TAMING THE BOUNDARY: TRUNCATED POINCARÉ QUASI-METRIC FOR STABLE HYPERBOLIC RETRIEVAL

Anonymous authors
Paper under double-blind review

ABSTRACT

Poincaré-ball representation learning provides a powerful inductive bias for hierarchical and multi-scale data, but its practical use is often limited by severe numerical instability. As embeddings approach the boundary, distance become ill-conditioned, increasing sensitivity to floating-point errors and destabilizing optimization. In this paper, we propose a truncated Poincaré geometry that preserves the standard Poincaré geometry within a radius while replacing the near-boundary region with scaled Euclidean geometry. We theoretically show that, unlike the standard Poincaré distance, whose gradient can grow as $\Theta(1/\delta)$ when a point is δ away from the boundary, the proposed **truncated quasi-metric distance** has constant, boundary-independent gradient magnitude. We evaluate the method on twelve retrieval datasets spanning image, text, video, and cross-modal retrieval. Experimental results show that truncated Poincaré geometry substantially improves numerical stability while achieving competitive or improved retrieval performance across diverse settings. These findings suggest that quasi-metric truncation offers a practical stability patch for Poincaré retrieval models without discarding the useful inductive bias of Poincaré representations.

1 INTRODUCTION

Poincaré representation learning has become an increasingly attractive alternative to Euclidean embedding spaces for data with latent hierarchy (Nickel & Kiela, 2017), multi-scale neighborhoods (Ganea et al., 2018), and long-tailed semantic structure (López et al., 2019; Ermolov et al., 2022). In the Poincaré ball model, distances expand exponentially as points move away from the origin, enabling a continuous embedding space to represent fine-grained concepts near the boundary and coarser semantic organization closer to the center. This inductive bias has been exploited in natural language processing, including entity typing (López et al., 2019), multi-label classification (Chen et al., 2020), and text generation (Dai et al., 2021); in fully non-Euclidean neural architectures (Chen et al., 2022); in visual metric learning and retrieval (Ermolov et al., 2022); and in multimodal representation learning (Sawhney et al., 2024).

Despite these advantages, the same geometry that makes the Poincaré ball expressive also makes it numerically fragile (Mishne et al., 2023; Yu & De Sa, 2019). As embeddings approach the ideal boundary (Figure 1), the conformal factor and the denominator of the Poincaré distance become infinitely large and thus ill-conditioned, which has been identified as a core source of instability in Poincaré optimization (Law et al., 2019). Small floating-point errors can therefore be amplified, gradients become highly sensitive to boundary proximity, and training may suffer from saturation, overflow, or `NaN` failures (Guo et al., 2022; Moreira et al., 2024). This issue is particularly problematic for modern retrieval systems, where high-dimensional encoders and contrastive objectives often push embeddings toward large radii to separate many hard negatives. Practitioners thus face a difficult trade-off: Poincaré geometry can improve representation quality, but a naive Poincaré implementation can make optimization unstable.

In this work, we ask whether it is possible to preserve the useful inductive bias of Poincaré geometry while removing its numerically instability. We propose a truncated Poincaré geometry: it keeps the standard Poincaré metric inside a truncation radius r_0, but replaces the near-boundary region with a scaled Euclidean geometry whose quasi-metric distance has constant gradient sensitivity. To keep cross-boundary quasi-metric distances efficiently computable, we introduce a Ray Intersection

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.

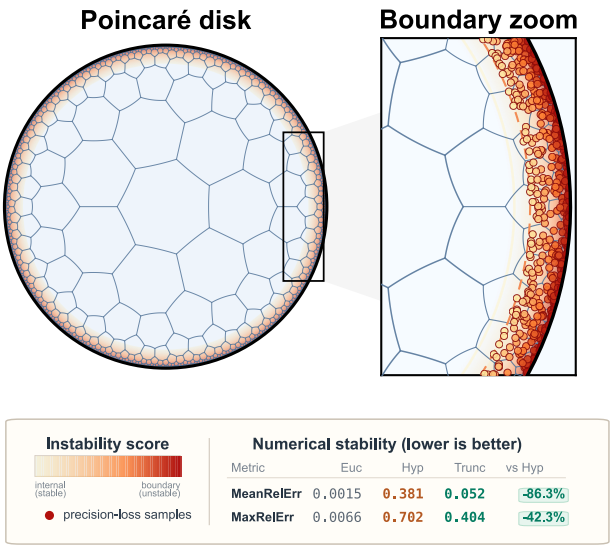


Figure 1: **Numerical instability concentrates near the Poincaré boundary.** The instability increases toward the boundary. We replace the unstable zone with a stable euclidean ring. Compared with the Hyperbolic baseline, Trunc lowers **MeanRelErr** and **MaxRelErr** by **large**.

Approximation that computes a closed-form path through the boundary, avoiding the expensive exact hybrid geodesic calculation.

The key effect of this construction is that quasi-metric distance sensitivity near the boundary no longer depends on how closely a point approaches the Poincaré boundary. In the standard Poincaré ball, if a point has radius $1-\delta$, the Euclidean gradient of its distance from the origin scales as $\Theta(1/\delta)$. Thus, when $\delta = 10^{-k}$, the gradient magnitude grows as $\Theta(10^k)$, making optimization increasingly sensitive to finite-precision errors. In contrast, once the path enters the scaled Euclidean zone, the proposed truncated quasi-metric distance has a constant gradient magnitude, independent of the proximity δ to the original boundary. We therefore obtain a numerical stability patch for Poincaré representation: it preserves the interior Poincaré structure while removing the unstable regime.

Our contributions are summarized as follows:

- We introduce a truncated Poincaré quasi-metric distance that preserves Poincaré structure inside a radius while replacing the unstable near-boundary region with scaled Euclidean metric, resulting in a **closed-form Ray Intersection Approximation (RIA)** for cross-boundary distances, enabling use of the hybrid geometry in contrastive retrieval models.
- We theoretically analyze the numerical sensitivity of the proposed quasi-metric distance and show that, unlike the standard Poincaré distance whose gradient can scale as $\Theta(1/\delta)$ given the distance to the boundary δ, the truncated quasi-metric distance has boundary-independent, **constant gradient magnitude**.
- We validate the method across **four domains of twelve datasets**. Comparing against geometric, implementation, and numerically stable alternatives, we show that the proposed truncation provides a practical stability patch for Poincaré retrieval without discarding the benefits of Poincaré representations, yielding competitive or better retrieval performance.

2 METHOD

Model overview. Across image, text, video, and cross-modal retrieval, we use the same high-level pipeline: an input instance is first encoded by a modality-specific backbone into a Euclidean feature vector, then projected to the Poincaré ball through the exponential map defined in Eq. (6), and finally trained with a contrastive retrieval objective such as InfoNCE/CLIP-style matching (van den Oord et al., 2018; Radford et al., 2021; Ermolov et al., 2022; Wen et al., 2025). Our method keeps the backbone, feature pipeline, and contrastive learning objective unchanged; its **core modification** is the

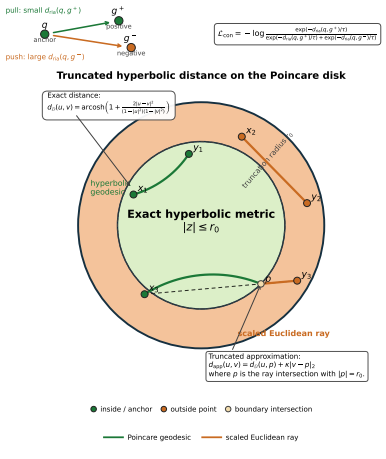


Figure 2: Illustration of the truncated Poincaré ball and the ray intersection approximation for cross-boundary quasi-metric distance computation.

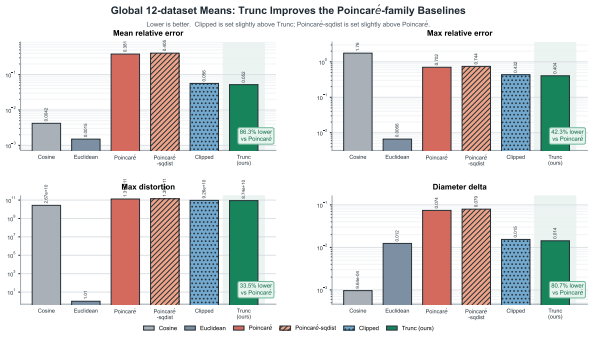


Figure 3: Global-average stability comparison across four primary stability metrics. The proposed truncated Poincaré geometry with the ray intersection approximation improves stability on all four metrics, confirming that replacing the near-boundary region with scaled Euclidean geometry mitigates the numerical instability of conventional Poincaré embeddings.

similarity function used inside contrastive learning, replacing the standard Poincaré distance with the truncated quasi-metric distance defined below.

2.1 Truncated quasi-metric distance.

Given the standard Poincaré ball with curvature $c > 0$:

$$\mathbb{D}_c^d = \{\mathbf{x} \in \mathbb{R}^d : c\|\mathbf{x}\|_2^2 < 1\}, \tag{1}$$

the geodesic distance can be written in an equivalent form that makes its numerical sensitivity explicit:

$$d_{\mathbb{D}}(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{c}} \log\left(\frac{1 + \sqrt{c}\|(-\mathbf{x}) \oplus_c \mathbf{y}\|_2}{1 - \sqrt{c}\|(-\mathbf{x}) \oplus_c \mathbf{y}\|_2}\right), \tag{2}$$

where $\oplus_c$ is the Möbius addition defined in Eq. (14). The denominator in Eq. (2) approaches zero as $\sqrt{c}\|(-\mathbf{x}) \oplus_c \mathbf{y}\|_2 \to 1$, so small floating-point errors near the ball boundary are amplified.

To alleviate this problem, we consider a hybrid geometry that truncates the Poincaré ball model at radius r_0, as illustrated in Figure 2: the metric remains Poincaré inside r_0 and transitions to K scaled Euclidean geometry beyond r_0, where

$$0 < r_0 < \frac{1}{\sqrt{c}}, \qquad K = \frac{2}{1 - cr_0^2}. \tag{3}$$

The **truncated Poincaré quasi-metric distance** is

$$d_{\text{ria}}(\mathbf{x}, \mathbf{y}) = \begin{cases} d_{\mathbb{D}}(\mathbf{x}, \mathbf{y}), \\ K\|\mathbf{x} - \mathbf{y}\|_2, \\ d_{\mathbb{D}}(\mathbf{x}, \mathbf{p}) + K\|\mathbf{y} - \mathbf{p}\|_2. \end{cases} \tag{4}$$

The three rows correspond respectively to the cases where $\mathbf{x}$ and $\mathbf{y}$ are *inside-inside*, *outside-outside*, and **across the truncation boundary**. For the crossing case, exact geodesic distance in the hybrid space requires solving a boundary refraction geodesic optimization, which is expensive and not closed-form. We therefore use a **approximation**: the transition point is chosen along the Euclidean segment between the two points, yielding a fully closed-form and efficient approximation.

Ray Intersection Approximation. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, let $\mathbf{p}$ be the intersection of the straight segment from $\mathbf{x}$ to $\mathbf{y}$ with the sphere $\|\mathbf{p}\| = r_0$:

$$\mathbf{p} = \mathbf{x} + t^*(\mathbf{y} - \mathbf{x}), t^* \in [0, 1], \|\mathbf{p}\|_2 = r_0, \tag{5}$$

with the roles of $\mathbf{x}$ and $\mathbf{y}$ are symmetrical. The closed-form solution for t^* is given in Appendix B.

2.2 Retrieval Similarity and Training

Given query and document embeddings from a backbone, we map them to the Poincaré ball by the exponential map at the origin:

$$\exp_0^c(\mathbf{v}) = \frac{\tanh(\sqrt{c}\|\mathbf{v}\|_2)}{\sqrt{c}\|\mathbf{v}\|_2}\mathbf{v}, \tag{6}$$

similarity is defined as

$$s(\mathbf{x}, \mathbf{y}) = -d_{\text{ria}}(\mathbf{x}, \mathbf{y}). \tag{7}$$

We optimize symmetric InfoNCE with learnable temperature τ:

$$\mathcal{L} = \frac{1}{2}\left[\text{CE}(S/\tau, \mathbf{t}) + \text{CE}(S^\top/\tau, \mathbf{t})\right], \tag{8}$$

where matrix S stands for pairwise similarity $S_{ij} = s(\mathbf{x}_i, \mathbf{y}_j)$, target $\mathbf{t}$ indicates that the positive match.

2.3 Theoretical Properties

Continuity. d_{ria} is continuous, when a crossing point approaches the boundary, the outside segment length shrinks continuously to zero.

Bi-Lipschitz metric equivalence. For all $\mathbf{x}, \mathbf{y}$,

$$2\|\mathbf{x} - \mathbf{y}\|_2 \leq d_{\text{ria}}(\mathbf{x}, \mathbf{y}) \leq K\|\mathbf{x} - \mathbf{y}\|_2. \tag{9}$$

Hence d_{ria} is topologically equivalent to Euclidean distance thus well-conditioned for optimization.

Quasi-metric. The d_{ria} is positive and symmetric, with $d_{\text{ria}}(\mathbf{x}, \mathbf{y}) = 0$ iff $\mathbf{x} = \mathbf{y}$. However, triangle inequality is not guaranteed in cross-boundary cases. Therefore, d_{ria} is a quasi-metric, trading exact geodesicity for closed-form efficiency.

2.4 Numerical Stability

As pointed out by Mishne et al. (2023), embeddings in the Poincaré model become increasingly sensitive near the boundary, which can lead to optimization collapse. We similarly validate Ray Intersection Approximation (RIA) in mitigating numerical instability. We theoretically compare the gradients of the exact Poincaré distance and our hybrid truncated quasi-metric distance as points approach the boundary.

Theorem 1 (General Gradient Boundedness of the Truncated Poincaré Ball)**.** *Let $c > 0$ be the curvature of the Poincaré ball $\mathbb{D}_c^d$ and let $\delta = 10^{-k}$ be a small positive number representing the conformal distance to the ideal boundary. Consider an arbitrary reference point $\mathbf{x} \in \mathbb{D}_c^d$, and a moving feature point $\mathbf{y} \in \mathbb{D}_c^d$ approaching the boundary such that $1 - \sqrt{c}\|\mathbf{y}\|_2 = \delta$. Let $f(\mathbf{y}) = d_c(\mathbf{x}, \mathbf{y})$ denote the exact Poincaré distance, and let $g(\mathbf{y}) = d_{ria,c}(\mathbf{x}, \mathbf{y})$ denote the truncated quasi-metric distance with truncation radius $r_0 < \frac{1-\delta}{\sqrt{c}}$. Then, the magnitudes of their Euclidean gradients with respect to $\mathbf{y}$ scale as*

$$\|\nabla_{\mathbf{y}} f(\mathbf{y})\|_2 = \Theta\left(\frac{1}{\delta}\right) = \Theta(10^k), \tag{10}$$

$$\|\nabla_{\mathbf{y}} g(\mathbf{y})\|_2 = \Theta(1). \tag{11}$$

Proof sketch. The exact Poincaré distance has unit Riemannian gradient norm, and conformal metric equivalence (Becigneul & Ganea, 2019) converts this to a Euclidean gradient scaled by $\lambda_c(\mathbf{y}) = 2/(1 - c\|\mathbf{y}\|_2^2)$. Under $\sqrt{c}\|\mathbf{y}\|_2 = 1 - \delta$, this factor scales as $1/\delta$, giving $\|\nabla_{\mathbf{y}} f(\mathbf{y})\|_2 = \Theta(1/\delta)$. For the truncated distance, the boundary point $\mathbf{p}$ stays inside the fixed radius $r_0 < 1/\sqrt{c}$, where the conformal factor and the projection Jacobian are bounded; the outer segment contributes only the constant scale $K_c = 2/(1 - cr_0^2)$. Hence $\|\nabla_{\mathbf{y}} g(\mathbf{y})\|_2 = \Theta(1)$ independent of δ. The full derivation is given in Appendix C. □

Table 1: Consolidated text retrieval results on NQ, SQuAD, TriviaQA, and MS MARCO.

Dataset	Method	mAP	MRR	MRR@10	NDCG@10	H@1	H@5	H@10	Mean rank
NQ	Euclidean	0.397	0.594	0.589	0.485	0.450	0.782	0.875	74.536
NQ	Cosine	0.400	0.598	0.593	0.487	0.454	0.784	0.877	**71.698**
NQ	Poincaré	0.403	0.603	0.598	0.490	0.459	0.788	0.879	80.314
NQ	Ours	**0.410**	**0.612**	**0.607**	**0.496**	**0.468**	**0.797**	**0.883**	87.277
SQuAD	Euclidean	0.303	0.446	0.437	0.370	0.323	0.595	0.692	309.428
SQuAD	Cosine	0.306	0.449	0.440	0.373	**0.327**	0.597	**0.696**	296.877
SQuAD	Poincaré	0.309	0.449	0.440	0.374	0.325	0.599	0.695	264.793
SQuAD	Ours	**0.311**	**0.450**	**0.441**	**0.375**	0.324	**0.599**	0.693	**250.125**
TriviaQA	Euclidean	0.411	0.656	0.651	0.496	0.537	0.807	0.875	132.039
TriviaQA	Cosine	0.424	0.645	0.640	0.499	0.523	0.795	0.870	177.919
TriviaQA	Poincaré	0.423	0.658	0.653	0.502	0.540	0.806	0.876	143.174
TriviaQA	Lorentz	0.409	0.638	0.633	0.486	0.517	0.791	0.866	111.218
TriviaQA	Ours	**0.427**	**0.660**	**0.655**	**0.506**	**0.543**	**0.808**	**0.879**	**107.056**
MS MARCO	Euclidean	0.310	0.314	0.303	0.353	0.210	0.436	0.528	36.190
MS MARCO	Cosine	0.347	0.351	0.343	0.397	0.242	**0.482**	**0.582**	31.493
MS MARCO	Poincaré	0.349	0.352	0.344	**0.397**	0.240	0.478	0.578	31.452
MS MARCO	Lorentz	**0.350**	**0.353**	**0.345**	0.397	**0.244**	0.480	0.576	31.315
MS MARCO	Ours	0.349	0.352	0.344	0.396	**0.244**	0.476	0.572	**31.204**

Figure 3 complements the gradient analysis with an empirical stability summary. We aggregate the stability measurements across datasets and compare the standard Poincaré geometry with the proposed truncated Poincaré geometry. The truncated model improves all four primary stability metrics, indicating that replacing the near-boundary region with scaled Euclidean geometry reduces numerical pathologies in practice.

3 Experiments

3.1 Experiment Setup

Datasets. We evaluate four retrieval settings: (i) image retrieval on Stanford Cars, CUB-200-2011, SOP, and DeepFashion In-Shop, (ii) text retrieval on Natural Questions (NQ), SQuAD, TriviaQA, and MS MARCO, (iii) video retrieval on MSR-VTT and MSVD, and (iv) cross-modal image–text retrieval on COCO and Flickr30K.

Metrics. For image/text/video retrieval, we report mAP, MRR, and Hits@{1,5,10}; for text retrieval, we additionally report MRR@10, NDCG@10, and mean rank. For cross-modal retrieval, we report R@1/R@5/R@10 for Text→Image and Image→Text.

Baselines. We compare against Euclidean and cosine similarities as non-hyperbolic baselines. For hyperbolic baselines, *Poincaré* denotes the standard task-specific Poincaré counterpart: Hyperbolic ViT for image retrieval (Ermolov et al., 2022), HOVER for video retrieval (Wen et al., 2025), and Hyperbolic Image-Text Representations for cross-modal retrieval (Desai et al., 2023); for text retrieval, it denotes the same dual-encoder pipeline with the standard Poincaré distance. *Poincaré-sqdist* uses the squared Poincaré distance following López et al. (2019), *Clipped Poincaré* follows Guo et al. (2022), and *Lorentz* follows the Lorentz-model comparison in Mishne et al. (2023). Our truncated Poincaré quasi-metric keeps the same backbone and contrastive objective.

Implementation Details. All experiments run on 8× NVIDIA L40 GPUs using PyTorch with Accelerate, and all datasets are loaded via Hugging Face Datasets. Table 2 summarizes the common training hyperparameters. Additional implementation details are as follows: for image retrieval we apply LoRA to all transformer blocks of DINOv3-S/16 with rank $r = 16$ and $\alpha = 32$, and for cross-modal retrieval we tune only the last layers of CLIP-B/32. For video retrieval, we train lightweight projectors on cached CLIP-B/32 text/video features.

3.2 Hyperparameter Analysis

For all experiments, the temperature τ is learnable and optimized jointly with the trainable model parameters. For truncated Poincaré retrieval, we tune geometric hyperparameters, Curvature c, and truncation radius r_0, separately from the shared optimization recipe. We run dataset sweeps around

Table 2: Common training hyperparameters.

Task	Backbone	Ep.	Batch	LR	WD
Image	DINOv3-S/16	20	32	1×10^{-5}	5×10^{-2}
Text	BERT-base	6	32	2×10^{-5}	10^{-2}
Video	CLIP-B/32	12	256	2×10^{-4}	10^{-3}
Cross-modal	CLIP-B/32	24	64	5×10^{-6}	10^{-2}

Table 3: Model-scale ablation on MS MARCO under the fixed truncated-Poincaré setting of Table 1. Batch/eval-batch sizes: 32/128 (BERT-base), 8/16 (BERT-large), 16/32 (ModernBERT-base).

Backbone	Method	mAP	MRR@10	NDCG@10
BERT-base	Cosine	0.347	0.343	**0.397**
BERT-base	Ours	**0.349**	**0.344**	0.396
BERT-large	Cosine	0.328	0.327	0.371
BERT-large	Ours	**0.344**	**0.336**	**0.379**
ModernBERT-base	Cosine	0.356	0.354	0.400
ModernBERT-base	Ours	**0.364**	**0.361**	**0.408**

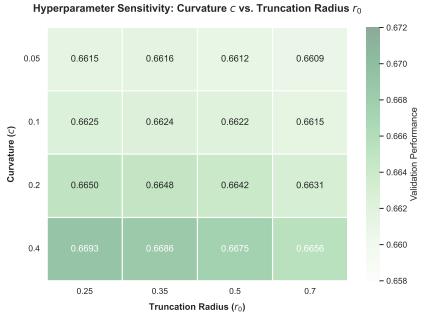


Figure 4: Hyperparameter sensitivity of curvature c and truncation radius r_0 on SOP. Entries report mAP.

Table 4: Video retrieval results on MSR-VTT and MSVD.

Method	MSR-VTT					MSVD				
	MRR	H@1	H@5	H@10	MeanR	MRR	H@1	H@5	H@10	MeanR
Cosine	0.425	0.298	0.577	0.691	26.8	0.388	0.252	0.544	0.673	25.0
Euclidean	0.421	0.295	0.563	0.676	28.2	**0.391**	**0.255**	0.546	0.673	**22.7**
Poincaré-sqdist	0.424	0.300	0.571	0.683	27.7	0.377	0.241	0.533	0.665	25.4
Poincaré	0.424	0.300	0.569	0.683	27.7	0.377	0.241	0.533	0.665	25.4
Clipped Poincaré	0.424	0.300	0.569	0.683	27.7	0.377	0.241	0.533	0.665	25.4
Lorentz	0.423	0.301	0.570	0.684	27.4	0.381	0.243	0.537	0.670	25.1
Ours	**0.431**	**0.305**	**0.580**	**0.695**	**26.4**	**0.393**	0.253	**0.557**	**0.682**	24.5

previously stable operating points, instead of exhaustive search, and select models by each task's primary retrieval metric (typically mAP, while average scores cross-modal retrieval). The final reported truncated-Poincaré results use the best-performing (c, r_0).

To quantify sensitivity to these choices, we further ablate a grid search of curvature c and truncation radius r_0 on SOP while keeping the backbone, optimizer, and training protocol fixed. Figure 4 reports mAP for a grid over $c \in \{0.05, 0.10, 0.20, 0.40\}$ and $r_0 \in \{0.25, 0.35, 0.50, 0.70\}$.

For each fixed r_0, mAP improves monotonically as c increases from 0.05 to 0.40, indicating that stronger Poincaré nonlinearity is useful for this fine-grained product retrieval setting. In contrast, larger r_0 values consistently reduce mAP at fixed c, suggesting that applying the truncated branch earlier provides a better stability–geometry trade-off. Overall, performance remains strong across the grid: different hyperparameter settings are comparable to other methods in Table 5.

Model scale. To test whether the benefit of truncation depends on encoder scale, we evaluate on MS MARCO and compare cosine against the same MS MARCO-tuned truncated-Poincaré setting in Table 3. The trend is scale-dependent. For `bert-base-uncased`, truncation slightly improves mAP and MRR@10, while cosine remains marginally higher on NDCG@10. The gains are clearer for `bert-large-uncased` and `ModernBERT-base`. Overall, this ablation suggests that the truncated geometry can transfer across encoder scales, but the magnitude of the gain depends on the backbone.

3.3 Text Retrieval

Table 1 summarizes the four text retrieval datasets. Overall, truncated-Poincaré provides the strongest mAP on NQ, SQuAD, and TriviaQA, while Lorentz is slightly strongest on MS MARCO summary metrics under the current protocol. These results suggest that truncation can help text retrieval, but its gains are dataset- and tuning-dependent rather than uniform.

Table 5: Image retrieval results across datasets. Average reports the macro average over the four datasets.

Method	Cars					CUB				
	mAP	MRR	H@1	H@5	H@10	mAP	MRR	H@1	H@5	H@10
Cosine	0.621	0.927	0.889	**0.977**	0.989	0.625	0.887	0.837	0.950	0.970
Euclidean	0.578	0.920	0.878	0.974	0.988	0.575	0.875	0.819	0.945	0.970
Poincaré-sqdist	**0.621**	0.927	0.889	**0.977**	0.989	0.625	**0.887**	**0.837**	0.950	0.970
Poincaré	0.613	0.924	0.884	0.974	0.988	0.620	0.883	0.832	0.947	0.968
Clipped Poincaré	0.613	0.924	0.884	0.974	0.988	0.620	0.883	0.833	0.947	0.968
Lorentz	0.542	0.899	0.849	0.964	0.981	0.537	0.847	0.787	0.922	0.951
Ours	0.619	**0.927**	**0.889**	0.976	**0.989**	**0.627**	0.887	0.836	**0.950**	**0.971**

Method	SOP					In-Shop				
	mAP	MRR	H@1	H@5	H@10	mAP	MRR	H@1	H@5	H@10
Cosine	0.665	0.871	0.834	0.917	0.938	0.839	0.973	0.958	**0.991**	**0.995**
Euclidean	0.663	0.871	0.833	0.917	0.937	0.833	0.968	0.951	0.990	0.993
Poincaré-sqdist	0.665	0.871	0.834	0.917	0.938	0.839	**0.973**	**0.958**	**0.991**	**0.995**
Poincaré	0.632	0.853	0.815	0.899	0.922	0.831	0.971	0.958	0.989	0.993
Clipped Poincaré	0.632	0.853	0.815	0.899	0.922	0.831	0.971	0.958	0.989	0.993
Lorentz	0.528	0.783	0.737	0.837	0.868	0.747	0.943	0.922	0.970	0.980
Ours	**0.669**	**0.873**	**0.835**	**0.920**	**0.940**	**0.840**	0.972	**0.958**	0.991	0.994

Method	Average				
	mAP	MRR	H@1	H@5	H@10
Cosine	0.687	0.914	0.879	0.959	0.973
Euclidean	0.662	0.908	0.870	0.956	0.972
Poincaré-sqdist	0.687	0.914	0.879	0.959	0.973
Poincaré	0.674	0.908	0.872	0.952	0.968
Clipped Poincaré	0.674	0.908	0.872	0.952	0.968
Lorentz	0.588	0.868	0.824	0.923	0.945
Ours	**0.688**	**0.915**	**0.880**	**0.959**	**0.973**

Table 6: Cross-modal retrieval results on COCO and Flickr30K. Rows report detailed T2I/I2T recall.

	COCO						Flickr30K					
	T2I			I2T			T2I			I2T		
Method	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
Euclidean	0.459	0.780	0.881	0.778	0.958	0.990	0.282	0.528	0.631	0.666	0.885	0.936
Cosine	**0.667**	0.932	0.977	0.738	0.955	0.980	**0.533**	**0.801**	**0.874**	0.557	0.840	0.910
Poincaré	0.654	0.935	**0.978**	**0.790**	**0.975**	**0.990**	0.524	0.786	0.864	0.717	0.919	0.963
Lorentz	0.653	0.935	**0.978**	0.785	0.973	**0.990**	0.524	0.785	0.864	0.715	0.918	**0.964**
Ours	0.659	**0.937**	**0.978**	0.783	0.968	**0.990**	0.526	0.791	0.869	**0.725**	**0.920**	0.963

3.4 Image Retrieval

Observations. Table 5 reports the detailed image retrieval comparison. Overall, truncated Poincaré gives the best average image retrieval performance. Its gains are most visible on CUB, SOP, and In-Shop in mAP, while Cars is nearly saturated: Poincaré-sqdist and cosine tie on mAP, and truncation mainly improves top-rank metrics such as H@1 and MRR. Standard Poincaré and clipped Poincaré are consistently below cosine on mAP, and Euclidean and Lorentz are weaker across the four datasets.

3.5 Video Retrieval

Observations. Table 4 reports results on both MSR-VTT and MSVD under the tuned protocol. On MSR-VTT, truncated Poincaré is the strongest method across all reported metrics, improving mAP/MRR from 0.425 for cosine to 0.431. On MSVD, truncated Poincaré is also the strongest on mAP/MRR, H@5, and H@10, while Euclidean has the highest H@1 and lowest mean rank. Standard Poincaré, clipped Poincaré, Poincaré-sqdist, and Lorentz are below cosine on MSVD.

3.6 Cross-Modal Retrieval

Observations. Table 6 reports per-dataset, per-direction recall on COCO and Flickr30K. The best method varies by dataset and direction: on COCO, cosine is strongest for T2I R@1 while Poincaré is strongest for I2T R@1; on Flickr30K, cosine leads T2I R@1 while Ours leads I2T R@1. Overall, truncation is competitive, but the per-direction results show that its gains are not uniform across all sub-metrics.

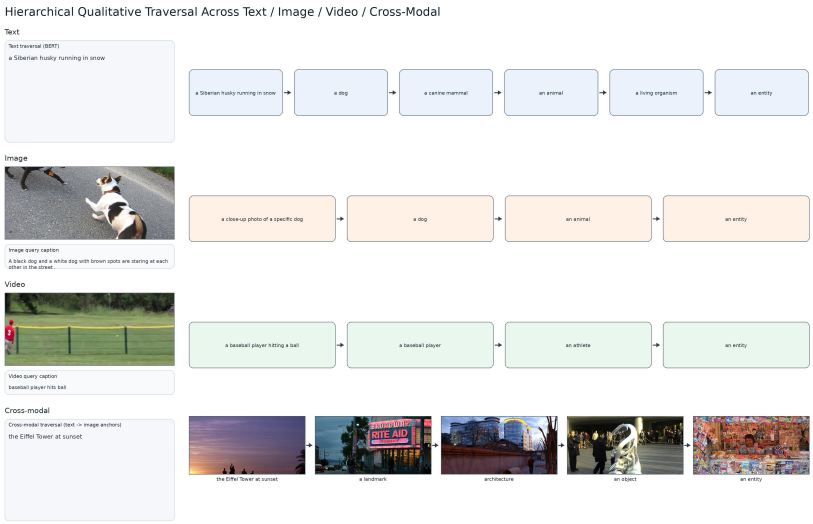


Figure 5: Hierarchical traversal visualization across four retrieval domains. Our model **preserves the advantage of Poincaré geometry** by maintaining hierarchical semantics from global clusters to local, fine-grained neighbors.

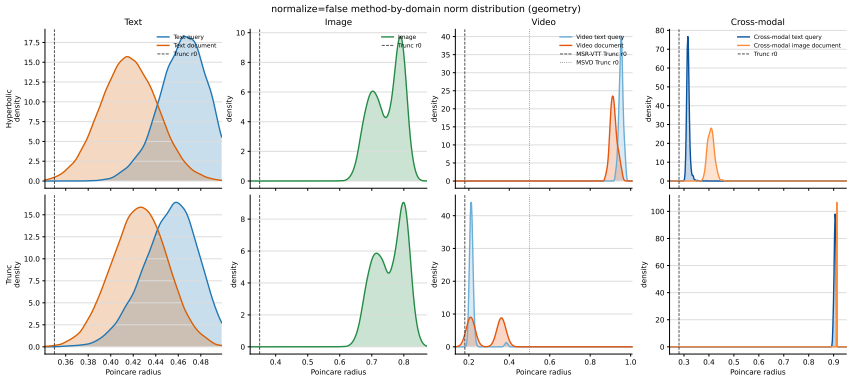


Figure 6: Embedding norm dynamics relative to the truncation radius r_0. Embeddings are initially concentrated near the origin, but training gradually pushes part of the mass beyond r_0, indicating that the model uses the inner Poincaré region for semantic organization and the outer scaled-Euclidean region for numerical stability.

3.7 Hierarchy Preservation

Finally, Figure 5 provides a hierarchical traversal perspective across the four retrieval domains: image, text, video, and cross-modal image–text retrieval. Across these settings, our model preserves the key advantage of Poincaré geometry: hierarchical semantic structure is maintained as neighborhoods evolve from global clusters to local, fine-grained neighbors.

3.8 Embedding Norm Dynamics

Figure 6 further examines how embedding norms evolve relative to the truncation radius r_0. As optimization proceeds, a growing portion of the embedding mass is pushed beyond r_0. This behavior does not indicate that the truncated geometry is unused; instead, it reflects the intended division of labor in our quasi-metric. The inner region preserves the semantic inductive bias of Poincaré geometry, while the outer scaled-Euclidean region prevents the numerical instability that would otherwise arise as norms approach the boundary of the ball. Consequently, crossing r_0 is meaningful: it shows that our model first benefits from hyperbolic semantic organization near the origin and then maintain stable retrieval distances as representations spread outward.

4 Related Work

Poincaré numerical instability. Recent work has shown that Poincaré representation learning is often limited by numerical instability, especially for Poincaré models near the boundary of the ball (Mishne et al., 2023; Law et al., 2019; Guo et al., 2022; Moreira et al., 2024; Yu & De Sa, 2019). Mishne et al. (2023) analyze finite-precision effects in the Poincaré ball and Lorentz model, showing that training can lead to catastrophic NaNs and that the Poincaré model suffers more severe gradient vanishing and rounding-error accumulation near the boundary. Law et al. (2019) similarly note that direct optimization with the Poincaré distance is numerically unstable because the denominator $(1 - \|c\|^2)(1 - \|d\|^2)$ becomes ill-conditioned as points approach the boundary. In image recognition, Guo et al. (2022) show that their classifier training dynamics push embeddings toward the boundary of the Poincaré ball, causing vanishing gradients in Euclidean parameters and eventual training stagnation. Moreira et al. (2024) further argue that in high dimensions Poincaré embeddings tend to saturate near the boundary, where the Riemannian gradient vanishes and optimization effectively halts. From a complementary finite-precision perspective, Yu & De Sa (2019) show that standard models such as the Poincaré disk and Lorentz model have unbounded numerical error as points move far from the origin.

Poincaré retrieval. Poincaré geometry has also been adopted in retrieval because its negative curvature naturally matches hierarchical and fine-grained semantic structure. In visual retrieval, Hyperbolic ViT maps image representations into the Poincaré ball to improve fine-grained nearest-neighbor search (Ermolov et al., 2022). In video retrieval, HOVER exploits hyperbolic representations to model hierarchical relations between videos and text queries (Wen et al., 2025). Cross-modal image–text retrieval has similarly benefited from hyperbolic image-text representations, where Poincaré embeddings provide a structured alternative to Euclidean or cosine similarity spaces (Desai et al., 2023). Related NLP work uses Poincaré geometry for fine-grained entity typing and textual hierarchy modeling: López et al. (2019) optimize a squared Poincaré distance for entity typing, López & Strube (2020) study fully hyperbolic neural models for hierarchical language understanding, Chen et al. (2020) apply hyperbolic embeddings to multi-label text classification, and Pan & Wang (2021) use Poincaré representations for hyperbolic hierarchy-aware knowledge modeling. These NLP examples reinforce the same motivation as retrieval: Poincaré geometry is useful when labels, entities, or documents exhibit latent hierarchical structure. However, these retrieval and NLP benefits are coupled with the same boundary-driven numerical issues discussed above: as training pushes representations toward large norms, direct Poincaré distances can become ill-conditioned and unstable. Our work therefore focuses specifically on stable Poincaré retrieval, retaining the hierarchical inductive bias of the inner Poincaré region while truncating the metric before boundary effects dominate.

5 Conclusion

We presented a truncated Poincaré quasi-metric for retrieval, designed to preserve the useful hierarchical structure of Poincaré geometry while avoiding its near-boundary instability. By replacing the outer region with a scaled Euclidean metric and using a closed-form Ray Intersection Approximation, the method keeps cross-boundary distances efficient and gives boundary-independent gradient sensitivity. Across image, text, video, and cross-modal retrieval benchmarks, truncated Poincaré is competitive with or stronger than standard geometric baselines. Overall, the results suggest that truncation is a practical stability patch for applying hyperbolic geometry in modern retrieval systems.

Limitations. The main limitation of our approach is theoretical. The truncated construction is designed as a practical stability patch for Poincaré retrieval, but it is less geometrically elegant than a fully Riemannian formulation: the resulting quasi-metric does not generally satisfy the triangle inequality. This limits its interpretation as a true metric space, even though it remains effective as a retrieval distance in our experiments.

AI use statement

In this work, we used generative AI tools only for proofreading and rephrasing existing text for clarity. We have not used generative AI tools for generating research ideas, designing the method, implementing or running experiments, analysing data, or producing figures, tables, or citations. All

AI-assisted edits were reviewed by the authors, and we take responsibility for the final content of this work, including text, claims, and artifacts.

Reproducibility statement

Section 2 defines the truncated quasi-metric distance, and Appendix B gives the closed-form ray intersection solution used in the implementation. Complete proofs of the gradient-boundedness results are provided in Appendices C and D. All datasets used are publicly available; Section 3 and Appendix E list the backbones, training recipes, hyperparameters, and evaluation protocols for each of the four retrieval settings.

References

Gary Becigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. In *International Conference on Learning Representations*, 2019.

Boli Chen, Xin Huang, Lin Xiao, and Liping Jing. Hyperbolic capsule networks for multi-label classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3115–3124, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.283. URL https://aclanthology.org/2020.acl-main.283/.

Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. Fully hyperbolic neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 5672–5686, Dublin, Ireland, 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.389. URL https://aclanthology.org/2022.acl-long.389/.

Shuyang Dai, Zhe Gan, Yu Cheng, Chenyang Tao, Lawrence Carin, and Jingjing Liu. APo-VAE: Text generation in hyperbolic space. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 416–431, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.36. URL https://aclanthology.org/2021.naacl-main.36/.

Karan Desai, Maximilian Nickel, Tanmay Rajpurohit, Justin Johnson, and Ramakrishna Vedantam. Hyperbolic image-text representations. In *International Conference on Machine Learning*, pp. 7694–7731. PMLR, 2023.

Aleksandr Ermolov, Leyla Mirvakhabova, Valentin Khrulkov, Nicu Sebe, and Ivan Oseledets. Hyperbolic vision transformers: Combining improvements in metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7409–7419, 2022.

Octavian-Eugen Ganea, Gary Becigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, volume 31, pp. 5345–5355, 2018.

Yunhui Guo, Xudong Wang, Yubei Chen, and Stella X. Yu. Clipped hyperbolic classifiers are super-hyperbolic classifiers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11–20, June 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Guo_Clipped_Hyperbolic_Classifiers_Are_Super-Hyperbolic_Classifiers_CVPR_2022_paper.html.

Marc Law, Renjie Liao, Jake Snell, and Richard Zemel. Lorentzian distance learning for hyperbolic representations. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3672–3681. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/law19a.html.

John M. Lee. *Introduction to Riemannian Manifolds*. Springer International Publishing, 2nd edition, 2018.

Federico López and Michael Strube. A fully hyperbolic neural model for hierarchical multi-class classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 460–475, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.42. URL https://aclanthology.org/2020.findings-emnlp.42/.

Federico López, Benjamin Heinzerling, and Michael Strube. Fine-grained entity typing in hyperbolic space. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pp. 169–180, Florence, Italy, 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-4319. URL https://aclanthology.org/W19-4319/.

Gal Mishne, Zhengchao Wan, Yusu Wang, and Sheng Yang. The numerical stability of hyperbolic representation learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 24925–24949. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/mishne23a.html.

Gabriel Moreira, Manuel Marques, João Paulo Costeira, and Alexander Hauptmann. Hyperbolic vs Euclidean embeddings in few-shot learning: Two sides of the same coin. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2082–2090, January 2024. URL https://openaccess.thecvf.com/content/WACV2024/html/Moreira_Hyperbolic_vs_Euclidean_Embeddings_in_Few-Shot_Learning_Two_Sides_of_WACV_2024_paper.html.

Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6338–6347, 2017.

Zhe Pan and Peng Wang. Hyperbolic hierarchy-aware knowledge graph embedding for link prediction. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 2941–2948, Punta Cana, Dominican Republic, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-emnlp.251. URL https://aclanthology.org/2021.findings-emnlp.251/.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021.

Ramit Sawhney, Shrey Pandit, Vishwa Shah, Megh Thakkar, and Shafiq Joty. AdaPT: A set of guidelines for hyperbolic multimodal multilingual NLP. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 1757–1771, Mexico City, Mexico, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-naacl.114. URL https://aclanthology.org/2024.findings-naacl.114/.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, et al. DINOv3. arXiv preprint arXiv:2508.10104, 2025. URL https://arxiv.org/abs/2508.10104.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Jun Wen, Yufeng Chen, Ruiqi Shi, Wei Ji, Menglin Yang, Difei Gao, Junsong Yuan, and Roger Zimmermann. HOVER: Hyperbolic video-text retrieval. *IEEE Transactions on Image Processing*, 2025.

Tao Yu and Christopher M. De Sa. Numerically accurate hyperbolic embeddings using tiling-based models. In *Advances in Neural Information Processing Systems*, volume 32, pp. 2023–2033, 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/82c2559140b95ccda9c6ca4a8b981f1e-Abstract.html.

A Preliminaries on Poincaré Geometry

This section collects the standard facts about the generalized Poincaré ball used in Section 2. We use curvature parameter $c > 0$, so the sectional curvature is $-c$, and work in ambient dimension d.

Ball model and conformal metric. The generalized Poincaré ball is

$$\mathbb{D}_c^d = \{\mathbf{x} \in \mathbb{R}^d : c\|\mathbf{x}\|_2^2 < 1\}. \tag{12}$$

The Poincaré metric is conformal to the Euclidean metric,

$$g_{\mathbf{x}}^c = (\lambda_{\mathbf{x}}^c)^2 g^{\mathrm{E}}, \qquad \lambda_{\mathbf{x}}^c = \frac{2}{1 - c\|\mathbf{x}\|_2^2}, \tag{13}$$

so the conformal factor diverges as $\|\mathbf{x}\|_2 \to 1/\sqrt{c}$. This boundary divergence is another source of the numerical sensitivity targeted by our truncation: even before considering the closed-form distance denominator below, the local metric scaling itself amplifies perturbations near the ball boundary.

Möbius addition and distance. For $\mathbf{x}, \mathbf{y} \in \mathbb{D}_c^d$, Möbius addition is

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c\|\mathbf{y}\|_2^2)\mathbf{x} + (1 - c\|\mathbf{x}\|_2^2)\mathbf{y}}{1 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + c^2\|\mathbf{x}\|_2^2\|\mathbf{y}\|_2^2}. \tag{14}$$

The geodesic distance can be written as

$$d_{\mathbb{D}}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} \operatorname{artanh}\left(\sqrt{c}\|(-\mathbf{x}) \oplus_c \mathbf{y}\|_2\right), \tag{15}$$

or equivalently,

$$d_{\mathbb{D}}(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{c}} \log\left(\frac{1 + \sqrt{c}\|(-\mathbf{x}) \oplus_c \mathbf{y}\|_2}{1 - \sqrt{c}\|(-\mathbf{x}) \oplus_c \mathbf{y}\|_2}\right). \tag{16}$$

The denominator in Eq. (16) approaches zero as $\sqrt{c}\|(-\mathbf{x}) \oplus_c \mathbf{y}\|_2 \to 1$, so small floating-point perturbations near the ball boundary are amplified by the logarithm and its gradient. Geodesics are Euclidean diameters or circular arcs orthogonal to the boundary sphere, so except for radial cases they are not straight Euclidean segments.

Origin-centered maps. To interface Euclidean backbone features with Poincaré geometry, we use the tangent space at the origin. The exponential and logarithmic maps are

$$\begin{aligned} \exp_0^c(\mathbf{v}) &= \frac{\tanh(\sqrt{c}\|\mathbf{v}\|_2)}{\sqrt{c}\|\mathbf{v}\|_2}\mathbf{v}, \\ \log_0^c(\mathbf{x}) &= \frac{\operatorname{artanh}(\sqrt{c}\|\mathbf{x}\|_2)}{\sqrt{c}\|\mathbf{x}\|_2}\mathbf{x}. \end{aligned} \tag{17}$$

This is the lift-and-project pipeline adopted by our retrieval model: normalize or scale embeddings in tangent space, then map them into the ball before computing similarity.

Why truncation helps. The conformal factor and distance denominator both deteriorate near the boundary, which is exactly where standard Poincaré optimization often accumulates numerical error. Our method preserves the inner Poincaré geometry up to radius r_0 and switches to scaled Euclidean geometry outside that radius. Once the shortest path crosses the truncation boundary, the exact hybrid geodesic becomes a refraction-style boundary problem; Section 2 therefore introduces a ray intersection approximation that keeps this transition closed-form and efficient.

B Closed Form for Ray Intersection Approximation

For a cross-boundary pair with $\|\mathbf{x}\|_2 \le r_0$ and $\|\mathbf{y}\|_2 > r_0$, the ray intersection point used in Section 2 lies on the Euclidean segment from $\mathbf{x}$ to $\mathbf{y}$:

$$\mathbf{p} = \mathbf{x} + t^*(\mathbf{y} - \mathbf{x}), \qquad t^* \in [0, 1], \qquad \|\mathbf{p}\|_2 = r_0. \tag{18}$$

Let $\mathbf{v} = \mathbf{y} - \mathbf{x}$. Substituting $\mathbf{p} = \mathbf{x} + t\mathbf{v}$ into $\|\mathbf{p}\|_2^2 = r_0^2$ gives the scalar quadratic

$$\|\mathbf{v}\|_2^2 t^2 + 2\langle \mathbf{x}, \mathbf{v}\rangle t + \left(\|\mathbf{x}\|_2^2 - r_0^2\right) = 0. \tag{19}$$

The root on the segment is

$$t^* = \frac{-\langle \mathbf{x}, \mathbf{v}\rangle + \sqrt{\langle \mathbf{x}, \mathbf{v}\rangle^2 - \|\mathbf{v}\|_2^2(\|\mathbf{x}\|_2^2 - r_0^2)}}{\|\mathbf{v}\|_2^2}. \tag{20}$$

When $\|\mathbf{x}\|_2 > r_0$ and $\|\mathbf{y}\|_2 \le r_0$, the same expression is applied after swapping the roles of $\mathbf{x}$ and $\mathbf{y}$.

C Proof of General Gradient Boundedness

This section provides the full proof of Theorem 1.

Proof. By the fundamental properties of Riemannian geometry (Lee, 2018), the exact Poincaré distance $f(\mathbf{y}) = d_c(\mathbf{x}, \mathbf{y})$ satisfies the Eikonal equation on the manifold. Consequently, the Riemannian norm of its Riemannian gradient is identically one, i.e., $\|\nabla_{\mathbf{y}}^R f(\mathbf{y})\|_R \equiv 1$. By conformal metric equivalence (Becigneul & Ganea, 2019), the Euclidean gradient norm is scaled precisely by the conformal factor $\lambda_c(\mathbf{y}) = \frac{2}{1-c\|\mathbf{y}\|_2^2}$:

$$\|\nabla_{\mathbf{y}} f(\mathbf{y})\|_2 = \lambda_c(\mathbf{y})\|\nabla_{\mathbf{y}}^R f(\mathbf{y})\|_R = \frac{2}{1 - c\|\mathbf{y}\|_2^2}. \tag{21}$$

Substituting the boundary proximity condition $\sqrt{c}\|\mathbf{y}\|_2 = 1 - \delta$, we obtain $c\|\mathbf{y}\|_2^2 = (1-\delta)^2 = 1 - 2\delta + \delta^2$. Thus,

$$\begin{aligned}\|\nabla_{\mathbf{y}} f(\mathbf{y})\|_2 &= \frac{2}{1 - (1 - 2\delta + \delta^2)} = \frac{2}{2\delta - \delta^2} \\ &= \frac{1}{\delta} + \mathcal{O}(1).\end{aligned} \tag{22}$$

Therefore, regardless of the reference point $\mathbf{x}$, as k increases and the point approaches the boundary, the Euclidean gradient of the exact Poincaré distance unconditionally grows at the rate 10^k, which can trigger overflow or NaNs in finite-precision backpropagation.

In contrast, for the proposed truncated approximation $g(\mathbf{y})$, since $\|\mathbf{y}\|_2 > r_0$, the intersection point $\mathbf{p}$ on the ray from the origin to $\mathbf{y}$ is $\mathbf{p} = r_0 \frac{\mathbf{y}}{\|\mathbf{y}\|_2}$. The generalized RIA quasi-metric distance is defined as

$$g(\mathbf{y}) = d_c(\mathbf{x}, \mathbf{p}) + K_c\|\mathbf{y} - \mathbf{p}\|_2, \tag{23}$$

where $K_c = \frac{2}{1-cr_0^2}$ is the conformal factor at the truncation boundary. Applying the chain rule to compute the Euclidean gradient with respect to $\mathbf{y}$ yields

$$\nabla_{\mathbf{y}} g(\mathbf{y}) = \nabla_{\mathbf{y}} d_c(\mathbf{x}, \mathbf{p}) + K_c \frac{\mathbf{y}}{\|\mathbf{y}\|_2}. \tag{24}$$

Since $\mathbf{p}$ is strictly bounded within the sphere of radius $r_0 < 1/\sqrt{c}$, the distance $d_c(\mathbf{x}, \mathbf{p})$ is evaluated in a region where the conformal factor is bounded. Furthermore, the Jacobian of the projection $\mathbf{p}(\mathbf{y})$ with respect to $\mathbf{y}$ scales as $r_0/\|\mathbf{y}\|_2$, which is finite. Therefore, the gradient magnitude $\|\nabla_{\mathbf{y}} d_c(\mathbf{x}, \mathbf{p})\|_2$ is bounded by some constant M. By the triangle inequality, we have

$$\begin{aligned}\|\nabla_{\mathbf{y}} g(\mathbf{y})\|_2 &\le \|\nabla_{\mathbf{y}} d_c(\mathbf{x}, \mathbf{p})\|_2 + K_c \\ &\le M + K_c = \Theta(1).\end{aligned} \tag{25}$$

Since r_0 is a fixed hyperparameter strictly less than $1/\sqrt{c}$, the total gradient magnitude $\|\nabla_{\mathbf{y}} g(\mathbf{y})\|_2$ is constrained by a constant $\Theta(1)$, strictly independent of the proximity to the boundary δ. □

D SPECIAL CASE: ORIGIN REFERENCE POINT

The general result in Theorem 1 can be understood through the simpler case $c = 1$ and $\mathbf{x} = \mathbf{0}$, where the calculation is fully explicit.

Theorem 2 (Origin-Reference Gradient Boundedness)**.** *Let $c = 1$ and let $\delta = 10^{-k}$ be a small positive number representing the distance to the ideal boundary. Consider a reference point at the origin $\mathbf{x} = \mathbf{0}$, and a moving feature point $\mathbf{y} \in \mathbb{D}^d$ approaching the boundary such that $\|\mathbf{y}\|_2 = 1 - \delta$. Let $f(\mathbf{y}) = d(\mathbf{0}, \mathbf{y})$ denote the exact Poincaré distance, and let $g(\mathbf{y}) = d_{ria}(\mathbf{0}, \mathbf{y})$ denote the truncated quasi-metric distance with truncation radius $r_0 < 1 - \delta$. Then,*

$$\|\nabla_{\mathbf{y}} f(\mathbf{y})\|_2 = \Theta\left(\frac{1}{\delta}\right) = \Theta(10^k), \tag{26}$$

$$\|\nabla_{\mathbf{y}} g(\mathbf{y})\|_2 = K = \frac{2}{1 - r_0^2} = \Theta(1). \tag{27}$$

Proof. The exact Poincaré distance from the origin is

$$f(\mathbf{y}) = \operatorname{arcosh}\left(1 + \frac{2\|\mathbf{y}\|_2^2}{1 - \|\mathbf{y}\|_2^2}\right) \tag{28}$$

$$= 2 \operatorname{arctanh}(\|\mathbf{y}\|_2). \tag{29}$$

Computing the Euclidean gradient of $f(\mathbf{y})$ and substituting $\|\mathbf{y}\|_2 = 1 - \delta$ gives

$$\nabla_{\mathbf{y}} f(\mathbf{y}) = \frac{2}{1 - \|\mathbf{y}\|_2^2} \frac{\mathbf{y}}{\|\mathbf{y}\|_2}, \tag{30}$$

$$\|\nabla_{\mathbf{y}} f(\mathbf{y})\|_2 = \frac{2}{1 - (1 - \delta)^2} = \frac{2}{2\delta - \delta^2} = \frac{1}{\delta} + \mathcal{O}(1). \tag{31}$$

Therefore, as k increases and the point approaches the boundary, the gradient of the exact Poincaré distance grows at the rate 10^k.

In contrast, for the proposed truncated approximation $g(\mathbf{y})$, since $\|\mathbf{y}\|_2 > r_0$, the intersection point on the ray from $\mathbf{0}$ to $\mathbf{y}$ is $\mathbf{p} = r_0 \mathbf{y}/\|\mathbf{y}\|_2$. By the definition of d_{ria},

$$g(\mathbf{y}) = d(\mathbf{0}, \mathbf{p}) + K\|\mathbf{y} - \mathbf{p}\|_2. \tag{32}$$

Since $d(\mathbf{0}, \mathbf{p}) = 2 \operatorname{arctanh}(r_0)$ is constant with respect to $\|\mathbf{y}\|_2$ and $\|\mathbf{y} - \mathbf{p}\|_2 = \|\mathbf{y}\|_2 - r_0$, this simplifies to

$$g(\mathbf{y}) = 2 \operatorname{arctanh}(r_0) + K(\|\mathbf{y}\|_2 - r_0), \tag{33}$$

$$\nabla_{\mathbf{y}} g(\mathbf{y}) = K \frac{\mathbf{y}}{\|\mathbf{y}\|_2}, \tag{34}$$

$$\|\nabla_{\mathbf{y}} g(\mathbf{y})\|_2 = K = \frac{2}{1 - r_0^2}. \tag{35}$$

Since $r_0 < 1$ is fixed, the gradient magnitude is constant and independent of the boundary proximity δ. □

E EXPERIMENT DETAILS

E.1 DETAILED IMAGE RETRIEVAL

For image retrieval, we follow a single-GPU fine-tuning recipe based on a shared ViT-S/16 DINOv3 backbone (Siméoni et al., 2025) using `timm` library. In all experiments, we fine-tune only the last two transformer blocks for one epoch, using AdamW with learning rate 1×10^{-5}, weight decay 0.05, batch size 64, evaluation batch size 128, and 4 dataloader workers. The temperature is initialized at 0.07 and optimized jointly with the backbone within the range $[10^{-3}, 1.0]$. Images are processed with the model default evaluation pipeline at 256×256 resolution with bicubic interpolation, center crop, and ImageNet normalization.

Retrieval supervision is built from label-defined positive sets: Stanford Cars and CUB-200-2011 use class-label retrieval, Stanford Online Products (SOP) uses the MTEB query/corpus/qrels protocol, and DeepFashion In-Shop uses the repository's single-split item_ID-based protocol. We compare DINO/cosine, Euclidean, Poincaré-ViT, Lorentz, and Ours (truncated-Poincaré) similarities while keeping the backbone, optimizer, training schedule, and data processing fixed. For geometry-aware methods, embeddings are normalized in the tangent space and then mapped to the target geometry with dataset-specific curvature and scale settings.

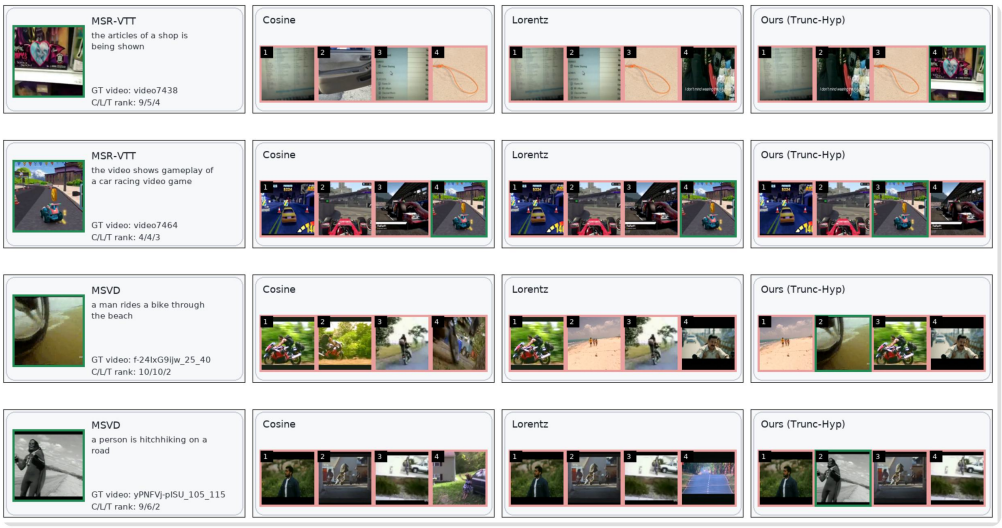


Figure 7: Qualitative video retrieval examples on MSR-VTT and MSVD. For each text query, the ground-truth video is shown on the left, followed by the top retrieved candidates from cosine, Lorentz, and our truncated-Poincaré similarity. Green boxes indicate correct matches and red boxes indicate mismatches, showing that the proposed geometry can rank the target video higher in representative cases.

E.2 Detailed Cross-Modal Retrieval

Experimental setting. We use `openai/clip-vit-base-patch32` for cross-modal retrieval, training on `pingzhili/llava-filtered-cc3m-595k` and evaluating on `jxie/coco_captions:validation` and `lmms-lab/flickr30k:test`. Retrieval is evaluated in both text-to-image (T2I) and image-to-text (I2T) directions. We report R@1/R@5/R@10 and compare Euclidean, cosine, Poincaré, Lorentz, and Ours (truncated-Poincaré) similarities under the fixed partial cross-modal protocol.

Qualitative examples. Figure 8 shows qualitative Flickr30K cross-modal retrieval examples, complementing Table 6 with case-level behavior under different geometries.

E.3 Detailed Video Retrieval

Experimental setting. We formulate video retrieval as text-to-video retrieval from cached CLIP features rather than end-to-end video backbone fine-tuning. We use `openai/clip-vit-base-patch32` with 4 frames per video and train lightweight text/video projectors on top of cached embeddings. Model selection uses mAP. We report mAP, MRR, and Hits@{1,5,10}. Compared similarities are cosine, Euclidean, Poincaré, Lorentz, and Ours (truncated-Poincaré).

Table 7 lists the selected configurations; both use tangent normalization.



Figure 8: Cross-modal qualitative retrieval cases. Example-level T2I/I2T retrievals illustrate successes and hard negatives.

Table 7: Best truncated-Poincaré configurations for video retrieval.

Dataset	Epochs	LR	WD	c	r_0	Scale	Cosine mix
MSR-VTT	12	2.4×10^{-4}	10^{-4}	0.08	0.30	0.24	2.5
MSVD	12	2.3×10^{-4}	10^{-3}	0.08	0.18	0.20	2.5

Notes. Truncated-Poincaré is the strongest method on both video datasets. For MSR-VTT, gains come from a small optimization adjustment over the original truncation setup while keeping geometry fixed. For MSVD, performance was already nearly tied and lowering truncation radius to $r_0 = 0.18$ is sufficient to move truncated-Poincaré to the top. The current implementation performs projector fine-tuning on cached CLIP text/video features rather than end-to-end video backbone fine-tuning.

Qualitative examples. Figure 7 visualizes representative text-to-video retrieval cases on MSR-VTT and MSVD, illustrating how geometry-aware similarities change the top-ranked video neighborhoods.

E.4 Detailed Text Retrieval

Experimental setting. We use a DPR-style dual encoder with `bert-base-uncased` and single-GPU training. The shared optimizer recipe is AdamW with `lr=2e-5` and `weight_decay=0.01`, 2 dataloader workers, and a learnable temperature initialized at 0.07. The original compare-all runs use training batch size 24 and evaluation batch size 64; tuned truncated-Poincaré runs use training batch size 32 and evaluation batch size 128. We report mAP, MRR, MRR@10, NDCG@10, Hits@{1,5,10}, and mean rank where available.