamic displacement (left) and spatial smoothness (right) distributions. c. Deformation field components (lateral, anterior-posterior, inferior-superior) projected onto XY plane; dominant inferior-superior expansion (mean: 12.66 mm) reflects diaphragm-driven motion. d. Multi-view 3D vector field with trajectory curves illustrating continuous displacement across respiratory phases. e. Bilateral lung reconstruction: expiratory/inspiratory point clouds and meshes (volume expansion: +7.50%, surface area: +4.90%) with transparent overlay confirming anatomical alignment and preserved fissure integrity.

Deformation field estimation. The discrete displacement field $\mathbf{U} = \{\mathbf{u}_i = \mathbf{q}_i - \mathbf{p}_i\}$ is extended to $\mathcal{S}_0$ via RBF interpolation Carr et al. (2001):

$$\mathbf{u}(\mathbf{x}) = \sum_{j=1}^{M} \mathbf{c}_j \phi(\|\mathbf{x} - \mathbf{p}_j\|) + \mathbf{P}(\mathbf{x}), \tag{26}$$

where $\phi(r) = r^2 \log r$ and $\mathbf{P}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$. For $N > 50{,}000$, control points are selected via FPS. Trajectories are visualised as Bézier curves with RGB directional encoding (Fig. 4d).

Physical consistency. The map $\phi_t(\mathbf{x}) = \mathbf{x} + t \cdot \mathbf{u}(\mathbf{x})$ is validated via $J(\mathbf{x}) = \det(\mathbf{I} + t \cdot \nabla\mathbf{u}(\mathbf{x}))$, with $\nabla\mathbf{u}$ estimated by Tikhonov-regularised least squares ($\lambda = 10^{-6}$). Physical plausibility requires $J \in [0.8, 1.3]$ Rueckert et al. (1999). $\mathcal{S}_0$ comprises 43,258 vertices and 79,195 faces; $\mathcal{S}_1$ exhibits volume expansion of 7.50% and surface area increase of 4.90% (physiological range: 5–20%). The mean Jacobian $\mathbb{E}[J] = 1.075$ confirms near-incompressible tissue behaviour, with 99.99% of vertices satisfying $J > 0$, and the absence of trajectory crossings in Fig. 4e further confirms the diffeomorphic property of ϕ_t.

C Complexity Analysis

Let $N{=}8192$ be the number of input points, L the number of pyramid levels with $N_l \in \{8192, 2048, 512, 256, 64\}$, $K_c{=}32$ the cost-volume neighborhood, $K_f{=}16$ the feature-aggregation neighborhood and $K_p{=}9$ the scene-flow-predictor neighborhood.

Parameter budget (measured). Loading the trained checkpoints and summing all parameter tensors gives **7,754,284** parameters for TopoReg ($\mathrm{topo_dim}{=}6$) and **7,720,236** for the topology-free backbone ($\mathrm{topo_dim}{=}0$, identical code path with the FiLM branch disabled). The proposed coupling thus costs **+34,048 parameters (+0.441%)**: 192 from widening the L_0 input from 3 to 9 channels, and 33,856 from the FiLM generator at L_4 (projection $6 \to 64$, LayerNorm, and the two $64 \to 256$ heads). Table 6 breaks the budget down per component; each row was verified against the checkpoint.

Time. All learnable operations are local and cost $\mathcal{O}(N_l K C_{\text{in}} C_{\text{out}})$; furthest-point sampling costs $\mathcal{O}(N_l N_{l-1})$ and each cost volume performs two k-NN lookups of $\mathcal{O}(N_l^2)$ distance evaluations. The forward pass therefore totals $\mathcal{O}\big(\sum_l N_l K C_l\big) + \mathcal{O}(N^2)$, with the quadratic term arising only from the k-NN at the finest level ($8192^2 = 67$ M distance evaluations). Counting matrix operations only, the model costs **14.15 G-MACs (28.31 GFLOPs)** per sample; of this matrix-operation total, the scene-flow predictors take 52.5% and the cost volumes 40.9%, while the added topology-coupled layer contributes 1.0%. The element-wise BN/FiLM affine maps add 0.11 G-MACs, and the geometric k-NN/FPS distance evaluations amount to 0.81 G evaluations ($\mathcal{O}(N^2)$, memory-bound rather than compute-bound). Section-level shares quoted below refer to the measured total and therefore use a different denominator.

Space (measured). A forward pass of the released checkpoint has a measured peak of **0.85 GB** at $B{=}1$ and **3.18 GB** at $B{=}4$; one training step (forward+backward) peaks at **6.33 GB** allocated (7.20 GB reserved) at $B{=}4$. These figures are per-process peak-memory measurements obtained after resetting the peak statistics. The largest single tensor is the finest-level k-NN distance matrix, $\mathcal{O}(BN^2)$ (268 MB per sample in fp32), which dominates the $B{=}1$ footprint; the next largest activations are the cost-volume MLP input (137 MB at L_0) and the L_4 grouped features (2.1 MB). Memory is thus governed by the point-cloud k-NN rather than by the topological coupling. All measurements use fp32, matching every reported run on both the KITTI and Lung benchmarks.

The efficiency figures are dataset-independent (verified). Repeating the identical measurement on the Lung SOTA checkpoint (6-d topological descriptor) reproduces every number exactly: 7,754,284 parameters, $0.848/3.176/6.333$ GB of peak memory, 101.4 ms latency and 32.899 GFLOPs—identical to the KITTI checkpoint to the last digit (latency differs by 0.1 ms, i.e. run-to-run noise). This is expected: both datasets feed $N{=}8192$ points through the same network, so the cost depends on N and on the architecture, not on the dataset. It also explains earlier whole-GPU instantaneous readings that appeared to disagree between the two datasets (e.g. 8.3–9.1 GB): those samples were taken while co-tenant jobs were running, whereas the numbers here are per-process peak measurements.

Compute (measured). An operator-level FLOP count of the released checkpoint gives **32.90 GFLOPs** per sample at $B{=}1$ and **131.60 GFLOPs** at $B{=}4$—exactly $4.0\times$, i.e. perfectly linear in the batch size. By operator class (shares of the measured 32.90 GFLOPs): fully connected layers 48.7%, convolutions 39.0% and batched matrix products (point aggregation plus the k-NN squared-distance products) 12.3%. Attributed to top-level modules (leaf entries only, so no entry is counted twice, again as shares of the measured total): the L_0 flow predictor alone takes 28.5%, followed by Cost3 11.7%, Cost0 9.4%, Flow1 8.7%, Cost1 7.1% and Cost2 6.8%. This measured total lies within 7% of the analytic accounting above (28.31 GFLOPs of matrix operations plus 2.42 GFLOPs of k-NN products); the residual is a matter of convention—the counter skips element-wise products (the inverse-distance weighted sums and the BN/FiLM affine maps that the analytic column counts) but includes the k-NN squared-distance products that the analytic column lists separately.

Table 6: Parameter and compute budget of TopoReg. Parameters are measured from the trained checkpoints (Table rows sum to 7,754,284); MACs are analytic, computed per sample at N=8192 with formulas validated module-by-module against the same checkpoints (1 MAC=2 FLOPs).

Component	Params (measured)	MACs (matrix ops)
Feature pyramid (L_0–L_3)	1,662,024	756.6 M
TopoCoupledPointConvD + FiLM (L_4)	2,143,864	143.7 M
Deconvolution layers ($\times 4$)	39,104	37.7 M
Cost volumes ($\times 4$)	330,464	5.78 G
Scene-flow predictors ($\times 4$)	3,578,828	7.43 G
Upsampling / warping (no parameters)	0	5.3 M
Total (topo_dim=6)	**7,754,284**	**14.15 G**
Backbone only (topo_dim=0, measured)	7,720,236	—

Algorithm 1 Topological Feature Extraction

Require: point cloud $S \in \mathbb{R}^{N\times 3}$
subsample size m (=500); neighborhood size k (=20)
Ensure: topological descriptor $T \in \mathbb{R}^6$

$S' \leftarrow \textsc{Subsample}(S, m)$ ▷ uniform random, without replacement
$A \leftarrow k\text{-}\textsc{NNGraph}(S', k)$ ▷ weighted by Euclidean distance
$D \leftarrow \textsc{Dijkstra}(A)$ ▷ geodesic (graph) distance matrix
$D[\infty] \leftarrow 2\max_{i,j} D_{ij} < \infty; \quad \mathrm{diag}(D) \leftarrow 0$
$\{(b_i^0, d_i^0)\}, \{(b_i^1, d_i^1)\} \leftarrow \textsc{Ripser}(D, \mathrm{dim}{=}1)$
for $\ell \in \{0, 1\}$ **do**
  $\mathcal{L}^\ell \leftarrow \{d_i^\ell - b_i^\ell\}_i$ ▷ finite lifetimes only
  $f^\ell \leftarrow \left[\ \mathrm{mean}(\mathcal{L}^\ell),\ \max(\mathcal{L}^\ell),\ -\sum_i p_i^\ell \log p_i^\ell\ \right], \quad p_i^\ell = \mathcal{L}_i^\ell / \sum_j \mathcal{L}_j^\ell$
end for
return $T \leftarrow [\, f^0, f^1\,]$

Latency and measured time breakdown. Forward latency is **101.5 ms** per sample (B=1, RTX 5090, mean of 10 runs after warm-up). A kernel-level profile shows that wall-clock time is dominated by the geometric operators rather than by matrix products: furthest-point sampling 34.9%, top-k (k-NN) 20.1% and element-wise kernels 13.4% (together $\approx 68\%$), whereas all GEMM kernels account for only $\approx 9\%$. The MACs distribution reported above (flow predictors 52.5%) therefore measures arithmetic volume, not latency: the k-NN/FPS stages are memory-bound, consistent with their $\mathcal{O}(N^2)$ distance evaluations, which are listed separately in the paragraph above.

Topological branch (offline). Topological descriptors are computed once per point cloud, before training: uniform subsampling to m=500 points, a k-NN graph (k=20) plus Dijkstra's algorithm in $\mathcal{O}(m^2 \log m)$ (a 2 MB fp64 distance matrix) and Ripser persistent homology (maxdim=1), whose cost is superquadratic in m but independent of N. The descriptors are cached on disk, so the branch adds no cost to training or inference.

Effect on complexity class. Because the coupling only (i) concatenates six extra input channels at L_0 and (ii) inserts an element-wise $\gamma \odot x + \beta$ map (with per-sample γ, β) inside L_4, it changes the constant factor ($+0.441\%$ parameters) while leaving the asymptotic order of both time and memory unchanged.

D Pseudocode of Our Proposed Model

Algorithm 2 TopoCoupledPointConvD (L4 coupling module)

Require: coordinates $X \in \mathbb{R}^{3\times N}$; features $F \in \mathbb{R}^{C_{in}\times N}$; topology $T \in \mathbb{R}^6$
output points N'; neighborhood size K
identity initialization $W_\gamma \leftarrow \mathbf{0}$, $b_\gamma \leftarrow \mathbf{1}$, $W_\beta \leftarrow \mathbf{0}$, $b_\beta \leftarrow \mathbf{0}$ ▷ $\Rightarrow \gamma \equiv 1,\ \beta \equiv \mathbf{0}$ at $t{=}0$
Ensure: downsampled coordinates X'; modulated features Z

idx $\leftarrow$ FPS(X, N'); $\quad X' \leftarrow$ GATHER(X, idx)
$G \leftarrow$ GROUPQUERY(K, X, X', F) ▷ $(N' \times K \times C_{in})$ neighbor features
$\Delta \leftarrow G_{xyz} - X'$ ▷ local coordinates, $(N' \times K \times 3)$
$W \leftarrow$ WEIGHTNET(Δ) ▷ MLP $3 \to 8 \to 8 \to C_{out}$, $(N' \times K \times C_{out})$
$Z \leftarrow$ LINEAR$\big(\sum_k W_k \odot G_k\big)$ ▷ weighted aggregation
$t \leftarrow$ LEAKYRELU(LAYERNORM(LINEAR$_{6\to 64}(T)$))
$\gamma \leftarrow$ LINEAR$_{64\to C_{out}}(t)$; $\quad \beta \leftarrow$ LINEAR$_{64\to C_{out}}(t)$ ▷ FiLM generator, shared t
$Z \leftarrow$ RELU$(\gamma \odot Z + \beta)$ ▷ pure multiplicative FiLM; no BN
return X', Z

Algorithm 3 TopoReg Forward Pass

Require: source P, target $Q \in \mathbb{R}^{N\times 3}$
per-point inputs C_P, C_Q; topological descriptors T_P, T_Q
pyramid sizes $\{N_l\}_{l=0}^{L} = \{8192, 2048, 512, 256, 64\}$
Ensure: multi-scale scene flows $\{SF_l\}_{l=0}^{L-1}$

Phase 1 — feature encoding (shared weights for P and Q)
$C_P \leftarrow [\,C_P; T_P\,]$ ▷ 9-dim input: append 6-dim topology
$F_P^0 \leftarrow$ CONV1D$_{9\to 32}(C_P)$; $F_P^0 \leftarrow$ CONV1D$_{32\to 32}(F_P^0)$; $E_P^0 \leftarrow$ CONV1D$_{32\to 64}(F_P^0)$
for $l = 1$ **to** $L-1$ **do**
  $X_P^l \leftarrow$ FPS(X_P^{l-1}, N_l)
  $F_P^l \leftarrow$ POINTCONVD(X_P^{l-1}, E_P^{l-1}) ▷ FPS + PointConv, $K{=}16$
  $E_P^l \leftarrow$ CONV1D$_{2\times}(\,F_P^l\,)$ ▷ expand channels for level $l{+}1$
end for
$F_P^L \leftarrow$ TOPOCOUPLEDPOINTCONVD$(X_P^{L-1}, E_P^{L-1}, T_P)$ ▷ Algorithm 2; only topology-coupled layer
▷ obtain $\{F_Q^l\}$ identically from Q and T_Q
Phase 2 — coarse-to-fine cost volume cascade
$\tilde{F}_P^l \leftarrow [\,F_P^l;$ DECONV(UPSAMPLE$(X^l, X^{l+1}, F_P^{l+1}))\,]$ ▷ enlarged receptive field, $l = L-1, \ldots, 0$
for $l = L-1$ **downto** 0 **do**
  if $l = L-1$ **then** ▷ coarsest level, no coarser flow available
    $V^l \leftarrow$ COSTVOLUME$(X_P^l, X_Q^l, \tilde{F}_P^l, \tilde{F}_Q^l)$
  else
    $SF_\uparrow^{l+1} \leftarrow$ UPSAMPLE(X^l, X^{l+1}, SF^{l+1}) ▷ inverse-distance interpolation, $K{=}3$
    $X_{Q,w}^l \leftarrow$ WARP$(X_P^l, X_Q^l, SF_\uparrow^{l+1})$ ▷ warp target back to source frame
    $V^l \leftarrow$ COSTVOLUME$(X_P^l, X_{Q,w}^l, \tilde{F}_P^l, \tilde{F}_Q^l)$
  end if
  $H^l \leftarrow [\,F_P^l;$ UPSAMPLE$(X^l, X^{l+1}, H^{l+1})\,]$, $\quad SF_\uparrow^{l+1}$ appended if $l \neq L-1$ ▷ predictor features reused from level $l{+}1$
  $H^l, SF^l \leftarrow$ FLOWPREDICTOR$(X_P^l, H^l, V^l, SF_\uparrow^{l+1})$ ▷ PointConv $\times 2$ + 1×1 conv, output $\mathbb{R}^3$
end for
return $\{SF^0, \ldots, SF^{L-1}\}$

Algorithm 4 Training of TopoReg

Require: dataset $\mathcal{D} = \{(P, Q, C_P, C_Q, GT, T_P, T_Q)\}$
epochs E; learning rate η; topology weight λ
warm-up epochs E_w; level weights $\{\alpha_l\}$
Ensure: parameters Θ

$\Theta \leftarrow$ random init (FiLM: $\gamma \equiv 1$, $\beta \equiv 0$); optimizer $\leftarrow$ ADAM(η); milestones at $[40, 50]$ (or $[1200, 1400]$)
for ep $= 1$ **to** E **do**
  $\lambda(\text{ep}) \leftarrow \lambda \cdot \min(1,\ \text{ep}/E_w)$ ▷ linear warm-up of the topology term
  for all mini-batch $(P, Q, C_P, C_Q, GT, T_P, T_Q)$ **in** $\mathcal{D}$ **do**
    $\{SF_l\} \leftarrow$ FORWARD$(P, Q, C_P, C_Q, T_P, T_Q)$ ▷ Algorithm 3
    $\mathcal{L}_{\text{flow}} \leftarrow \sum_{l=0}^{L-1} \alpha_l \left\| SF_l - GT_l \right\|_2^2$ ▷ multi-scale supervised loss
    $\mathcal{L}_{\text{topo}} \leftarrow$ TOPOCONSISTENCY(SF_0, P, Q) ▷ local-distance / edge-length alignment, k=20
    $\mathcal{L} \leftarrow \mathcal{L}_{\text{flow}} + \lambda(\text{ep})\, \mathcal{L}_{\text{topo}}$
    $\Theta \leftarrow$ ADAMUPDATE$(\Theta, \nabla_\Theta \mathcal{L})$; clip $\|\nabla_\Theta \mathcal{L}\|_2 \le 1.0$
  end for
  decay η at milestones
end for
return Θ