Augmented Equivariant Mesh Networks for Anatomical Mesh Segmentation

Daniel Saragih [1]

Abstract

Triangle meshes are a structured yet underexplored clinical data modality: as the native output of many surgical-planning, dental, and vascular reconstruction pipelines, they carry rich surface topology that flat point clouds and volumetric grids cannot directly represent. Robust segmentation on such meshes requires models that respect arbitrary patient pose and varying mesh resolution—properties that current task-specific methods lack. We present EAMS, a lightweight (<2M parameter) equivariant mesh segmentation framework augmented with anatomy-aware intrinsic features and two long-range context modules. Evaluated across three clinically distinct tasks spanning vertex-, face-, and edge-level supervision: intracranial aneurysm, intraoral tooth, and liver-surface segmentation, EAMS variants are competitive with strong task-specific baselines on canonical inputs while delivering markedly stronger robustness under test-time geometric perturbation. These results position anatomical meshes as a practically important structured-data regime for healthcare ML and demonstrate that compact equivariant models can achieve robustness without task-specific architecture redesign.

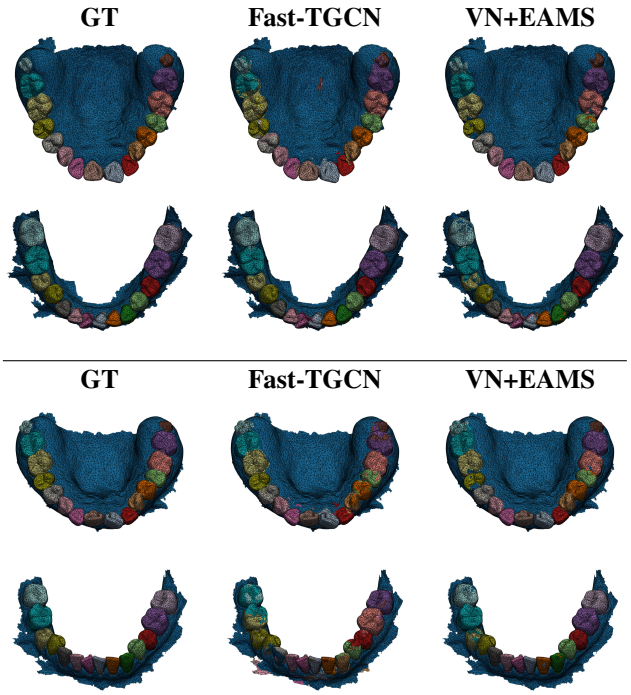


Figure 1. Qualitative tooth-segmentation comparisons. The top half shows the canonical orientation and the bottom half shows a 40° z-axis rotation. Fast-TGCN is competitive at baseline but degrades near posterior teeth and gingival boundaries under rotation, whereas VN+EAMS remains stable across both settings.

1. Introduction

Anatomical surface meshes represent an important but less-studied form of structured data: they are native to surgical planning, dental, and vascular reconstruction pipelines, and they are *graph-structured* with a topology that connects vertices through an explicit face-edge graph (Martin et al., 2015; Koo et al., 2017; Yang et al., 2020). Segmenting landmarks on these surfaces (e.g., identifying individual teeth, aneurysm boundaries, or liver ligaments) is a direct clinical downstream task.

Effective mesh segmentation must satisfy two competing demands. First, predictions should be invariant to rigid transforms: a model that relabels the same anatomy differently after a change in patient orientation provides no reliable clinical signal. Second, mesh resolution varies dramatically across datasets and scanners, requiring features that are stable under resampling. Existing point-cloud architectures (Qi et al., 2017; Wang et al., 2019; Wu et al., 2024) discard the topological structure of meshes. Current mesh segmentation methods (Yang et al., 2020; Ben-Hamadou et al., 2023; Li et al., 2024; Zhang et al., 2025) are not equivariant and can degrade sharply under test-time perturbation.

We address both challenges with EAMS, an **E**quivariant **A**natomical **M**esh **S**egmentor built on Equivariant Mesh Neural Networks (EMNN) (Trang et al., 2024), which extends E(n)-equivariant graph networks (Satorras et al., 2021) to triangle meshes by exploiting invariant geometric scalars

[1]Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada. Correspondence to: Daniel Saragih <danielg.saragih@gmail.com>.

Accepted to the ICML 2026 Workshop on Structured Data for Health.

and equivariant face normals. EAMS adds anatomy-aware intrinsic features and two lightweight long-range context modules (soft regional aggregators and virtual nodes) that expand the receptive field without breaking equivariance.

Contributions.

- **Anatomical mesh segmentation as a structured-data robustness benchmark.** We evaluate equivariant mesh segmentation across liver surfaces, intraoral scans, and intracranial aneurysms, covering edge-, face-, and vertex-level tasks under both clean and perturbed conditions.
- **Anatomy-aware featurization.** We combine intrinsic shape descriptors (HKS, dihedral angles, local area) with task-aware global priors (PCA-derived dental and liver frames) whose canonical axes transform equivariantly with the mesh under SE(3) motions; in practice, any residual approximation enters only through the dataset-specific sign-resolution heuristic.
- **Augmented equivariant message passing.** We extend EMNN's receptive field with soft regional aggregators, which impose a structured regional bottleneck, and virtual nodes, which provide a lightweight global memory for flexible long-range exchange. This compact equivariant pipeline achieves strong robustness across clinically distinct tasks without task-specific redesign.

2. Method

2.1. Problem Setting

We consider semantic segmentation on triangle meshes $\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ with vertex-, edge-, or face-level supervision. The central requirement is robustness to rigid transforms and mesh resolution changes. For intracranial aneurysm meshes, where orientation is arbitrary, we target full E(3) invariance. For dental and liver meshes, where left–right chirality matters clinically, we target SE(3)-robust representations that preserve chirality; in these datasets, the only practical deviation from exact invariance comes from the heuristic sign disambiguation used to orient the canonical frame. Further background is given in Appendix B.

2.2. Anatomy-Aware Featurization

All tasks receive *intrinsic* node and edge descriptors that are stable under rigid transforms: the Heat Kernel Signature (Sun et al., 2009) (8-dimensional, computed from the Laplace–Beltrami spectrum), pointwise local area, and per-edge dihedral angles. For dental and liver tasks we additionally derive *PCA-based anatomical frames*: the mesh is centred, principal axes are estimated from vertex or area-weighted covariance, an anterior-posterior sign is resolved from the arch or lobe geometry, and vertex positions are expressed in cylindrical coordinates (r, θ, z) relative to that frame. Because the canonical frame co-rotates with the mesh under SE(3) transforms, the resulting coordinates are SE(3)-invariant at the level of the PCA construction itself; the only non-exact step is the sign-resolution heuristic used to orient the frame. Full feature tables, per-dataset configurations, and the appendix equivariance argument are in Appendix C.

2.3. Equivariant Mesh Encoder

EMNN (Trang et al., 2024) extends EGNN (Satorras et al., 2021) to triangle meshes by coupling invariant scalar messages with equivariant coordinate updates. In this paper we adopt EMNN as the mesh-native equivariant backbone introduced for this setting, while leaving a direct EGNN-versus-EMNN comparison to future work so the benefit of equivariance per se can be separated from the benefit of explicit mesh structure. The edge message is $\boldsymbol{m}_{ij} = \phi_e(\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{e}_{ij}, \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)$, and the surface-aware face message to vertex i from face (i, j, k) is $\boldsymbol{m}_{ijk} = \phi_f\big(\boldsymbol{h}_i,\ \boldsymbol{h}_j + \boldsymbol{h}_k,\ \|(\boldsymbol{x}_j - \boldsymbol{x}_i) \times (\boldsymbol{x}_k - \boldsymbol{x}_i)\|\big)$. Vertex features aggregate both message types:

$$\boldsymbol{h}_i \leftarrow \phi_h\Big(\boldsymbol{h}_i,\ \textstyle\sum_{j \in N(i)} \boldsymbol{m}_{ij}, \sum_{(j,k) \in \tau(i)} \boldsymbol{m}_{ijk}\Big). \quad (1)$$

The coordinate update is equivariant through edge displacements and face normals:

$$\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i + \sum_{j \in N(i)} (\boldsymbol{x}_i - \boldsymbol{x}_j)\phi_x(\boldsymbol{m}_{ij}) + \sum_{(j,k) \in \tau(i)} \big((\boldsymbol{x}_j - \boldsymbol{x}_i) \times (\boldsymbol{x}_k - \boldsymbol{x}_i)\big)\phi_n(\boldsymbol{m}_{ijk}). \quad (2)$$

The first term propagates pairwise displacement information, while the second injects oriented local surface geometry through face normals without breaking equivariance.[1] All EAMS-family encoders add learned graph-level aggregation (i.e. a global node).

2.4. Augmented Message Passing

Local EMNN layers aggregate only within the one-ring neighbourhood, limiting the receptive field. We study two lightweight augmentations that inject long-range context while preserving equivariance. They differ in how strongly they constrain the global summaries they create: SRA biases the model toward a small set of semantically distinct regional summaries, whereas VN provides a more flexible graph-level memory for diffuse long-range context.

[1] For the face-normal branch, our E(3) claim follows the oriented-surface convention used in EMNN: each triangular face is ordered so that $(\boldsymbol{x}_j - \boldsymbol{x}_i) \times (\boldsymbol{x}_k - \boldsymbol{x}_i)$ points outward. Under a reflection, the embedded geometry changes handedness, so the reflected face must be rewound by swapping j and k to preserve the outward normal.

Soft regional aggregators (SRA). Each vertex i computes a soft assignment $\boldsymbol{a}_i = \mathrm{softmax}(\phi_a(\boldsymbol{h}_i)) \in \mathbb{R}^K$ over K region prototypes, forming the assignment matrix $[\boldsymbol{A}]_{ik} = [\boldsymbol{a}_i]_k$. Region tokens $\boldsymbol{r}_k = \sum_i [\boldsymbol{A}]_{ik} \boldsymbol{h}_i$ are mixed by a small transformer encoder and scattered back to vertices: $\boldsymbol{h}_i \leftarrow \boldsymbol{h}_i + \alpha \sum_k [\boldsymbol{A}]_{ik} \phi_{\mathrm{proj}}(\hat{\boldsymbol{r}}_k)$. This creates a structured bottleneck useful for compact, localised anatomy, where a small number of coherent regional summaries is a natural inductive bias. SRA is regularised with diversity and equipartition objectives on $\boldsymbol{A}$ so that these summaries remain distinct and balanced rather than collapsing onto a single region.

Virtual nodes (VN) (Zhang et al., 2026). V virtual-node features $\{\boldsymbol{v}_k\}$ and coordinates $\{\boldsymbol{u}_k\}$ are maintained as a global memory bank. At each EMNN layer, real nodes exchange messages with virtual nodes as additional graph neighbours; virtual nodes aggregate from all real nodes. This design is more flexible across tasks with heterogeneous spatial layouts. VN is regularised with a Gaussian kernel-based energy that prevents token collapse and keeps virtual positions near the surface. This makes VN a better fit when useful context is spatially diffuse and forcing the mesh into a small number of regions would be unnecessarily restrictive.

Regularization objectives. The shared task loss $\mathcal{L}_{\mathrm{task}} = \mathcal{L}_{\mathrm{pred}} + \lambda_{\mathrm{cbl}} \mathcal{L}_{\mathrm{cbl}}$ uses boundary-aware bDoU predictions (Sun et al., 2023) together with a contrastive boundary loss on latent embeddings (Tang et al., 2022) to sharpen discrimination at class boundaries. For SRA we add a diversity term $\mathcal{L}_{\mathrm{div}} = \|\tilde{\boldsymbol{A}}^\top \tilde{\boldsymbol{A}} - \boldsymbol{I}\|_F^2$ (penalising correlated region tokens) and an equipartition term $\mathcal{L}_{\mathrm{eq}}$ (penalising imbalanced region mass). Together these losses preserve the intended structured bottleneck by discouraging degenerate reuse of the same token and preventing all vertices from collapsing into one or two regions. For VN we add a Gaussian kernel energy $\mathcal{L}_{\mathrm{VN}} = w_{\mathrm{vv}} k_{\mathrm{vv}} - w_{\mathrm{rv}} k_{\mathrm{rv}}$ that repels virtual nodes from collapsing together while attracting them toward real mesh vertices. This keeps the virtual nodes as distributed relay tokens that cover different parts of the shape while remaining anchored near the geometry they summarize. On liver meshes we additionally use a local continuity loss $\mathcal{L}_{\mathrm{cont}}$ on adjacent edge-prediction pairs to encourage spatial smoothness near landmarks. Full loss details are in Appendix D.

3. Experiments

3.1. Tasks and Baselines

We evaluate three anatomical mesh settings spanning different supervision granularities and clinical contexts: intracranial aneurysm segmentation (vertex-level) on IntrA (Yang et al., 2020) with 103 meshes; intraoral tooth segmentation (face-level) on 3D-IOSSeg (Li et al., 2024) (180 direct scans) and Teeth3DS (Ben-Hamadou et al., 2023) (1800 indirect scans); and liver landmark segmentation (edge-level) on annotated liver surfaces (Zhang et al., 2025) with 200 meshes. We compare against DGCNN (Wang et al., 2019), PointTransformerV3 (PTv3) (Wu et al., 2024), Fast-TGCN (Li et al., 2024), PointNet++ (Qi et al., 2017), and MeshGraphCNN (Zhang et al., 2025). Standard deviations are over five folds except for Teeth3DS (single designated split).

Table 1. IntrA results (mean ± std, 5 folds). PV: parent vessel; AN: aneurysm Dice (%). EAMS variants are E(3)-invariant: Rot-40 equals baseline (†). **Bold**: best; underline: second-best.

	PV Dice (%)		AN Dice (%)	
Method	Baseline	Rot-40	Baseline	Rot-40
DGCNN (Wang et al., 2019)	97.11 ± 1.4	94.34 ± 0.9	87.77 ± 5.3	76.61 ± 2.9
PTv3 (Wu et al., 2024)	94.86 ± 1.2	93.06 ± 1.3	79.80 ± 4.6	72.67 ± 4.2
EAMS	98.31 ± 1.4	98.31 ± 1.4†	93.59 ± 4.7	93.59 ± 4.7†
SRA+EAMS	98.69 ± 0.9	98.69 ± 0.9†	**95.81 ± 2.9**	**95.81 ± 2.9**†
VN+EAMS	**98.90 ± 1.0**	**98.90 ± 1.0**†	95.31 ± 4.6	95.31 ± 4.6†

3.2. Intracranial Aneurysm Segmentation

Table 1 is the clearest demonstration of why E(3) invariance matters in this paper: IntrA meshes arrive in arbitrary orientation with no reliable pose prior, so any drop under rigid transforms is purely a robustness failure rather than useful pose sensitivity. All three EAMS variants improve over both baselines on every reported metric—including parent vessel (PV) and aneurysm (AN) Dice—and remain exactly stable under rotation, whereas DGCNN loses 11 points on AN Dice at Rot-40 and PTv3 remains well below the mesh-based variants even before perturbation. Figure 4 shows the same stability qualitatively for the invariant mesh variants. Within the EAMS family, SRA+EAMS achieves the best AN Dice (95.81 ± 2.9), suggesting that a structured regional bottleneck is especially helpful for isolating this compact, localised abnormality, while VN+EAMS leads on PV Dice (98.90 ± 1.0). Appendix Table 8 extends these results with IoU, Rot-15°, and Refl-x conditions.

3.3. Intraoral Tooth Segmentation

Table 2 shows a different regime from IntrA: here the best non-equivariant model, Fast-TGCN, is highly competitive on canonical inputs, but its performance collapses once the jaw is rotated away from the training orientation. VN+EAMS approaches Fast-TGCN on clean inputs despite being nearly **10× smaller** (1.9M vs. 19.8M parameters), then substantially outperforms all baselines under rotation—DGCNN drops 25 Average-IoU points at Rot-40 and Fast-TGCN drops 28, while VN+EAMS stays stable within measurement noise across both direct (3D-IOSSeg) and indirect (Teeth3DS) scan protocols. Figure 1 shows this pattern qual-

Table 2. Intraoral tooth segmentation: Average IoU (%). 3D-IOSSeg: mean ± std, 5 folds. Average IoU averages each mesh over the classes present in its ground truth and then averages those per-mesh values across the evaluation set; it is not the same as averaging the final per-class IoU summaries. Teeth3DS: single designated split. **Bold**: best; underline: second-best.

	3D-IOSSeg		Teeth3DS	
Method	Baseline	Rot-40	Baseline	Rot-40
DGCNN (Wang et al., 2019)	67.12 ± 0.4	42.02 ± 2.5	—	—
PTv3 (Wu et al., 2024)	80.08 ± 0.4	39.05 ± 2.4	—	—
Fast-TGCN (Li et al., 2024)	**81.05** ± **0.7**	52.33 ± 1.6	**84.31**	58.53
EAMS	71.43 ± 1.6	72.05 ± 1.6	80.54	81.11
SRA+EAMS	75.28 ± 1.2	75.98 ± 1.2	82.31	83.05
VN+EAMS	79.74 ± 0.9	**80.09** ± **1.0**	83.16	**83.40**

itatively: Fast-TGCN is visually competitive in the canonical pose but deteriorates once the jaw is rotated, whereas VN+EAMS remains consistent. We also tested an EAMS variant without the dental frame coordinates and observed a **more than 50% drop**, with frequent left/right tooth confusions, confirming the value of this anatomy-aware feature design. Appendix Table 9 adds Rot-15° results, and Tables 10–13 break performance down by FDI tooth number; the wisdom-tooth classes (T18/38 and T28/48) are the main driver of variance across methods, owing to greater morphological variation and fewer training examples.

3.4. Liver Surface Segmentation

All three EAMS variants operate directly on the mesh graph and target SE(3)-invariance; for liver, SRA+EAMS uses $K = 32$ regional tokens and VN+EAMS uses $V = 8$ virtual nodes. This task highlights a balanced trade-off between canonical-pose accuracy and rotation robustness: the baselines can exploit the near-canonical scan alignment in the training set as an orientation prior, whereas EAMS must identify landmarks from intrinsic geometry alone.

Table 3 focuses on the main segmentation metric, Dice, together with Chamfer distance (CD) for the ligament and ridge classes on unperturbed meshes and under a 40° z-axis rotation. We report CD after multiplying by 100 for readability. EAMS, SRA+EAMS, and VN+EAMS remain nearly unchanged under rotation, while the non-equivariant baselines degrade sharply. Figure 2 therefore focuses on the strongest baseline-versus-invariant comparison, MeshGraphCNN against VN+EAMS, while Appendix Figure 5 adds the remaining PointNet++, DGCNN, EAMS, and SRA+EAMS qualitative results for the same task. Full IoU, CD×100, and Hausdorff distance (HD) results for the baseline, 15°, and 40° conditions are given in Table 14.

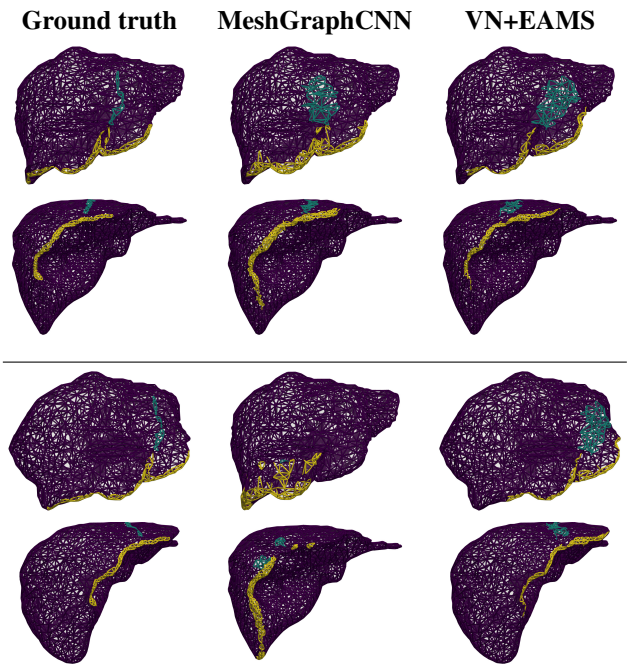


Figure 2. Qualitative liver-surface comparisons on two representative meshes, with the top half in the canonical orientation and the bottom half after a 40° z-axis rotation. Appendix Figure 5 adds the extra PointNet++, DGCNN, EAMS, and SRA+EAMS results.

3.5. Ablation Studies

Table 4 reports single-fold IntrA ablations using IoU only, isolating the shared global node, the SRA geometry term, and the number of learned global tokens. For base EAMS, removing the global node already hurts both parent-vessel and aneurysm IoU, and removing both dihedrals and HKS causes the largest collapse, indicating that HKS remains the dominant non-coordinate cue. SRA+EAMS is stable across the tested region counts, while VN+EAMS is more sensitive to the number of virtual nodes.

Table 5 gives a simple liver ablation for the base encoder. Adding the anatomical frame gives the main gain on both classes, and the continuity loss provides a smaller follow-up gain, mainly on the ligament metrics. Taken together with the full-task results, these ablations support the qualitative design split used throughout the paper: anatomy-aware canonical features matter most when labels have strong spatial priors, while the SRA and VN modules trade off structured summarization against flexible long-range exchange.

4. Discussion

Limitations. The clearest limitation is the liver trade-off: when global orientation is predictive, strict invariance sacrifices clean-set accuracy, particularly for ligament and ridge landmarks whose anatomical positions correlate strongly with body orientation. More precisely, the PCA-derived dental and liver canonical frames are not themselves approx-

Table 3. Liver surface segmentation on the dataset of (Zhang et al., 2025). Results are mean ± std over five cross-validation folds. CD is reported after multiplying by 100. **Bold**: best per column (all rows); underline: second-best. †: MeshGraphCNN results are our own reproduction. Full IoU, CD×100, and HD results are in Table 14.

	Ligament				Ridge			
	Baseline		Rot-z 40°		Baseline		Rot-z 40°	
Method	Dice (%) ↑	CD×100 ↓	Dice (%) ↑	CD×100 ↓	Dice (%) ↑	CD×100 ↓	Dice (%) ↑	CD×100 ↓
PointNet++ (Qi et al., 2017)	35.14 ± 2.38	**0.202 ± 0.051**	6.27 ± 1.97	2.842 ± 0.947	60.61 ± 0.91	**0.067 ± 0.007**	44.73 ± 2.53	1.352 ± 0.235
DGCNN (Wang et al., 2019)	34.25 ± 3.92	0.293 ± 0.032	2.77 ± 0.62	4.331 ± 1.134	60.31 ± 1.17	0.080 ± 0.021	42.11 ± 1.10	1.619 ± 0.126
MeshGraphCNN† (Zhang et al., 2025)	**47.00 ± 0.75**	0.677 ± 0.046	4.66 ± 0.66	7.677 ± 0.865	**60.92 ± 1.11**	0.253 ± 0.056	39.64 ± 0.80	2.484 ± 0.073
EAMS	20.46 ± 1.25	3.928 ± 0.668	20.47 ± 1.26	4.307 ± 0.193	56.36 ± 0.82	0.873 ± 0.064	56.35 ± 0.82	0.876 ± 0.065
SRA+EAMS	26.33 ± 2.16	1.722 ± 0.651	**26.32 ± 2.16**	**1.722 ± 0.650**	54.54 ± 1.11	1.071 ± 0.372	54.52 ± 1.14	1.069 ± 0.371
VN+EAMS	20.98 ± 1.38	2.013 ± 0.592	20.99 ± 1.38	2.012 ± 0.591	58.50 ± 2.07	0.610 ± 0.160	**58.51 ± 2.08**	**0.609 ± 0.159**

Table 4. Single-fold IntrA ablations using IoU only. PA = parent vessel, AN = aneurysm.

Config.	PA IoU (%) ↑	AN IoU (%) ↑
EAMS	93.39	79.90
- no global node	92.29	76.90
- no dihedrals	93.05	77.12
- no dihed. & HKS	78.89	49.86
SRA+EAMS ($K = 32$)	94.77	84.12
- no reg. geom.	94.06	83.22
- $K = 8$	94.42	83.52
- $K = 16$	94.55	83.59
- $K = 64$	94.14	82.95
VN+EAMS ($V = 16$)	95.02	83.05
- $V = 8$	94.67	81.62
- $V = 32$	93.31	80.40
- $V = 64$	92.01	75.38

Table 5. Simple liver ablation for the base EAMS encoder.

Class	Config.	Dice ↑	CD×100 ↓
Ligament	Base	12.76	5.78
Ligament	+ frame	18.78	4.38
Ligament	+ frame + cont	19.51	3.81
Ridge	Base	51.06	1.70
Ridge	+ frame	56.17	0.81
Ridge	+ frame + cont	56.08	0.88

imate: the covariance-based frame construction is equivariant, and only the dataset-specific sign-resolution heuristic introduces a small practical deviation from exact SE(3) invariance. The SRA/VN split should likewise be read as an architectural trade-off rather than only a leaderboard difference: SRA is better matched to anatomies that can be compressed into a few coherent regional summaries, while VN is more natural when informative context is spatially diffuse. The augmented variants also incur modest runtime overhead relative to base EAMS (SRA: 1.45×, VN: 1.64×; see Appendix D). Future work should therefore focus on richer intrinsic descriptors, more principled sign disambiguation for canonical frames, learned pose-aware aggregation that preserves equivariance, and direct EGNN-versus-EMNN comparisons that separate the benefit of equivariance per se from the benefit of explicit mesh structure.

SD4H relevance. For the structured health data community, the central point is that anatomical meshes form a practically important clinical data modality with its own robustness challenges. EAMS addresses one directly: sensitivity to rigid transforms at deployment. More broadly, multimodal fusion of meshes with other structured clinical inputs (time series, images) is a natural next step, as is evaluating equivariant graph models on other structured health data regimes such as vascular networks or multi-organ systems.

Impact Statement

This work studies robust anatomical mesh segmentation as a decision-support component for clinical and research workflows built on structured health data. More stable predictions under pose variation could reduce manual correction and make downstream geometric analysis less sensitive to acquisition differences, but incorrect segmentations could still propagate errors into measurements, visualisation, or treatment-planning systems. Any deployment should therefore remain clinician-supervised, validated for the target anatomy and acquisition pipeline, and paired with task-specific quality control rather than treated as a fully autonomous clinical tool.

References

Basu, S., Gallego-Posada, J., Viganò, F., Rowbottom, J., and Cohen, T. Equivariant mesh attention networks. *arXiv preprint arXiv:2205.10662*, 2022.

Ben-Hamadou, A., Smaoui, O., Rekik, A., Pujades, S., Boyer, E., Lim, H., Kim, M., Lee, M., Chung, M., Shin, Y.-G., et al. 3dteethseg'22: 3d teeth scan segmentation and labeling challenge. *arXiv preprint arXiv:2305.18277*, 2023.

De Haan, P., Weiler, M., Cohen, T., and Welling, M. Gauge equivariant mesh cnns: Anisotropic convolutions on geometric graphs. *arXiv preprint arXiv:2003.05425*, 2020.

Eijkelboom, F., Hesselink, R., and Bekkers, E. J. E (n) equivariant

message passing simplicial networks. In *International Conference on Machine Learning*, pp. 9071–9081. PMLR, 2023.

Feng, Y., Feng, Y., You, H., Zhao, X., and Gao, Y. Meshnet: Mesh neural network for 3d shape representation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 8279–8286, 2019.

Gong, S., Chen, L., Bronstein, M., and Zafeiriou, S. Spiralnet++: A fast and highly efficient mesh convolution operator. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pp. 0–0, 2019.

Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., and Cohen-Or, D. Meshcnn: a network with an edge. *ACM Transactions on Graphics (ToG)*, 38(4):1–12, 2019.

Koo, B., Özgür, E., Le Roy, B., Buc, E., and Bartoli, A. Deformable registration of a preoperative 3d liver volume to a laparoscopy image using contour and shading cues. In *International conference on medical image computing and computer-assisted intervention*, pp. 326–334. Springer, 2017.

Li, J., Cheng, B., Niu, N., Gao, G., Ying, S., Shi, J., and Zeng, T. A fine-grained orthodontics segmentation model for 3d intraoral scan data. *Computers in Biology and Medicine*, 168:107821, 2024.

Lim, I., Dielen, A., Campen, M., and Kobbelt, L. A simple approach to intrinsic correspondence learning on unstructured 3d meshes. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.

Martin, C. B., Chalmers, E. V., McIntyre, G. T., Cochrane, H., and Mossey, P. A. Orthodontic scanners: what's available? *Journal of orthodontics*, 42(2):136–143, 2015.

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

Satorras, V. G., Hoogeboom, E., and Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Sun, F., Luo, Z., and Li, S. Boundary difference over union loss for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pp. 292–301. Springer, 2023.

Sun, J., Ovsjanikov, M., and Guibas, L. J. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, volume 28, pp. 1383–1392. Wiley Online Library, 2009.

Tang, L., Zhan, Y., Chen, Z., Yu, B., and Tao, D. Contrastive boundary learning for point cloud segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8489–8499, 2022.

Trang, T. A., Ngo, N. K., Levy, D. T., Vo, T. N., Ravanbakhsh, S., and Hy, T. S. E (3)-equivariant mesh neural networks. In *International Conference on Artificial Intelligence and Statistics*, pp. 748–756. PMLR, 2024.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5):1–12, 2019.

Wu, X., Jiang, L., Wang, P.-S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., and Zhao, H. Point transformer v3: Simpler faster stronger. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4840–4851, 2024.

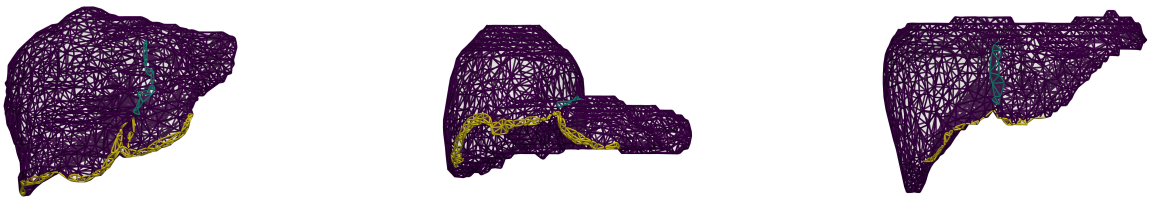
Yang, X., Xia, D., Kin, T., and Igarashi, T. Intra: 3d intracranial aneurysm dataset for deep learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2656–2666, 2020.

Zhang, X., Feng, J., Liu, P., Han, M., Kang, Y., Zhu, J., Wang, L., Wang, X., Ali, S., and Zhang, L. Nested resolution mesh-graph cnn for automated extraction of liver surface anatomical landmarks. *Medical Image Analysis*, pp. 103825, 2025.

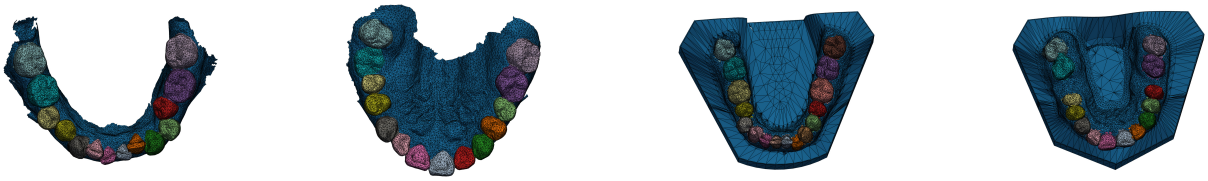
Zhang, Y., Cen, J., Han, J., and Huang, W. Fast and distributed equivariant graph neural networks by virtual node learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

A. Data and Preprocessing

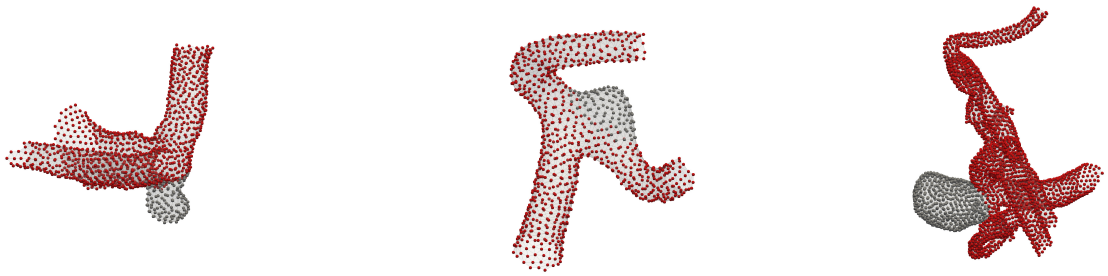
Datasets. The three benchmark datasets span different clinical domains and supervision types. **IntrA** (Yang et al., 2020) contains 103 intracranial aneurysm meshes with per-vertex binary labels (aneurysm sac vs. parent vessel). **3D-IOSSeg** (Li et al., 2024) contains 180 direct intraoral scans with per-face labels covering gingiva and 16 tooth classes; **Teeth3DS** (Ben-Hamadou et al., 2023) provides 1800 indirect scans with the same FDI labelling scheme. The **liver** benchmark (Zhang et al., 2025) provides 200 annotated liver surface meshes from three public CT datasets (3Dircadb, LiTS, Amos) with per-edge labels for the falciform ligament and liver ridge. Representative ground-truth segmentations are shown in Figure 3.



(a) Liver surfaces. Edge colours: falciform ligament (blue), liver ridge (yellow).



(b) Intraoral scans. Left pair: 3D-IOSSeg (direct). Right pair: Teeth3DS (indirect).



(c) Intracranial aneurysm surfaces. Vertex colours: aneurysm vs. vessel wall.

Figure 3. Representative ground-truth segmentations from the three benchmark datasets.

Pipeline overview. The data pipeline separates raw mesh ingestion, reusable geometric preprocessing, and late feature assembly. Expensive quantities—surface normals, adjacency structures, and the Laplace–Beltrami spectral basis—are computed once and stored in per-mesh caches. Model input tensors are assembled after batching, so the same cached geometry supports multiple feature configurations without rerunning the full pipeline.

Mesh cleanup. Each mesh undergoes a defensive cleanup pass: non-finite vertex coordinates and faces referencing removed vertices are discarded; degenerate triangles (repeated vertex indices) and near-zero-area faces are filtered; unreferenced vertices are pruned; connectivity is rebuilt. This cleanup acts as a numerical regularization pass—the Laplace–Beltrami eigensystem is sensitive to degenerate elements—without remeshing or vertex merging, so label alignment is preserved.

Coordinate normalisation. After cleanup, vertex coordinates are centred and isotropically scaled:

$$\boldsymbol{x}_i^{\text{norm}} = (\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \,/\, (s + \varepsilon), \qquad \bar{\boldsymbol{x}} = \tfrac{1}{N} \textstyle\sum_i \boldsymbol{x}_i, \tag{3}$$

where s is the mesh extent. Normalisation reduces sensitivity to acquisition scale and stabilises spectral decomposition.

Spectral preprocessing and HKS caching. The pipeline solves the generalised eigenproblem $\boldsymbol{W}\phi_k = \lambda_k \boldsymbol{M}\phi_k$ (discrete stiffness $\boldsymbol{W}$, mass $\boldsymbol{M}$) and stores the eigenpairs $\{(\lambda_k, \phi_k)\}$. Heat Kernel Signatures (Sun et al., 2009) are then evaluated as $\text{HKS}(i,t) = \sum_k e^{-\lambda_k t}\phi_k(i)^2$ at 8 logarithmically spaced time scales. Caching avoids repeating the expensive spectral decomposition across training runs.

Dataset-specific annotations. Intraoral annotations follow the FDI World Dental Federation numbering system; we squash mirrored FDI pairs (e.g. T11/31) for evaluation. IntrA provides per-vertex binary annotations. Liver annotations are provided per-edge with three classes (background, ligament, ridge).

B. Background and Related Work

Triangle meshes A triangle mesh is a tuple $\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$, where $\mathcal{V} = \{1, \ldots, n\}$ indexes the vertices, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the undirected edge set, and $\mathcal{F} \subseteq \mathcal{V}^3$ is the set of triangular faces, each an ordered triplet of vertices. Each vertex $i \in \mathcal{V}$ carries a coordinate $x_i \in \mathbb{R}^3$. A valid triangle mesh is a manifold: every edge belongs to at most two faces, no two distinct faces intersect except along a shared edge or vertex, and the faces around every vertex form a topological disk. We write $N(i) = \{j : (i,j) \in \mathcal{E}\}$ for the one-ring neighbourhood of vertex i and $\tau(i) = \{(j,k) : (i,j,k) \in \mathcal{F}\}$ for its adjacent faces.

Geometric features Every face $(i,j,k) \in \mathcal{F}$ induces a normal vector $n_{ijk} = (x_j - x_i) \times (x_k - x_i) \in \mathbb{R}^3$, whose magnitude equals twice the face area: $a_{ijk} = \|n_{ijk}\|/2$. Vertex-level normals can be obtained by aggregating over adjacent faces; an area-weighted average gives $n_i = \sum_{(j,k)\in\tau(i)} a_{ijk} n_{ijk} \,/\, \| \sum_{(j,k)\in\tau(i)} a_{ijk} n_{ijk} \|$. These quantities play a central role in our model: the cross product $(x_j - x_i) \times (x_k - x_i)$ appears directly in the surface-aware face messages and in the equivariant coordinate update, encoding face geometry in a way that is invariant to translation and equivariant to rotation.

Deep learning on meshes Working with mesh data requires methods that incorporates the graph topology and geometric features of the mesh. Equivariant methods commonly rely on the invariants of their input features, which are then updated with learnable functions: prominent examples include E(n)-equivariant graph neural networks (Satorras et al., 2021), their extension to meshes EMNN (Trang et al., 2024), and simplices (Eijkelboom et al., 2023). In this work we adopt EMNN as the mesh-native equivariant backbone while leaving a direct EGNN-versus-EMNN comparison to future work so the contribution of equivariance can be disentangled from the contribution of explicit mesh structure. EMNN also has the upside of simplicity, as opposed to more complex convolutional (Lim et al., 2018; Hanocka et al., 2019; Gong et al., 2019; Feng et al., 2019; De Haan et al., 2020) and attention-based methods (Basu et al., 2022).

Medical mesh segmentation These span numerous clinically relevant applications, including liver surface segmentation (Zhang et al., 2025), intracranial aneurysm segmentation (Yang et al., 2020), and intraoral scan tooth segmentation (Ben-Hamadou et al., 2023; Li et al., 2024). We focus on these three tasks as they involve labels on each of the components of the mesh (edges, vertices, and faces, respectively), allowing us to evaluate the versatility of our method across different segmentation paradigms.

C. Feature Construction Details

Table 6 lists all features with their dimensionalities and computation methods. Table 7 summarises the per-dataset configuration.

Equivariance of the dental frame. Under a rigid-body transformation $\boldsymbol{x}_i \mapsto \boldsymbol{R}\boldsymbol{x}_i + \boldsymbol{t}$ with $\boldsymbol{R} \in \text{SO}(3)$, the centred coordinates become $\boldsymbol{R}(\boldsymbol{x}_i - \boldsymbol{c})$, so the covariance transforms as $\boldsymbol{S} \mapsto \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top$ and the PCA eigenvectors as $\boldsymbol{U} \mapsto \boldsymbol{R}\boldsymbol{U}$. Thus the canonical PCA frame itself is exactly SE(3)-equivariant; the only heuristic element is line 3, which chooses between the two sign-equivalent orientations of the leading axis. The AP sign resolution (line 3) depends only on the projected tooth-arch shape, not on the external orientation. Consequently the cylindrical coordinates (r_i, θ_i, z_i) are *identical before*

Table 6. Feature definitions. All features are computed post-batching. Node features are concatenated into $\boldsymbol{h}_i^{(0)}$; edge features into $\boldsymbol{e}_{ij}$.

Feature	Type	Dim	Description
`pointwise_area`	node	1	Mean area of incident triangles per vertex.
`hks`	node	8	Heat Kernel Signature (Sun et al., 2009) from Laplace–Beltrami spectrum; 8 normalised time scales.
`dental_cyl`	node	3	Cylindrical coords (r, θ, z) in a PCA-derived dental frame (Alg. 1); a sign heuristic selects one of the two equivalent axis orientations, while the frame construction itself is SE(3)-equivariant. Teeth3DS and 3D-IOSSeg only.
`canonical_cyl`	node	3	Cylindrical coords in an area-weighted PCA liver frame; a sign heuristic selects one of the two equivalent axis orientations, while the frame construction itself is SE(3)-equivariant. Liver only.
`com_fps`	node	4	Distances to centroid and 3 furthest-point-sampled anchors. Liver only.
`degree_weight`	edge	1	Reciprocal target-node degree.
`copy_weight`	edge	1	Pre-computed pipeline edge weight (or small constant).
`dihedrals`	edge	1	Dihedral angle between adjacent face pair normals (π for boundary edges).

Table 7. Per-dataset default feature configurations and input dimensionalities. The coordinate channels column counts the number of $\mathbb{R}^3$ vector channels propagated alongside the scalar node features.

Dataset	Node features	Edge features	Node dim	Edge dim	Coord. channels
All datasets	`area`, `hks`	`deg`, `copy`, `dihedrals`	9	3	2 (pos. + normal)
Liver	`canonical_cyl` `canonical_cart` `com_fps_anchor_distances`	—	19	3	2
Teeth3DS	`dental_cyl`	—	12	3	2
3D-IOSSeg	`dental_cyl`	—	12	3	2
IntrA	—	—	9	3	2

and after any SE(3) transformation. Reflections (improper rotations) are not compensated: they flip $\det(\boldsymbol{U})$ and would swap left and right arch sides, which is intentional since dental anatomy has left–right chirality.

Liver frame. The liver uses the same PCA-to-cylindrical construction as the dental frame, but with area-weighted quantities to reduce sensitivity to irregular tessellation. If A_i is the local vertex area, the area-weighted centroid is

$$\boldsymbol{\mu} = \frac{1}{\sum_i A_i} \sum_i A_i \, \boldsymbol{x}_i, \tag{4}$$

and the covariance is

$$\boldsymbol{C} = \frac{1}{\sum_i A_i} \sum_i A_i (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top. \tag{5}$$

Weighting by area makes the canonical frame depend on the physical surface distribution rather than vertex density: densely triangulated nuisance regions near vessels or thin folds contribute proportionally less, keeping the principal axes aligned with the dominant right-lobe versus tapering left-lobe mass distribution. The same SE(3)-equivariance argument as above applies, so here too the only approximation comes from the sign-disambiguation heuristic rather than from the PCA frame construction itself.

Algorithm 1 Dental cylindrical frame

Require: Vertex coordinates $\{\boldsymbol{x}_i\}$
Centre: $\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i - \frac{1}{N}\sum_j \boldsymbol{x}_j$
PCA: $[\boldsymbol{U}, \boldsymbol{\Lambda}] \leftarrow \mathrm{eig}(\sum_i \boldsymbol{x}_i \boldsymbol{x}_i^\top)$
Resolve AP sign: project vertices as $y_i \leftarrow [\boldsymbol{U}^\top \boldsymbol{x}_i]_1$; fit a quadratic $y \mapsto p(y)$ to the arch; if p is minimised at the positive end, negate column 1 of $\boldsymbol{U}$
Rotate: $\boldsymbol{y}_i \leftarrow \boldsymbol{U}^\top \boldsymbol{x}_i$
Shift posterior baseline: $[\boldsymbol{y}_i]_2 \leftarrow [\boldsymbol{y}_i]_2 - \min_j [\boldsymbol{y}_j]_2$
Convert to cylindrical: $(r_i, \theta_i, z_i) \leftarrow \mathrm{cart2cyl}(\boldsymbol{y}_i)$
Ensure: Per-vertex (r_i, θ_i, z_i)

Algorithm 2 Soft regional aggregator (SRA) update

Require: Node features $\{\boldsymbol{h}_i\}$, coordinates $\{\boldsymbol{x}_i\}$, number of regions K, learned residual weight α
for each node i **do**
  $\boldsymbol{a}_i \leftarrow \mathrm{softmax}(\phi_a(\boldsymbol{h}_i)) \in \mathbb{R}^K$ (optionally augmented with dist. to CoM $\boldsymbol{u}_i = \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|$)
end for
Stack into assignment matrix: $[\boldsymbol{A}]_{ik} \leftarrow [\boldsymbol{a}_i]_k$, so $\boldsymbol{A} \in \mathbb{R}^{N \times K}$
for each region k **do**
  $\boldsymbol{r}_k \leftarrow \sum_i [\boldsymbol{A}]_{ik}\, \boldsymbol{h}_i$ (optionally augmented with geom. feature $\boldsymbol{u}_k^r = \sum_i [\boldsymbol{A}]_{ik}\, \boldsymbol{u}_i$)
end for
Mix: $[\hat{\boldsymbol{r}}_1, \ldots, \hat{\boldsymbol{r}}_K] \leftarrow \mathrm{TransformerEncoder}([\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K])$
for each node i **do**
  $\boldsymbol{h}_i \leftarrow \boldsymbol{h}_i + \alpha \sum_k [\boldsymbol{A}]_{ik} \phi_{\mathrm{proj}}(\hat{\boldsymbol{r}}_k)$
end for

D. Training and Implementation Details

Optimiser and schedule. AdamW (β_1=0.9, β_2=0.999, weight decay 0.01), initial LR 10^{-3}, plateau schedule (factor 0.6, min LR 10^{-5}), gradient clipping at 1.0. 100 epochs, validation every 5 epochs.

Batch size and cross-validation. Batch size 1 (liver: 8). Five-fold cross-validation for liver, 3D-IOSSeg, and IntrA. Single split for Teeth3DS following Ben-Hamadou et al. (2023).

Loss details. Here we describe the different loss types we use during training.

- **Prediction loss ($\mathcal{L}_{\text{pred}}$).** We use the boundary difference over union (bDoU) (Sun et al., 2023) loss for the segmentation predictions.
- **Contrastive boundary loss ($\mathcal{L}_{\text{cbl}}$).** We add a contrastive boundary objective (CBL) (Tang et al., 2022) on the latent embeddings, encouraging nearby points across class boundaries to become more discriminative.
- **Local continuity loss ($\mathcal{L}_{\text{cont}}$).** Segmentation of landmarks on the liver surface is particularly challenging due to the subtle geometric cues needed to identify the folds defining the ligament. Moreover, the lack of absolute coordinates makes it challenging to identify the "front" of the liver (where both landmarks are located), the "bottom" of the liver (where the ridge is located), and the "middle" of the liver (where the ligament is located). To address this, we introduce a local continuity loss that encourages adjacent edges to have similar predictions. Let $\mathcal{A}$ denote the set of adjacent edge pairs. The loss $\mathcal{L}_{\text{cont}} = \frac{1}{|\mathcal{A}|} \sum_{(e_i, e_j) \in \mathcal{A}} \|\hat{y}_{e_i} - \hat{y}_{e_j}\|^2$ encourages smoothness in the predicted labels across neighboring edges.
- **SRA regularization losses ($\mathcal{L}_{\text{SRA}}$).** We regularise the assignment matrix $\boldsymbol{A} \in \mathbb{R}^{N \times K}$ (whose rows are the per-vertex assignment vectors $\boldsymbol{a}_i$) with two unsupervised objectives. A *diversity loss* penalises off-diagonal entries of the Gram matrix of the column-normalised assignments: $\mathcal{L}_{\text{div}} = \|\tilde{\boldsymbol{A}}^\top \tilde{\boldsymbol{A}} - \boldsymbol{I}\|_F^2$, where $\tilde{\boldsymbol{A}}$ denotes column-wise ℓ_2-normalised $\boldsymbol{A}$. An *equipartition loss* penalises imbalance in the total mass assigned to each region: $\mathcal{L}_{\text{eq}} = \mathrm{Var}(\mathbf{1}^\top \boldsymbol{A})/(\bar{m}^2 + \epsilon)$, where $\bar{m}$ is the mean column mass.

Algorithm 3 Virtual node update (per EMNN layer)

Require: Node features $\{\boldsymbol{h}_i\}$, coordinates $\{\boldsymbol{x}_i\}$, mesh edges/faces; virtual features $\{\boldsymbol{v}_k\}_{k=1}^{V}$, virtual coordinates $\{\boldsymbol{u}_k\}$

Compute EMNN edge messages $\boldsymbol{m}_i^{\text{edge}}$, face messages $\boldsymbol{m}_i^{\text{face}}$, and coordinate updates $\Delta\boldsymbol{x}_i^{\text{edge}}, \Delta\boldsymbol{x}_i^{\text{face}}$
Compute graph centroid $\bar{\boldsymbol{x}}$; set $\boldsymbol{m}_k^{v} \leftarrow \|\boldsymbol{u}_k - \bar{\boldsymbol{x}}\|^2$ (virtual correlation feature)
for each node i and virtual node k in same graph **do**
    $\Delta\boldsymbol{x}_{ik} \leftarrow \boldsymbol{u}_k - \boldsymbol{x}_i$
    $\boldsymbol{m}_{ik}^{\text{virt}} \leftarrow \phi_{rv}(\boldsymbol{h}_i,\, \boldsymbol{v}_k,\, \|\Delta\boldsymbol{x}_{ik}\|^2,\, \boldsymbol{m}_k^{v})$
end for
Aggregate to nodes: $\boldsymbol{m}_i^{\text{virt}} \leftarrow \frac{1}{V}\sum_k \boldsymbol{m}_{ik}^{\text{virt}}$; $\Delta\boldsymbol{x}_i^{\text{virt}} \leftarrow \frac{1}{V}\sum_k \phi_{\text{v2n}}(\boldsymbol{m}_{ik}^{\text{virt}})(\boldsymbol{x}_i - \boldsymbol{u}_k)$
Aggregate to virtual nodes: $\bar{\boldsymbol{m}}_k^{\text{virt}} \leftarrow \frac{1}{N}\sum_i \boldsymbol{m}_{ik}^{\text{virt}}$; $\Delta\boldsymbol{u}_k \leftarrow \frac{1}{N}\sum_i \phi_{\text{n2v}}(\boldsymbol{m}_{ik}^{\text{virt}})(\boldsymbol{u}_k - \boldsymbol{x}_i)$
Update virtual: $\boldsymbol{v}_k \leftarrow \boldsymbol{v}_k + \phi_v(\boldsymbol{v}_k,\, \bar{\boldsymbol{m}}_k^{\text{virt}})$; $\boldsymbol{u}_k \leftarrow \boldsymbol{u}_k + \Delta\boldsymbol{u}_k$
Update nodes: $\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i + \Delta\boldsymbol{x}_i^{\text{edge}} + \Delta\boldsymbol{x}_i^{\text{face}} + \Delta\boldsymbol{x}_i^{\text{virt}}$; $\boldsymbol{h}_i \leftarrow \phi_h([\boldsymbol{h}_i,\, \boldsymbol{m}_i^{\text{edge}},\, \boldsymbol{m}_i^{\text{face}},\, \boldsymbol{m}_i^{\text{virt}}])$

Ensure: Updated $\{\boldsymbol{h}_i,\, \boldsymbol{x}_i,\, \boldsymbol{v}_k,\, \boldsymbol{u}_k\}$

- **Virtual node regularization losses ($\mathcal{L}_{\text{VN}}$).** We use a kernel-based energy objective on the virtual-node coordinates $\{\boldsymbol{u}_k\}$, following (Zhang et al., 2026). Let $k(\boldsymbol{p}, \boldsymbol{q}) = \exp(-\|\boldsymbol{p} - \boldsymbol{q}\|^2/2\sigma^2)$ be a Gaussian kernel, which is E(3)-invariant. We compute a virtual–virtual repulsion term k_{vv} as the mean off-diagonal pairwise kernel among $\{\boldsymbol{u}_k\}$, and a real–virtual attraction term k_{rv} as the mean kernel between a random subsample of $\{\boldsymbol{x}_i\}$ and $\{\boldsymbol{u}_k\}$. The loss $\mathcal{L}_{\text{VN}} = w_{vv}\, k_{vv} - w_{rv}\, k_{rv}$ simultaneously discourages virtual nodes from collapsing together and encourages them to remain close to the mesh surface.

The full objective is obtained by adding the task-specific auxiliary terms to $\mathcal{L}_{\text{task}}$. For the base EAMS model, we optimise $\mathcal{L}_{\text{task}}$, and on liver meshes we additionally add the continuity term $\lambda_{\text{cont}}\, \mathcal{L}_{\text{cont}}$. For SRA+EAMS, we optimise $\mathcal{L}_{\text{task}} + \mathcal{L}_{\text{SRA}}$, while for VN+EAMS we optimise $\mathcal{L}_{\text{task}} + \mathcal{L}_{\text{VN}}$.

Loss weights. The prediction loss coefficient is 1.0 for all tasks. The contrastive boundary loss coefficient is 10.0 for IntrA, 3D-IOSSeg, and Teeth3DS, and 1.0 for liver. The local continuity term is used only for liver, with $\lambda_{\text{cont}} = 1.0$. For SRA runs we set $\lambda_{\text{div}} = \lambda_{\text{eq}} = 1.0$, and for VN runs we use $w_{\text{vv}} = w_{\text{rv}} = 1.0$.

Global-token configuration. For the liver experiments, SRA+EAMS uses $K = 32$ soft regional aggregators and VN+EAMS uses $V = 8$ virtual nodes.

Architecture. Encoder: 3-layer EMNN with hidden dim 128, SiLU activations, batch normalisation, 2 equivariant vector channels. Decoder: MLP with hidden dims $(128, 128, 128)$, SiLU, dropout 0.1, applied per-node. Liver uses an edge-level decoder that forms $\boldsymbol{z}_{ji} = [\boldsymbol{h}_j \| \boldsymbol{h}_i \| \boldsymbol{e}_{ji} \| \psi(\|\boldsymbol{x}_j - \boldsymbol{x}_i\|)]$ and aggregates edge predictions to vertices, where ψ is a 2-layer MLP with output dim 16.

Training runtimes. On a single H100 GPU, the base EAMS model takes about 1.2 s per epoch on IntrA, 7 s on liver, 9.2 s on 3D-IOSSeg, and 72 s on Teeth3DS. Relative to EAMS, SRA+EAMS increases runtime by $1.45\times$ and VN+EAMS $1.64\times$.

E. Full Experimental Results

E.1. Intracranial Aneurysm: Full Robustness

Table 8 extends Table 1 with IoU scores and additional perturbation conditions for the two non-equivariant baselines. EAMS, SRA+EAMS, and VN+EAMS are E(3)-invariant, so their scores are identical to the unperturbed results already shown in the main text and are omitted here. The purpose of this table is therefore to make the failure mode of the baselines explicit: once the aneurysm mesh is rotated or reflected, both parent-vessel and aneurysm quality degrade together.

E.2. Intraoral Scans: Full Results

Table 9 extends Table 2 with Rot-15° for both benchmarks. The more important standalone detail, however, is the per-class behaviour: the same aggregate mIoU can hide large differences between anterior teeth, premolars, molars, and wisdom teeth.

Table 8. Full robustness results for IntrA (mean ± std over 5 folds). EAMS variants are E(3)-invariant; their scores equal their baseline values under all perturbations and are omitted.

		Parent vessel		Aneurysm	
Method	Condition	Dice (%) ↑	IoU (%) ↑	Dice (%) ↑	IoU (%) ↑
DGCNN (Wang et al., 2019)	Baseline	97.11 ± 1.35	94.62 ± 2.38	87.77 ± 5.33	80.79 ± 7.45
	Rot-z 15°	96.75 ± 1.44	93.95 ± 2.52	86.24 ± 5.95	78.42 ± 7.88
	Rot-z 40°	94.34 ± 0.88	89.59 ± 1.53	76.61 ± 2.85	65.20 ± 3.73
	Refl-x	95.10 ± 0.22	90.90 ± 0.32	77.18 ± 1.32	66.77 ± 1.17
PTv3 (Wu et al., 2024)	Baseline	94.86 ± 1.17	90.54 ± 2.11	79.80 ± 4.63	70.59 ± 5.76
	Rot-z 15°	94.25 ± 1.32	89.49 ± 2.33	78.51 ± 3.93	68.54 ± 5.18
	Rot-z 40°	93.06 ± 1.33	87.40 ± 2.29	72.67 ± 4.23	62.18 ± 5.05
	Refl-x	93.02 ± 1.52	87.40 ± 2.56	72.68 ± 4.44	62.61 ± 5.29

Table 9. 3D-IOSSeg and Teeth3DS intraoral tooth segmentation. Metric: mean IoU (%) over all classes. 3D-IOSSeg: mean ± std over 5 folds. Teeth3DS: single fold (no std). **Bold**: best; underline: second-best.

	3D-IOSSeg			Teeth3DS		
Method	Baseline ↑	Rot-15° ↑	Rot-40° ↑	Baseline ↑	Rot-15° ↑	Rot-40° ↑
DGCNN (Wang et al., 2019)	67.12 ± 0.36	63.25 ± 0.81	42.02 ± 2.46	—	—	—
PTv3 (Wu et al., 2024)	80.08 ± 0.44	75.87 ± 0.45	39.05 ± 2.35	—	—	—
Fast-TGCN (Li et al., 2024)	**81.05 ± 0.74**	76.45 ± 1.14	52.33 ± 1.62	**84.31**	81.84	58.53
EAMS	71.43 ± 1.64	71.71 ± 1.53	72.05 ± 1.60	80.54	80.94	81.11
SRA+EAMS	75.28 ± 1.21	75.49 ± 1.44	75.98 ± 1.23	82.31	82.98	83.05
VN+EAMS	79.74 ± 0.86	**79.92 ± 0.98**	**80.09 ± 1.01**	83.16	**83.33**	**83.40**

For that reason we also include the per-tooth tables from the reference paper below. We use squashed FDI notation, so a heading such as T11/31 averages the mirrored upper and lower classes.

Tables 10–13 provide the full per-class IoU breakdowns by FDI tooth number for both intraoral benchmarks. These appendix rows are class-wise summaries, whereas the main-table Average IoU averages each mesh over its ground-truth-present classes and then averages over meshes. For 3D-IOSSeg, the EAMS-family models are shown only at baseline because their clean and rotated scores remain close at the aggregate level; for Teeth3DS we retain all conditions because the final sign-resolution heuristic in the PCA-derived frame introduces a small residual sensitivity in practice. Across both datasets, the highest variance concentrates in the rearmost wisdom-tooth classes (T18/38 and T28/48), which are both morphologically diverse and relatively data-sparse.

E.3. Liver Surface: Full Results

Table 14 extends Table 3 with IoU and Hausdorff distance (HD), and reports Chamfer distance after multiplying by 100 for readability, across all listed test-time conditions. The EAMS-family rows show only minor variation across rotations, in contrast to the large degradation of the non-equivariant baselines. Among the EAMS variants, SRA+EAMS gives the strongest ligament distance metrics, while VN+EAMS gives the strongest ridge distance metrics.

E.4. Qualitative Results

Figures 4 and 5 restore qualitative comparisons that are omitted from the main paper for space. Figure 4 focuses on the invariant mesh variants for IntrA, while Figure 5 provides extra liver qualitative results beyond the main-text MeshGraphCNN-versus-VN comparison.

Table 10. Per-class IoU (%) for the left jaw half on 3D-IOSSeg (Li et al., 2024) (Gingiva and T11/31–T18/38). Results are mean ± std over five cross-validation folds. Baseline scores are mostly unchanged under rotation.

Method	Condition	Gingiva ↑	T11/31 ↑	T12/32 ↑	T13/33 ↑	T14/34 ↑	T15/35 ↑	T16/36 ↑	T17/37 ↑	T18/38 ↑
DGCNN (Wang et al., 2019)	Baseline	81.69 ± 0.68	60.02 ± 0.41	60.00 ± 0.67	71.68 ± 0.82	71.55 ± 0.40	64.33 ± 0.32	74.95 ± 0.36	75.33 ± 1.68	50.93 ± 1.10
	Rot-z 15°	79.36 ± 0.32	54.56 ± 1.24	52.68 ± 1.53	67.77 ± 0.71	65.95 ± 1.60	57.08 ± 0.56	73.71 ± 0.89	74.89 ± 1.83	47.29 ± 1.68
	Rot-z 40°	72.41 ± 1.43	29.86 ± 2.96	28.00 ± 3.09	46.15 ± 4.36	42.99 ± 2.40	27.02 ± 0.74	53.82 ± 2.38	41.85 ± 5.00	33.36 ± 5.50
PTv3 (Wu et al., 2024)	Baseline	91.09 ± 0.22	79.95 ± 0.62	77.55 ± 0.45	83.99 ± 0.64	81.31 ± 0.57	80.88 ± 1.51	83.77 ± 0.70	83.33 ± 0.97	54.23 ± 2.71
	Rot-z 15°	90.50 ± 0.20	76.48 ± 0.98	71.93 ± 1.55	80.42 ± 0.86	77.25 ± 0.46	73.87 ± 1.25	82.05 ± 0.98	81.90 ± 1.13	34.27 ± 6.71
	Rot-z 40°	80.92 ± 0.61	54.92 ± 2.23	28.71 ± 2.46	53.94 ± 4.05	28.53 ± 3.75	21.91 ± 6.41	34.62 ± 5.91	35.23 ± 3.65	10.79 ± 3.65
Fast-TGCN (Li et al., 2024)	Baseline	94.80 ± 0.66	83.96 ± 0.93	77.78 ± 2.07	84.70 ± 0.42	77.58 ± 1.79	80.20 ± 1.06	82.09 ± 0.93	84.80 ± 0.87	56.12 ± 3.26
	Rot-z 15°	94.00 ± 0.78	81.52 ± 1.66	72.83 ± 1.80	79.39 ± 0.61	71.63 ± 2.85	75.02 ± 1.87	79.96 ± 1.22	82.70 ± 1.10	52.54 ± 1.79
	Rot-z 40°	84.90 ± 2.32	54.67 ± 3.50	38.50 ± 3.15	46.09 ± 2.71	42.71 ± 1.42	49.19 ± 0.76	63.91 ± 2.41	67.19 ± 2.51	43.16 ± 4.78
EAMS (ours)	Baseline	93.15 ± 0.75	66.18 ± 3.22	69.83 ± 1.93	73.80 ± 2.51	71.46 ± 2.22	67.08 ± 2.06	76.23 ± 1.75	74.93 ± 2.71	26.04 ± 5.75
SRA+EAMS (ours)	Baseline	94.29 ± 0.38	77.19 ± 2.03	75.33 ± 1.07	78.78 ± 1.28	72.40 ± 1.86	71.99 ± 1.50	75.01 ± 1.40	77.00 ± 1.65	19.52 ± 4.43
VN+EAMS (ours)	Baseline	93.87 ± 0.71	80.16 ± 1.35	80.00 ± 0.68	84.85 ± 0.90	77.94 ± 0.91	78.20 ± 0.82	82.90 ± 0.71	82.17 ± 1.36	44.90 ± 7.71

Table 11. Per-class IoU (%) for the right jaw half on 3D-IOSSeg (Li et al., 2024) (T21/41–T28/48). Results are mean ± std over five cross-validation folds. Baseline scores are mostly unchanged under rotation.

Method	Condition	T21/41 ↑	T22/42 ↑	T23/43 ↑	T24/44 ↑	T25/45 ↑	T26/46 ↑	T27/47 ↑	T28/48 ↑
DGCNN (Wang et al., 2019)	Baseline	61.89 ± 1.22	60.57 ± 1.79	70.44 ± 1.05	70.10 ± 0.28	66.39 ± 0.48	70.69 ± 0.39	69.18 ± 1.18	55.80 ± 0.96
	Rot-z 15°	54.77 ± 1.37	55.35 ± 1.25	65.21 ± 0.46	66.75 ± 0.22	63.95 ± 0.28	68.53 ± 0.38	68.94 ± 1.52	53.08 ± 2.62
	Rot-z 40°	28.21 ± 1.41	29.98 ± 3.45	46.87 ± 1.92	55.17 ± 1.96	49.64 ± 1.56	52.65 ± 3.06	48.98 ± 4.36	24.86 ± 8.75
PTv3 (Wu et al., 2024)	Baseline	80.04 ± 0.90	77.58 ± 0.84	83.15 ± 0.38	81.32 ± 0.59	78.34 ± 0.98	79.24 ± 1.52	78.97 ± 1.22	59.51 ± 3.31
	Rot-z 15°	74.91 ± 1.16	73.93 ± 0.90	79.77 ± 1.05	77.48 ± 0.77	74.84 ± 1.27	74.70 ± 1.53	75.54 ± 1.07	48.60 ± 8.30
	Rot-z 40°	53.85 ± 1.48	44.78 ± 1.32	46.37 ± 3.93	35.08 ± 3.54	23.32 ± 8.13	30.94 ± 7.96	28.99 ± 4.76	24.90 ± 2.15
Fast-TGCN (Li et al., 2024)	Baseline	85.53 ± 0.88	82.43 ± 1.41	87.28 ± 1.02	82.70 ± 0.68	74.43 ± 1.67	76.35 ± 0.73	78.14 ± 2.28	46.50 ± 9.52
	Rot-z 15°	81.77 ± 1.42	76.38 ± 2.04	77.91 ± 0.76	71.81 ± 1.39	66.08 ± 1.69	75.08 ± 1.23	78.09 ± 2.97	53.29 ± 6.24
	Rot-z 40°	57.53 ± 3.55	45.88 ± 4.40	41.52 ± 2.48	38.50 ± 1.54	41.32 ± 3.51	57.72 ± 2.94	64.30 ± 3.02	49.42 ± 6.71
EAMS (ours)	Baseline	68.50 ± 3.20	67.63 ± 2.31	73.54 ± 2.04	72.32 ± 1.89	67.50 ± 1.59	72.28 ± 1.31	73.61 ± 1.96	36.78 ± 3.26
SRA+EAMS (ours)	Baseline	79.57 ± 1.77	77.09 ± 1.69	77.78 ± 1.57	76.06 ± 1.38	68.19 ± 1.88	71.39 ± 1.84	77.80 ± 2.29	27.73 ± 8.92
VN+EAMS (ours)	Baseline	80.87 ± 2.94	78.84 ± 2.25	82.66 ± 1.08	82.39 ± 1.28	75.29 ± 1.04	77.55 ± 1.52	77.68 ± 0.71	43.31 ± 5.77

Table 12. Per-class IoU (%) for the left jaw half on Teeth3DS (Ben-Hamadou et al., 2023) (Gingiva and T11/31–T18/38). Single-fold results; no standard deviation reported. All conditions listed for all methods due to PCA-induced imperfect SE(3)-invariance.

Method	Condition	Gingiva ↑	T11/31 ↑	T12/32 ↑	T13/33 ↑	T14/34 ↑	T15/35 ↑	T16/36 ↑	T17/37 ↑	T18/38 ↑
Fast-TGCN (Li et al., 2024)	Baseline	90.63	85.73	84.02	82.55	87.25	85.90	85.55	75.35	55.69
	Rot-z 15°	90.37	83.27	81.32	79.72	84.16	82.17	81.73	71.64	36.88
	Rot-z 40°	86.20	57.21	55.35	56.87	50.97	53.54	65.15	55.06	19.28
EAMS (ours)	Baseline	90.53	80.80	80.55	82.28	82.41	78.22	79.81	75.27	62.87
	Rot-z 15°	90.68	81.17	81.33	82.97	82.98	78.55	79.86	75.48	70.10
	Rot-z 40°	90.73	81.48	81.74	83.37	83.39	78.75	79.88	75.52	70.11
SRA+EAMS (ours)	Baseline	90.42	84.34	83.17	83.78	85.41	82.39	82.15	73.64	57.64
	Rot-z 15°	90.74	84.32	83.58	84.50	85.67	82.65	83.01	75.86	69.54
	Rot-z 40°	90.76	84.42	83.58	84.46	85.69	82.70	83.09	75.89	69.51
VN+EAMS (ours)	Baseline	91.18	85.11	84.07	83.88	84.99	81.79	82.04	76.32	36.04
	Rot-z 15°	91.26	84.89	83.79	83.52	85.03	81.73	82.19	76.93	48.14
	Rot-z 40°	91.28	84.99	83.95	83.75	85.15	81.86	82.28	77.05	48.14

Table 13. Per-class IoU (%) for the right jaw half on Teeth3DS (Ben-Hamadou et al., 2023) (T21/41–T28/48). Single-fold results; no standard deviation reported. All conditions listed for all methods due to PCA-induced imperfect SE(3)-invariance.

Method	Condition	T21/41 ↑	T22/42 ↑	T23/43 ↑	T24/44 ↑	T25/45 ↑	T26/46 ↑	T27/47 ↑	T28/48 ↑
Fast-TGCN (Li et al., 2024)	Baseline	84.62	84.04	83.12	87.66	87.66	85.78	75.22	57.28
	Rot-z 15°	83.68	81.76	82.40	86.02	83.74	82.83	73.26	61.85
	Rot-z 40°	61.81	54.16	56.86	66.17	54.32	49.73	57.37	46.00
EAMS (ours)	Baseline	80.04	79.40	81.66	81.49	80.33	81.83	75.27	63.04
	Rot-z 15°	80.17	79.51	81.64	81.96	81.15	82.53	75.73	68.62
	Rot-z 40°	80.41	79.69	81.70	81.96	81.10	82.55	75.83	68.58
SRA+EAMS (ours)	Baseline	83.36	82.53	83.37	84.41	82.46	81.26	72.71	58.28
	Rot-z 15°	83.67	83.02	83.75	84.86	82.92	82.38	74.76	68.98
	Rot-z 40°	83.78	83.07	83.67	84.87	82.93	82.59	75.32	69.00
VN+EAMS (ours)	Baseline	85.05	84.63	84.72	84.84	82.59	82.01	75.19	44.90
	Rot-z 15°	84.86	84.78	85.00	85.07	82.84	82.41	76.28	57.18
	Rot-z 40°	85.02	84.80	85.00	85.06	82.81	82.41	76.18	57.16

Table 14. Full robustness results for liver surface segmentation on the dataset of (Zhang et al., 2025). Results are mean ± std over five cross-validation folds. CD is reported after multiplying by 100. Table 3 shows the corresponding Dice and CD summaries for the baseline and 40° conditions.

		Ligament			Ridge		
Method	Condition	IoU (%) ↑	CD×100 ↓	HD ↓	IoU (%) ↑	CD×100 ↓	HD ↓
PointNet++ (Qi et al., 2017)	Baseline	23.65 ± 1.76	0.202 ± 0.051	0.0536 ± 0.0041	44.65 ± 0.94	0.067 ± 0.007	0.0498 ± 0.0028
	Rot-z 15°	14.36 ± 1.38	0.444 ± 0.099	0.0717 ± 0.0061	43.24 ± 0.39	0.176 ± 0.037	0.0704 ± 0.0077
	Rot-z 40°	3.94 ± 1.40	2.842 ± 0.947	0.1490 ± 0.0173	29.51 ± 2.14	1.352 ± 0.235	0.2037 ± 0.0169
DGCNN (Wang et al., 2019)	Baseline	22.60 ± 2.80	0.293 ± 0.032	0.0629 ± 0.0032	44.42 ± 1.10	0.080 ± 0.021	0.0534 ± 0.0055
	Rot-z 15°	17.31 ± 1.48	0.735 ± 0.177	0.0943 ± 0.0033	42.70 ± 1.06	0.189 ± 0.075	0.0826 ± 0.0129
	Rot-z 40°	1.58 ± 0.41	4.331 ± 1.134	0.1929 ± 0.0163	27.37 ± 0.93	1.619 ± 0.126	0.2229 ± 0.0116
MeshGraphCNN (Zhang et al., 2025)	Baseline	34.57 ± 0.99	0.677 ± 0.046	0.1127 ± 0.0066	46.96 ± 0.87	0.253 ± 0.056	0.0760 ± 0.0104
	Rot-z 15°	20.59 ± 0.31	1.424 ± 0.036	0.1431 ± 0.0063	41.32 ± 0.85	0.507 ± 0.053	0.1143 ± 0.0069
	Rot-z 40°	2.59 ± 0.38	7.677 ± 0.865	0.2810 ± 0.0083	25.52 ± 0.55	2.484 ± 0.073	0.2840 ± 0.0056
EAMS (ours)	Baseline	13.02 ± 0.75	3.928 ± 0.668	0.1353 ± 0.0153	41.12 ± 0.96	0.873 ± 0.064	0.1510 ± 0.0046
	Rot-z 15°	13.02 ± 0.75	4.118 ± 0.527	0.1376 ± 0.0129	41.12 ± 0.94	0.874 ± 0.065	0.1512 ± 0.0044
	Rot-z 40°	13.03 ± 0.75	4.307 ± 0.193	0.1399 ± 0.0094	41.11 ± 0.96	0.876 ± 0.065	0.1511 ± 0.0044
SRA+EAMS (ours)	Baseline	17.10 ± 1.54	1.722 ± 0.651	0.0953 ± 0.0051	39.35 ± 0.63	1.071 ± 0.372	0.1416 ± 0.0267
	Rot-z 15°	17.13 ± 1.52	1.722 ± 0.650	0.0954 ± 0.0051	39.35 ± 0.65	1.070 ± 0.371	0.1416 ± 0.0267
	Rot-z 40°	17.10 ± 1.54	1.722 ± 0.650	0.0953 ± 0.0051	39.34 ± 0.66	1.069 ± 0.371	0.1416 ± 0.0267
VN+EAMS (ours)	Baseline	13.42 ± 0.76	2.013 ± 0.592	0.1161 ± 0.0260	42.74 ± 2.15	0.610 ± 0.160	0.1283 ± 0.0168
	Rot-z 15°	13.43 ± 0.77	2.014 ± 0.593	0.1161 ± 0.0260	42.75 ± 2.16	0.608 ± 0.159	0.1282 ± 0.0166
	Rot-z 40°	13.43 ± 0.77	2.012 ± 0.591	0.1161 ± 0.0260	42.75 ± 2.16	0.609 ± 0.159	0.1282 ± 0.0168

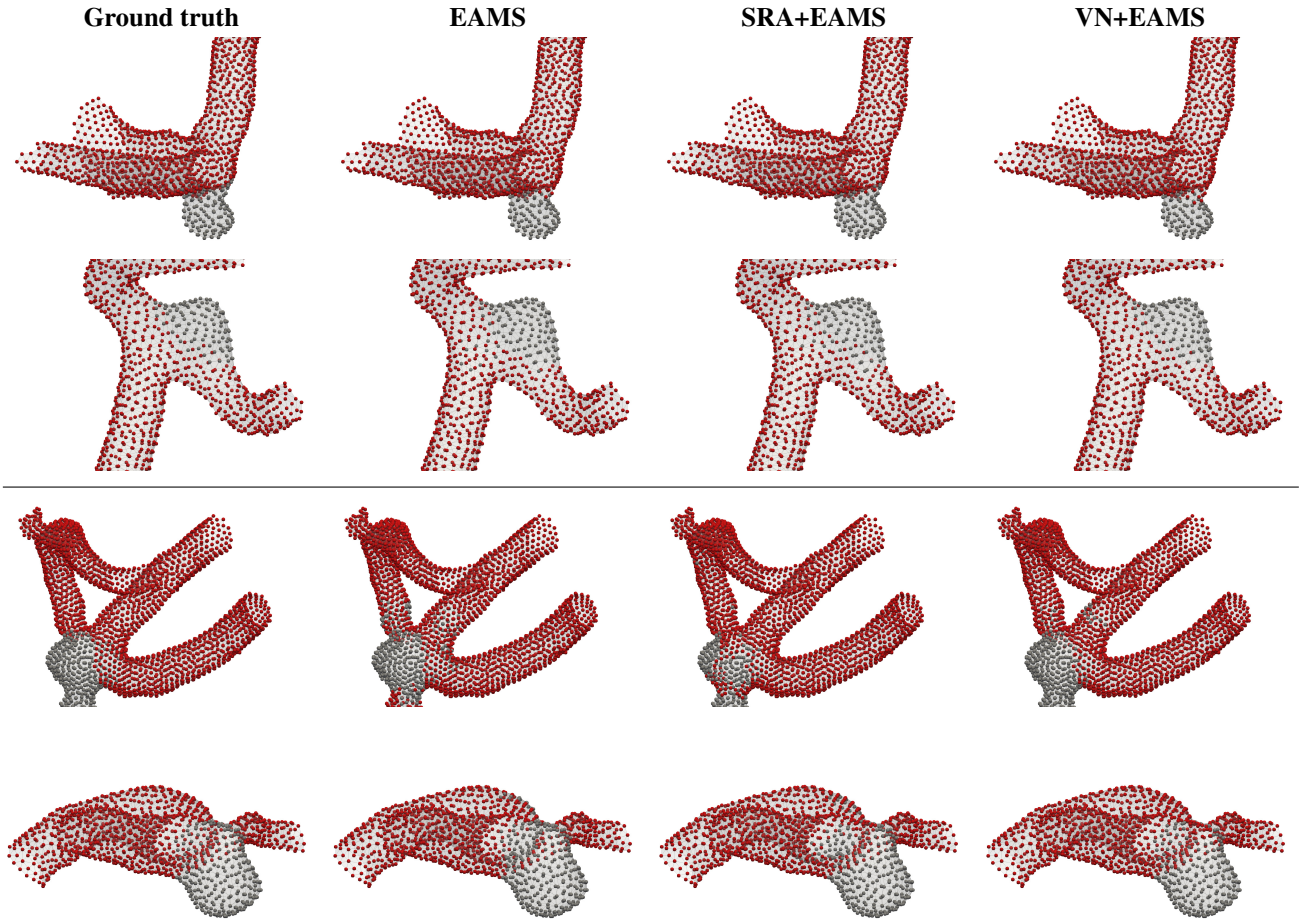


Figure 4. Additional qualitative IntrA comparison for the invariant mesh variants. The top half shows the canonical orientation and the bottom half shows Rot-40°. All three EAMS-family models remain visually stable under rotation, with SRA+EAMS and VN+EAMS usually producing the cleanest aneurysm boundaries.

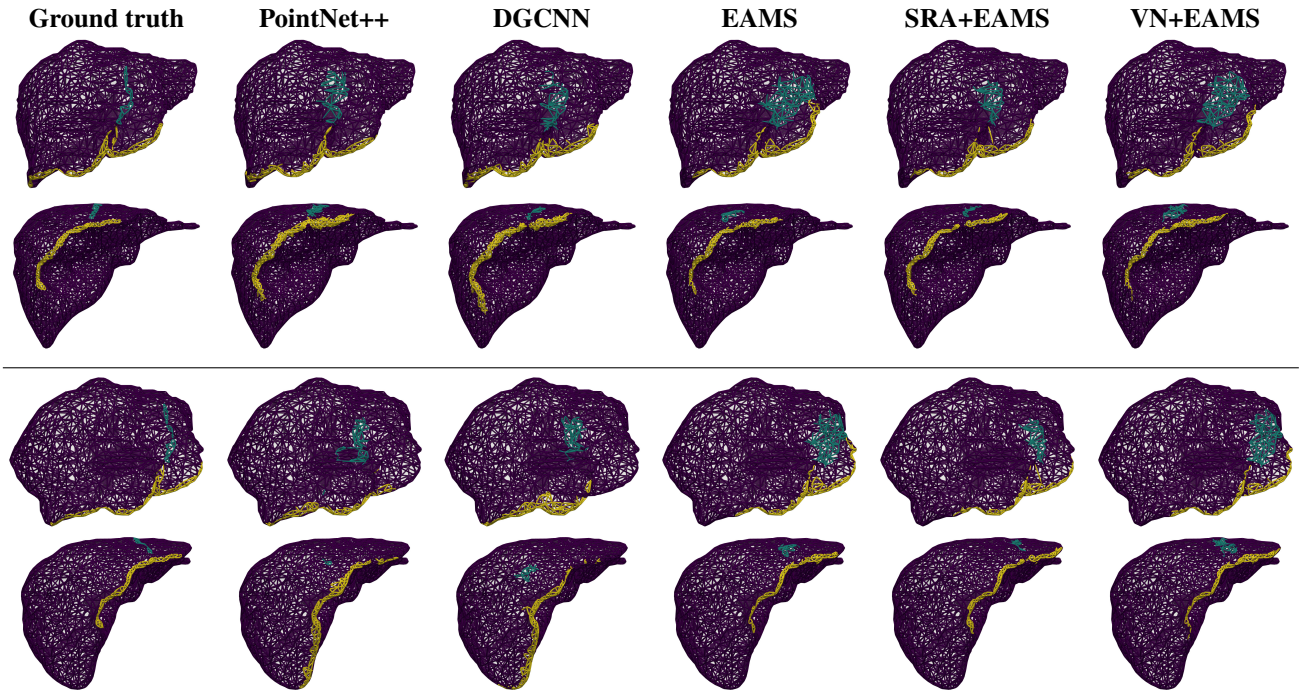


Figure 5. Additional qualitative liver results beyond the main-text figure. The extra columns add PointNet++, DGCNN, and the full EAMS family on representative cases from the liver task, with the top half in the canonical orientation and the bottom half after a 40° z-axis rotation. These extra results make the same robustness pattern explicit across both non-equivariant point baselines and invariant mesh variants.