Beyond Euclidean - Hyperbolic Representation Learning

Joakim Bruslund Haurum
Assistant Professor
Center for Software Technology - SDU Vejle

SDU

Isn’t the world Euclidean?

Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

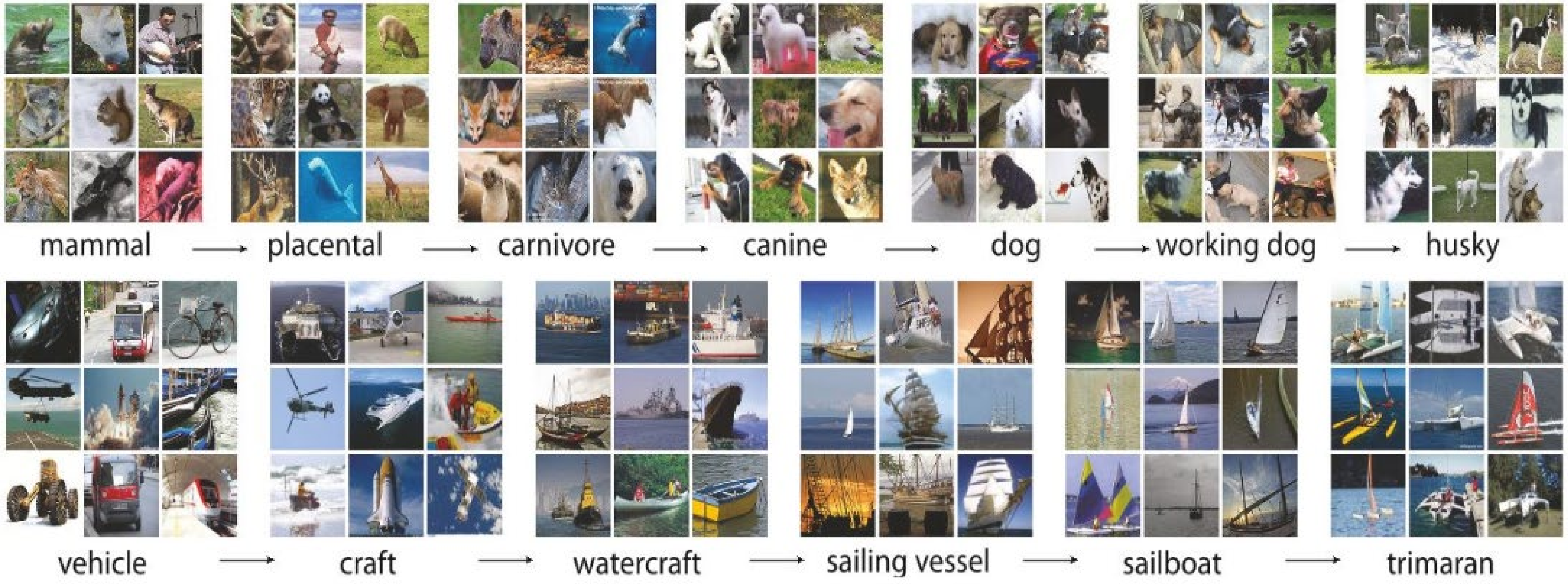
→ Vast majority of Neural Networks are implemented in Euclidean Geometry

Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

→ Vast majority of Neural Networks are implemented in Euclidean Geometry

→ However, our domain knowledge / data is not Euclidean. It is hierarichal!



Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

→ Vast majority of Neural Networks are implemented in Euclidean Geometry

→ However, our domain knowledge / data is not Euclidean. It is hierarichal!

→ But Euclidean Spaces don’t embed hiearchies/tree graphs nicely

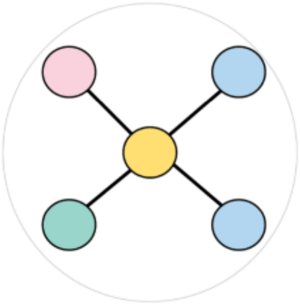
Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

→ Vast majority of Neural Networks are implemented in Euclidean Geometry

→ However, our domain knowledge / data is not Euclidean. It is hierarichal!

→ But Euclidean Spaces don’t embed hiearchies/tree graphs nicely



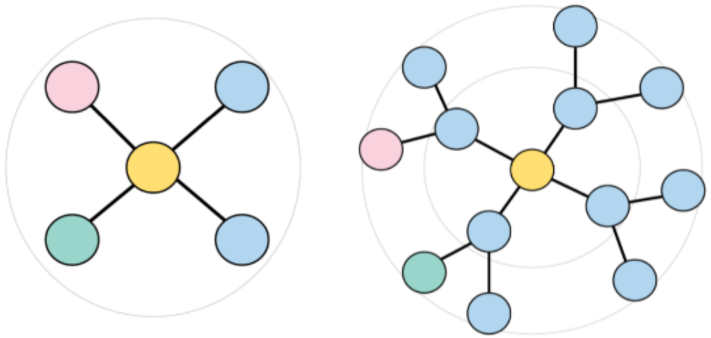
Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

→ Vast majority of Neural Networks are implemented in Euclidean Geometry

→ However, our domain knowledge / data is not Euclidean. It is hierarichal!

→ But Euclidean Spaces don’t embed hiearchies/tree graphs nicely



Constant Curvature Graph Convolutional Networks, Bachmann et al. 2020

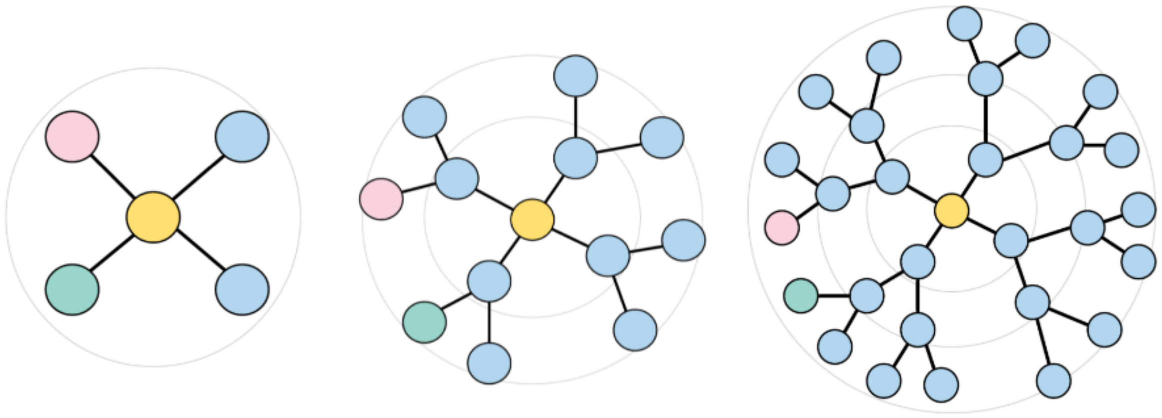
Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

→ Vast majority of Neural Networks are implemented in Euclidean Geometry

→ However, our domain knowledge / data is not Euclidean. It is hierarichal!

→ But Euclidean Spaces don’t embed hiearchies/tree graphs nicely



Constant Curvature Graph Convolutional Networks, Bachmann et al. 2020

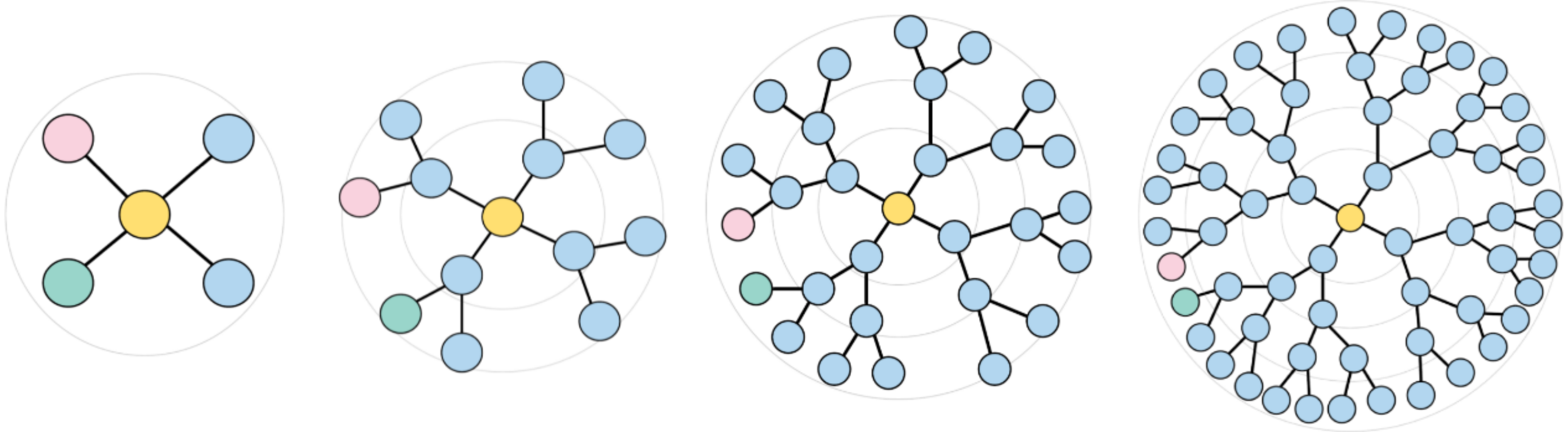
Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

→ Vast majority of Neural Networks are implemented in Euclidean Geometry

→ However, our domain knowledge / data is not Euclidean. It is hierarichal!

→ But Euclidean Spaces don’t embed hiearchies/tree graphs nicely



Constant Curvature Graph Convolutional Networks, Bachmann et al. 2020

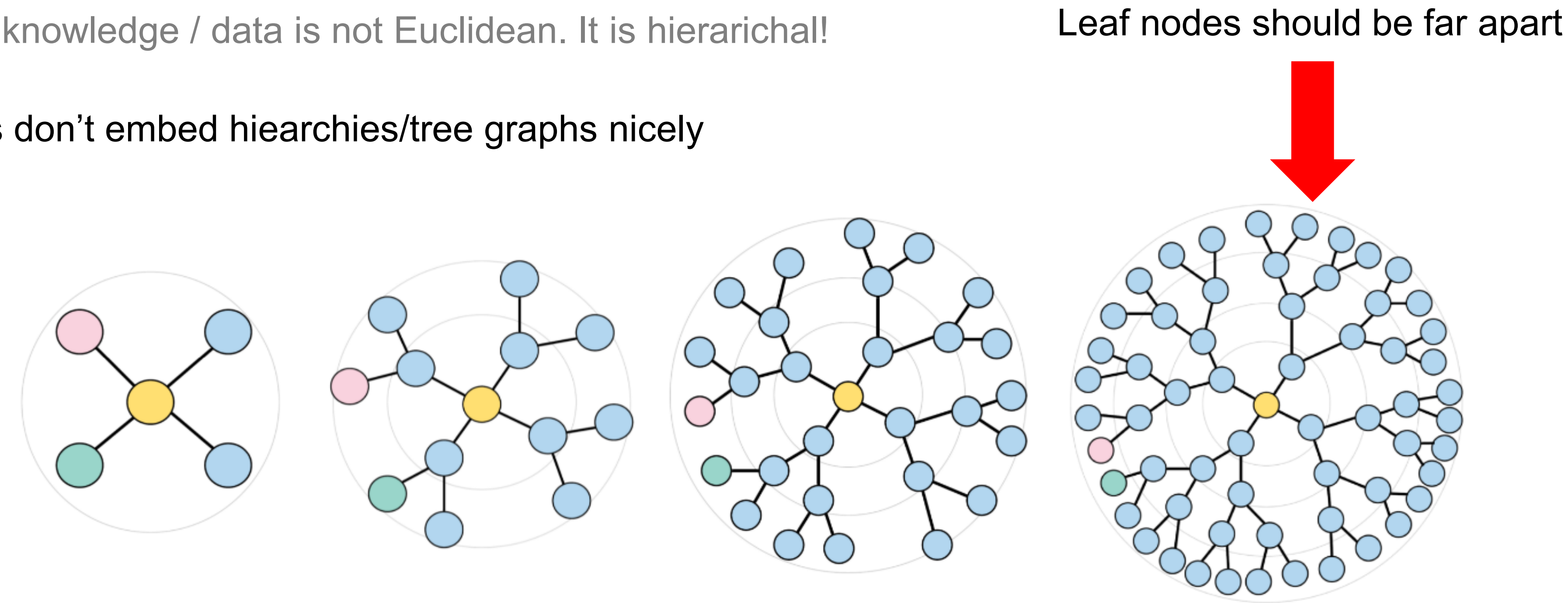
Isn’t the world Euclidean?

→ Euclidean Geometry is the ”default” space for most settings

→ Vast majority of Neural Networks are implemented in Euclidean Geometry

→ However, our domain knowledge / data is not Euclidean. It is hierarichal!

→ But Euclidean Spaces don’t embed hiearchies/tree graphs nicely



Constant Curvature Graph Convolutional Networks, Bachmann et al. 2020

Hyperbolic Geometry

Euclid’s postulates

Euclid’s postulates

1. To draw a straight line from any point to any point.
2. To produce (extend) a finite straight line continuously in a straight line.
3. To describe a circle with any centre and distance (radius).
4. That all right angles are equal to one another.

Euclid's postulates

1. To draw a straight line from any point to any point.
2. To produce (extend) a finite straight line continuously in a straight line.
3. To describe a circle with any centre and distance (radius).
4. That all right angles are equal to one another.
5. That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

Euclid's postulates

1. To draw a straight line from any point to any point.
2. To produce (extend) a finite straight line continuously in a straight line.
3. To describe a circle with any centre and distance (radius).
4. That all right angles are equal to one another.
5. That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

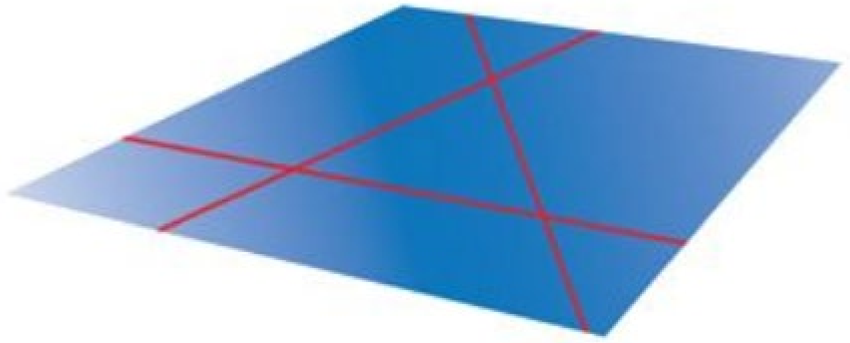
Huh?

Euclid's postulates

1. To draw a straight line from any point to any point.
2. To produce (extend) a finite straight line continuously in a straight line.
3. To describe a circle with any centre and distance (radius).
4. That all right angles are equal to one another.
5. That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

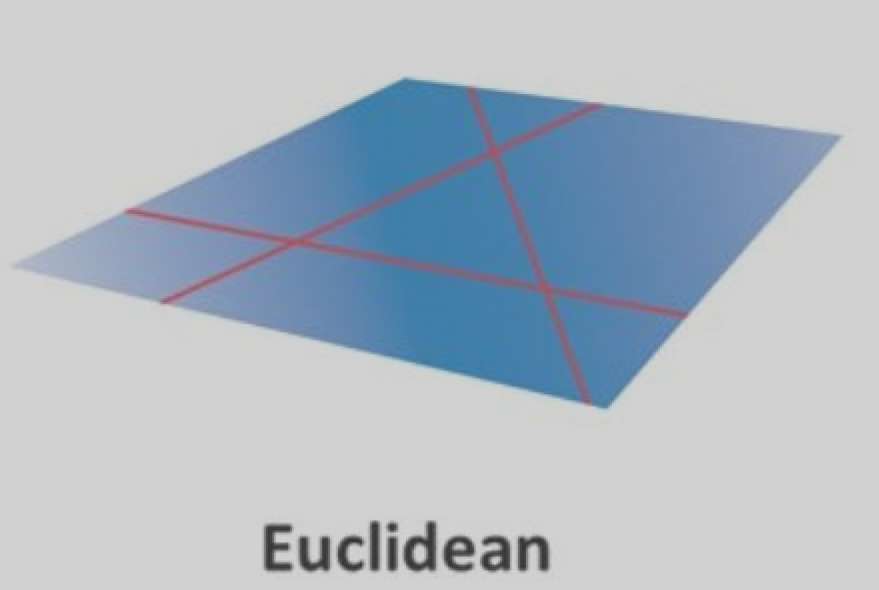
What if this is not true?

What does Non-Euclidean Geometry look like?



Euclidean

Constant 0 curvature
Exactly one set of parallel lines



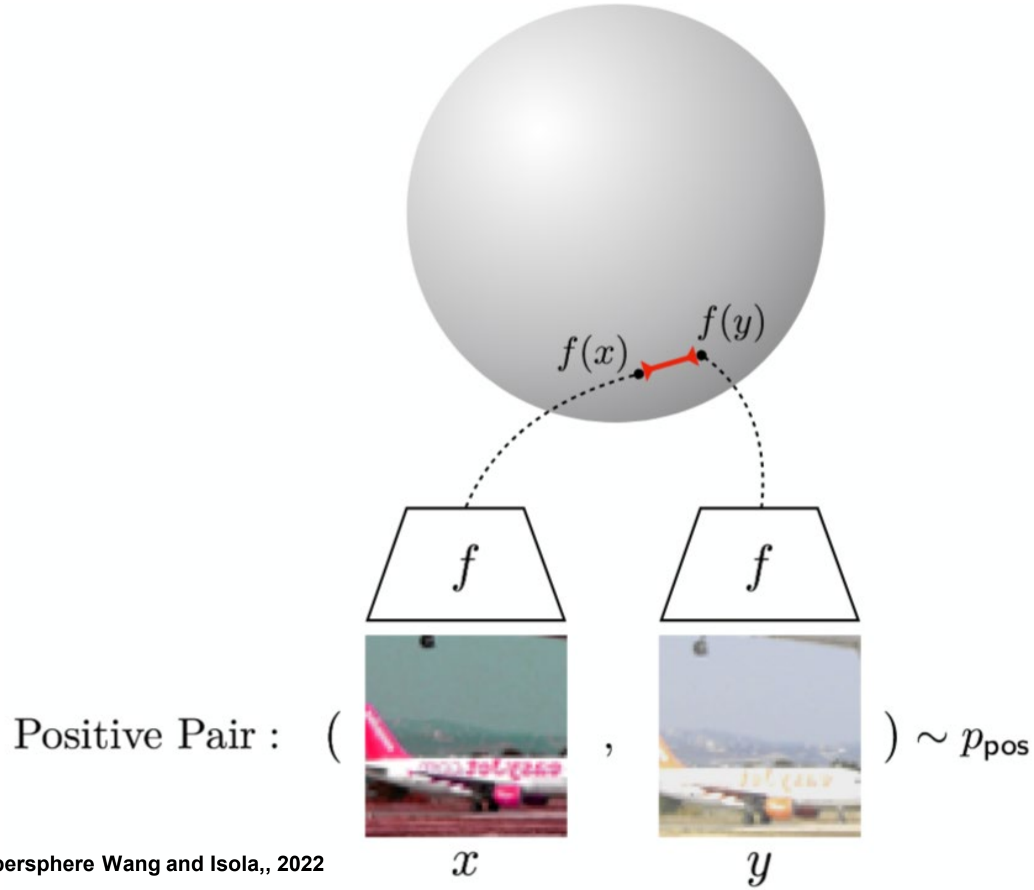
Constant 0 curvature
Exactly one set of parallel lines

Constant positive curvature
No sets of parallel lines

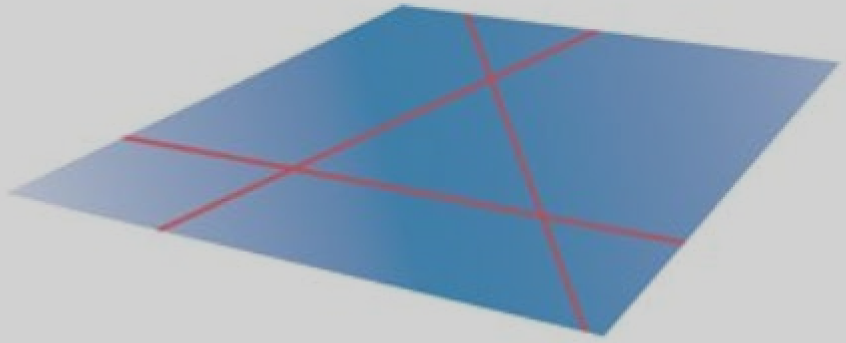
Contrastive Learning is Hyperspherical Learning

→ Contrastive Learning is often applied by measuring cosine similarity of unit vectors

→ All embeddings lie on a unit hypersphere



Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere Wang and Isola,, 2022

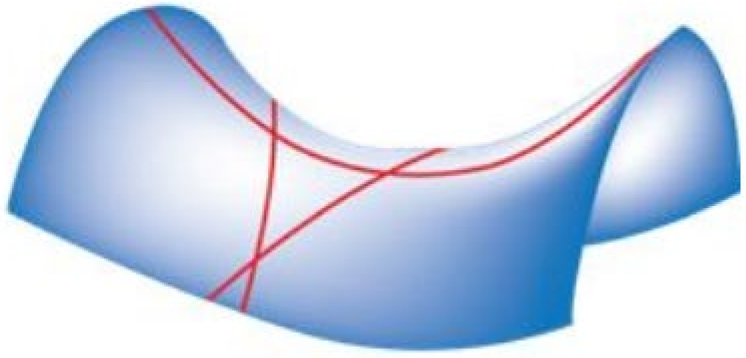


Euclidean

Constant 0 curvature
Exactly one set of parallel lines

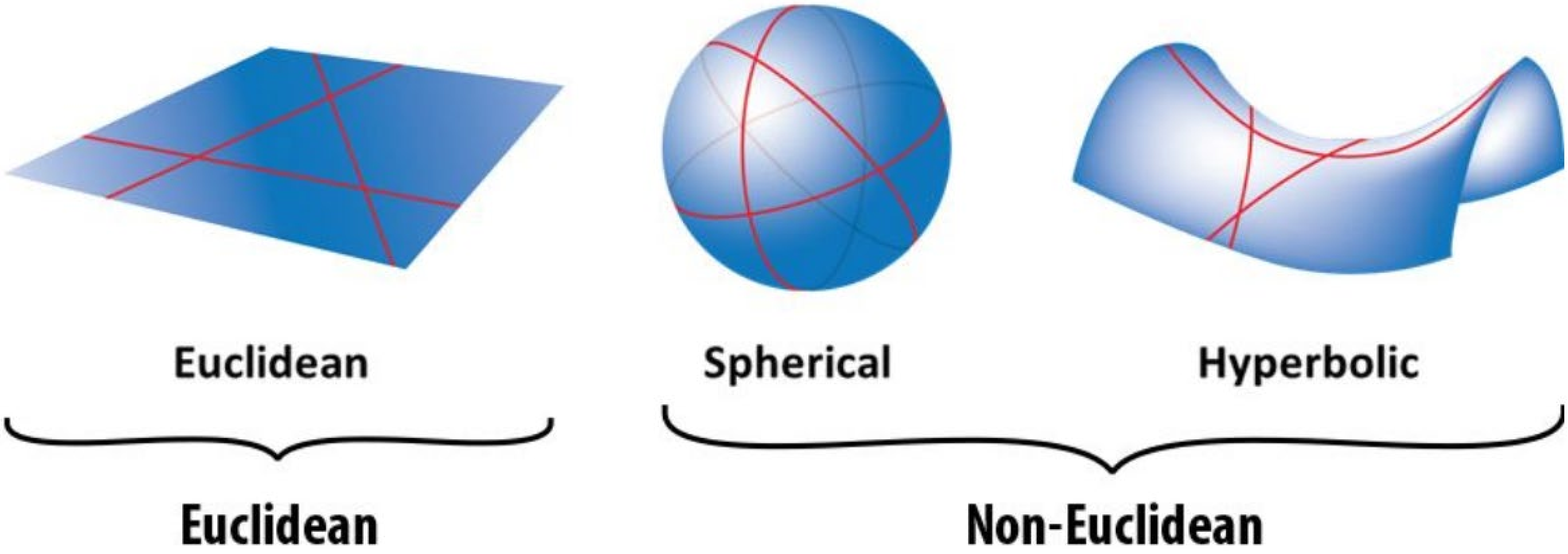
Spherical

Constant positive curvature
No sets of parallel lines



Hyperbolic

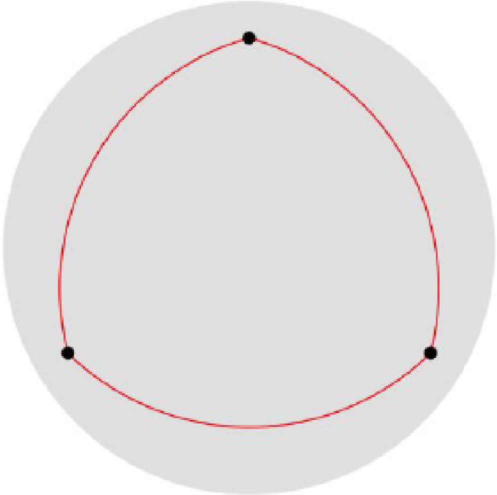
Constant negative curvature
Infinite sets of parallel lines


Euclidean
Spherical
Hyperbolic
Euclidean
Non-Euclidean

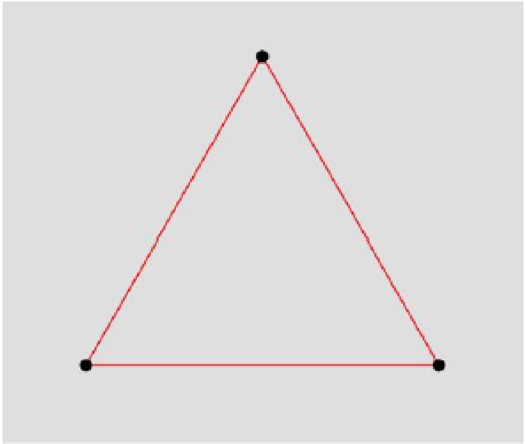
But why is Hyperbolic Geometry interesting?

But why is Hyperbolic Geometry interesting?

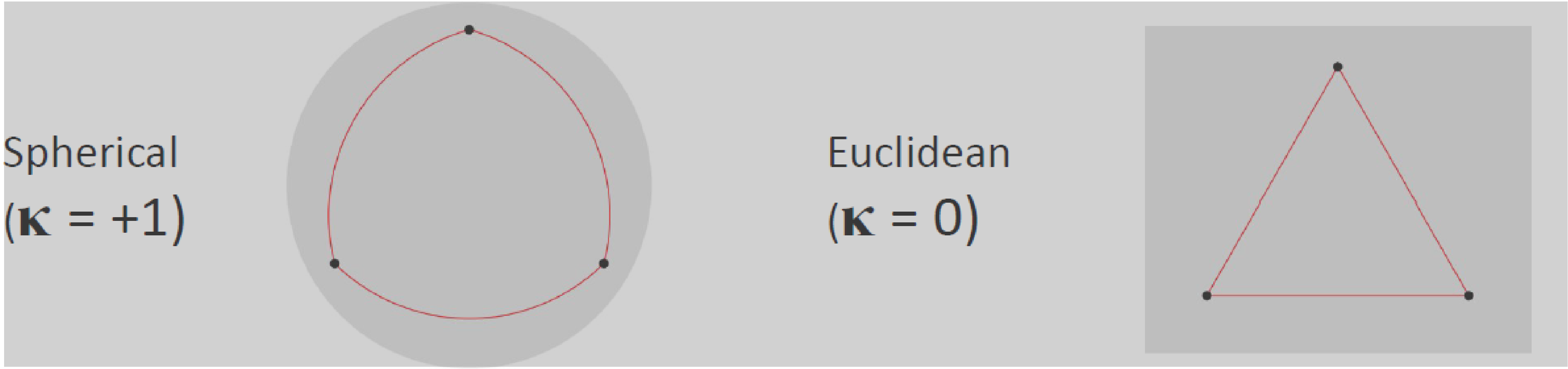
Spherical
($\kappa = +1$)



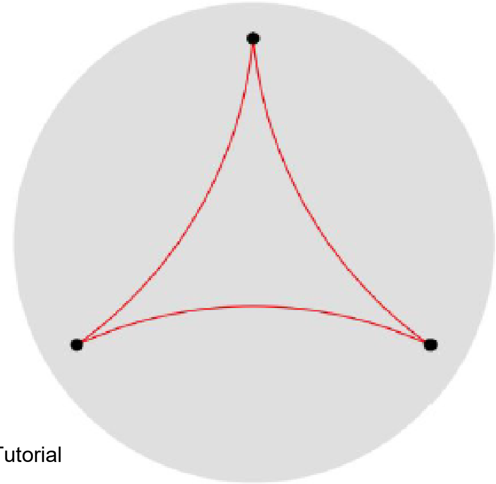
Euclidean
($\kappa = 0$)



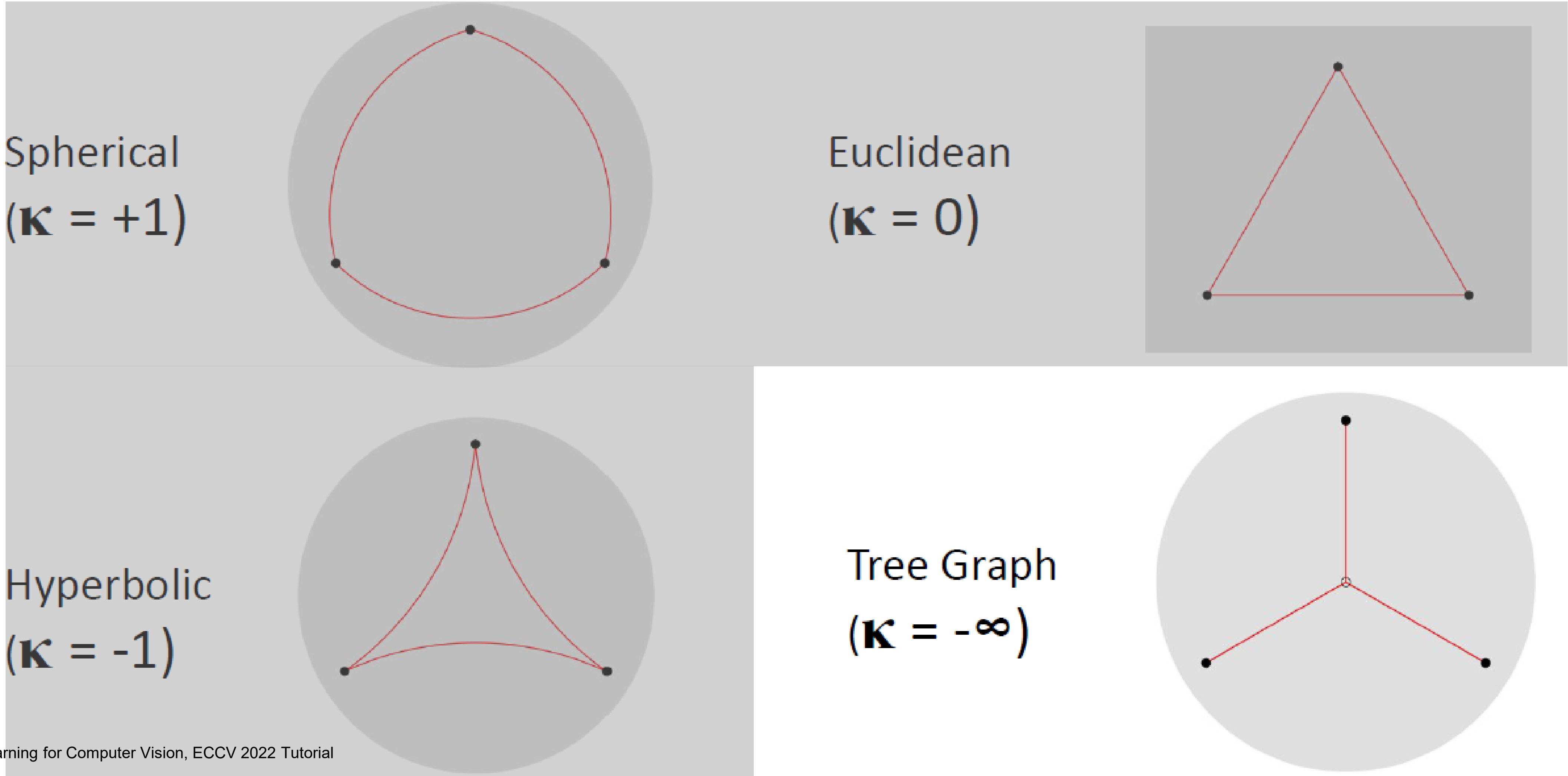
But why is Hyperbolic Geometry interesting?



Hyperbolic
(κ = -1)



But why is Hyperbolic Geometry interesting?



How do we represent Hyperbolic spaces?

How do we represent Hyperbolic spaces?

→ There are several options!

→ The two most commonly used are:

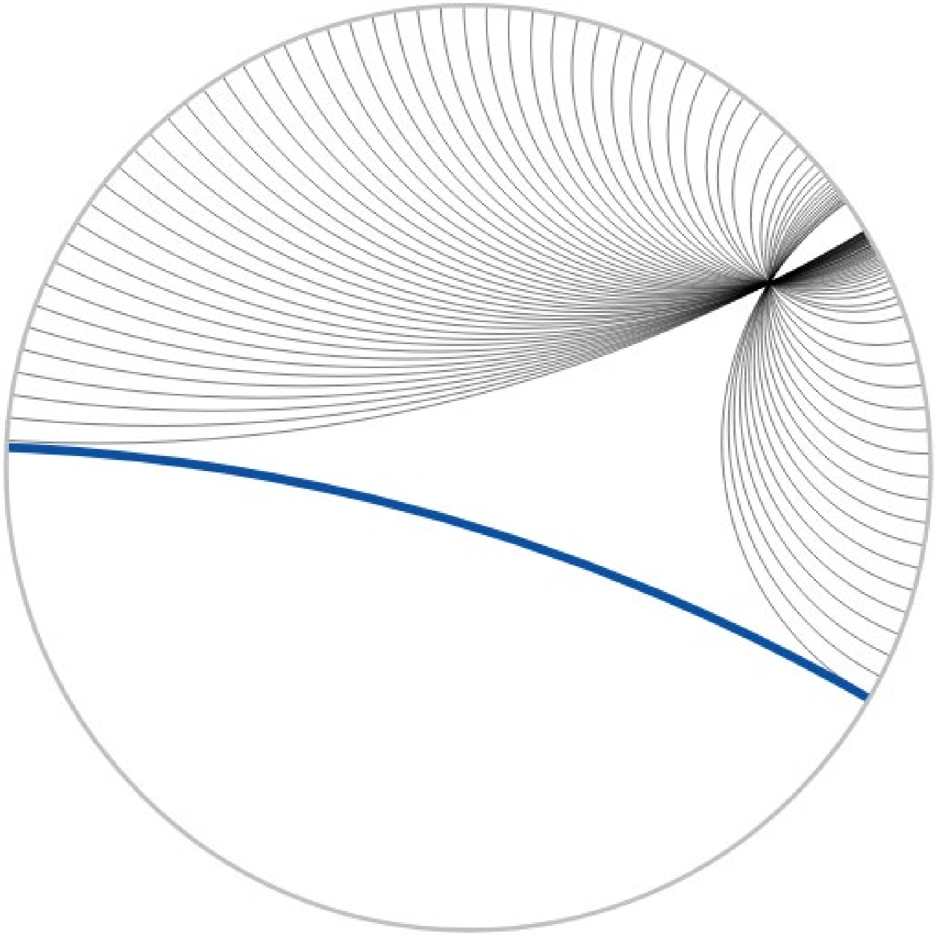
How do we represent Hyperbolic spaces?

→ There are several options!

→ The two most commonly used are:

→ The **Poincaré disk model**
- → Conformal to Euclidean space
- → Restricted to a unit disc

$$\mathbb{D}_d = \{p \in \mathbb{R}^d : p_1^2 + \cdots + p_d^2 < 1\}$$



How do we represent Hyperbolic spaces?

→ There are several options!

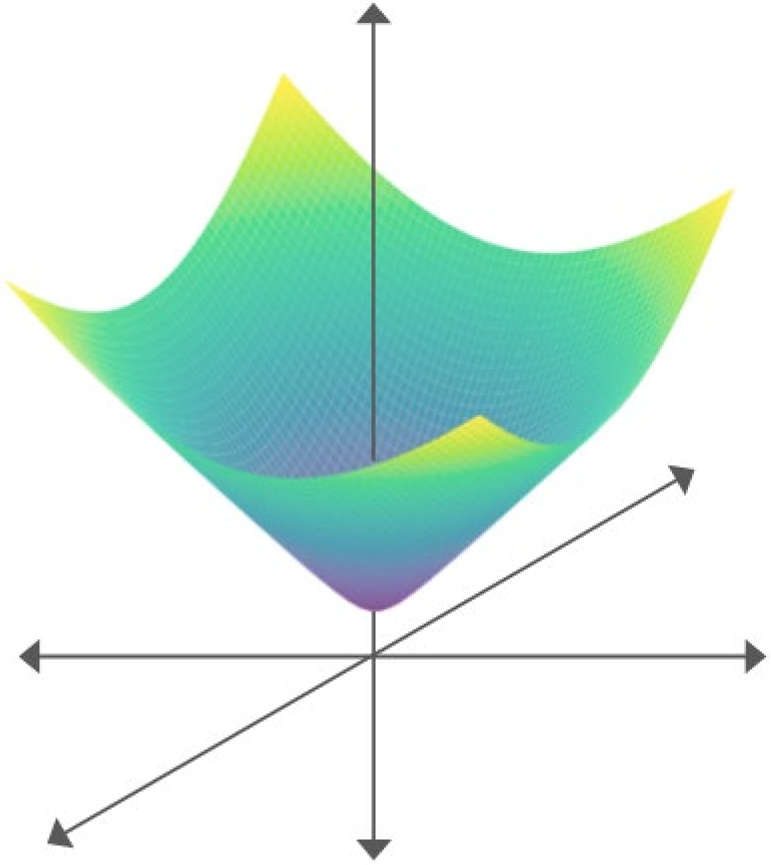
→ The two most commonly used are:

→ The **Poincaré disk model**

→ The **Hyperboloid / Lorentzian model**

  → Embeds a d-dimensional Euclidean space into a d+1 hyperbolic space

$$\mathbb{H}_d = \{x \in \mathbb{R}^{d+1} : x_0^2 - \left(x_1^2 + \cdots + x_d^2\right) = 1, x_0 > 0\}$$



Key concepts of Hyperbolic Geometry

Key concepts of Hyperbolic Geometry

Distance: Measure how far two points and are from each other in hyperbolic space.

Key concepts of Hyperbolic Geometry

Distance: Measure how far two points and are from each other in hyperbolic space.

Poincaré

$$d_{\mathbb{D}}(p,q) = \frac{1}{\sqrt{\kappa}} \operatorname{arcosh}\left(1 + \frac{2|p-q|^2}{(1-|p|^2)(1-|q|^2)}\right)$$

Lorentz

$$d_{\mathbb{H}}(x,y) = \frac{1}{\sqrt{\kappa}} \operatorname{arcosh}\left(x \circ y\right)$$

$$x \circ y = x_0 y_0 - (x_1 y_1 + \cdots + x_d y_d)$$

Key concepts of Hyperbolic Geometry

Distance: Measure how far two points and are from each other in hyperbolic space.

Geodesic arc: The shortest (=distance-minimizing) curve from x to y.

Geodesic: A geodesic arc, extended as far as possible.

Key concepts of Hyperbolic Geometry

Distance: Measure how far two points and are from each other in hyperbolic space.

Geodesic arc: The shortest (=distance-minimizing) curve from *x* to *y*.

Geodesic: A geodesic arc, extended as far as possible.

Exponential map: From a point *x* follow a geodesic in direction *v*, at speed *r*.

Key concepts of Hyperbolic Geometry

Distance: Measure how far two points and are from each other in hyperbolic space.

Geodesic arc: The shortest (=distance-minimizing) curve from *x* to *y*.

Geodesic: A geodesic arc, extended as far as possible.

Exponential map: From a point *x* follow a geodesic in direction *v*, at speed *r*.

This is done by mapping a tangent space vector, *v*, onto the manifold at point *x,* with unit speed.

Key concepts of Hyperbolic Geometry

Distance: Measure how far two points and are from each other in hyperbolic space.

Geodesic arc: The shortest (=distance-minimizing) curve from *x* to *y*.

Geodesic: A geodesic arc, extended as far as possible.

Exponential map: From a point *x* follow a geodesic in direction *v*, at speed *r*.

This is done by mapping a tangent space vector, *v*, onto the manifold at point *x*, with unit speed.

Poincaré

$$\exp_x^\kappa(v) = x \oplus_\kappa \left(\tanh\left(\frac{\sqrt{\kappa}\,||v||}{1-\kappa||x||^2}\right) \frac{v}{\sqrt{\kappa}\,||v||} \right)$$

Lorentz

$$\exp_x^\kappa(v) = \cosh\left(\sqrt{\kappa}\,||v||_L\right) x + \frac{\sinh\left(\sqrt{\kappa}\,||v||_L\right)}{\sqrt{\kappa}\,||v||_L} v$$

Key concepts of Hyperbolic Geometry

Distance: Measure how far two points and are from each other in hyperbolic space.

Geodesic arc: The shortest (=distance-minimizing) curve from *x* to *y*.

Geodesic: A geodesic arc, extended as far as possible.

Exponential map: From a point *x* follow a geodesic in direction *v*, at speed *r*.

This is done by mapping a tangent space vector, *v*, onto the manifold at point *x,* with unit speed.

This is how you map from Euclidean spaces to Hyperbolic space!

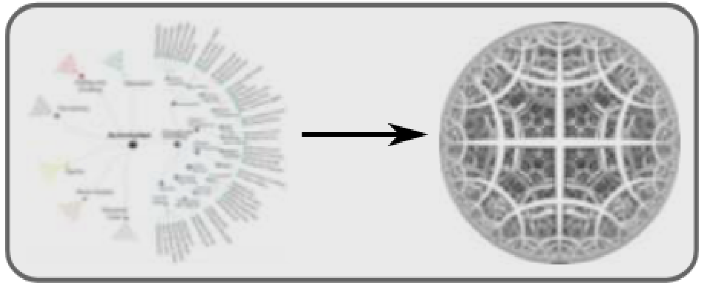
Hyperbolic Representation Learning

So what has been done?

Hyperbolic Action Recognition

Hyperbolic Action Recognition

Hyperbolic action embedding



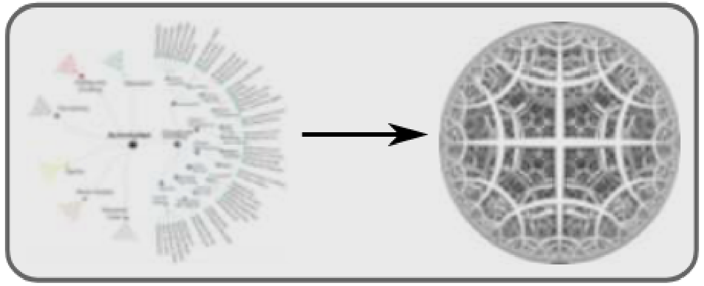
Action hierarchy

Discriminative hyperbolic embedding

Searching for Actions on the Hyperbole, Long et al., 2020

Hyperbolic Action Recognition

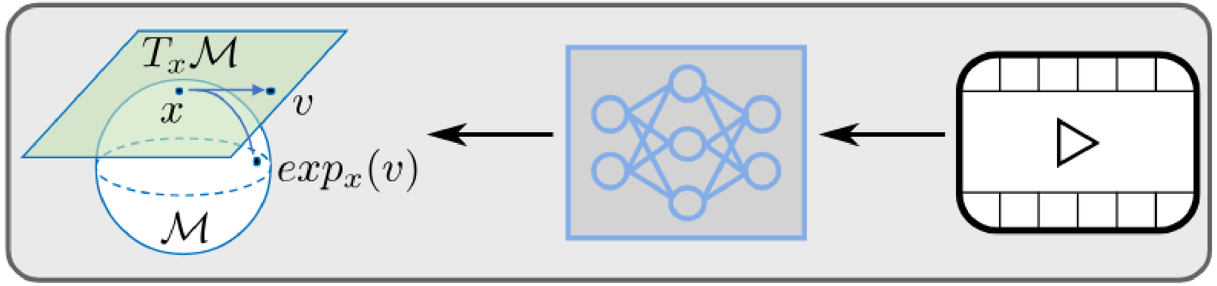
Hyperbolic action embedding



Action hierarchy

Discriminative hyperbolic embedding

Hyperbolic video embedding



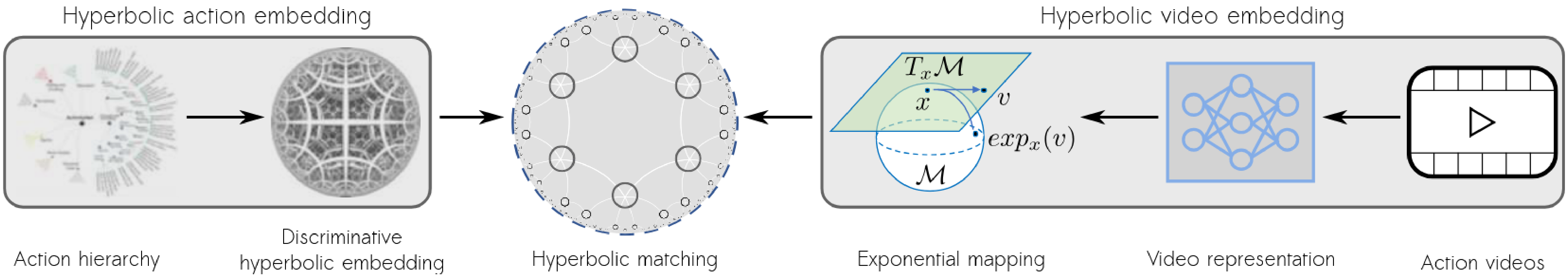
Exponential mapping

Video representation

Action videos

Searching for Actions on the Hyperbole, Long et al., 2020

Hyperbolic Action Recognition



Searching for Actions on the Hyperbole, Long et al., 2020

Hyperbolic Image Segmentations

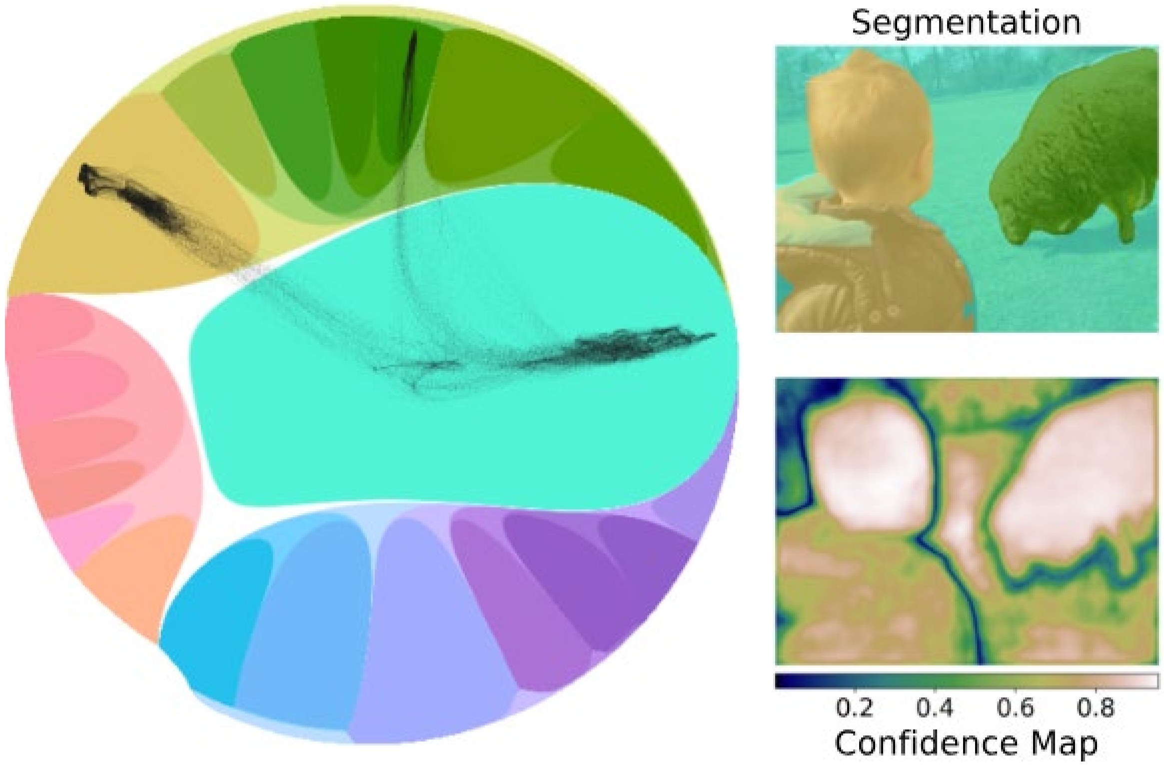
Hyperbolic Image Segmentations

→ Key insight: The closer a Poincaré embedding norm is to 1, the more certain the prediction is

Hyperbolic Image Segmentation, GhadimiAtigh et al., 2022

Hyperbolic Image Segmentations

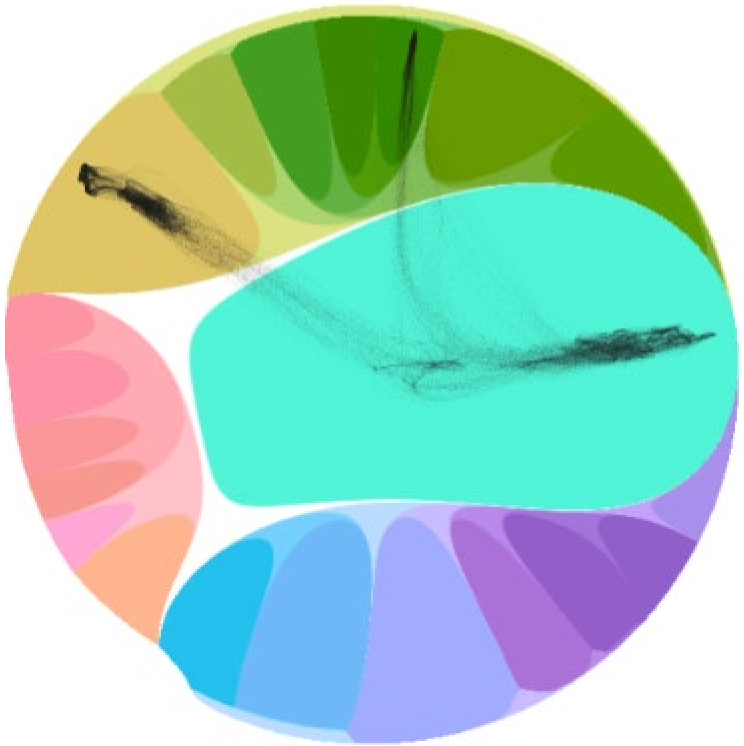
→ Key insight: The closer a Poincaré embedding norm is to 1, the more certain the prediction is

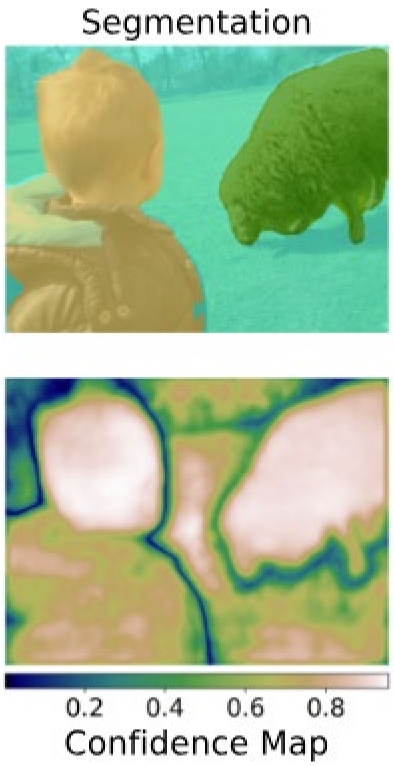


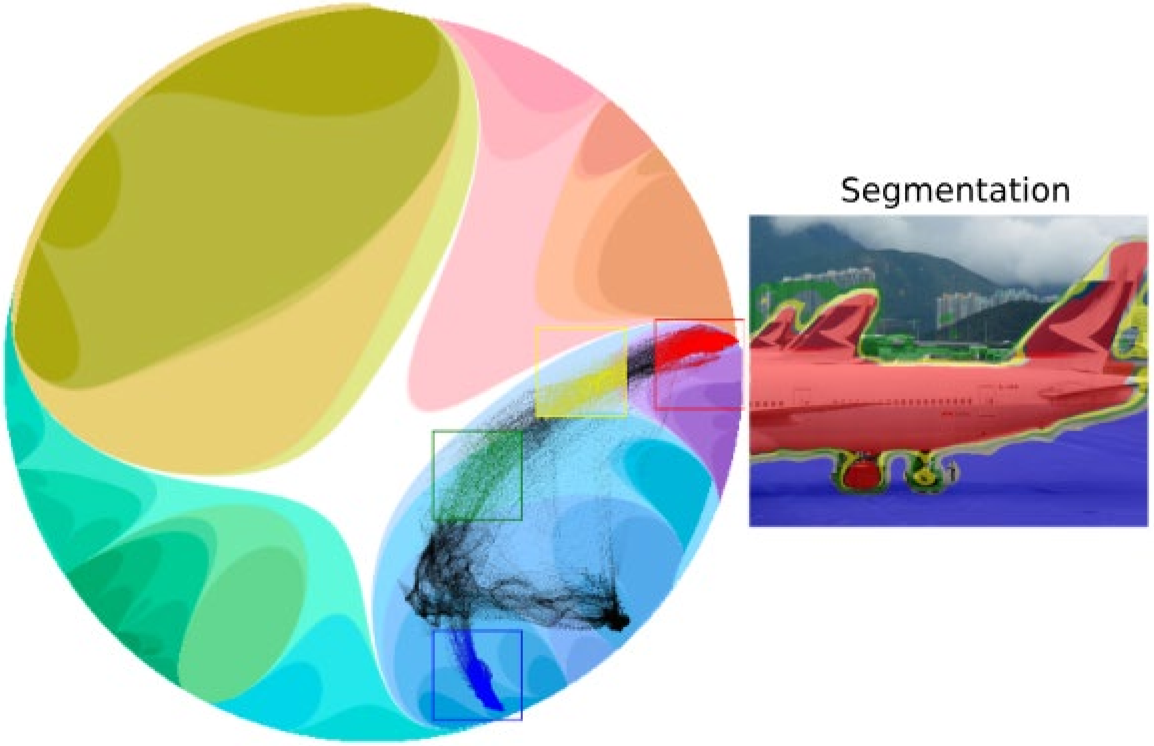
Hyperbolic Image Segmentation, GhadimiAtigh et al., 2022

Hyperbolic Image Segmentations

→ Key insight: The closer a Poincaré embedding norm is to 1, the more certain the prediction is







Hyperbolic Image Segmentation, GhadimiAtigh et al., 2022

Hyperbolic Image-Text Embeddings

Hyperbolic Image-Text Representations

Karan Desai [1] Maximilian Nickel [2] Tanmay Rajpurohit [3] Justin Johnson [1 2] Ramakrishna Vedantam [4]

Abstract

Visual and linguistic concepts naturally organize themselves in a hierarchy, where a textual concept "dog" entails all images that contain dogs. Despite being intuitive, current large-scale vision and language models such as CLIP (Radford et al., 2021) do not explicitly capture such hierarchy. We propose MERU, a contrastive model that yields hyperbolic representations of images and text. Hyperbolic spaces have suitable geometric properties to embed tree-like data, so MERU can better capture the underlying hierarchy in image-text datasets. Our results show that MERU learns a highly interpretable and structured representation space while being competitive with CLIP's performance on standard multi-modal tasks like image classification and image-text retrieval. Our code and models are available at: https://github.com/facebookresearch/meru

Figure 1. **Hyperbolic image-text representations. Left:** Images and text depict *concepts* and can be jointly viewed in a *visual-semantic hierarchy*, wherein text *'exhausted doggo'* is more generic than an image (which might have more details like a cat or snow). Our method MERU embeds images and text in a hyperbolic space that is well-suited to embed tree-like data. **Right:** Representation manifolds of CLIP (*hypersphere*) and MERU (*hyperboloid*) illustrated in 3D. MERU assumes the origin to represent the *most generic concept*, and embeds text closer to the origin than images.

1. Introduction

Visual-semantic hierarchy. It is commonly said that *'an image is worth a thousand words'* – consequently, images contain a lot more information than the sentences which typically describe them. For example, given the middle image in Figure 1 one might describe it as *'a cat and a dog playing in the street'* or with a less specific sentence like *'exhausted doggo'* or *'so cute <3'*. These are not merely diverse descriptions but contain varying levels of detail about the underlying semantic contents of the image.

As humans, we can reason about the relative detail in each caption, and can organize such concepts into a meaningful visual-semantic hierarchy (Vendrov et al., 2016), namely, *'exhausted doggo'* → *'a cat and a dog playing in the street'* → (Figure 1 middle image). Providing multimodal models access to this inductive bias about vision and language has the potential to improve generalization (Radford et al., 2021), interpretability (Selvaraju et al., 2017) and enable better exploratory data analysis of large-scale datasets (Radford et al., 2021; Schuhmann et al., 2022).

Vision-language representation learning. Approaches such as CLIP (Radford et al., 2021) and ALIGN (Jia et al., 2021) have catalyzed a lot of recent progress in computer vision by showing that Transformer-based (Vaswani et al., 2017) models trained using large amounts of image-text data from the internet can yield transferable representations, and such models can perform *zero-shot* recognition and retrieval using natural language queries. All these models represent images and text as vectors in a high-dimensional Euclidean, affine space and normalize the embeddings to unit L^2 norm. However, such a choice of geometry can find it hard to capture the visual-semantic hierarchy.

An affine Euclidean space treats all embedded points in the same manner, with the same distance metric being applied to all points (Murphy, 2013). Conceptually, this can cause issues when modeling hierarchies – a *generic* concept (closer to the *root node* of the hierarchy) is close to many other concepts compared to a *specific* concept (which is only close to its immediate neighbors). Thus, a Euclidean space can find it hard to pack all the images that say a generic

KD and Rama did this work while at Meta. [1]University of Michigan [2]Meta AI [3]Independent Researcher [4]New York University. Correspondence to: Karan Desai <kdexd@umich.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

Hyperbolic Image-Text Representations, Desai et al., 2023

Hyperbolic Image-Text Embeddings

→ Proposes a CLIP style setup in Hyperbolic space

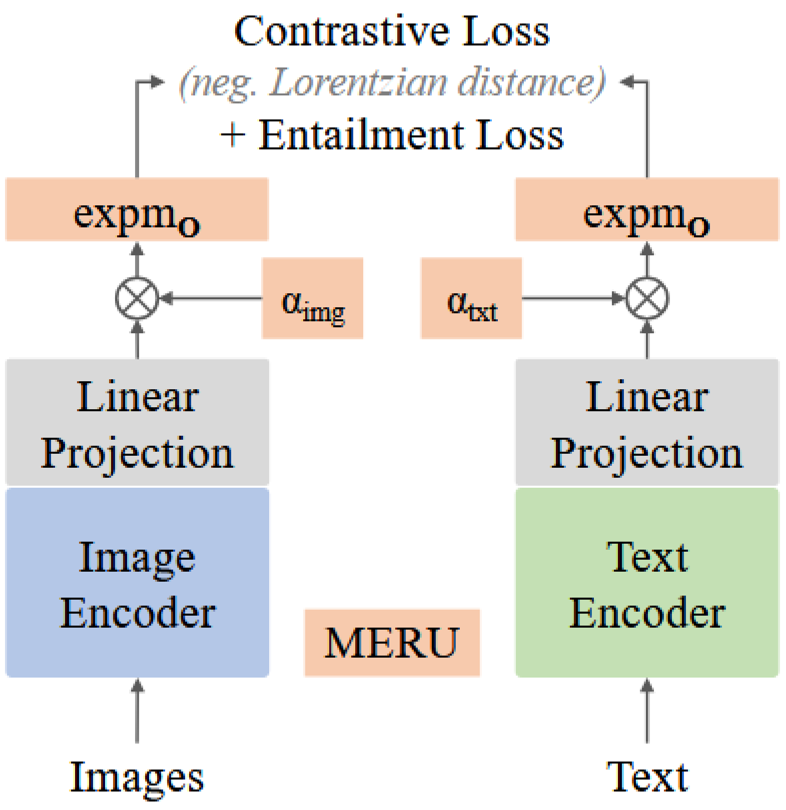
→ Uses the Lorentzian space due to *numerical instabilities* in Poincaré space

Hyperbolic Image-Text Representations, Desai et al., 2023

Hyperbolic Image-Text Embeddings

→ Proposes a CLIP style setup in Hyperbolic space

→ Uses the Lorentzian space due to *numerical instabilities* in Poincaré space



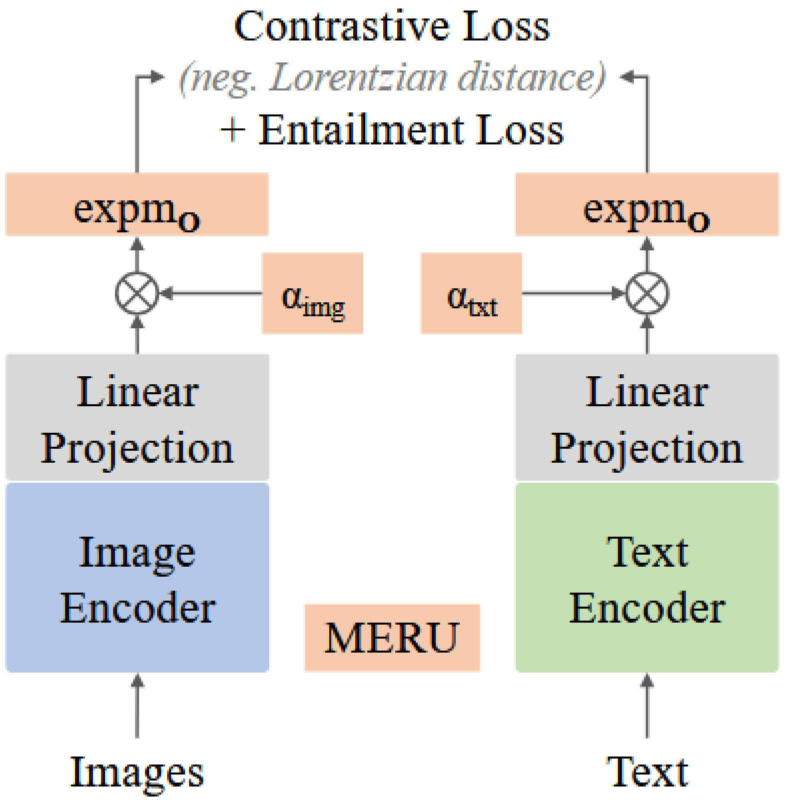
Hyperbolic Image-Text Representations, Desai et al., 2023

Hyperbolic Image-Text Embeddings

→ Proposes a CLIP style setup in Hyperbolic space

→ Uses the Lorentzian space due to *numerical instabilities* in Poincaré space

→ Two losses: Contrastive and *Entailment loss*



Hyperbolic Image-Text Representations, Desai et al., 2023

Contrastive Loss

Contrastive Loss

→ Standard contrastive loss (like in CLIP)

→ However, uses Lorentzian distance instead of cosine similarity

Hyperbolic Image-Text Representations, Desai et al., 2023

Contrastive Loss

→ Standard contrastive loss (like in CLIP)

→ However, uses Lorentzian distance instead of cosine similarity

→ Lorentzian inner product: $\langle x, y \rangle_L = x \circ y = x_0 y_0 - (x_1 y_1 + \cdots + x_d y_d)$

→ Lorentzian distance: $d_L(x, y) = \sqrt{\frac{1}{c}} \cosh^{-1}(c \langle x, y \rangle_L)$

Contrastive Loss

→ Standard contrastive loss (like in CLIP)

→ However, uses Lorentzian distance instead of cosine similarity

→ Lorentzian inner product: $\langle x, y \rangle_L = x \circ y = x_0 y_0 - (x_1 y_1 + \cdots + x_d y_d$

→ Lorentzian distance: $d_L(x, y) = \sqrt{\frac{1}{c}} \cosh^{-1}(c \, \langle x, y \rangle_L)$

Minimize distance for positive pairs
Maximize distance for negative pairs

Entailment Loss

→ Key concept: Fine-grained concepts should be embedded *deeper* into the space

Hyperbolic Image-Text Representations, Desai et al., 2023

Entailment Loss

→ Key concept: Fine-grained concepts should be embedded *deeper* into the space

→ Enforced by making sure the fine-grained concept is within the *entailment cone* of a parent concept

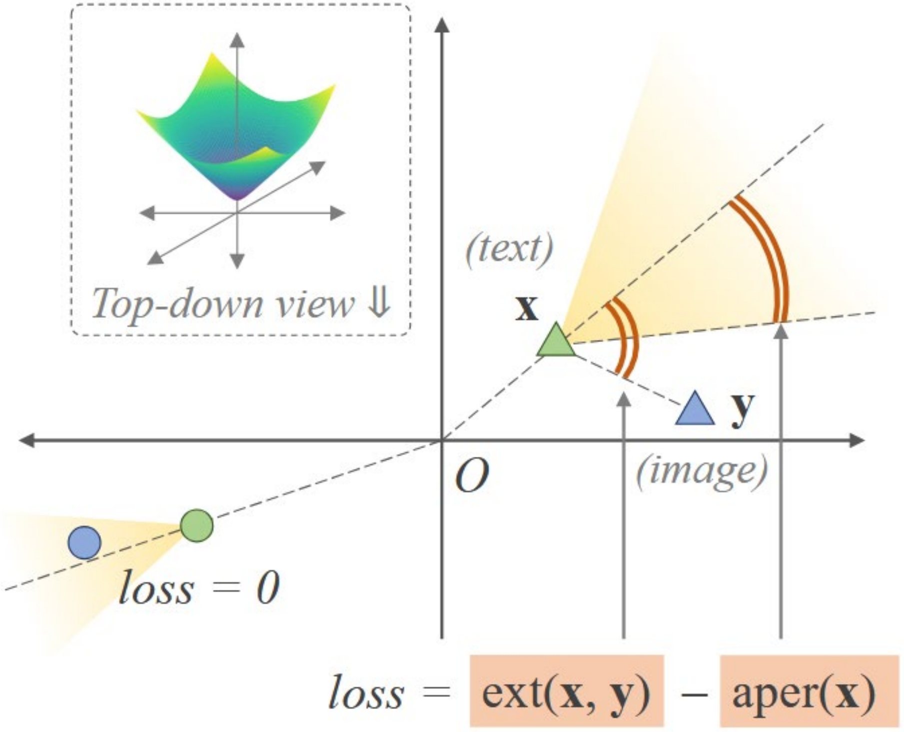
Hyperbolic Image-Text Representations, Desai et al., 2023

Entailment Loss

→ Key concept: Fine-grained concepts should be embedded *deeper* into the space

→ Enforced by making sure the fine-grained concept is within the *entailment cone* of a parent concept

$$\mathcal{L}_{entail}(\mathbf{x}, \mathbf{y}) = \max(0,\ \mathrm{ext}(\mathbf{x}, \mathbf{y}) - \mathrm{aper}(\mathbf{x}))$$



Hyperbolic Image-Text Representations, Desai et al., 2023

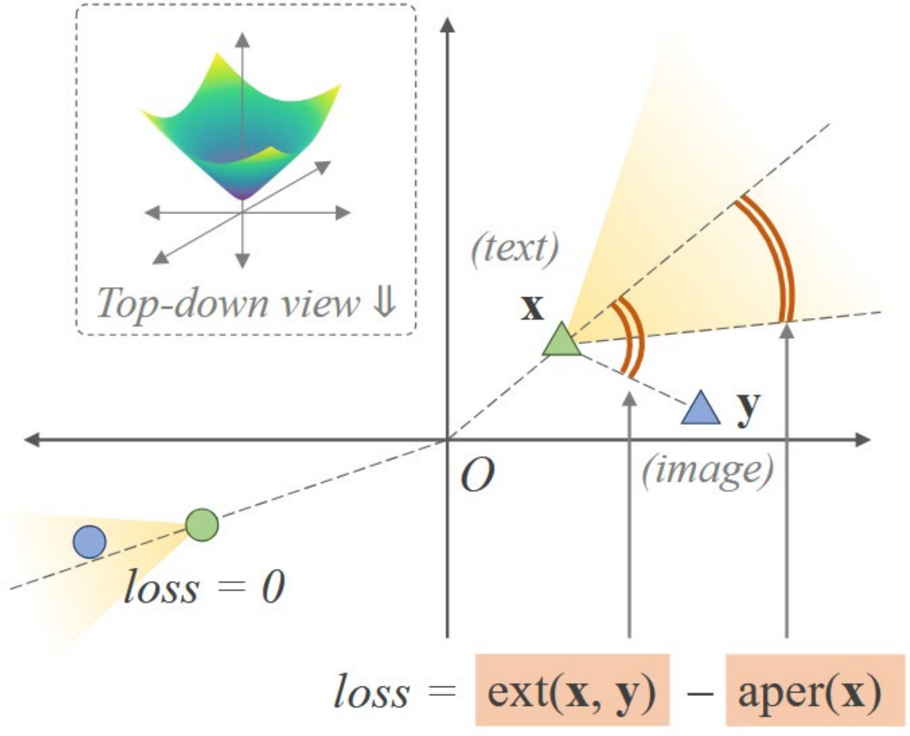
Entailment Loss

→ Key concept: Fine-grained concepts should be embedded *deeper* into the space

→ Enforced by making sure the fine-grained concept is within the *entailment cone* of a parent concept

$$\mathcal{L}_{entail}(\mathbf{x}, \mathbf{y}) = \max(0,\ \text{ext}(\mathbf{x}, \mathbf{y}) - \text{aper}(\mathbf{x}))$$

Angle between parent and child nodes



Hyperbolic Image-Text Representations, Desai et al., 2023

Entailment Loss

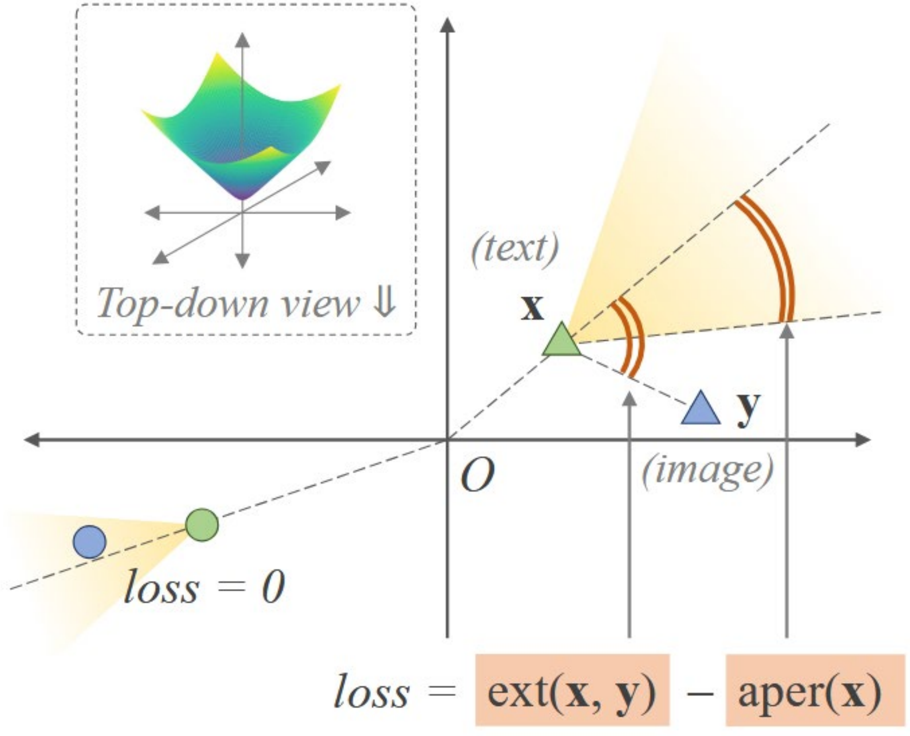
→ Key concept: Fine-grained concepts should be embedded *deeper* into the space

→ Enforced by making sure the fine-grained concept is within the *entailment cone* of a parent concept

$$\mathcal{L}_{entail}(\mathbf{x}, \mathbf{y}) = \max(0,\ \mathrm{ext}(\mathbf{x}, \mathbf{y}) - \mathrm{aper}(\mathbf{x}))$$

Angle between parent and child nodes

Entailment cone of parent node



Hyperbolic Image-Text Representations, Desai et al., 2023

Entailment Loss Effect

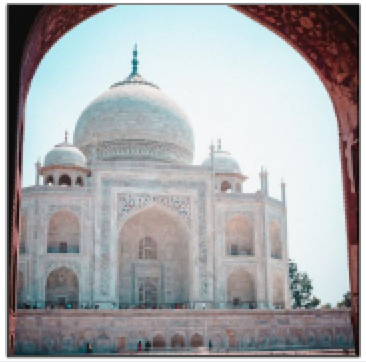
Entailment Loss Effect

MERU	CLIP
brooklyn bridge	*photo of brooklyn bridge, new york*
new york city	*new york city*
city	*new york*
outdoors	↓
day	↓
[ROOT]	[ROOT]

Hyperbolic Image-Text Representations, Desai et al., 2023

Entailment Loss Effect

MERU	CLIP
brooklyn bridge	*photo of brooklyn bridge, new york*
new york city	*new york city*
city	*new york*
outdoors	↓
day	↓
[ROOT]	[ROOT]

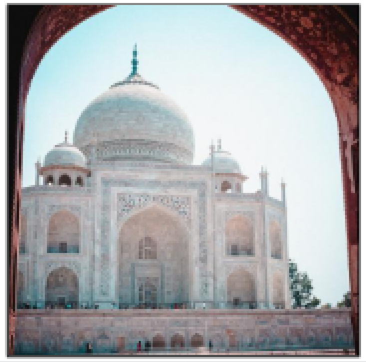


MERU	CLIP
taj mahal	*taj mahal through an arch*
monument	*travel*
architecture	*inspiration*
travel	↓
day	↓
[ROOT]	[ROOT]

Hyperbolic Image-Text Representations, Desai et al., 2023

Entailment Loss Effect

MERU	CLIP
brooklyn bridge	*photo of brooklyn bridge, new york*
new york city	*new york city*
city	*new york*
outdoors	↓
day	↓
[ROOT]	[ROOT]



MERU	CLIP
taj mahal	*taj mahal through an arch*
monument	*travel*
architecture	*inspiration*
travel	↓
day	↓
[ROOT]	[ROOT]

MERU	CLIP
sydney opera house	*sydney opera house*
opera house	*opera house*
holiday	*gift*
day	*beauty*
[ROOT]	[ROOT]

Hyperbolic Image-Text Representations, Desai et al., 2023

Stacked Entailment Loss

How is this used for Taxonomic Classification?

How is this used for Taxonomic Classification?

→ Assume paired multi-modal data, with expert textual labels

→ The textual labels contains the full hierarchy!

How is this used for Taxonomic Classification?

→ Assume paired multi-modal data, with expert textual labels

→ The textual labels contains the full hierarchy!

Diptera Psychodidae Psychoda Psychoda grisescens

How is this used for Taxonomic Classification?

→ Assume paired multi-modal data, with expert textual labels

→ The textual labels contains the full hierarchy!

Diptera	Psychodidae	Psychoda	Psychoda grisescens
Order	***Family***	***Genus***	***Species***

Stacked Entailment Loss

Hyperbolic Multimodal Representation Learning for Biological Taxonomies

ZeMing Gong[1] Chuanqi Tang[1] Xiaoliang Huo[1] Nicholas Pellegrino[2]
Austin T. Wang[1] Graham W. Taylor[3,4] Angel X. Chang[1,5]
Scott C. Lowe[3†] Joakim Bruslund Haurum[6†]
Simon Fraser University[1] University of Waterloo[2] Vector Institute[3]
University of Guelph[4] Alberta Machine Intelligence Institute (Amii)[5] Aalborg University[6]
{zmgong, ctal56, xiaoliang_huo, atw7, angelx}@sfu.ca,{nicholas.pellegrino}@uwaterloo.ca,
{gwtaylor}@uoguelph.ca {scott.lowe}@vectorinstitute.ai, {joha}@create.aau.dk

Abstract

Taxonomic classification in biodiversity research involves organizing biological specimens into structured hierarchies based on evidence, which can come from multiple modalities such as images and genetic information. We investigated whether hyperbolic networks provide a better embedding space for such hierarchical models. Our method embeds multimodal inputs into a shared hyperbolic space using contrastive and novel entailment-based objectives. Experiments on the BIOSCAN-1M dataset show that hyperbolic embedding gs achieve competitive performance with Euclidean baselines, and outperforms all other models on unseen species classification using DNA barcodes. However, fine-grained classification and open-world generalization remain challenging. This framework offers a scalable and structure-aware foundation for biodiversity modelling, with potential applications to species discovery, ecological monitoring, and conservation efforts.

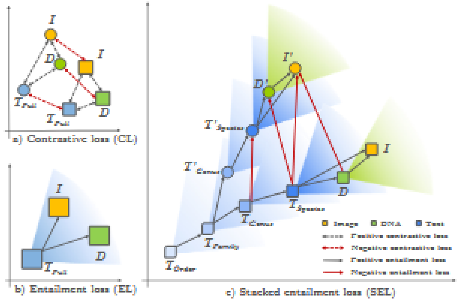


Figure 1. (a) **Contrastive loss**: instance-level alignment between modalities. (b) **Entailment loss**: enforces hierarchy within the text modality using entailment cones. (c) **Stacked entailment loss**: combines EL and cross-modal constraints by aligning image and DNA embeddings to multiple levels of the text hierarchy.

1. Introduction

Taxonomic classification is essential for monitoring and mitigating biodiversity loss, requiring accurate identification of specimens across diverse ecosystems. DNA barcodes [1, 7] provide a way to classify specimens to known taxa or identify them as novel to science, but classification to the species level remains challenging when barcodes are unavailable. To tackle this, Gong et al. [5] showed that using contrastive learning to align DNA barcode embeddings to image embeddings can improve classification at the species level even when only using images as input at inference.

However, a key limitation of CLIBD [5] and other recent biodiversity-focused multimodal methods [17] is that the methods do not utilize the known taxonomic hierarchy of the input data. To address this, we explore whether embeddings in hyperbolic space can better capture the hierarchical structure of taxonomic relationships, enabling better fine-grained classification. While training, the model takes inputs from multiple modalities—DNA barcodes, specimen images, and hierarchical taxonomic labels—and co-aligns their embeddings into a shared hyperbolic space to promote taxonomic alignment across modalities.

Our experimental results show that our hyperbolic multimodal learning framework achieves strong performance in taxonomic classification and retrieval, especially at higher taxonomic ranks. The approach consistently matches or outperforms Euclidean baselines and better preserves the hierarchical relationships among modalities. However, all methods—including ours—face challenges in fine-grained species classification, particularly for previously unseen taxa. These results highlight both the potential of hyperbolic learning for hierarchical biological data, and the ongoing difficulty of open-world classification for biodiversity.

Stacked Entailment Loss

→ Core ideas:

1. Enforce entailment loss between consecutive taxonomic ranks
2. Also apply a negative entailment loss

Hyperbolic Multimodal Representation Learning for Biological Taxonomies

ZeMing Gong[1] Chuanqi Tang[1] Xiaoliang Huo[1] Nicholas Pellegrino[2]
Austin T. Wang[1] Graham W. Taylor[3,4] Angel X. Chang[1,5]
Scott C. Lowe[3†] Joakim Bruslund Haurum[6†]
Simon Fraser University[1] University of Waterloo[2] Vector Institute[3]
University of Guelph[4] Alberta Machine Intelligence Institute (Amii)[5] Aalborg University[6]
{zmgong, ctal56, xiaoliang_huo, atw7, angelx}@sfu.ca,{nicholas.pellegrino}@uwaterloo.ca,
{gwtaylor}@uoguelph.ca {scott.lowe}@vectorinstitute.ai, {joha}@create.aau.dk

Abstract

Taxonomic classification in biodiversity research involves organizing biological specimens into structured hierarchies based on evidence, which can come from multiple modalities such as images and genetic information. We investigated whether hyperbolic networks provide a better embedding space for such hierarchical models. Our method embeds multimodal inputs into a shared hyperbolic space using contrastive and novel entailment-based objectives. Experiments on the BIOSCAN-1M dataset show that hyperbolic embedding gs achieve competitive performance with Euclidean baselines, and outperforms all other models on unseen species classification using DNA barcodes. However, fine-grained classification and open-world generalization remain challenging. This framework offers a scalable and structure-aware foundation for biodiversity modelling, with potential applications to species discovery, ecological monitoring, and conservation efforts.

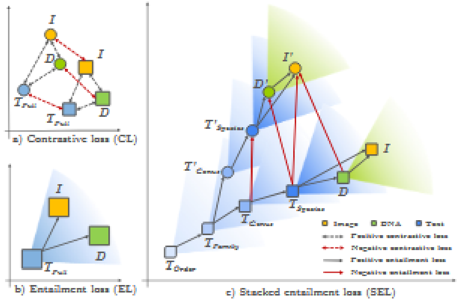


Figure 1. (a) **Contrastive loss**: instance-level alignment between modalities. (b) **Entailment loss**: enforces hierarchy within the text modality using entailment cones. (c) **Stacked entailment loss**: combines EL and cross-modal constraints by aligning image and DNA embeddings to multiple levels of the text hierarchy.

1. Introduction

Taxonomic classification is essential for monitoring and mitigating biodiversity loss, requiring accurate identification of specimens across diverse ecosystems. DNA barcodes [1, 7] provide a way to classify specimens to known taxa or identify them as novel to science, but classification to the species level remains challenging when barcodes are unavailable. To tackle this, Gong et al. [5] showed that using contrastive learning to align DNA barcode embeddings to image embeddings can improve classification at the species level even when only using images as input at inference.

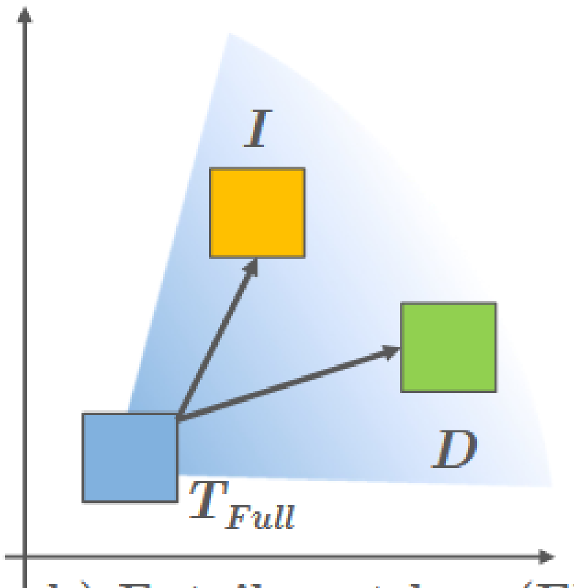
However, a key limitation of CLIBD [5] and other recent biodiversity-focused multimodal methods [17] is that the methods do not utilize the known taxonomic hierarchy of the input data. To address this, we explore whether embeddings in hyperbolic space can better capture the hierarchical structure of taxonomic relationships, enabling better fine-grained classification. While training, the model takes inputs from multiple modalities—DNA barcodes, specimen images, and hierarchical taxonomic labels—and co-aligns their embeddings into a shared hyperbolic space to promote taxonomic alignment across modalities.

Our experimental results show that our hyperbolic multimodal learning framework achieves strong performance in taxonomic classification and retrieval, especially at higher taxonomic ranks. The approach consistently matches or outperforms Euclidean baselines and better preserves the hierarchical relationships among modalities. However, all methods—including ours—face challenges in fine-grained species classification, particularly for previously unseen taxa. These results highlight both the potential of hyperbolic learning for hierarchical biological data, and the ongoing difficulty of open-world classification for biodiversity.

Stacked Entailment Loss

→ Core ideas:

1. Enforce entailment loss between consecutive taxonomic ranks
2. Also apply a negative entailment loss

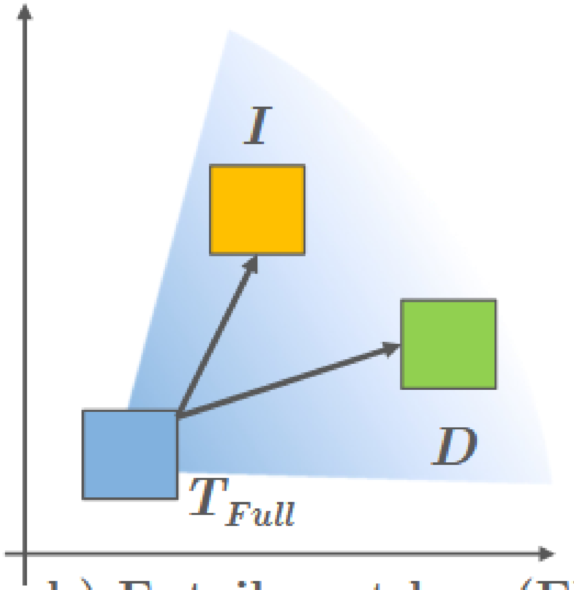


b) Entailment loss (EL)

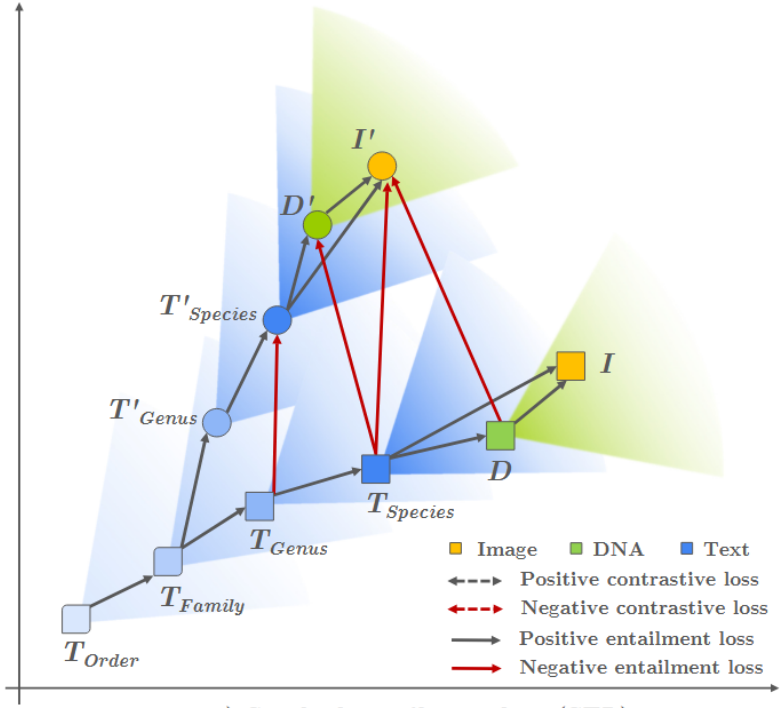
Stacked Entailment Loss

→ Core ideas:

1. Enforce entailment loss between consecutive taxonomic ranks
2. Also apply a negative entailment loss



b) Entailment loss (EL)



c) Stacked entailment loss (SEL)

Postive and Negative Entailment loss:

$$L_{\text{ent}}^{+} = \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \max\left(0,\ \text{ext}(x_i, y_j) - \text{aper}(x_i)\right)$$

$$L_{\text{ent}}^{-} = \frac{1}{|\mathcal{N}|} \sum_{(i,j)\in\mathcal{N}} \max\left(0,\ \text{aper}(x_i) - \text{ext}(x_i, y_j) + m\right)$$

Postive and Negative Entailment loss:

$$L_{\text{ent}}^{+} = \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \max\left(0,\ \text{ext}(x_i, y_j) - \text{aper}(x_i)\right)$$

$$L_{\text{ent}}^{-} = \frac{1}{|\mathcal{N}|} \sum_{(i,j)\in\mathcal{N}} \max\left(0,\ \text{aper}(x_i) - \text{ext}(x_i, y_j) + m\right)$$

$$L_{\text{ent}} = {}^{1}\!/_{2}\left(L_{\text{ent}}^{+} + L_{\text{ent}}^{-}\right)$$

Inter- and Intra-modal Entailment Losses

Inter- and Intra-modal Entailment Losses

$$L_{\text{SEL-intra}} = \frac{1}{\sum_{r=2}^{R} \mathbb{1}_r} \sum_{r=2}^{R} \mathbb{1}_r \times L_{\text{ent}}(T_r, T_{r-1})$$

Inter- and Intra-modal Entailment Losses

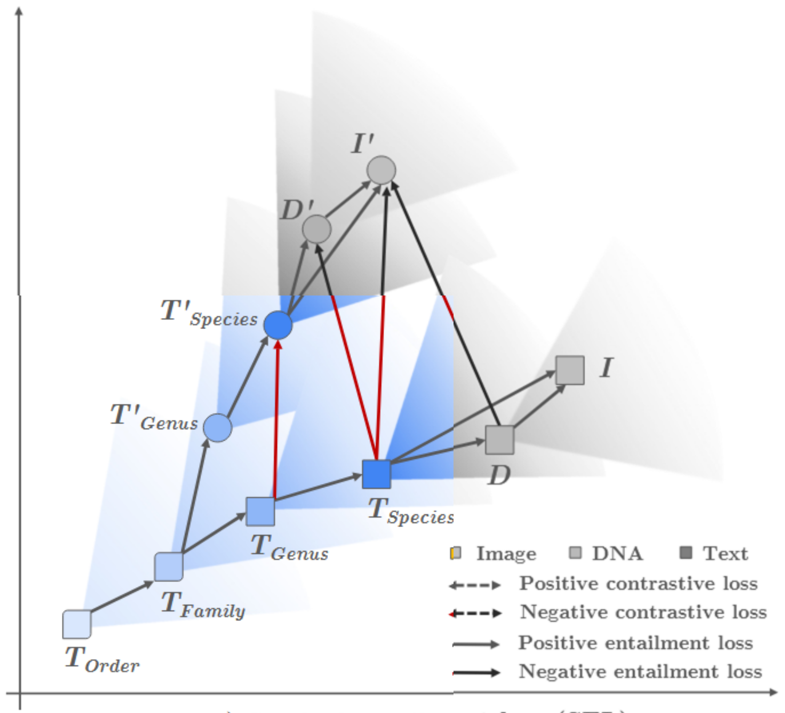
$$L_{\text{SEL-intra}} = \frac{1}{\sum_{r=2}^{R} \mathbb{1}_r} \sum_{r=2}^{R} \mathbb{1}_r \times L_{\text{ent}}(T_r, T_{r-1})$$

Indicates whether the rank is annotated

Inter- and Intra-modal Entailment Losses

$$L_{\text{SEL-intra}} = \frac{1}{\sum_{r=2}^{R} \mathbb{1}_r} \sum_{r=2}^{R} \mathbb{1}_r \times L_{\text{ent}}(T_r, T_{r-1})$$

Indicates whether the rank is annotated



c) Stacked entailment loss (SEL)

Inter- and Intra-modal Entailment Losses

$$L_{\text{SEL-intra}} = \frac{1}{\sum_{r=2}^{R} \mathbb{1}_r} \sum_{r=2}^{R} \mathbb{1}_r \times L_{\text{ent}}(T_r, T_{r-1})$$

$$L_{\text{SEL-inter}} = \frac{1}{3}\Big(L_{\text{ent}}(I, T_{R'}) + L_{\text{ent}}(D, T_{R'}) + L_{\text{ent}}(I, D)\Big)$$

Inter- and Intra-modal Entailment Losses

$$L_{\text{SEL-intra}} = \frac{1}{\sum_{r=2}^{R} \mathbb{1}_r} \sum_{r=2}^{R} \mathbb{1}_r \times L_{\text{ent}}(T_r, T_{r-1})$$

$$L_{\text{SEL-inter}} = \frac{1}{3}\Big(L_{\text{ent}}(I, T_{R'}) + L_{\text{ent}}(D, T_{R'}) + L_{\text{ent}}(I, D) \Big)$$

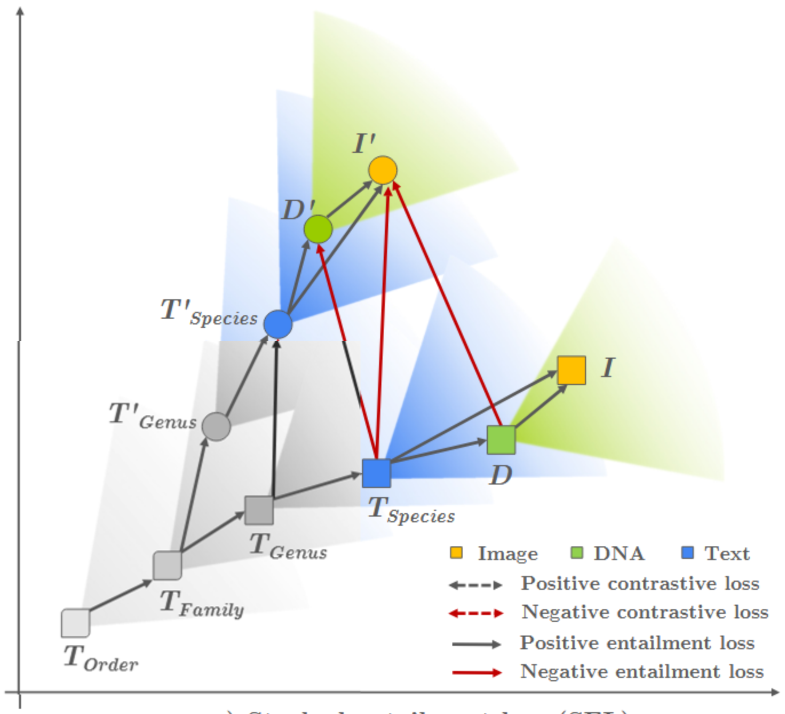
Deepest annotated taxonomic rank

Inter- and Intra-modal Entailment Losses

$$L_{\text{SEL-intra}} = \frac{1}{\sum_{r=2}^{R} \mathbb{1}_r} \sum_{r=2}^{R} \mathbb{1}_r \times L_{\text{ent}}(T_r, T_{r-1})$$

$$L_{\text{SEL-inter}} = \frac{1}{3}\left(L_{\text{ent}}(I, T_{R'}) + L_{\text{ent}}(D, T_{R'}) + L_{\text{ent}}(I, D)\right)$$

Deepest annotated taxonomic rank



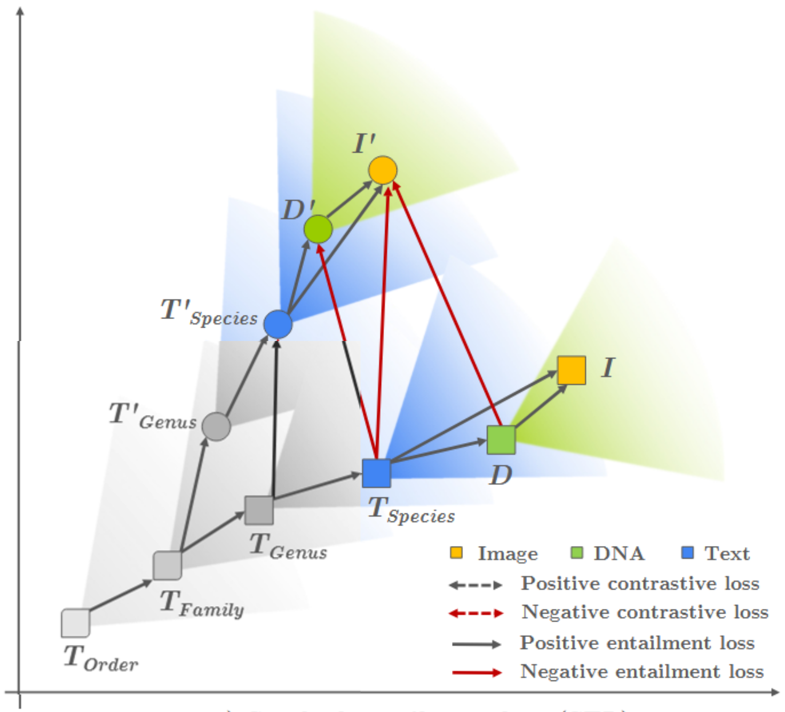
c) Stacked entailment loss (SEL)

Inter- and Intra-modal Entailment Losses

$$L_{\text{SEL-intra}} = \frac{1}{\sum_{r=2}^{R} \mathbb{1}_r} \sum_{r=2}^{R} \mathbb{1}_r \times L_{\text{ent}}(T_r, T_{r-1})$$

$$L_{\text{SEL-inter}} = \frac{1}{3}\left(L_{\text{ent}}(I, T_{R'}) + L_{\text{ent}}(D, T_{R'}) + L_{\text{ent}}(I, D) \right)$$

$$L_{\text{SEL}} = L_{\text{SEL-intra}} + L_{\text{SEL-inter}}$$



c) Stacked entailment loss (SEL)

Preliminary Results

BIOSCAN-1M Results

					DNA-to-DNA			Image-to-Image			Image-to-DNA		
Rank	Method	EL config.	Full Text	Space	Seen	Unseen	H.M.	Seen	Unseen	H.M.	Seen	Unseen	H.M.
Order	CLIBD	–	✓	$\mathbb{R}^n$	**89.1**	87.8	88.4	**99.5**	**66.4**	**79.6**	**98.7**	**49.5**	**65.9**
	CL	–	✓	$\mathbb{H}^n_L$	**89.1**	85.6	87.3	98.5	61.2	75.5	89.1	47.8	62.2
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	88.6	86.5	87.5	98.6	56.9	72.1	77.8	48.4	59.7
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.4	**90.8**	**89.6**	79.3	62.3	69.8	**98.7**	<u>48.9</u>	<u>65.4</u>
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.7	86.3	87.5	<u>99.4</u>	<u>65.9</u>	<u>79.3</u>	78.6	48.2	59.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	88.9	<u>88.2</u>	<u>88.5</u>	99.0	60.9	75.4	78.6	<u>48.9</u>	60.3

Rank	Method	EL config.	Full Text	Space	DNA-to-DNA			Image-to-Image			Image-to-DNA		
					Seen	Unseen	H.M.	Seen	Unseen	H.M.	Seen	Unseen	H.M.
Order	CLIBD	–	✓	$\mathbb{R}^n$	**89.1**	87.8	88.4	**99.5**	**66.4**	**79.6**	**98.7**	**49.5**	**65.9**
	CL	–	✓	$\mathbb{H}^n_L$	**89.1**	85.6	87.3	98.5	61.2	75.5	89.1	47.8	62.2
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	88.6	86.5	87.5	98.6	56.9	72.1	77.8	48.4	59.7
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.4	**90.8**	**89.6**	79.3	62.3	69.8	**98.7**	48.9	65.4
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.7	86.3	87.5	99.4	65.9	79.3	78.6	48.2	59.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	88.9	88.2	88.5	99.0	60.9	75.4	78.6	48.9	60.3
Family	CLIBD	–	✓	$\mathbb{R}^n$	90.8	75.8	82.6	**89.2**	**52.2**	**65.9**	**83.6**	**19.3**	**31.4**
	CL	–	✓	$\mathbb{H}^n_L$	90.3	76.6	82.9	83.9	48.5	61.4	79.6	18.8	30.4
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	89.3	74.9	81.4	81.9	37.6	51.5	76.7	16.8	27.6
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	86.8	**78.8**	82.6	79.0	41.8	54.7	78.9	18.4	29.9
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	89.0	76.9	82.5	79.6	46.6	58.8	78.7	17.3	28.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	**91.2**	77.0	**83.6**	82.4	41.5	55.2	78.1	17.4	28.4

Rank	Method	EL config.	Full Text	Space	DNA-to-DNA			Image-to-Image			Image-to-DNA		
					Seen	Unseen	H.M.	Seen	Unseen	H.M.	Seen	Unseen	H.M.
Order	CLIBD	–	✓	$\mathbb{R}^n$	**89.1**	87.8	88.4	**99.5**	**66.4**	**79.6**	**98.7**	**49.5**	**65.9**
	CL	–	✓	$\mathbb{H}^n_L$	**89.1**	85.6	87.3	98.5	61.2	75.5	89.1	47.8	62.2
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	88.6	86.5	87.5	98.6	56.9	72.1	77.8	48.4	59.7
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.4	**90.8**	**89.6**	79.3	62.3	69.8	**98.7**	48.9	65.4
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.7	86.3	87.5	99.4	65.9	79.3	78.6	48.2	59.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	88.9	88.2	88.5	99.0	60.9	75.4	78.6	48.9	60.3
Family	CLIBD	–	✓	$\mathbb{R}^n$	90.8	75.8	82.6	**89.2**	**52.2**	**65.9**	**83.6**	**19.3**	**31.4**
	CL	–	✓	$\mathbb{H}^n_L$	90.3	76.6	82.9	83.9	48.5	61.4	79.6	18.8	30.4
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	89.3	74.9	81.4	81.9	37.6	51.5	76.7	16.8	27.6
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	86.8	**78.8**	82.6	79.0	41.8	54.7	78.9	18.4	29.9
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	89.0	76.9	82.5	79.6	46.6	58.8	78.7	17.3	28.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	**91.2**	77.0	**83.6**	82.4	41.5	55.2	78.1	17.4	28.4
Genus	CLIBD	–	✓	$\mathbb{R}^n$	85.2	64.3	73.3	**71.3**	**35.0**	**47.0**	**70.8**	**7.1**	**12.9**
	CL	–	✓	$\mathbb{H}^n_L$	**86.4**	64.9	74.1	65.6	32.4	43.4	66.9	6.5	11.8
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	84.7	63.1	72.3	63.0	22.8	33.5	64.2	6.6	11.9
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	82.7	65.9	73.4	62.1	29.2	39.7	63.1	6.6	12.0
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	83.6	**66.9**	**74.3**	63.3	33.1	43.5	67.6	6.4	11.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	85.8	64.8	73.9	64.8	27.5	38.6	64.8	6.2	11.4

Rank	Method	EL config.	Full Text	Space	DNA-to-DNA			Image-to-Image			Image-to-DNA		
					Seen	Unseen	H.M.	Seen	Unseen	H.M.	Seen	Unseen	H.M.
Order	CLIBD	–	✓	$\mathbb{R}^n$	**89.1**	87.8	88.4	**99.5**	**66.4**	**79.6**	**98.7**	**49.5**	**65.9**
	CL	–	✓	$\mathbb{H}^n_L$	**89.1**	85.6	87.3	98.5	61.2	75.5	89.1	47.8	62.2
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	88.6	86.5	87.5	98.6	56.9	72.1	77.8	48.4	59.7
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.4	**90.8**	**89.6**	79.3	62.3	69.8	**98.7**	48.9	65.4
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.7	86.3	87.5	99.4	65.9	79.3	78.6	48.2	59.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	88.9	88.2	88.5	99.0	60.9	75.4	78.6	48.9	60.3
Family	CLIBD	–	✓	$\mathbb{R}^n$	90.8	75.8	82.6	**89.2**	**52.2**	**65.9**	**83.6**	**19.3**	**31.4**
	CL	–	✓	$\mathbb{H}^n_L$	90.3	76.6	82.9	83.9	48.5	61.4	79.6	18.8	30.4
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	89.3	74.9	81.4	81.9	37.6	51.5	76.7	16.8	27.6
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	86.8	**78.8**	82.6	79.0	41.8	54.7	78.9	18.4	29.9
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	89.0	76.9	82.5	79.6	46.6	58.8	78.7	17.3	28.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	**91.2**	77.0	**83.6**	82.4	41.5	55.2	78.1	17.4	28.4
Genus	CLIBD	–	✓	$\mathbb{R}^n$	85.2	64.3	73.3	**71.3**	**35.0**	**47.0**	**70.8**	**7.1**	**12.9**
	CL	–	✓	$\mathbb{H}^n_L$	**86.4**	64.9	74.1	65.6	32.4	43.4	66.9	6.5	11.8
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	84.7	63.1	72.3	63.0	22.8	33.5	64.2	6.6	11.9
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	82.7	65.9	73.4	62.1	29.2	39.7	63.1	6.6	12.0
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	83.6	**66.9**	**74.3**	63.3	33.1	43.5	67.6	6.4	11.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	85.8	64.8	73.9	64.8	27.5	38.6	64.8	6.2	11.4
Species	CLIBD	–	✓	$\mathbb{R}^n$	81.8	60.6	69.7	**55.1**	**24.3**	**33.7**	**55.8**	0.7	1.4
	CL	–	✓	$\mathbb{H}^n_L$	**84.4**	61.8	**71.4**	48.2	22.6	30.8	53.7	0.9	1.7
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	82.5	60.1	69.6	45.4	14.3	21.8	50.5	0.9	1.8
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	79.5	62.3	69.9	45.5	20.0	27.8	52.0	**1.1**	**2.1**
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	80.5	**63.2**	70.8	46.8	22.8	30.7	54.2	0.7	1.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	82.6	62.0	70.8	47.8	19.0	27.2	51.4	1.0	**2.1**

Rank	Method	EL config.	Full Text	Space	DNA-to-DNA			Image-to-Image			Image-to-DNA		
					Seen	Unseen	H.M.	Seen	Unseen	H.M.	Seen	Unseen	H.M.
Order	CLIBD	–	✓	$\mathbb{R}^n$	**89.1**	87.8	88.4	**99.5**	**66.4**	**79.6**	**98.7**	**49.5**	**65.9**
	CL	–	✓	$\mathbb{H}^n_L$	**89.1**	85.6	87.3	98.5	61.2	75.5	89.1	47.8	62.2
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	88.6	86.5	87.5	98.6	56.9	72.1	77.8	48.4	59.7
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.4	**90.8**	**89.6**	79.3	62.3	69.8	**98.7**	48.9	65.4
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.7	86.3	87.5	99.4	65.9	79.3	78.6	48.2	59.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	88.9	88.2	88.5	99.0	60.9	75.4	78.6	48.9	60.3
Family	CLIBD	–	✓	$\mathbb{R}^n$	90.8	75.8	82.6	**89.2**	**52.2**	**65.9**	**83.6**	**19.3**	**31.4**
	CL	–	✓	$\mathbb{H}^n_L$	90.3	76.6	82.9	83.9	48.5	61.4	79.6	18.8	30.4
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	89.3	74.9	81.4	81.9	37.6	51.5	76.7	16.8	27.6
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	86.8	**78.8**	82.6	79.0	41.8	54.7	78.9	18.4	29.9
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	89.0	76.9	82.5	79.6	46.6	58.8	78.7	17.3	28.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	**91.2**	77.0	**83.6**	82.4	41.5	55.2	78.1	17.4	28.4
Genus	CLIBD	–	✓	$\mathbb{R}^n$	85.2	64.3	73.3	**71.3**	**35.0**	**47.0**	**70.8**	**7.1**	**12.9**
	CL	–	✓	$\mathbb{H}^n_L$	**86.4**	64.9	74.1	65.6	32.4	43.4	66.9	6.5	11.8
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	84.7	63.1	72.3	63.0	22.8	33.5	64.2	6.6	11.9
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	82.7	65.9	73.4	62.1	29.2	39.7	63.1	6.6	12.0
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	83.6	**66.9**	**74.3**	63.3	33.1	43.5	67.6	6.4	11.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	85.8	64.8	73.9	64.8	27.5	38.6	64.8	6.2	11.4
Species	CLIBD	–	✓	$\mathbb{R}^n$	81.8	60.6	69.7	**55.1**	**24.3**	**33.7**	**55.8**	0.7	1.4
	CL	–	✓	$\mathbb{H}^n_L$	**84.4**	61.8	**71.4**	48.2	22.6	30.8	53.7	0.9	1.7
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	82.5	60.1	69.6	45.4	14.3	21.8	50.5	0.9	1.8
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	79.5	62.3	69.9	45.5	20.0	27.8	52.0	**1.1**	**2.1**
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	80.5	**63.2**	70.8	46.8	22.8	30.7	54.2	0.7	1.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	82.6	62.0	70.8	47.8	19.0	27.2	51.4	1.0	**2.1**

Rank	Method	EL config.	Full Text	Space	DNA-to-DNA			Image-to-Image			Image-to-DNA		
					Seen	Unseen	H.M.	Seen	Unseen	H.M.	Seen	Unseen	H.M.
Order	CLIBD	–	✓	$\mathbb{R}^n$	**89.1**	87.8	88.4	**99.5**	**66.4**	**79.6**	**98.7**	**49.5**	**65.9**
	CL	–	✓	$\mathbb{H}^n_L$	**89.1**	85.6	87.3	98.5	61.2	75.5	89.1	47.8	62.2
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	88.6	86.5	87.5	98.6	56.9	72.1	77.8	48.4	59.7
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.4	**90.8**	**89.6**	79.3	62.3	69.8	**98.7**	48.9	65.4
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.7	86.3	87.5	99.4	65.9	79.3	78.6	48.2	59.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	88.9	88.2	88.5	99.0	60.9	75.4	78.6	48.9	60.3
Family	CLIBD	–	✓	$\mathbb{R}^n$	90.8	75.8	82.6	**89.2**	**52.2**	**65.9**	**83.6**	**19.3**	**31.4**
	CL	–	✓	$\mathbb{H}^n_L$	90.3	76.6	82.9	83.9	48.5	61.4	79.6	18.8	30.4
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	89.3	74.9	81.4	81.9	37.6	51.5	76.7	16.8	27.6
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	86.8	**78.8**	82.6	79.0	41.8	54.7	78.9	18.4	29.9
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	89.0	76.9	82.5	79.6	46.6	58.8	78.7	17.3	28.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	**91.2**	77.0	**83.6**	82.4	41.5	55.2	78.1	17.4	28.4
Genus	CLIBD	–	✓	$\mathbb{R}^n$	85.2	64.3	73.3	**71.3**	**35.0**	**47.0**	**70.8**	**7.1**	**12.9**
	CL	–	✓	$\mathbb{H}^n_L$	**86.4**	64.9	74.1	65.6	32.4	43.4	66.9	6.5	11.8
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	84.7	63.1	72.3	63.0	22.8	33.5	64.2	6.6	11.9
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	82.7	65.9	73.4	62.1	29.2	39.7	63.1	6.6	12.0
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	83.6	**66.9**	**74.3**	63.3	33.1	43.5	67.6	6.4	11.7
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	85.8	64.8	73.9	64.8	27.5	38.6	64.8	6.2	11.4
Species	CLIBD	–	✓	$\mathbb{R}^n$	81.8	60.6	69.7	**55.1**	**24.3**	**33.7**	**55.8**	0.7	1.4
	CL	–	✓	$\mathbb{H}^n_L$	**84.4**	61.8	**71.4**	48.2	22.6	30.8	53.7	0.9	1.7
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	82.5	60.1	69.6	45.4	14.3	21.8	50.5	0.9	1.8
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	79.5	62.3	69.9	45.5	20.0	27.8	52.0	**1.1**	**2.1**
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	80.5	**63.2**	70.8	46.8	22.8	30.7	54.2	0.7	1.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	82.6	62.0	70.8	47.8	19.0	27.2	51.4	1.0	**2.1**

BIOSCAN-1M Results Summary

BIOSCAN-1M Results Summary

→ Across all ranks Hypebolic models match or outperform Euclidean CLIBD

→ SEL methods consistently perform best at unseen DNA retrieval

→ But Euclidean CLIBD is better at image retrieval tasks

BIOSCAN-1M Results Summary

→ Across all ranks Hypebolic models match or outperform Euclidean CLIBD

→ SEL methods consistently perform best at unseen DNA retrieval

→ But Euclidean CLIBD is better at image retrieval tasks

→ This is still ongoing research, so we hope to further improve results within the coming months