TAXONOMY-ALIGNED HYPERBOLIC LEARNING FOR 3D POINT CLOUDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Hierarchical 3D segmentation requires accurate point-wise predictions and coherent representations across semantic levels. However, enforcing consistency among predicted labels does not by itself ensure that feature geometry reflects the underlying taxonomy. We introduce Hy3D, a hierarchical hyperbolic prototype network that learns multi-level point features in a shared Poincaré ball for 3D semantic and part segmentation. A shared hyperbolic projection layer maps features from all hierarchy levels into a common space with a unified scale, while an online prototype bank maintains per-level, per-class prototypes through cloud-balanced exponential moving averages in the tangent space. During training, a parent-prototype triplet loss and a depth-aware radius loss encourage parent affinity and radial progression, aligning the learned representations with the label hierarchy. Möbius classification heads produce class logits from the hyperbolic features without modifying the underlying Euclidean backbone. Experiments on Campus3D and PartNeXt demonstrate the effectiveness of Hy3D in improving segmentation performance across both scene-level semantic hierarchies and fine-grained part hierarchies.

1 INTRODUCTION

A point on a roof belongs simultaneously to a roof, a building, and a broader construction category. Hierarchical 3D understanding uses these relationships to predict semantic labels at several granularities, in both outdoor scenes and object parts (Li et al., 2020; Wang et al., 2025). The problem requires fine-grained discrimination together with a coherent account of each prediction's ancestry. Variable branch depths and incomplete annotations further complicate the alignment between semantic categories and network outputs.

Existing methods couple multilevel predictions through output-consistency losses, feature fusion, or taxonomy-constrained decoding (Li et al., 2020; 2025). These mechanisms address different aspects of the problem. A decoder can always return a valid path, even when the underlying features do not encode parent relationships. Conversely, representations can favor the correct parent without forcing independently trained heads to predict compatible labels. This distinction motivates our question: *how can a supplied semantic taxonomy directly supervise the geometry of point representations across prediction levels?*

Hyperbolic geometry offers a useful prior because its intrinsic volume grows exponentially with radius, as does the number of nodes in a branching tree (Nickel & Kiela, 2017). Hyperbolic neural networks make this geometry accessible to input-dependent representations (Ganea et al., 2018). In 3D learning, HyCoRe and HyCoRe-seg already regularize compositional relationships, with the latter combining hyperbolic heads and radial and contrastive objectives on pooled object and part features (Montanaro et al., 2022; 2023). Our contribution concerns a different supervision structure: point-aligned representations at the levels of a known semantic taxonomy, anchored by feature-derived parent-class prototypes.

Hy3D maps all prediction levels into one Poincaré ball with shared curvature, coordinates, and projection parameters. A detached prototype memory aggregates tangent-space features within each cloud and then across clouds containing the class. The parent-margin loss favors the correct parent prototype over active alternatives. The radial loss relates the representations of the same point across adjacent levels, with a margin proportional to the underlying semantic-depth gap (Figure 1). Repeated

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.

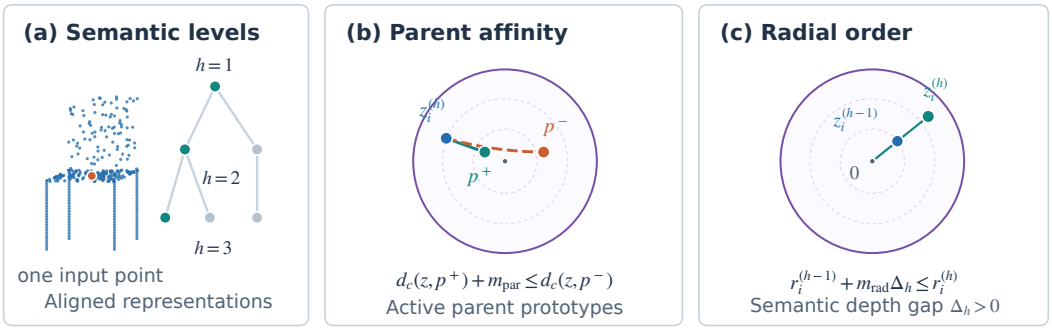

Figure 1: **Taxonomy-aligned supervision.** (a) The same point has representations at multiple semantic levels. (b) The parent-margin target compares a child representation with its correct and nearest competing parent prototypes. (c) The radial target scales outward progression by semantic depth. Diagrams are schematic; both inequalities are soft training targets. Valid output paths are supplied separately by decoding.

terminal labels add no artificial depth. The interface accommodates any number of prediction levels and any feature extractor satisfying the stated equivariance assumptions.

We make the scope of these objectives explicit. Their guarantees concern included geometric constraints under a frozen prototype bank, rather than convergence, test-set generalization, or arbitrary classification heads. Bounded embeddings also impose a finite budget on feasible radial margins. For valid predictions, we retain the probability-sum hierarchical ensemble decoder introduced by Li et al. (2020). This separates the effect of representation learning from consistency imposed at inference.

Our contributions are threefold:

- A taxonomy-aligned framework coupling point representations through an cloud-balanced parent-prototype memory and semantic-depth-dependent radial margins.
- Conditional certificates and feasibility results that identify what the geometric objectives enforce, including their limits under clipping, missing labels, and detached memory.
- Evaluation on scene and part hierarchies with two point-cloud encoders, component ablations, geometry comparisons, and taxonomy-based diagnostics. The strongest directly controlled component result is the improvement from adding the geometric losses to the same Möbius-head model.

2 Related Work

Hierarchical prediction and 3D understanding. Hierarchical classification distinguishes predictive accuracy from consistency with a supplied label structure (Silla Jr. & Freitas, 2011). Campus3D provides multilevel scene annotations, a multitask predictor, an output-consistency loss, and hierarchical ensemble decoding (Li et al., 2020). Deep hierarchical learning further combines probabilistic coherence with hierarchical embedding fusion (Li et al., 2025). PartNeXt supplies fine-grained, variable-depth part annotations (Wang et al., 2025). PointNet++ and PointNeXt provide spatially hierarchical feature extractors (Qi et al., 2017; Qian et al., 2022); their spatial abstraction levels need not coincide with semantic prediction levels. We build on this multilevel prediction setting by defining direct metric constraints between point-aligned semantic representations.

Hyperbolic representations and regularization. Poincaré embeddings learn symbolic hierarchies (Nickel & Kiela, 2017), and hyperbolic neural networks provide differentiable operations on the ball (Ganea et al., 2018). HIER learns hierarchical proxies without explicit hierarchy annotations (Kim et al., 2023); our prototypes instead summarize observed features under known parent maps. HyCoRe studies point-cloud compositionality (Montanaro et al., 2022). Its segmentation extension, HyCoRe-seg, is particularly close: it uses Möbius heads and radial and contrastive penalties on pooled whole-object and feature-neighborhood representations (Montanaro et al., 2023). Hy3D

supervises pointwise semantic levels through parent-class memory and recorded depth gaps. Thus, the distinction lies in the supervised relationships and memory process, rather than the use of negative curvature or hinge losses alone. HiR^2 combines taxonomy-aware hyperbolic entailment and angular separation for multimodal models (He et al., 2026), in a different prediction setting. Our empirical comparisons assess the stated 3D configurations; they do not establish superiority to all these related methods.

3 HY3D

Let $\mathcal{T} = (\mathcal{V}, \mathcal{E}, o)$ be a finite rooted semantic tree, and let $D : \mathcal{V} \to \mathbb{R}_{\geq 0}$ denote its semantic depth function, with $D(o) = 0$. For each node v, $D(v)$ is the sum of positive edge lengths along the unique path from o to v; semantic edges have unit length by default. For $H \geq 1$ prediction levels, let $\mathcal{C}_h$ be a label set, $\pi_h : \mathcal{C}_h \to \mathcal{C}_{h-1}$ a parent map, and $\iota_h : \mathcal{C}_h \to \mathcal{V}$ a map assigning each label occurrence to its semantic node. Set $\mathcal{C}_0 = \{o\}$ and $\iota_0(o) = o$. For every $h \in \{1, \dots, H\}$ and $k \in \mathcal{C}_h$, require

$$\iota_{h-1}(\pi_h(k)) \preceq \iota_h(k), \qquad \Delta_h(k) = D(\iota_h(k)) - D(\iota_{h-1}(\pi_h(k))) \geq 0, \tag{1}$$

where $\preceq$ denotes ancestry, including equality. When multiple semantic edges are compressed into one prediction transition, $\Delta_h(k)$ equals their total length; repeating a terminal node gives $\Delta_h(k) = 0$. Thus, h indexes prediction outputs, not necessarily semantic depth. Alignment must preserve a unique parent for each label occurrence.

Cloud s contains N_s points $\mathcal{P}_s = ((\boldsymbol{x}_{si}, \boldsymbol{a}_{si}))_i$, with coordinates in $\mathbb{R}^3$ and optional attributes in $\mathbb{R}^C$. A point has a valid target path $\boldsymbol{y}_{si} \in \Omega_s$, where

$$\Omega_s = \{(k_1, \dots, k_H) : k_h \in \mathcal{C}_{s,h},\ \pi_h(k_h) = k_{h-1}\ (h \geq 2)\} \neq \varnothing. \tag{2}$$

The default is $\mathcal{C}_{s,h} = \mathcal{C}_h$. Any restricted sets must use context available independently of unknown point labels at inference, and retain only labels that lie on at least one complete admissible path; ignored IDs are excluded. Observed labels must agree with the taxonomy. Partial labels are supported by the formulation, but are not an experimental claim of this paper.

3.1 A SHARED BOUNDED REPRESENTATION

A permutation-equivariant backbone with level-specific, point-aligned readouts and adapters produces $F_s^{(h)} \in \mathbb{R}^{N_s \times d_f}$. A shared projector ϕ_ψ maps these features to dimension $d \geq 2$ and satisfies $\phi_\psi(PF) = P\phi_\psi(F)$ for a point permutation P. All levels use the same Poincaré ball $\mathbb{B}_c^d = \{\boldsymbol{z} : c\|\boldsymbol{z}\|^2 < 1\}$, with fixed curvature $-c < 0$ and metric $g_{\boldsymbol{z}} = 4(1 - c\|\boldsymbol{z}\|^2)^{-2} I$. Standard maps and distances are detailed in Appendix A (Ganea et al., 2018). The representation is

$$\boldsymbol{u}_{si}^{(h)} = \mathrm{clip}_R\Big(\tau[\phi_\psi(F_s^{(h)})]_i\Big), \qquad \boldsymbol{z}_{si}^{(h)} = \Pi_{c,\varepsilon}\Big(\exp_0^c(\boldsymbol{u}_{si}^{(h)})\Big), \tag{3}$$

where $\tau > 0$, clip_R caps the tangent norm at R, and $\Pi_{c,\varepsilon}$ projects radially to coordinate norm at most $(1-\varepsilon)/\sqrt{c}$, with $0 < \varepsilon < 1$. Write d_c for geodesic distance and $r_c(\boldsymbol{z}) = d_c(\boldsymbol{0}, \boldsymbol{z})$. Under this metric convention,

$$r_c(\exp_0^c(\boldsymbol{u})) = 2\|\boldsymbol{u}\|, \qquad r_c(\boldsymbol{z}_{si}^{(h)}) \leq R_* := 2\overline{R}, \quad \overline{R} = \min\{R, \operatorname{artanh}(1-\varepsilon)/\sqrt{c}\}. \tag{4}$$

The implemented region is therefore bounded despite the unbounded intrinsic extent of the full ball.

3.2 OBJECT-BALANCED PROTOTYPE MEMORY

Each level–class occurrence (h, k) has a tangent state $\boldsymbol{a}_{h,k}^t$ and an active flag $v_{h,k}^t$, initially zero. Active prototypes are $\boldsymbol{p}_{h,k}^t = \exp_0^c(\boldsymbol{a}_{h,k}^t)$. For observed class-$k$ points $I_{s,h,k}$ in cloud s, compute detached tangent means, then average across the set $B_{h,k}$ of clouds containing that class:

$$\widehat{\boldsymbol{a}}_{h,k}^t = \frac{1}{|B_{h,k}|} \sum_{s \in B_{h,k}} \frac{1}{|I_{s,h,k}|} \sum_{i \in I_{s,h,k}} \mathrm{sg}\Big(\log_0^c(\boldsymbol{z}_{si}^{(h)})\Big). \tag{5}$$

Here sg denotes the stop-gradient operator. The loss uses the frozen bank at step t. After updating network parameters, initialize an unseen class with its cached mean, or update an active state as:

$$\boldsymbol{a}_{h,k}^{t+1} = \beta_t \boldsymbol{a}_{h,k}^t + (1 - \beta_t)\widehat{\boldsymbol{a}}_{h,k}^t, \qquad 0 \leq \beta_t < 1. \tag{6}$$

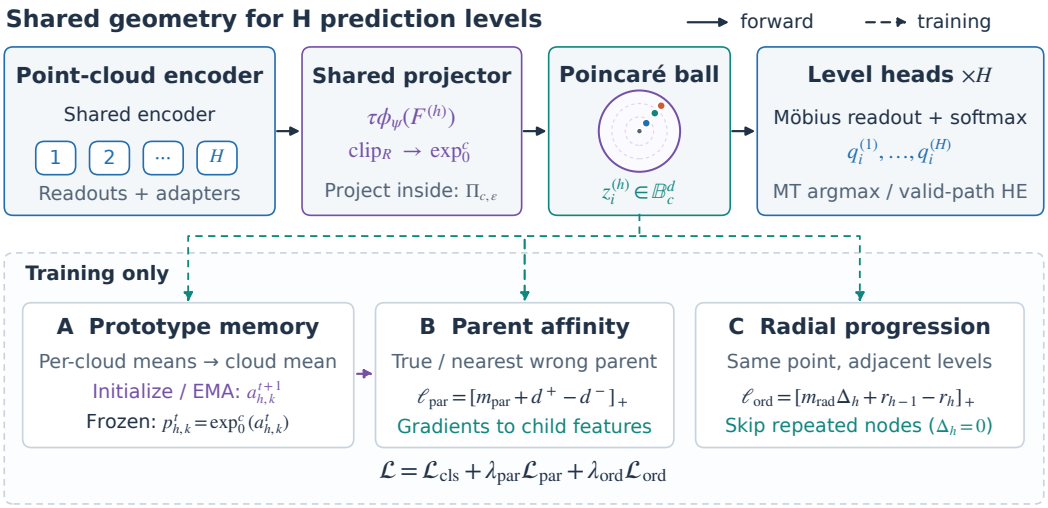

Figure 2: **Hy3D architecture.** Shared projection embeds point-aligned features for H semantic levels in one metric space. Pointwise heads support independent predictions or valid-path HE decoding. During training, detached cloud-balanced means update an initialized prototype memory. The frozen memory supplies parent candidates; adjacent point representations supply radial constraints. Dashed connections are used only during training.

Absent classes retain their state; newly observed classes become active for the next iteration. Means use the current forward pass, before the parameter update. The estimator weights clouds equally conditional on containing a class. It is an origin-tangent mean, not a Fréchet mean. Separate level occurrences retain separate states, including repeated semantic nodes.

3.3 Taxonomy-aligned objectives

Parent affinity. For $j = (s, i, h)$ with $h \geq 2$ and observed child label, let $k_j^+ = \pi_h(y_{si}^{(h)})$. Include j only if the correct parent is active and the set $\mathcal{N}_j^t$ of other active admissible parent labels is nonempty. Define

$$d_j^+ = d_c(\boldsymbol{z}_{si}^{(h)}, \boldsymbol{p}_{h-1,k_j^+}^t), \qquad d_j^- = \min_{k \in \mathcal{N}_j^t} d_c(\boldsymbol{z}_{si}^{(h)}, \boldsymbol{p}_{h-1,k}^t), \tag{7}$$

$$\ell_{\mathrm{par},j} = [m_{\mathrm{par}} + d_j^+ - d_j^-]_+, \qquad [x]_+ = \max\{0, x\}, \quad m_{\mathrm{par}} > 0. \tag{8}$$

The nearest incorrect parent is the maximally violating candidate. Prototypes are detached; uninitialized states never act as candidates. These constraints separate parent branches, without directly separating siblings that share a parent.

Radial progression. For adjacent observed, compatible labels with $\Delta_h(y_{si}^{(h)}) > 0$, use

$$\ell_{\mathrm{ord},j} = [m_{\mathrm{rad}}\Delta_h(y_{si}^{(h)}) + r_c(\boldsymbol{z}_{si}^{(h-1)}) - r_c(\boldsymbol{z}_{si}^{(h)})]_+, \qquad m_{\mathrm{rad}} > 0. \tag{9}$$

Repeated-node edges are skipped. Their radii need not coincide, so a zero loss does not imply monotonicity across skipped edges.

Training objective. Let A_h and R_h be the sampled eligible parent and radial tuples. For declared positive normalizers,

$$\mathcal{L}_{\mathrm{par}} = \sum_{h=2}^{H} \frac{\sum_{j \in A_h} \ell_{\mathrm{par},j}}{Z_h^{\mathrm{par}}}, \qquad \mathcal{L}_{\mathrm{ord}} = \frac{\sum_{h=2}^{H} \sum_{j \in R_h} \ell_{\mathrm{ord},j}}{Z^{\mathrm{ord}}},$$

$$\mathcal{L}_t = \mathcal{L}_{\mathrm{cls}} + \lambda_{\mathrm{par}}(t)\mathcal{L}_{\mathrm{par}} + \lambda_{\mathrm{ord}}(t)\mathcal{L}_{\mathrm{ord}}. \tag{10}$$

Empty terms contribute zero before division. $\mathcal{L}_{\mathrm{cls}}$ sums class-weighted cross-entropies over observed labels (Appendix B). Per-level parent means set $Z_h^{\mathrm{par}} = |A_h|$; a radial sum sets $Z^{\mathrm{ord}} = 1$. A radial mean changes the effective loss weight. Sampling and reductions are part of the model specification. Because the memory evolves outside backpropagation, training is a coupled parameter–state process.

3.4 Prediction and valid-path decoding

At level h, define $K_h = |\mathcal{C}_h|$ and $W_h \in \mathbb{R}^{K_h \times d}$, with output coordinates indexed by a fixed ordering of $\mathcal{C}_h$. Each level has a continuous pointwise head followed by a masked softmax over the nonempty admissible set $\mathcal{C}_{s,h} \subseteq \mathcal{C}_h$. For point i in cloud s, the reported Möbius-coordinate head defines

$$\boldsymbol{s}_{si}^{(h)} = \alpha_h \exp_0^{c,K_h}\Big(W_h \log_0^{c,d}(\boldsymbol{z}_{si}^{(h)})\Big),$$
$$q_{si}^{(h)}(k) = \begin{cases} \dfrac{\exp(s_{si,k}^{(h)})}{\sum_{\ell \in \mathcal{C}_{s,h}} \exp(s_{si,\ell}^{(h)})}, & k \in \mathcal{C}_{s,h}, \\ 0, & k \in \mathcal{C}_h \setminus \mathcal{C}_{s,h}. \end{cases} \tag{11}$$

The superscripts d and K_h indicate the dimensions of the logarithmic and exponential maps, respectively. When $\mathcal{C}_{s,h} = \mathcal{C}_h$, the masked softmax reduces to ordinary softmax over all K_h classes. For $\alpha_h > 0$, this bias-free head has the same argmax as $W_h \log_0^{c,d}(\boldsymbol{z}_{si}^{(h)})$ when both are restricted to $\mathcal{C}_{s,h}$, although their softmax probabilities and gradients generally differ. It is not a nearest-prototype classifier.

Independent multitask prediction (MT) takes the argmax at each level. Hierarchical ensemble decoding (HE) (Li et al., 2020) instead returns

$$\widehat{\boldsymbol{y}}_{si} = \arg\max_{\boldsymbol{k} \in \Omega_s} \sum_{h=1}^{H} q_{si}^{(h)}(k_h). \tag{12}$$

Tree dynamic programming computes the optimum in $O(\sum_h |\mathcal{C}_{s,h}|)$ time per point (Appendix B.1). Fixed label-order tie breaking preserves equivariance. HE guarantees a valid path, not a correct path; its 100% consistency is a decoder property. Both inference modes discard the prototype bank.

4 What the Geometric Objectives Guarantee

The following results assume the specified feature interface and update rules. Loss certificates use a frozen bank and fixed eligible sets; perturbations preserve candidate identities. Proofs appear in Appendix C. None of the statements assumes zero-loss optimization or establishes generalization.

Proposition 1 (Equivariance and boundedness)**.** *With permutation-equivariant features and projection, pointwise adapters and heads, and order-independent admissible sets and tie breaking, representations and predictions are point-permutation equivariant. Prototype aggregation and losses are invariant under consistent permutations of points, masks, labels, and sampled tuples. If active initial states have norm at most $\overline{R}$, the prescribed clipping and convex memory updates ensure*

$$\|\log_0^c(\boldsymbol{z}_{si}^{(h)})\| \le \overline{R}, \qquad r_c(\boldsymbol{p}_{h,k}^t) \le R_* \quad \textit{for every active prototype.} \tag{13}$$

Proposition 2 (Parent-margin certificate)**.** *For a nonempty eligible set S, let $\overline{\ell}_S = |S|^{-1} \sum_{j \in S} \ell_{\mathrm{par},j}$. A positive weighted sum of its hinges is zero if and only if $d_j^- - d_j^+ \ge m_{\mathrm{par}}$ for every included tuple. Counting ties as failures,*

$$\frac{1}{|S|} \sum_{j \in S} \mathbf{1}[d_j^- \le d_j^+] \le \min\{1, \overline{\ell}_S / m_{\mathrm{par}}\}. \tag{14}$$

Proposition 3 (Perturbation stability)**.** *Let $g = \min_k\{d_c(\boldsymbol{z}, \boldsymbol{p}_k^-) - d_c(\boldsymbol{z}, \boldsymbol{p}^+)\}$ over fixed incorrect-parent candidates. If the anchor, correct prototype, and each incorrect prototype move by at most e_z, e_+, e_- in geodesic distance, the new gap satisfies*

$$g' \ge g - 2e_z - e_+ - e_-. \tag{15}$$

Thus a gap $g \ge m_{\mathrm{par}}$ preserves the unique nearest parent when $2e_z + e_+ + e_- < m_{\mathrm{par}}$.

These two certificates concern auxiliary parent retrieval among active candidates. They impose no guarantee on a separate classification head or on perturbations of input coordinates without further assumptions.

Proposition 4 (Finite radial budget)**.** *On a contiguous eligible segment $h = a+1, \dots, b$, write $r_h = r_c(\boldsymbol{z}^{(h)}) \le R_*$ and $e_h = [m_{\mathrm{rad}}\Delta_h + r_{h-1} - r_h]_+$. Then*

$$\sum_{h=a+1}^{b} e_h \ge \left[m_{\mathrm{rad}} \sum_{h=a+1}^{b} \Delta_h + r_a - r_b \right]_+ \ge \left[m_{\mathrm{rad}} \sum_{h=a+1}^{b} \Delta_h + r_a - R_* \right]_+ . \tag{16}$$

For fixed r_a, zero radial loss is feasible in representation space if and only if $m_{\mathrm{rad}} \sum_{h=a+1}^{b} \Delta_h \le R_* - r_a$.

This is a radial-only existence result, not joint feasibility with parent constraints or realizability by a particular network. It gives a direct check on margin choices after taxonomy compression. Parent certificates and perturbation bounds follow from hinge losses and metric inequalities; they do not, by themselves, establish a special advantage of hyperbolic space.

5 Experiments

We evaluate segmentation across scene and part hierarchies, ablate prediction heads and geometric objectives, and analyze prototype–taxonomy alignment. Results are organized by these questions across datasets and encoders.

5.1 Experimental Setup

Datasets. **Campus3D** contains approximately 937 million points from six outdoor regions covering 1.58 km^2, with a five-level semantic hierarchy and 24 semantic classes (Li et al., 2020). We follow its region-level train/validation/test split. **PartNeXt** contains 23,519 textured objects, 350,187 part instances, and 50 object categories (Wang et al., 2025). We align its variable-depth part trees to five prediction levels with reported output widths $[63, 264, 638, 919, 1260]$, use a custom 8:1:1 train/validation/test split, and sample 2,048 colored points per object. Results on PartNeXt therefore refer to this hierarchical segmentation protocol.

Models and evaluation. The default encoder is PointNet++, with five prediction branches and a shared 128-dimensional projector. On PartNeXt, we also evaluate PointNeXt by replacing the encoder while retaining the five PointNet++-style decoders and training design. We report overall accuracy (OA), class-averaged intersection-over-union (mIoU), and their arithmetic means across the five levels. Raw path consistency is the percentage of independently predicted label sequences that form valid taxonomy paths. Independent multi-task decoding (MT) and hierarchical ensemble decoding (HE) use the same trained model; HE guarantees a valid path, so its consistency is 100% by construction.

Our Euclidean control retains the prototype memory, parent and radial objectives, architecture dimensions, and training schedule, using Euclidean distances, norms, and heads. The corresponding hyperbolic model uses Poincaré operations and Möbius heads. This comparison evaluates the complete geometry-dependent design, including head behavior and metric scale. Published Campus3D scores and a reproduced PartNeXt baseline with Euclidean heads and no PCT provide additional reference points. PartNeXt geometry comparisons and the M1–Hy3D ablation report three-seed means and sample standard deviations; comparable uncertainty is unavailable for the remaining rows. Detailed results appear in Appendix E.

5.2 Main Results Across Datasets and Encoders

Scene and part segmentation. Table 1 shows higher aggregate OA and mIoU for Hy3D than for the Euclidean control in all evaluated configurations. On Campus3D, Hy3D reaches 87.74 OA and 59.52 mIoU under MT, improving over the internal control by 7.92 and 9.47 percentage points. With HE, the corresponding scores reach 88.19 and 60.14. On PartNeXt with PointNet++, the three-seed

MT gains over the Euclidean control are 0.54 OA and 0.53 mIoU; with HE, they are 0.91 and 0.62 points. The published Campus3D scores provide benchmark context; the internal controls support component comparisons.

Transfer across encoders. With PointNeXt, Hy3D reaches 57.70 OA and 16.16 mIoU under MT, exceeding the corresponding Euclidean model by 0.58 and 0.45 points. The advantage remains positive under HE, although the mIoU gap narrows to 0.10 points relative to seed standard deviations of 0.66–0.69. These results support transfer to a second encoder, with configuration-dependent gains. Three-seed summaries alone do not establish statistical significance for every comparison.

Interaction with hierarchical decoding. The gains under MT establish that the improvement is already present before valid-path decoding. HE further improves the full model's aggregate OA and mIoU in every configuration in Table 1. Representation learning and hierarchical decoding therefore make distinct contributions: the former improves the reported segmentation scores, whereas the latter enforces valid label paths. Figure 3 illustrates Campus3D predictions; PartNeXt examples are provided in Appendix E.4.

Table 1: Main segmentation results. OA and mIoU are percentages averaged over five levels. PartNeXt geometry variants report three-seed mean ± sample standard deviation. Campus3D entries and the reproduced PartNeXt baseline are point estimates. The Euclidean control includes the analogous parent and radial objectives. † denotes published Campus3D reference values, not matched reruns.

	MT		MT+HE	
Model	OA	mIoU	OA	mIoU
Campus3D / PointNet++				
Published MT / MT+HE†	82.10	48.40	82.18	48.34
Euclidean control	79.82	50.05	79.97	50.31
Hy3D	**87.74**	**59.52**	**88.19**	**60.14**
PartNeXt / PointNet++				
Reproduced MT baseline	52.07	12.33	54.10	13.27
Euclidean control	55.48 ± 0.31	13.79 ± 0.60	56.13 ± 0.34	14.29 ± 0.89
Hy3D	$\mathbf{56.02} \pm 0.07$	$\mathbf{14.32} \pm 0.25$	$\mathbf{57.05} \pm 0.12$	$\mathbf{14.91} \pm 0.19$
PartNeXt / PointNeXt				
Euclidean control	57.11 ± 0.35	15.71 ± 0.42	57.85 ± 0.44	16.48 ± 0.66
Hy3D	$\mathbf{57.70} \pm 0.31$	$\mathbf{16.16} \pm 0.60$	$\mathbf{58.32} \pm 0.27$	$\mathbf{16.58} \pm 0.69$

5.3 Ablation Study

Table 2 groups the ablations across datasets to separate the roles of prediction heads and geometric supervision. M0 maps features into the Poincaré ball but applies coordinate Conv1d heads. M1 uses Möbius heads without PCT. Here, PCT denotes the parent-prototype triplet and depth-aware radial objectives; the full model uses both. The PartNeXt Radius-only and Triplet-only variants retain the same backbone and Möbius heads.

Prediction heads. On PartNeXt, M0 improves mean OA only slightly over the reproduced baseline while reducing mean mIoU from 12.33 to 9.59. M1 reaches 13.31 mIoU, showing that the way hyperbolic features are classified matters in this configuration. The ordering reverses on Campus3D, where M0 outperforms M1 in aggregate OA and mIoU. These results do not support a universal advantage from changing the head alone; the complete model achieves the best internal mean segmentation scores on both datasets.

Geometric objectives. The M1–Hy3D comparison isolates adding PCT to the same hyperbolic predictor. On PartNeXt, MT mean OA/mIoU improve by 0.58/1.01 points; the improvements persist under HE. On Campus3D, the corresponding MT gains are 1.65/2.30 points. For PartNeXt, combining Triplet and Radius also gives higher reported aggregate scores than either single loss: MT mIoU exceeds Radius-only by 0.55 points and Triplet-only by 0.36 points. These estimates are consistent

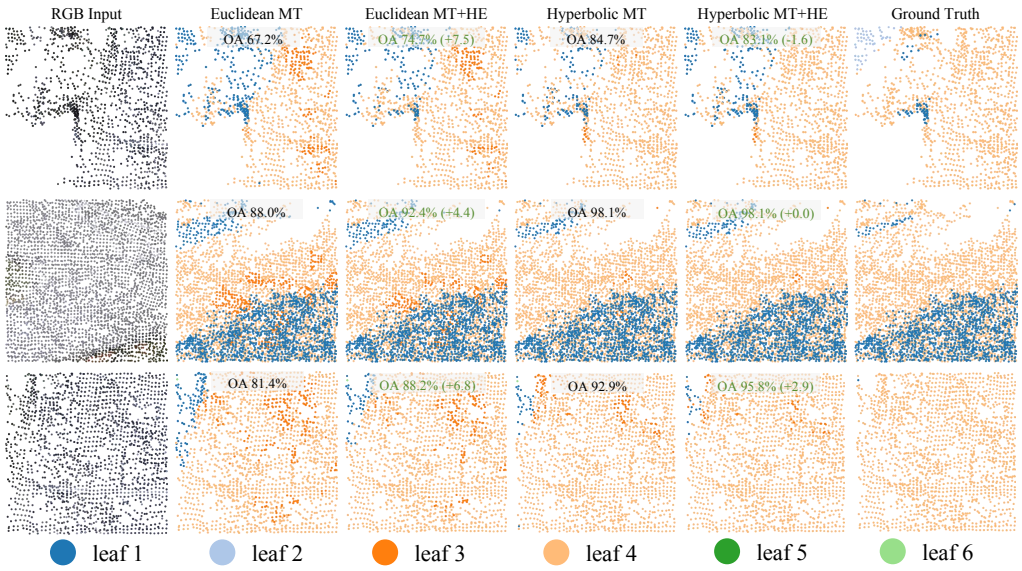

Figure 3: Campus3D leaf-level predictions. Columns show RGB input, Euclidean MT and MT+HE, hyperbolic MT and MT+HE, and ground truth. Each row is one test block; colors denote semantic labels. OA values describe the displayed blocks, and parenthesized changes compare HE with the corresponding MT prediction. These examples illustrate local predictions rather than test-set averages.

with complementary effects of parent affinity and radial ordering, although the single-loss variants lack comparable seed-level uncertainty.

Table 2: Component ablations with PointNet++. OA and mIoU are means across five levels; Cons. is raw MT path consistency. All values are percentages. PartNeXt M1 and Hy3D report three-seed mean ± sample standard deviation. Other rows are reported point estimates; uncertainty for the single-loss variants is unavailable. A dash indicates an unreported value.

	MT			MT+HE	
Variant	OA	mIoU	Cons.	OA	mIoU
Campus3D					
M0	87.51	58.30	89.45	88.04	59.44
M1	86.09	57.22	89.03	86.51	58.05
Hy3D	**87.74**	**59.52**	**90.41**	**88.19**	**60.14**
PartNeXt					
M0	52.50	9.59	54.76	53.23	9.84
M1	55.44 ± 0.38	13.31 ± 0.29	$\mathbf{58.62} \pm 1.28$	56.17 ± 0.47	13.94 ± 0.40
M1 + Radius	55.84	13.77	–	56.95	14.50
M1 + Triplet	55.45	13.96	–	56.21	14.42
Hy3D	$\mathbf{56.02} \pm 0.07$	$\mathbf{14.32} \pm 0.25$	58.57 ± 0.98	$\mathbf{57.05} \pm 0.12$	$\mathbf{14.91} \pm 0.19$

Accuracy and path consistency. Adding PCT increases Campus3D raw consistency from 89.03% to 90.41%, but leaves PartNeXt consistency essentially unchanged at 58.62% versus 58.57%. The controlled PartNeXt ablation therefore supports an accuracy benefit without an accompanying consistency gain. This distinction agrees with the method: feature-level geometric constraints and validity of independently predicted label paths are different properties.

5.4 Prototype–Taxonomy Alignment

We assess whether the learned prototype geometry reflects the semantic taxonomy by comparing prototype distances with semantic-tree shortest-path distances on identical valid node pairs. Table 3

reports four pair subsets and one level-wise macro average on PartNeXt with PointNet++. Definitions, additional dataset and encoder results, and bootstrap analyses appear in Appendix F, particularly Tables 9 and 10.

Table 3: Prototype–taxonomy Spearman correlation on PartNeXt with PointNet++. The table reports four pair subsets and the arithmetic mean of the L2–L5 correlations. Higher values indicate stronger rank agreement with taxonomy distances; $\Delta\rho = \rho_H - \rho_E$.

Pair subset / summary	Euclidean (ρ_E)	Hyperbolic (ρ_H)	$\Delta\rho$
All level instances	0.1427	**0.1968**	+0.0540
Unique taxonomy nodes	0.2298	**0.2622**	+0.0324
Same-level pairs	0.1525	**0.1944**	+0.0419
Leaf-only (L5)	0.0793	**0.1639**	+0.0846
L2–L5 macro average	0.0788	**0.1068**	+0.0279

Hyperbolic prototypes achieve higher correlation in all five reported summaries. The largest gain occurs among leaf categories, where correlation increases from 0.0793 to 0.1639 ($\Delta\rho = +0.0846$). Improvements also hold for same-level pairs and after collapsing repeated occurrences into unique taxonomy nodes. The L2–L5 macro average increases from 0.0788 to 0.1068. Together, these results provide complementary evidence of improved taxonomy alignment under the hyperbolic design, with the strongest observed gain at the finest semantic level.

The positive unique-node trend also appears with PointNeXt, where correlation increases from 0.1918 to 0.3051. Its paired branch-bootstrap 95% interval for $\Delta\rho$ is $[+0.0201, +0.1615]$, whereas the PointNet++ interval includes zero (Appendix F). These aggregate gains coexist with decreases for some individual relations, including PartNeXt L2 and Campus3D same-level pairs. The evidence therefore supports relation-dependent structural improvements; bootstrap intervals condition on fitted prototypes and do not establish training-seed stability or directly evaluate the local parent-retrieval and radial constraints.

5.5 Discussion

How do the geometric objectives constrain feature representations? In Hy3D, triplet gradients encourage eligible child features to favor the correct parent over the nearest active alternative by a margin, with prototypes detached. The radius loss encourages depth-scaled radial progression between the same point's representations at adjacent levels. Both directly supervise feature geometry; logit consistency constrains output distributions and influences features through backpropagation.

Why does the parent triplet help? Triplet supervision promotes parent-branch separation in shared features, potentially aiding Möbius classification without guaranteeing logit separation. Ablations show that combining Triplet and Radius outperforms either objective alone.

Why hyperbolic rather than Euclidean geometry? The parent triplet and radius losses presuppose that distances and radii reflect the tree structure. In Euclidean space, embedding an L-level tree with exponentially many leaves forces either dimension or distortion to grow; in the Poincaré ball the exponentially growing metric volume near the boundary accommodates the 1,260-class leaf level of PartNeXt in only 128 dimensions (Nickel & Kiela, 2017; Ganea et al., 2018).

6 Limitations and Conclusion

Hy3D learns taxonomy-aligned point representations through a shared hyperbolic space, object-balanced prototypes, and parent and radial supervision. Experiments on Campus3D and PartNeXt show improvements in aggregate segmentation accuracy across two encoders, with relation-dependent structural gains. Evaluation covers only five-level hierarchies and a custom PartNeXt split; general-depth and partial-label settings remain unevaluated, and fine-level mIoU remains low. Isolating curvature and testing additional taxonomies require further controlled evaluation. The framework separates representation supervision, conditional geometric guarantees, and valid-path decoding.

AI USE STATEMENT

Generative AI tools were used to improve the readability of the manuscript, including language refinement and stylistic polishing. They were also used to inspect and make limited adjustments to the implementation code. All AI-assisted text and code changes, as well as their effects on the experimental pipeline, were carefully reviewed, tested, and verified by the authors. The research methodology, experiments, findings, and scientific conclusions were developed and validated by the authors. The authors take full responsibility for every part of the manuscript.

REPRODUCIBILITY STATEMENT

Section 3 and Appendices A–B provide a complete specification of the mathematical operations, eligibility rules, memory mechanism, and decoding procedure. Appendix C presents the corresponding proofs, while Appendix D documents the training and implementation settings. The source code will be released upon acceptance.

REFERENCES

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, 2018. URL https://arxiv.org/abs/1805.09112.

Hulingxiao He, Zhi Tan, and Yuxin Peng. Learning taxonomic trees with hierarchical representation regularization for large multimodal models. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026. URL https://github.com/PKU-ICST-MIPL/HiR2_ICML2026.

Sungyeon Kim, Boseung Jeong, and Suha Kwak. Hier: Metric learning beyond class labels via hierarchical regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19903–19912, 2023. doi: 10.1109/CVPR52729.2023.01906. URL https://doi.org/10.1109/CVPR52729.2023.01906.

Chongshou Li, Yuheng Liu, Xinke Li, Yuning Zhang, Tianrui Li, and Junsong Yuan. Deep hierarchical learning for 3d semantic segmentation. *International Journal of Computer Vision*, 133: 4420–4441, 2025. doi: 10.1007/s11263-025-02387-6. URL https://doi.org/10.1007/s11263-025-02387-6.

Xinke Li, Chongshou Li, Zekun Tong, Andrew Lim, Junsong Yuan, Yuwei Wu, Jing Tang, and Raymond Huang. Campus3d: A photogrammetry point cloud benchmark for hierarchical understanding of outdoor scene. In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 238–246, 2020. doi: 10.1145/3394171.3413661. URL https://doi.org/10.1145/3394171.3413661.

Antonio Montanaro, Diego Valsesia, and Enrico Magli. Rethinking the compositionality of point clouds through regularization in the hyperbolic space. In *Advances in Neural Information Processing Systems*, volume 35, pp. 33741–33753, 2022. doi: 10.52202/068431-2445. URL https://doi.org/10.52202/068431-2445.

Antonio Montanaro, Diego Valsesia, and Enrico Magli. Towards hyperbolic regularizers for point cloud part segmentation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, 2017. URL https://arxiv.org/abs/1705.08039.

Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017. URL https://arxiv.org/abs/1706.02413.

Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Abed Al Kader Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. PointNeXt: Revisiting PointNet++ with improved training and scaling strategies. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://papers.nips.cc/paper_files/paper/2022/hash/9318763d049edf9a1f2779b2a59911d3-Abstract-Conference.html.

Carlos N. Silla Jr. and Alex A. Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22:31–72, 2011. doi: 10.1007/s10618-010-0175-9. URL https://doi.org/10.1007/s10618-010-0175-9.

Penghao Wang, Yiyang He, Xin Lv, Yukai Zhou, Lan Xu, Jingyi Yu, and Jiayuan Gu. Partnext: A next-generation dataset for fine-grained and hierarchical 3d part understanding. *Advances in Neural Information Processing Systems*, 38, 2025.

APPENDIX

A GEOMETRIC AND TAXONOMY DEFINITIONS

Poincaré operations. For $c > 0$, the metric is $g_{\boldsymbol{z}} = \lambda_{\boldsymbol{z}}^2 I$ with $\lambda_{\boldsymbol{z}} = 2/(1 - c\|\boldsymbol{z}\|^2)$. With continuous extensions at zero, the origin maps are

$$\exp_0^c(\boldsymbol{u}) = \frac{\tanh(\sqrt{c}\|\boldsymbol{u}\|)}{\sqrt{c}\|\boldsymbol{u}\|}\boldsymbol{u}, \qquad \log_0^c(\boldsymbol{z}) = \frac{\operatorname{artanh}(\sqrt{c}\|\boldsymbol{z}\|)}{\sqrt{c}\|\boldsymbol{z}\|}\boldsymbol{z}. \tag{17}$$

The distance and intrinsic radius are

$$d_c(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\sqrt{c}} \operatorname{arcosh}\left(1 + \frac{2c\|\boldsymbol{x} - \boldsymbol{y}\|^2}{(1 - c\|\boldsymbol{x}\|^2)(1 - c\|\boldsymbol{y}\|^2)}\right), \tag{18}$$

$$r_c(\boldsymbol{z}) = \frac{2}{\sqrt{c}} \operatorname{artanh}(\sqrt{c}\|\boldsymbol{z}\|). \tag{19}$$

These are standard Poincaré formulas (Nickel & Kiela, 2017; Ganea et al., 2018). The coordinate norm $\|\boldsymbol{z}\|$ is not the intrinsic radius. At fixed coordinates, $d_c(\boldsymbol{x}, \boldsymbol{y}) \to 2\|\boldsymbol{x} - \boldsymbol{y}\|$ as $c \to 0$, because the limiting metric is $4I$. For nonzero arguments, define

$$\operatorname{clip}_R(\boldsymbol{u}) = \boldsymbol{u} \min\{1, R/\|\boldsymbol{u}\|\}, \qquad \Pi_{c,\varepsilon}(\boldsymbol{z}) = \boldsymbol{z} \min\{1, (1 - \varepsilon)/(\sqrt{c}\|\boldsymbol{z}\|)\}; \tag{20}$$

both maps take zero to zero. Scaling precedes clipping in the specified model. Numerical evaluation should protect the inverse hyperbolic functions from floating-point excursions outside their domains. Such numerical safeguards do not change the exact-arithmetic propositions.

Semantic depth and alignment. Set $D(o) = 0$ and use unit semantic edge lengths by default. Positive edge lengths may instead define real-valued semantic depths. The maps π_h are functions on level-specific class occurrences. Labels with identical names but different ancestors need different lineage identifiers. A forest can be attached to an unpredicted dummy root; multiple-parent directed acyclic graphs are outside this formulation. Compression must preserve a unique parent at every output level, splitting occurrences if necessary. A repeated terminal occurrence has the same semantic image as its parent and therefore zero depth gap.

The alignment, ignored-label mask, admissible label sets, and terminal-repetition rule are part of a dataset's protocol. An observed fine label determines a parent through π_h, which is used by the parent-margin loss. The radial loss requires both adjacent labels to be observed, unless an explicitly documented ancestor-completion rule supplies them. An annotated per-object part-presence tree cannot be used to restrict test predictions unless it is supplied by the evaluation task. The default inference space is the global aligned taxonomy. For $H = 1$, all cross-level losses are empty and the system reduces to pointwise classification.

Feature-interface assumptions. The backbone and readouts must return features aligned to the original points. Shared pointwise adapters can reconcile different readout widths. A projector can include normalization across the point axis provided the normalization commutes with point permutations. Randomized sampling only gives samplewise equivariance under appropriately coupled randomness, or equivariance in distribution. Independently sampled runs need not agree. No rotation-invariance assumption is made.

B TRAINING AND DECODING DETAILS

For $O_h = \{(s, i) : M_{sih} = 1\}$, class weights $w_{h,k} > 0$, and level weights $\omega_h > 0$, define

$$\mathcal{L}_{\text{cls}} = \sum_{h=1}^{H} \omega_h \mathcal{L}_{\text{CE}}^{(h)}, \qquad \mathcal{L}_{\text{CE}}^{(h)} = -\frac{1}{Z_h^{\text{CE}}} \sum_{(s,i) \in O_h} w_{h, y_{si}^{(h)}} \log q_{si}^{(h)}(y_{si}^{(h)}). \tag{21}$$

Empty levels contribute zero. One weighted-mean convention is $Z_h^{\text{CE}} = \sum_{(s,i) \in O_h} w_{h, y_{si}^{(h)}}$; a point-count denominator is a different reduction. Likewise, $Z^{\text{ord}} = 1$ gives a radial sum, whereas $Z^{\text{ord}} = \sum_h |R_h|$ gives a mean. The supplied formulation records both possibilities; a reported checkpoint must be tied to one actual choice. Leaf-class-balanced subsampling changes the empirical objective and is not equivalent to uniform point sampling without importance correction.

Complete memory update. For nonempty class supports, use the detached mean in equation 5. The state transition is

$$\boldsymbol{a}_{h,k}^{t+1} = \begin{cases} \boldsymbol{a}_{h,k}^{t}, & B_{h,k} = \varnothing, \\ \widehat{\boldsymbol{a}}_{h,k}^{t}, & B_{h,k} \neq \varnothing,\ v_{h,k}^{t} = 0, \\ \beta_t \boldsymbol{a}_{h,k}^{t} + (1-\beta_t)\widehat{\boldsymbol{a}}_{h,k}^{t}, & \text{otherwise.} \end{cases} \tag{22}$$

Set $v_{h,k}^{t+1} = 1$ once observed; absent classes retain the old flag. Preinitialized active states, if used, must have norm at most $\overline{R}$. Distributed implementations must aggregate per-cloud class means and the counts of contributing clouds, rather than averaging worker means without their counts. The prescribed iteration is:

1. Sample clouds and supervised tuples; compute all point representations and logits.
2. Form eligible sets with the frozen active bank. Cache detached means of the current features.
3. Evaluate the declared classification and geometric losses; update network parameters.
4. Initialize or update memory states from the cached means. Newly activated classes participate at the next iteration.

Warm-up may suppress geometric loss weights while populating memory, but does not replace explicit activation flags.

B.1 Exact probability-sum decoding

After pruning labels that occur on no complete admissible path, define

$$F_H(k) = q^{(H)}(k), \qquad F_h(k) = q^{(h)}(k) + \max_{u \in \mathcal{C}_{s,h+1}:\pi_{h+1}(u)=k} F_{h+1}(u). \tag{23}$$

Select a maximizing level-one label and follow stored maximizing children. Use a fixed order on labels to resolve ties. Replacing q by $\log q$ maximizes a product of level probabilities and is a different rule from the probability-sum HE of Li et al. (2020).

Proposition 5 (Exact valid-path decoding)**.** *For finite nonempty Ω_s, the recurrence in equation 23 solves the maximization problem in equation 12 and returns an admissible path in $O(\sum_h |\mathcal{C}_{s,h}|)$ time and storage per point, apart from score computation.*

Proof. At the final level, F_H is the score of the only suffix at each retained label. If F_{h+1} is the optimal suffix value at each child, adding $q^{(h)}(k)$ to the best child value gives exactly the optimal suffix at k. Backward induction and maximization at level one therefore produce the global maximum. Each retained child has one parent, so the algorithm visits each label and interlevel edge once. Fixed tie breaking does not affect optimality. Validity follows from selecting only permitted parent–child edges; it does not imply semantic correctness. □

B.2 Structural diagnostics

For $n = \sum_s N_s$ evaluation points, full-path consistency of independent predictions is

$$\mathrm{Cons} = \frac{1}{n} \sum_{s,i} \mathbf{1}[(\widetilde{y}_{si}^{(1)}, \ldots, \widetilde{y}_{si}^{(H)}) \in \Omega_s]. \tag{24}$$

Consistency is multiplied by 100 as percentage value when reported in tables. OA counts correct point predictions at a level. Class IoU is $\mathrm{TP}/(\mathrm{TP}+\mathrm{FP}+\mathrm{FN})$, and mIoU averages over the evaluated classes; ignored classes and zero-union handling must be specified by the implementation. Means in the tables average the five level scores. Full-path consistency corresponds to the full-consistency threshold in Campus3D (Li et al., 2020).

Two useful complementary diagnostics, not supplied as completed experimental results, are strict parent-retrieval accuracy (PRA) and radial-violation rate (RVR):

$$\mathrm{PRA} = \frac{1}{|S_{\mathrm{par}}|} \sum_{j \in S_{\mathrm{par}}} \mathbf{1}[d_j^{+} < d_j^{-}], \qquad \mathrm{RVR} = \frac{1}{|S_{\mathrm{ord}}|} \sum_{j \in S_{\mathrm{ord}}} \mathbf{1}[\ell_{\mathrm{ord},j} > 0]. \tag{25}$$

C PROOFS OF THE CONDITIONAL GUARANTEES

Proof of Proposition 1. Let Q_s be a permutation matrix acting on the points of cloud s. By assumption, the feature interface and shared projector commute with Q_s. Shared pointwise adapters, clipping, exponential mapping, interior projection, and heads also commute with point permutations. The admissible label sets do not change, so softmax probabilities are permuted in the same way. Independent prediction and valid-path decoding operate separately on each point; fixed label-order tie-breaking preserves this correspondence even when scores tie.

Consistently permuting labels, masks, and sampled tuples only reindexes the supports used for prototype means and scalar losses. Their sums, counts, and stated reductions are unchanged. Hence prototype aggregation and loss values are invariant. For stochastic feature extraction or sampling, the samplewise assertion uses randomness coupled under the permutation; distributional equivariance gives the corresponding distributional assertion.

Under the metric convention used in the Method, for $\boldsymbol{u} \neq \mathbf{0}$,

$$\begin{aligned} r_c(\exp_0^c(\boldsymbol{u})) &= \frac{2}{\sqrt{c}} \operatorname{artanh}\big(\tanh(\sqrt{c}\|\boldsymbol{u}\|_2)\big) \\ &= 2\|\boldsymbol{u}\|_2. \end{aligned}$$

The identity also holds at zero by continuity. Tangent clipping ensures $\|\boldsymbol{u}\|_2 \le R$, while the radial interior projection gives

$$\|\log_0^c(\Pi_{c,\varepsilon}(\exp_0^c(\boldsymbol{u})))\|_2 = \min\left\{\|\boldsymbol{u}\|_2, \frac{\operatorname{artanh}(1-\varepsilon)}{\sqrt{c}}\right\} \le \overline{R}.$$

Thus every tangent feature entering the prototype estimator has norm at most $\overline{R}$. Per-cloud class means and means across clouds are convex combinations of these vectors and remain in the closed Euclidean ball of radius $\overline{R}$.

Any supplied active initial state satisfies the same bound by assumption. First-observation initialization uses a bounded mean, an absent class retains its state, and each EMA update is a convex combination of the previous state and a bounded mean. Induction on t therefore yields $\|\boldsymbol{a}_{h,k}^t\|_2 \le \overline{R}$ for every active state. Since $\boldsymbol{p}_{h,k}^t = \exp_0^c(\boldsymbol{a}_{h,k}^t)$, the radius identity gives $r_c(\boldsymbol{p}_{h,k}^t) \le 2\overline{R} = R_*$; it gives the same bound for every point representation.

The intrinsic ball of radius R_* is a Euclidean closed ball with coordinate radius $\tanh(\sqrt{c}R_*/2)/\sqrt{c} < 1/\sqrt{c}$, and is therefore compact within $\mathbb{B}_c^d$. Distances between its points are finite in exact arithmetic; stable numerical evaluation remains an implementation requirement. □

Proof of Proposition 2. Each hinge is nonnegative and satisfies

$$\ell_{\mathrm{par},j} = 0 \quad \Longleftrightarrow \quad d_j^- - d_j^+ \ge m_{\mathrm{par}}.$$

Since every w_j is strictly positive, their weighted sum is zero exactly when every included hinge is zero. As $m_{\mathrm{par}} > 0$, this places the correct parent strictly closer than every included incorrect candidate.

If $d_j^- \le d_j^+$, then $\ell_{\mathrm{par},j} \ge m_{\mathrm{par}}$. Therefore, for every $j \in S$,

$$\mathbf{1}[d_j^- \le d_j^+] \le \frac{\ell_{\mathrm{par},j}}{m_{\mathrm{par}}}.$$

Averaging over S and using that an error fraction is at most one proves equation 14. The argument uses only the included prototype distances and imposes no restriction on the logits of a separate classification head. □

Proof of Proposition 3. For every $k \in \mathcal{N}$, the triangle inequality gives

$$\begin{aligned} d_c(\boldsymbol{z}', \boldsymbol{p}_k'^{-}) &\ge d_c(\boldsymbol{z}, \boldsymbol{p}_k^-) - e_z - e_-, \\ d_c(\boldsymbol{z}', \boldsymbol{p}'^{+}) &\le d_c(\boldsymbol{z}, \boldsymbol{p}^+) + e_z + e_+. \end{aligned}$$

Subtracting yields

$$d_c(\boldsymbol{z}', \boldsymbol{p}_k'^{-}) - d_c(\boldsymbol{z}', \boldsymbol{p}'^{+}) \geq d_c(\boldsymbol{z}, \boldsymbol{p}_k^{-}) - d_c(\boldsymbol{z}, \boldsymbol{p}^{+}) - 2e_z - e_+ - e_-.$$

Taking the minimum over the same finite index set $\mathcal{N}$ proves equation 15. Under the stated margin and perturbation conditions, $g' > 0$, so the correct parent is strictly closer than every incorrect candidate after perturbation. □

Proof of Proposition 4. Set $x_h = m_{\mathrm{rad}}\Delta_h + r_{h-1} - r_h$. Because $\sum_h [x_h]_+ \geq [\sum_h x_h]_+$ and the included edges are contiguous, telescoping gives

$$\sum_{h=a+1}^{b} e_h \geq \left[\sum_{h=a+1}^{b} (m_{\mathrm{rad}}\Delta_h + r_{h-1} - r_h) \right]_+ = \left[m_{\mathrm{rad}} \sum_{h=a+1}^{b} \Delta_h + r_a - r_b \right]_+ .$$

Using $r_b \leq R_*$ proves the second inequality in equation 16. If all included hinges vanish, that inequality implies $m_{\mathrm{rad}} \sum_{h=a+1}^{b} \Delta_h \leq R_* - r_a$.

Conversely, suppose this budget condition holds. Define

$$r_h = r_a + m_{\mathrm{rad}} \sum_{j=a+1}^{h} \Delta_j, \qquad h = a+1, \ldots, b.$$

Since each $\Delta_j > 0$, all these radii belong to $[0, R_*]$. Choose a Euclidean unit vector $\boldsymbol{v}$ and set

$$\boldsymbol{z}^{(h)} = \exp_0^c\Big(\frac{r_h}{2}\boldsymbol{v}\Big), \qquad h = a, \ldots, b.$$

Proposition 1 gives $r_c(\boldsymbol{z}^{(h)}) = r_h$. Each tangent norm is at most $R_*/2 = \overline{R}$, so these points respect both representation caps. If the initial embedding is also fixed and nonzero, choose $\boldsymbol{v}$ in the direction of $\log_0^c(\boldsymbol{z}^{(a)})$; if it is zero, any direction suffices. Every included edge then satisfies $r_h - r_{h-1} = m_{\mathrm{rad}}\Delta_h$, and hence $e_h = 0$. This proves sufficiency for the radial constraints in representation space, without imposing parent-margin constraints or restrictions from a particular network parameterization.

Contiguity is essential for this bound on the included hinges. For example, with $R_* \geq 1$, radii $(0, 1, 0, 1)$ satisfy unit margins on edges $(1, 2)$ and $(3, 4)$ but decrease across the skipped edge $(2, 3)$. Thus separate eligible segments do not guarantee global radial monotonicity. □

D Implementation and Reproducibility Details

Table 4 summarizes the training settings.

Table 4: Implementation settings.

Component	Setting
Reference encoder	PointNet++; four set-abstraction stages
Readouts	Five feature-propagation decoders; 128-dimensional point features
Transfer encoder	PointNeXt; PointNet++-style readouts retained
Shared projector	1×1 convolution, GroupNorm, ReLU; output dimension 128
Geometry	Curvature $-c = -0.2$; tangent scale $\tau = 0.2$; tangent-norm cap $R = 2.3$
Prediction head	Möbius-coordinate readout; fixed logit scale 10; Kaiming-uniform weights
Geometric margins	$m_{\mathrm{par}} = m_{\mathrm{rad}} = 0.05$
Target loss weights	$\lambda_{\mathrm{par}} = 0.10$; $\lambda_{\mathrm{ord}} = 0.02$ after the ramps
Optimizer	SGD; initial learning rate 0.01; momentum 0.9; weight decay 10^{-4}
Training duration	80 epochs; learning rate halved every 20 epochs
Batch / devices	Batch size 32; Nvidia 4090 GPUs
Memory warm-up	5 epochs
Loss curriculum	Parent term starts at epoch 13; radial term at epoch 25; separate 10-epoch linear ramps
Prototype momentum	0.5 initially; 0.7 at epoch 12; 0.85 at epoch 30
Regularization sampling	Leaf-class-balanced; 2–16 points per present leaf class; at most 512 points per batch
PartNeXt input	2,048 colored points per object; custom 8:1:1 train/validation/test split

E Additional Experimental Results

This appendix gives the level-wise scores underlying the aggregate comparisons in Section 5. OA, mIoU, and consistency are percentages. The means average five prediction levels; HE uses the same trained models as MT and returns valid taxonomy paths by construction. Standard deviations, where provided, describe three training seeds. Aggregate means were computed before rounding level-wise scores, so averaging the displayed entries can introduce small rounding differences. In particular, the full-model PartNeXt HE mean OA is reported as 57.05 throughout, rounded from 57.0453 in Table 8.

E.1 Campus3D Level-wise Results

Table 5 separates published reference scores from the internally evaluated variants. The full model has the highest internal mean OA and mIoU in both decoding modes, but not the best value at every level: M0 has higher L5 OA under MT and HE, and higher L5 mIoU under HE.

Table 5: Campus3D ablation results. We report per-level overall accuracy (OA, %) and mean IoU (mIoU, %) for independent multi-task decoding (MT) and hierarchical enforcement (MT+HE). Rows marked † are external reference values from Li et al. (2020); other rows use the internal protocol.

	Per-level scores (%)						
Model	L1	L2	L3	L4	L5	Mean	Cons
MT: OA (no enforcement)							
Paper MT†	90.70	83.10	81.70	79.80	75.20	82.10	–
M0	94.12	88.77	87.61	85.96	81.10	87.51	89.45
M1	92.31	86.26	86.25	85.22	80.44	86.09	89.03
Hy3D	94.20	88.97	88.06	86.90	80.55	87.74	90.41
MT: mIoU							
Paper MT†	82.30	68.60	41.30	29.70	20.10	48.40	–
M0	86.73	77.64	60.68	41.00	25.43	58.30	–
M1	82.45	73.73	63.19	41.00	25.72	57.22	–
Hy3D	86.92	77.96	64.30	42.28	26.12	59.52	–
MT+HE: OA (hierarchical enforcement)							
Paper MT+HE†	90.70	83.10	81.70	80.00	75.40	82.18	100.00
M0	94.37	89.19	88.16	86.94	81.54	88.04	100.00
M1	92.95	87.47	86.39	85.32	80.44	86.51	100.00
Hy3D	94.58	89.55	88.51	87.36	80.97	88.19	100.00
MT+HE: mIoU							
Paper MT+HE†	82.20	68.50	41.10	29.80	20.10	48.34	–
M0	87.18	78.40	63.34	41.48	26.77	59.44	–
M1	83.77	75.76	63.26	41.13	26.35	58.05	–
Hy3D	87.70	79.08	64.94	42.37	26.58	60.14	–

E.2 PARTNEXT LEVEL-WISE RESULTS

Tables 6 and 7 report the PointNet++ component variants under MT and HE. M1 and the full model are three-seed summaries. The Baseline and M0 rows provide architectural reference points without comparable uncertainty. Adding both geometric objectives improves all five reported level-wise mean OA and mIoU values relative to M1, while raw full-path consistency remains essentially unchanged.

Table 6: PartNeXt results without hierarchical enforcement (MT). OA, mIoU, and consistency are percentages. M1 and Hy3D report three-seed means ± sample standard deviations.

	OA per level (%)						
Model	L1	L2	L3	L4	L5	Mean	Cons
Baseline	76.33	57.77	45.30	41.73	39.22	52.07	44.31
M0	77.48	58.17	46.27	42.04	38.53	52.50	54.76
M1	78.47 ± 0.2505	61.22 ± 0.5777	49.17 ± 0.3367	45.76 ± 0.2296	42.60 ± 0.5795	55.44 ± 0.3811	**58.62** ± 1.2796
Hy3D	**79.10** ± 0.1762	**62.02** ± 0.1572	**49.71** ± 0.2511	**46.18** ± 0.2237	**43.09** ± 0.2836	**56.02** ± 0.0727	58.57 ± 0.9790

	mIoU per level (%)					
Model	L1	L2	L3	L4	L5	Mean
Baseline	32.52	12.99	6.45	5.24	4.42	12.33
M0	31.29	8.59	3.69	2.56	1.84	9.59
M1	34.92 ± 0.1761	14.51 ± 0.5225	7.19 ± 0.4177	5.46 ± 0.2783	4.48 ± 0.3212	13.31 ± 0.2914
Hy3D	**37.25** ± 0.4941	**15.95** ± 0.1973	**7.73** ± 0.2194	**5.89** ± 0.2450	**4.79** ± 0.1159	**14.32** ± 0.2488

Table 7: PartNeXt results with hierarchical enforcement (MT+HE). HE guarantees valid paths for all models (consistency 100%). M1 and Hy3D report three-seed means ± sample standard deviations.

	OA per level (%)						
Model	L1	L2	L3	L4	L5	Mean	Cons
Baseline	77.46	59.71	47.89	44.34	41.11	54.10	100.0
M0	77.56	58.90	47.05	43.23	39.42	53.23	100.0
M1	79.51 ± 0.4317	61.81 ± 0.5933	50.01 ± 0.4826	46.43 ± 0.4528	43.10 ± 0.4190	56.17 ± 0.4732	100.0
Hy3D	**80.39** ± 0.1231	**62.97** ± 0.0802	**50.77** ± 0.1931	**47.18** ± 0.1716	**43.92** ± 0.1212	**57.05** ± 0.1218	100.0

	mIoU per level (%)					
Model	L1	L2	L3	L4	L5	Mean
Baseline	34.30	14.34	7.26	5.73	4.70	13.27
M0	30.16	9.43	4.25	3.06	2.29	9.84
M1	36.89 ± 1.1167	14.95 ± 0.4406	7.45 ± 0.3168	5.75 ± 0.1904	4.66 ± 0.1271	13.94 ± 0.3991
Hy3D	**39.26** ± 0.6385	**16.01** ± 0.1644	**8.00** ± 0.1601	**6.20** ± 0.1721	**5.07** ± 0.1210	**14.91** ± 0.1894

E.3 Full-precision Geometry Comparisons

Table 8 retains the available higher-precision aggregate results. These comparisons preserve the reported architecture dimensions, prototype-memory design, sampling, and optimization schedule, while changing geometry-dependent operations. The head parameterization and distance scale also change; equal numerical margins do not necessarily impose equal geometric penalties. The results therefore evaluate the two model designs rather than isolate curvature alone.

Table 8: Euclidean–hyperbolic model comparison with the same reported architecture dimensions and training design. Values are means over the five levels (%). Δ is Hyperbolic minus Euclidean. PartNeXt values report three-seed means $\pm$ sample standard deviations; Campus3D values are the corresponding internal checkpoints. Heads, distance functions, and metric scale depend on the geometry.

Backbone	Dataset	Mode/metric	Euclidean	Hyperbolic	Δ
PointNet++	Campus3D	MT OA	79.82	**87.74**	**+7.92**
PointNet++	Campus3D	MT mIoU	50.05	**59.52**	**+9.47**
PointNet++	Campus3D	HE OA	79.97	**88.19**	**+8.22**
PointNet++	Campus3D	HE mIoU	50.31	**60.14**	**+9.83**
PointNet++	PartNeXt	MT OA	55.4776 $\pm$ 0.3065	**56.0200** $\pm$ 0.0727	**+0.5424**
PointNet++	PartNeXt	MT mIoU	13.7901 $\pm$ 0.5950	**14.3227** $\pm$ 0.2488	**+0.5326**
PointNet++	PartNeXt	HE OA	56.1313 $\pm$ 0.3372	**57.0453** $\pm$ 0.1218	**+0.9140**
PointNet++	PartNeXt	HE mIoU	14.2920 $\pm$ 0.8928	**14.9107** $\pm$ 0.1894	**+0.6187**
PointNeXt	PartNeXt	MT OA	57.1128 $\pm$ 0.3531	**57.6972** $\pm$ 0.3050	**+0.5844**
PointNeXt	PartNeXt	MT mIoU	15.7120 $\pm$ 0.4162	**16.1649** $\pm$ 0.6018	**+0.4529**
PointNeXt	PartNeXt	HE OA	57.8460 $\pm$ 0.4382	**58.3205** $\pm$ 0.2694	**+0.4745**
PointNeXt	PartNeXt	HE mIoU	16.4815 $\pm$ 0.6554	**16.5845** $\pm$ 0.6934	**+0.1030**

E.4 Additional Qualitative Results

Figure 4 shows PartNeXt leaf predictions with PointNeXt under the two geometries and decoding modes. Each row uses the same semantic color mapping across models. The displayed OA values describe individual test objects; they do not estimate aggregate performance or quantify the frequency of similar improvements. HE guarantees cross-level path validity, while the leaf maps visualize the resulting semantic assignments. The dataset-level comparison is reported in Table 1.

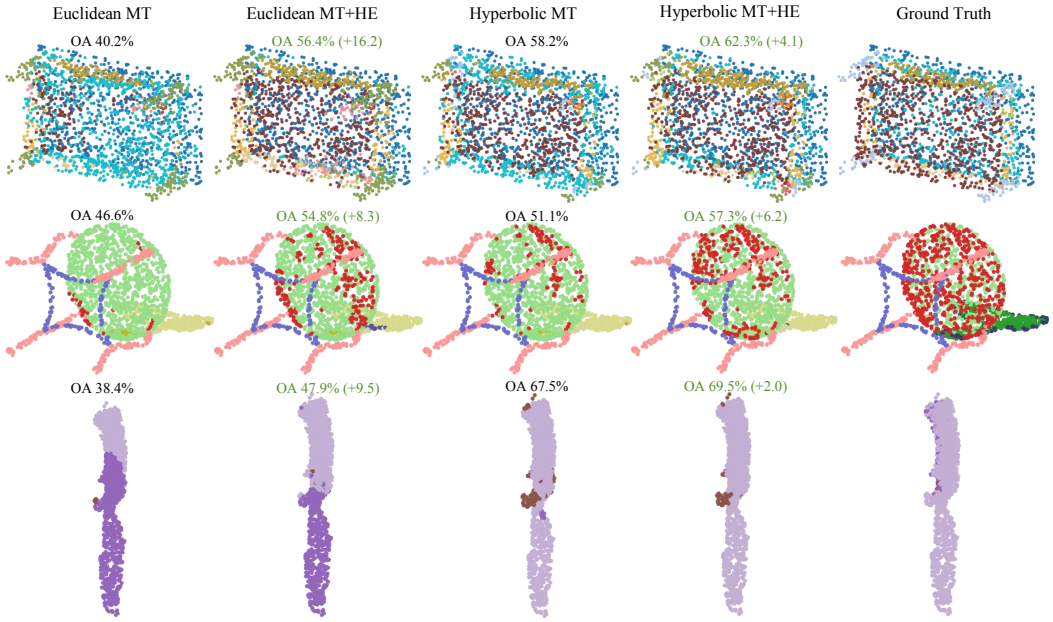

Figure 4: PartNeXt qualitative results with PointNeXt. Columns show Euclidean MT, Euclidean MT+HE, hyperbolic MT, hyperbolic MT+HE, and ground truth. OA is measured on the individual displayed object, and parenthesized changes compare HE with MT. Colors are consistent within each row.

F PROTOTYPE–TAXONOMY ANALYSIS

F.1 TREE DISTANCES AND PAIR SUBSETS

For semantic nodes u, v, we use the unweighted shortest-path distance

$$d_{\mathrm{tree}}(u, v) = D(u) + D(v) - 2D(\mathrm{LCA}(u, v)), \tag{26}$$

where LCA denotes the lowest common ancestor and D denotes semantic depth with unit-length edges. For a selected set of valid pairs $\mathcal{P}$, the alignment statistic is

$$\rho_g(\mathcal{P}) = \mathrm{corr}\big(\mathrm{rank}[d_g(p_u, p_v)]_{(u,v)\in\mathcal{P}}, \mathrm{rank}[d_{\mathrm{tree}}(u, v)]_{(u,v)\in\mathcal{P}}\big), \tag{27}$$

using Euclidean distance for $g = E$, Poincaré geodesic distance for $g = H$, and midranks for ties. Both geometries use the same valid node pairs.

PartNeXt pair subsets and level-wise results. Table 3 in Section 5.4 reports the PointNet++ results for four pair subsets and the arithmetic mean of the L2–L5 correlations. All five summaries show positive gains, with the largest gain for leaf pairs. Across levels L2–L5, the Euclidean correlations are $[0.0783, 0.0851, 0.0728, 0.0793]$, and the hyperbolic correlations are $[0.0580, 0.1039, 0.1013, 0.1639]$. Thus, L3–L5 improve while L2 decreases: positive aggregate gains do not imply improvement at every level. Deltas retain the reported precision and can differ by 0.0001 from subtracting the rounded entries.

Campus3D sibling ranking. To evaluate whether sibling categories are closer in the prototype space than categories with different parents, let $\mathcal{S}_h$ and $\mathcal{D}_h$ denote the valid same-level node pairs at level h that share a parent and have different parents, respectively. For a pair $s = (u, v)$, write $d_g(s) = d_g(p_u, p_v)$. Both geometries use the same valid pair sets. We pool the pairs across levels as

$$\mathcal{S} = \bigcup_h \mathcal{S}_h, \qquad \mathcal{D} = \bigcup_h \mathcal{D}_h,$$

and, when both pooled sets are nonempty, compute

$$\mathrm{AUC}_g = \frac{1}{|\mathcal{S}|\,|\mathcal{D}|} \sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{D}} \left[\mathbf{1}\{d_g(s) < d_g(t)\} + \frac{1}{2}\mathbf{1}\{d_g(s) = d_g(t)\}\right]. \tag{28}$$

This statistic is the probability that a uniformly selected sibling pair has a smaller prototype distance than a uniformly selected different-parent pair, with ties contributing one half. A value of 0.5 indicates chance-level ordering; larger values indicate stronger sibling proximity. The reported Sibling AUC in Table 9 pools pairs before comparison and is not the arithmetic mean of per-level AUCs. If either pooled set is empty, the statistic is undefined.

Additional relations and uncertainty. Table 9 complements the main-text results with additional relations, a second encoder, and branch-bootstrap intervals. On Campus3D, hyperbolic prototypes have higher all-pair and leaf correlations and sibling AUC, while Euclidean prototypes have higher same-level and different-parent correlations. On PartNeXt, the positive PointNet++ unique-node estimate is not decisive under branch resampling because its interval includes zero. For PointNeXt, the interval is strictly positive, supporting improved unique-node alignment under this resampling analysis.

The bootstrap resamples taxonomy units without retraining; its intervals are distinct from the training-seed standard deviations reported for segmentation. These distance-based diagnostics assess prototype–taxonomy agreement and do not directly measure the local parent-retrieval and radial-violation properties of the training objectives. The full bootstrap procedure and results are provided in the following subsection and Table 10.

F.2 PAIRED TAXONOMY BOOTSTRAP

To assess sensitivity to the sampled taxonomy units, the unique-node analysis uses 2,000 paired bootstrap replicates. The primary analysis resamples top-level branches as intact clusters; node

Table 9: Additional prototype–taxonomy alignment results and uncertainty estimates. Top: Spearman correlation (ρ) and pooled sibling-ranking AUC, with $\Delta = H - E$; larger values indicate stronger agreement for the specified relation. Bottom: unique-node $\Delta\rho$ and its percentile 95% interval from 2,000 paired branch-bootstrap replicates on PartNeXt. Intervals condition on fitted prototypes and do not measure training-seed variation.

Dataset / encoder	Relation / statistic	Euclidean	Hyperbolic	Δ
PartNeXt / PointNet++	Unique-node ρ	0.2298	**0.2622**	+0.0324
	Leaf-only ρ	0.0793	**0.1639**	+0.0846
	Level-two ρ	**0.0783**	0.0580	-0.0203
PartNeXt / PointNeXt	Unique-node ρ	0.1918	**0.3051**	+0.1133
Campus3D / PointNet++	All-pair ρ	0.2219	**0.3941**	+0.1722
	Same-level ρ	**0.3417**	0.3300	-0.0117
	Different-parent ρ	**0.2861**	0.2253	-0.0608
	Leaf-only ρ	0.0296	**0.1009**	+0.0713
	Sibling AUC	0.6923	**0.7348**	+0.0425

PartNeXt encoder	Observed $\Delta\rho$	Bootstrap mean $\Delta\rho$	95% interval
PointNet++	+0.0324	+0.0212	$[-0.0971, +0.1064]$
PointNeXt	+0.1133	+0.0997	$[+0.0201, +0.1615]$

resampling provides a sensitivity analysis. Each replicate uses the same resampled units for both geometries and records $\Delta\rho^{(b)} = \rho_H^{(b)} - \rho_E^{(b)}$. If node i appears with multiplicity m_i, each distinct pair (i, j) receives multiplicity $m_i m_j$, and self-pairs from repeated draws are excluded. Multiplicities must be applied to ranking and correlation consistently, as in the corresponding expanded repeated-pair sample.

Table 10: PartNeXt paired taxonomy bootstrap. The confidence intervals are percentile intervals for $\Delta\rho$ conditional on the fitted prototypes. $\Pr(\Delta\rho > 0)$ is the empirical positive-replicate fraction, not a p-value.

Encoder	Unit	Obs. ρ_E	Obs. ρ_H	Mean $\Delta\rho$	95% CI / $\Pr(\Delta\rho > 0)$
PointNet++	branch	0.2298	0.2622	+0.0212	$[-0.0971, +0.1064]$ / 0.6765
PointNet++	node	0.2298	0.2622	+0.0322	$[-0.0064, +0.0695]$ / 0.9465
PointNeXt	branch	0.1918	0.3051	+0.0997	$[+0.0201, +0.1615]$ / 0.9960
PointNeXt	node	0.1918	0.3051	+0.1133	$[+0.0896, +0.1365]$ / 1.0000

Both PointNet++ intervals include zero, while both PointNeXt intervals are positive. This qualifies the positive observed correlation gaps. The bootstrap does not retrain the networks and therefore does not measure training-seed variability. Branch resampling preserves within-branch structure but does not remove every dependence among pairs, particularly pairs spanning branches.