Hyperbolic Multimodal Representation Learning for Biological Taxonomies

ZeMing Gong[1], Chuanqi Tang[1], Xiaoliang Huo[1], Nicholas Pellegrino[2], Austin 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]

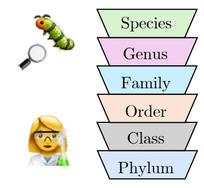
Alberta Machine Intelligence Institute (Amii)[5] Aalborg University[6]

Abstract

We address the challenge of representing hierarchical biological taxonomy from multimodal inputs — images, DNA, and text. Our approach embeds these modalities in a shared ***hyperbolic space***, capturing taxonomic structure and cross-modal alignment. This enables scalable biodiversity modeling and improved recognition of unseen species.

Background

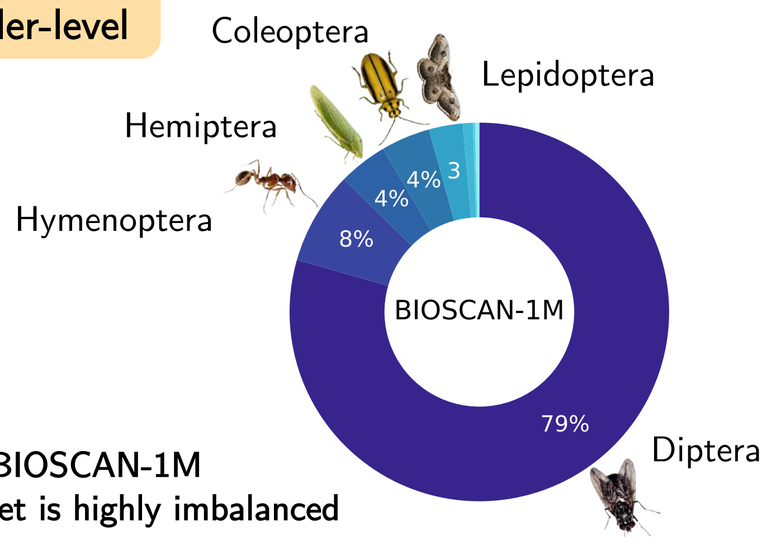
Environmental change requires scalable, automated tools for biodiversity monitoring.



Taxonomic classification is expert-driven, fine-grained, and long-tailed in distribution.

Dataset: BIOSCAN-1M

Order-level



The BIOSCAN-1M dataset is highly imbalanced

Inference

Task: Image to DNA

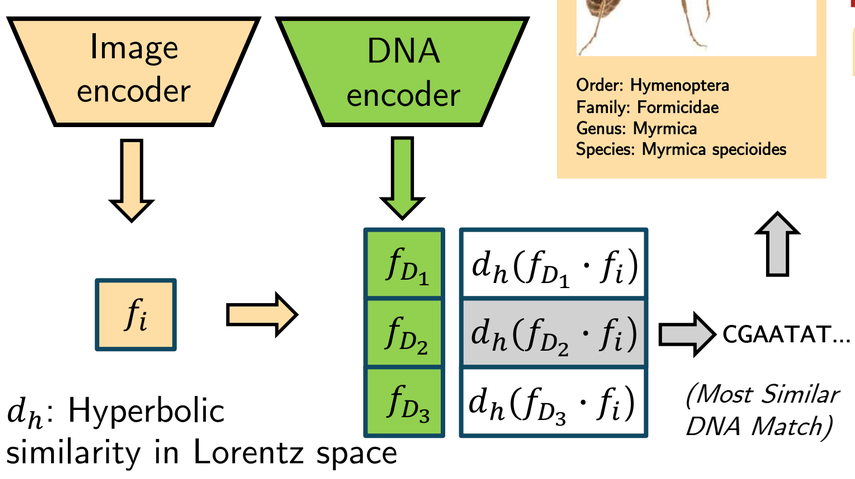
- Lorentz inner product
- Lorentz distance
- Hyperbolic angle

Query:

Keys:

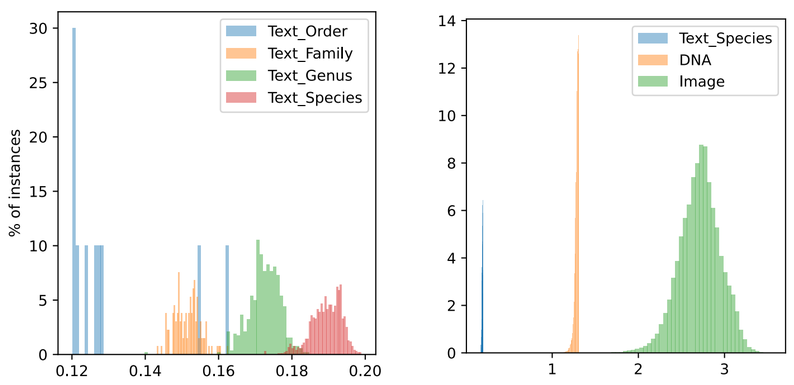
ATGCCCCA...
CGAATATG...
GAAATGCA...

Order: Hymenoptera
Family: Formicidae
Genus: Myrmica
Species: Myrmica specioides



d_h: Hyperbolic similarity in Lorentz space

Distance Distributions



Distribution of embedding distances $d(z) = ||z_{space}||$ from the root across modalities. Different modalities occupy distinct distance ranges, indicating clear separation in hyperbolic space.

Reference

[1] CLIBD: Contrastive Learning of Biodiversity. Gong et al. ICLR. 2025.
[2] MERU: Hyperbolic Image-Text Representations. Desai et al. ICML. 2023.
[3] BIOSCAN-1M Insect Dataset. Gharaee et al. NeurIPS. 2023.

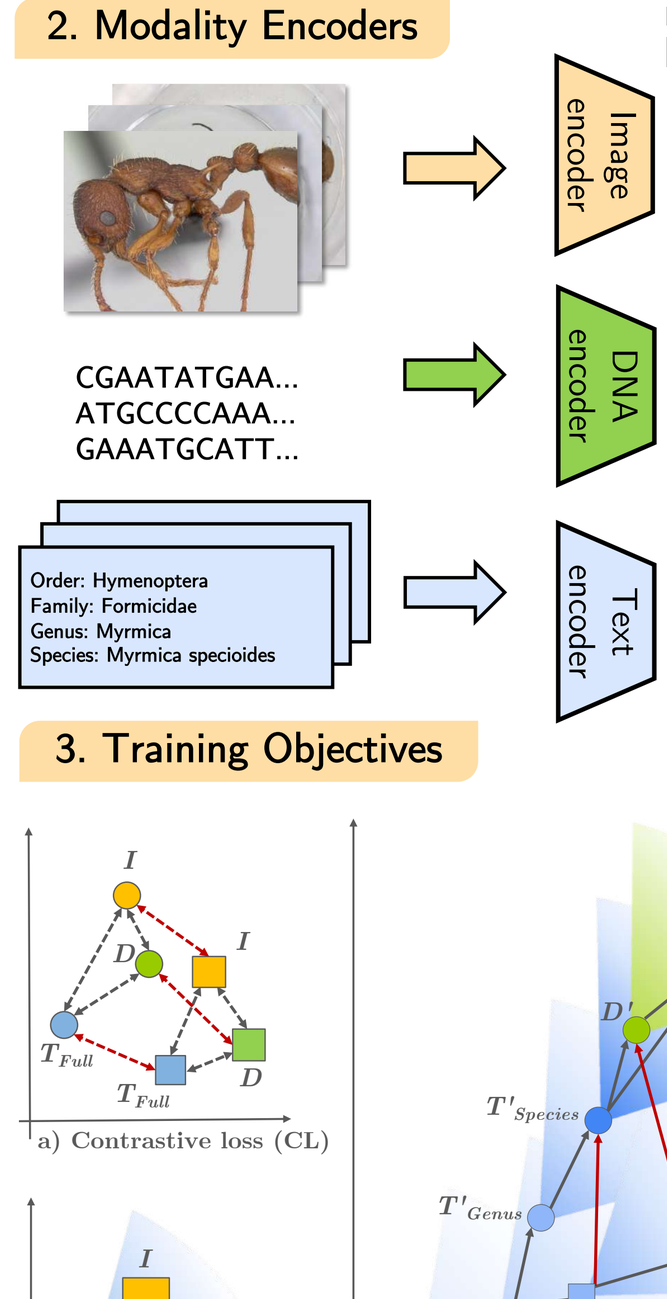
Acknowledgements. This work was supported by the Government of Canada's New Frontiers in Research Fund (NFRF) [NFRFT-2020-00073], Canada CIFAR AI chair.

Method

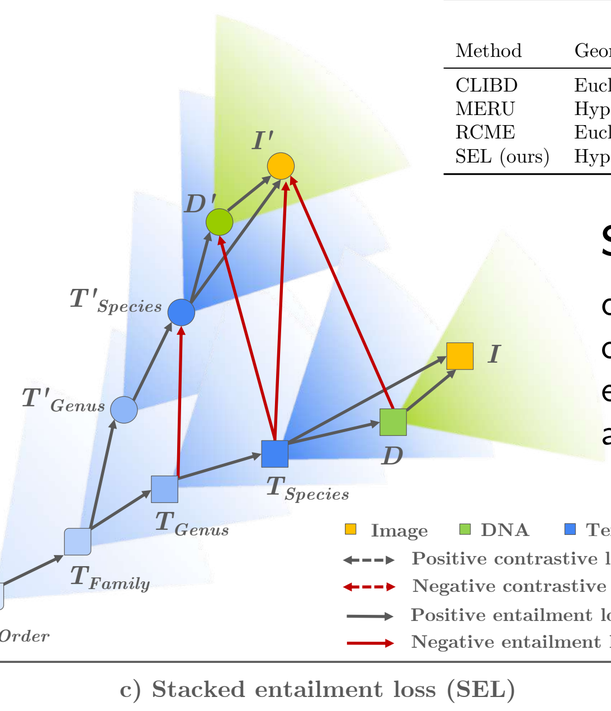
1. Problem Setup

We embed image, DNA barcodes, and text into a shared hyperbolic space. We use **stacked entailment loss** to train the embedding to reflect the biological taxonomy (Order to Species). This space supports species classification, cross-modal matching, and generalization to unseen taxa.

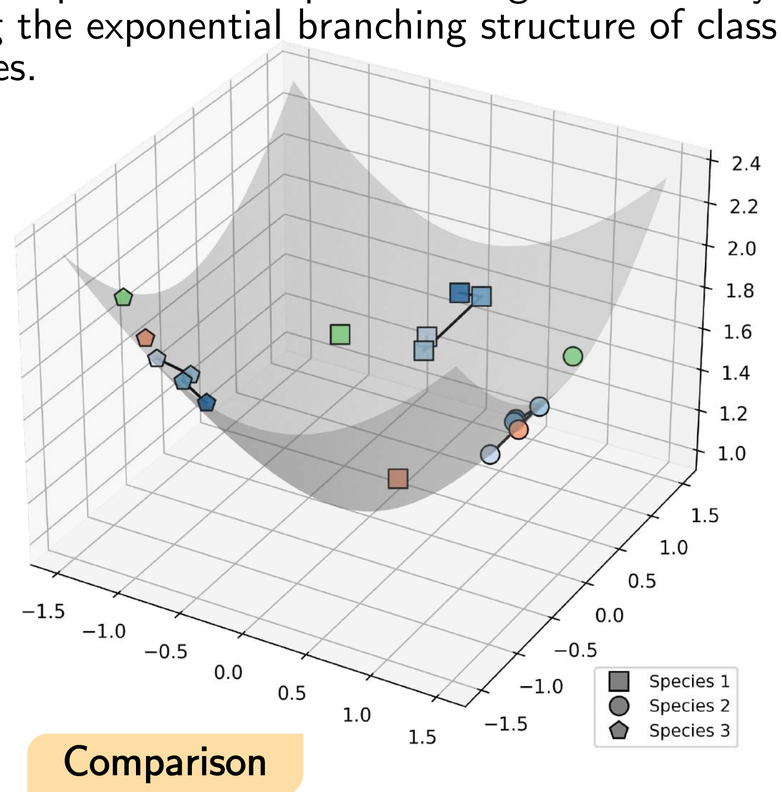
2. Modality Encoders



3. Training Objectives



Hyperbolic space better captures biological taxonomy by matching the exponential branching structure of class hierarchies.



Comparison

Method	Geometry	Modalities: Image	DNA	Text	Objective	Hierarchy Handling
CLIBD	Euclidean	✓	✓	✓	CL	Flat labels
MERU	Hyperbolic	✓	✗	✓	EL + CL	Implicit
RCME	Euclidean	✓	✗	✓	EL + CL	Global/local entailment
SEL (ours)	Hyperbolic	✓	✓	✓	EL + CL	Explicit ranks

Stacked entailment loss (SEL)

combines contrastive and entailment objectives across hierarchical levels, enforcing both inter-modality alignment and taxonomy consistency.

$$\alpha(u) = \sin^{-1}\left(\frac{K}{\|u\|_{\mathbb{H}}}\right)$$

Half-aperture of entailment cone centered at embedding u

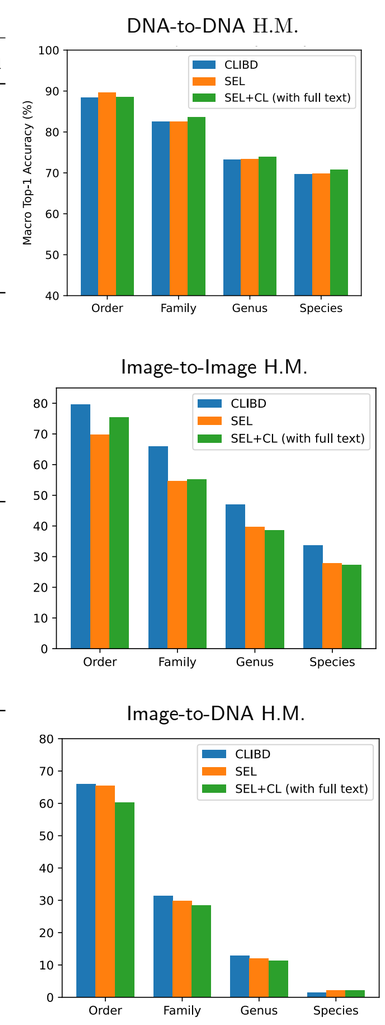
Result

Retrieval Performance

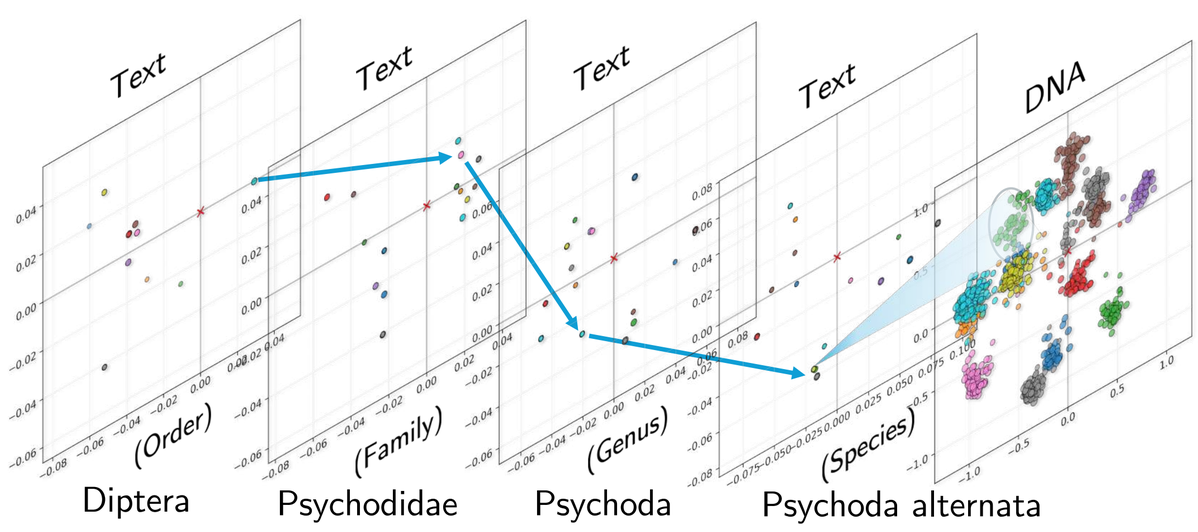
Macro top-1 accuracy across taxonomic levels on BIOSCAN-1M, comparing different training objectives for multimodal retrieval (DNA, Image, Text).

Rank	Method	EL config.	Full Text	Space	DNA-to-DNA Seen	DNA-to-DNA Unseen	Image-to-Image Seen	Image-to-Image Unseen	Image-to-DNA Seen	Image-to-DNA Unseen
Order	CLIBD	–	✓	$\mathbb{R}^n$	**89.1**	87.8	**99.5**	**66.4**	**98.7**	**49.5**
	CL	–	✓	$\mathbb{H}^n_L$	**89.1**	85.6	98.5	61.2	89.1	47.8
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	88.6	86.5	98.6	56.9	77.8	48.4
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.4	**90.8**	79.3	62.3	**98.7**	48.9
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	88.7	86.3	99.4	65.9	78.6	48.2
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	88.9	88.2	99.0	60.9	78.6	48.9
Family	CLIBD	–	✓	$\mathbb{R}^n$	90.8	75.8	**89.2**	**52.2**	**83.6**	**19.3**
	CL	–	✓	$\mathbb{H}^n_L$	90.3	76.6	83.9	48.5	79.6	18.8
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	89.3	74.9	81.9	37.6	76.7	16.8
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	86.8	**78.8**	79.0	41.8	78.9	18.4
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	89.0	76.9	79.6	46.6	78.7	17.3
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	**91.2**	77.0	82.4	41.5	78.1	17.4
Genus	CLIBD	–	✓	$\mathbb{R}^n$	85.2	64.3	**71.3**	**35.0**	**70.8**	**7.1**
	CL	–	✓	$\mathbb{H}^n_L$	**86.4**	64.9	65.6	32.4	66.9	6.5
	EL+CL	Pos.	✓	$\mathbb{H}^n_L$	84.7	63.1	63.0	22.8	64.2	6.6
	SEL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	82.7	65.9	62.1	29.2	63.1	6.6
	SEL+CL	Pos.+Neg.	✗	$\mathbb{H}^n_L$	83.6	**66.9**	63.3	33.1	67.6	6.4
	SEL+CL	Pos.+Neg.	✓	$\mathbb{H}^n_L$	85.8	64.8	64.8	27.5	64.8	6.2

Retrieval Plots



Embedding Plots



HoroPCA projection of hyperbolic embeddings across taxonomy levels (Order → Species). Representations expand radially and form aligned clusters, showing hierarchical and cross-modal consistency.

- Hyperbolic models achieve performance comparable to Euclidean baselines while providing stronger generalization to unseen taxa.
- Our stacked entailment loss (SEL) consistently improves unseen retrieval—especially for DNA queries and fine-grained taxonomic levels.