HIER: Metric Learning Beyond Class Labels via Hierarchical Regularization

Sungyeon Kim Boseung Jeong Suha Kwak

{sungyeon.kim, boseung01, suha.kwak}@postech.ac.kr

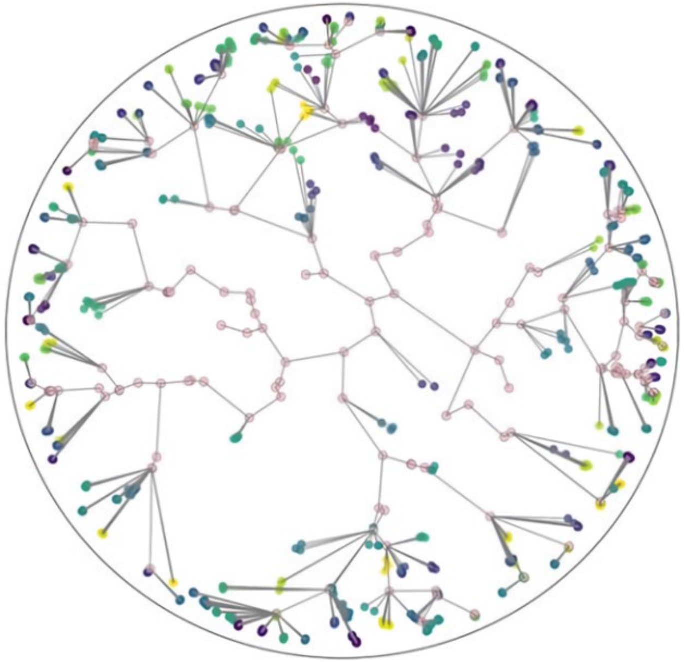
POSTECH

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Overview of Our Work

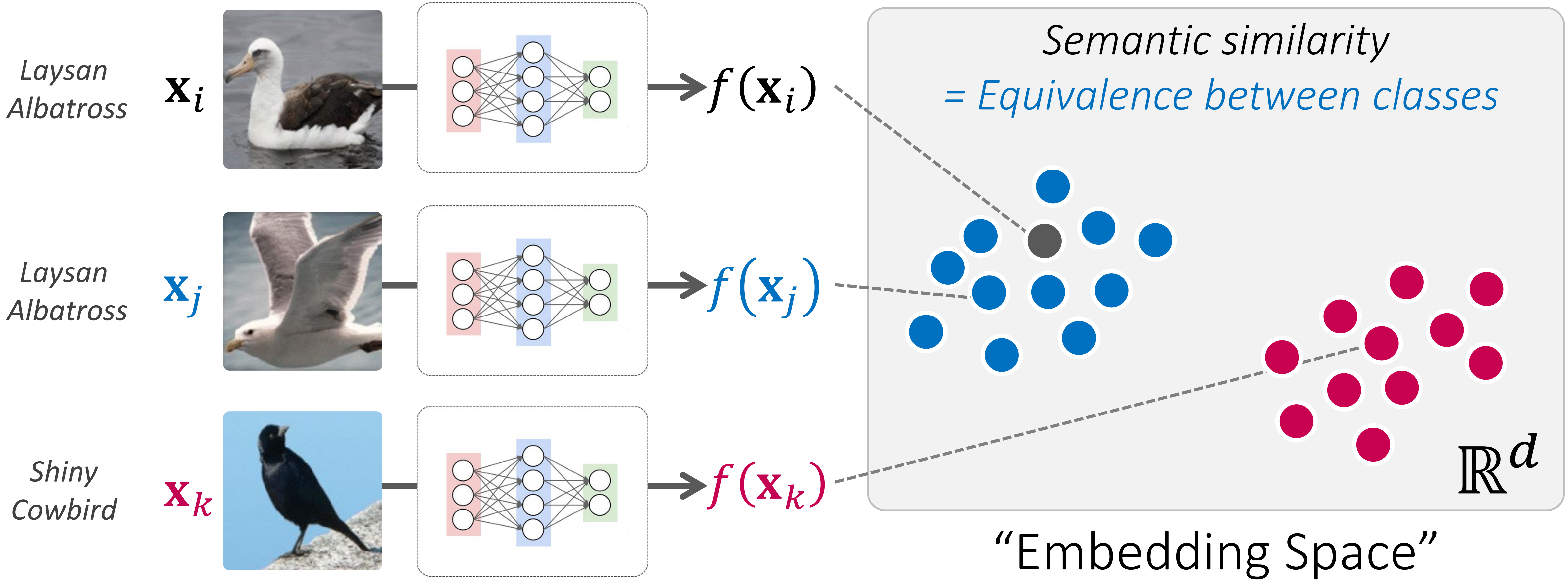
- HIErarchical Regularization (HIER)

- Providing richer and more fine-grained supervision beyond inter-class separability induced by common metric learning losses.
- Discovering and deploying the *latent semantic hierarchy* of data by approximating soft hierarchical clustering.
- *No need for extra annotation* for the semantic hierarchy.
- Utilize hyperbolic space as the embedding space to effectively represent hierarchical structures of data.
- Enable the use of conventional metric learning losses in conjunction with the hyperbolic embedding space.



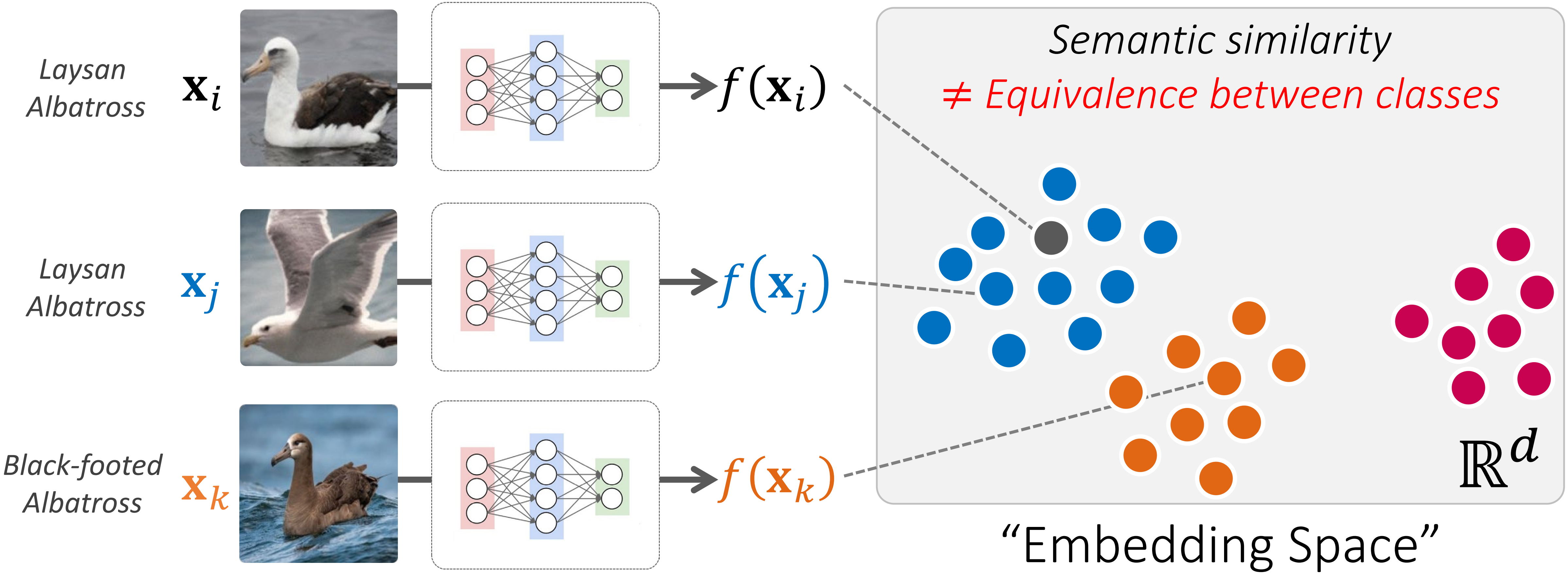
Deep Metric Learning

Learning a deep embedding network f so that
semantically similar images are closely grouped together



Deep Metric Learning

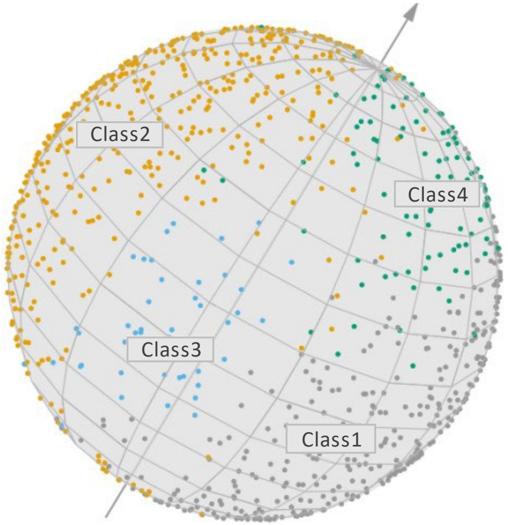
The equivalence of human-labeled classes deals with only a **tiny subset of possible relations** between samples.



Euclidean Space vs. Hyperbolic Space

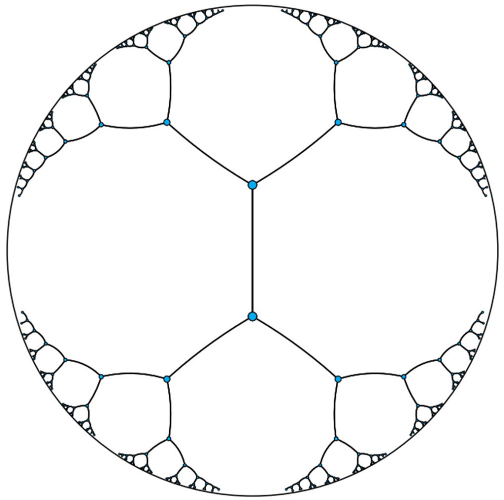
- Euclidean Embedding Space

- Most of deep metric learning model operate with spherical (cosine distance) or Euclidean embeddings.
- Euclidean space cannot embed large and complex graphs without distortion or loss of information.
  - e.g., Hierarchical tree



- Hyperbolic Embedding Space

- Space of *negative curvature* possess geometric properties that make embeddings *well-suited for modeling hierarchical relationships*.



Poincaré Ball Model

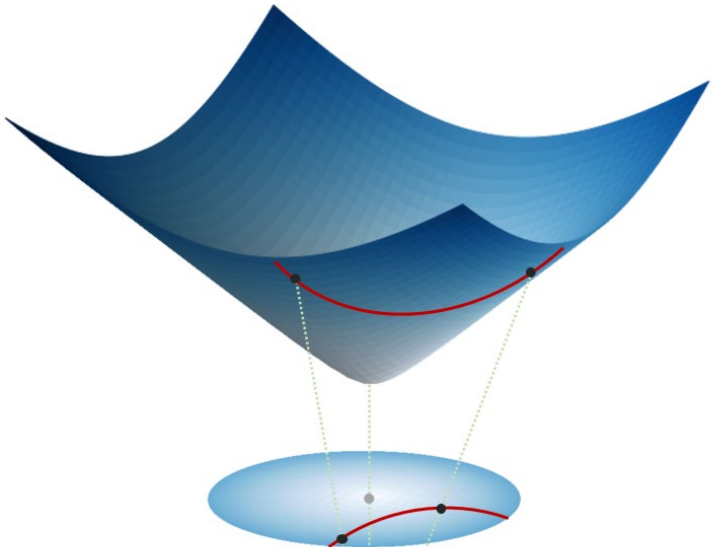
- There exist several models of hyperbolic space (e.g., Klein, Hyperboloid), but we take the popular one, *Poincaré ball model*.

- Poincare ball model is defined as $(\mathbb{D}_c^n, g^{\mathbb{D}})$ where the manifold $\mathbb{D}_c^n = \{\mathbf{x} \in \mathbb{R}^n : c||\mathbf{x}|| < 1\}$ Riemannian metric $g^{\mathbb{D}} = \lambda_c^2 g^E$ and c is curvature.

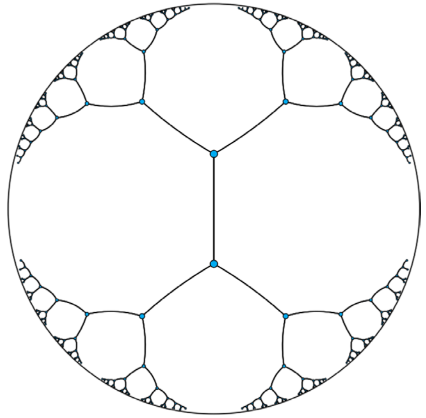
- The *hyperbolic distance* between the two vectors u and v is formulated as

$$d_H(u,v) = arccosh(1 + 2\,\frac{c\|u-v\|^2}{(1-c\|u\|^2)(1-c\|v\|^2)}$$

Hyperboloid model

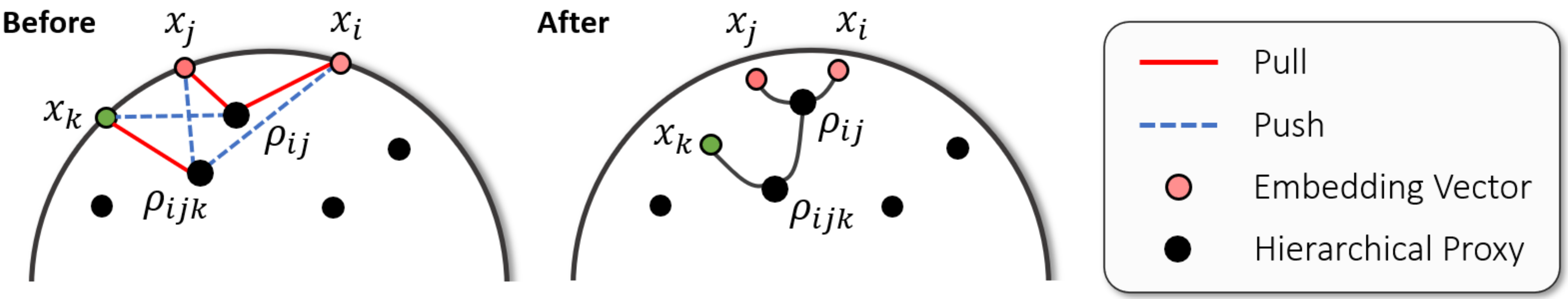


Poincaré Model

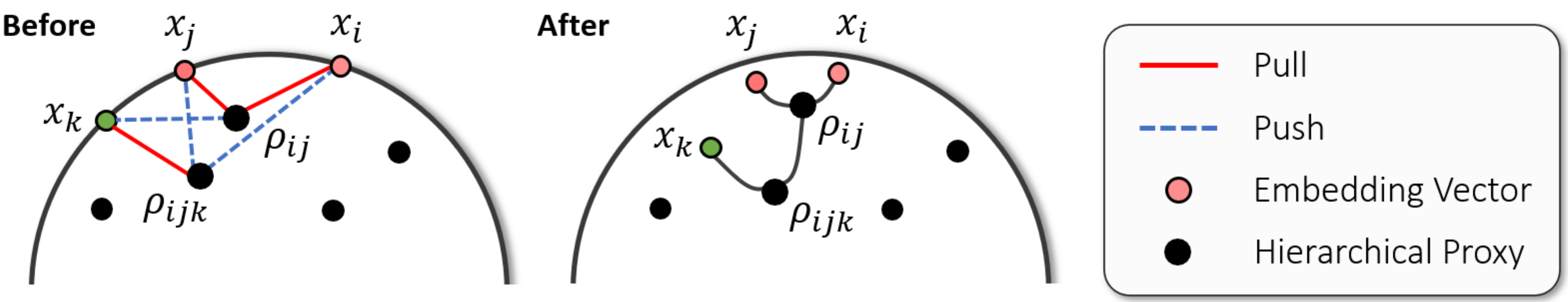


HIErarchical Regularization (HIER)

- Aiming to discover the latent semantic hierarchy of data *with no additional annotation for the hierarchy*.
- Employing *hierarchical proxies* that are learnable parameters serving as a virtual ancestor of data points in the hierarchy.
- Encouraging that the pair of related samples have the same lowest common ancestor (LCA) and the rest has a different LCA.



HIErarchical Regularization (HIER)



Set of feasible triplets $(\boldsymbol{x_i}, \boldsymbol{x_j}, \boldsymbol{x_k})$

$$\mathcal{T} = \{(x_i, x_j, x_k) \mid (x_j \in R_K(x_i) \wedge (x_k \notin R_K(x_i))\}$$

$$\text{where } R_K(x) = \{x' | x' \in N_K(x)) \wedge x \in N_K(x'))\}$$

***Probability that a hierarchical proxy is the LCA*:**

$$\pi_{ij}(\rho) = \exp(-\max\{d_H(x_i, \rho), d_H(x_j, \rho)\})$$

$$\pi_{ijk}(\rho) = \exp(-\max\{d_H(x_i, \rho), d_H(x_j, \rho), d_H(x_k, \rho)\})$$

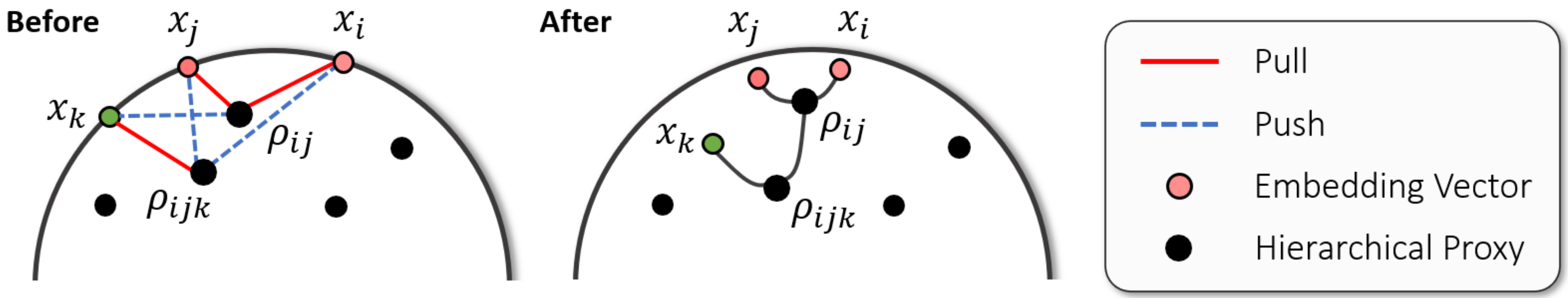
***Sampling LCAs*:**

$$\rho_{ij} = \text{argmax}_\rho(\pi_{ij}(\rho) + g_{ij})$$

$$\rho_{ijk} = \text{argmax}_\rho(\pi_{ijk}(\rho) + g_{ij})$$

where $g_{ij} \sim \text{Gumbel}(0,1)$

HIErarchical Regularization (HIER)



Objective of HIER

$$\begin{aligned}\mathcal{L}_{\mathrm{HIER}}(t) = &\left[d_H(x_i,\rho_{ij}) - d_H(x_i,\rho_{ijk}) + \delta\right]_+ \\ &+ \left[d_H(x_j,\rho_{ij}) - d_H(x_j,\rho_{ijk}) + \delta\right]_+ \\ &+ \left[d_H(x_k,\rho_{ijk}) - d_H(x_k,\rho_{ij}) + \delta\right]_+\end{aligned}$$

attracting (first two terms)

→ *repelling* (third term)

Total training loss

$$\mathcal{L} = \mathcal{L}_{\mathrm{ML}} + \lambda \sum_{t\in\{\mathcal{T}_x,\mathcal{T}_\rho\}} \mathcal{L}_{\mathrm{HIER}}(t)$$

$\mathcal{L}_{\mathrm{ML}}$: only controlling the **angles** between the normalized embedding vectors

$\mathcal{L}_{\mathrm{HIER}}$: adjusting hyperbolic distances based on **positions and norms**.

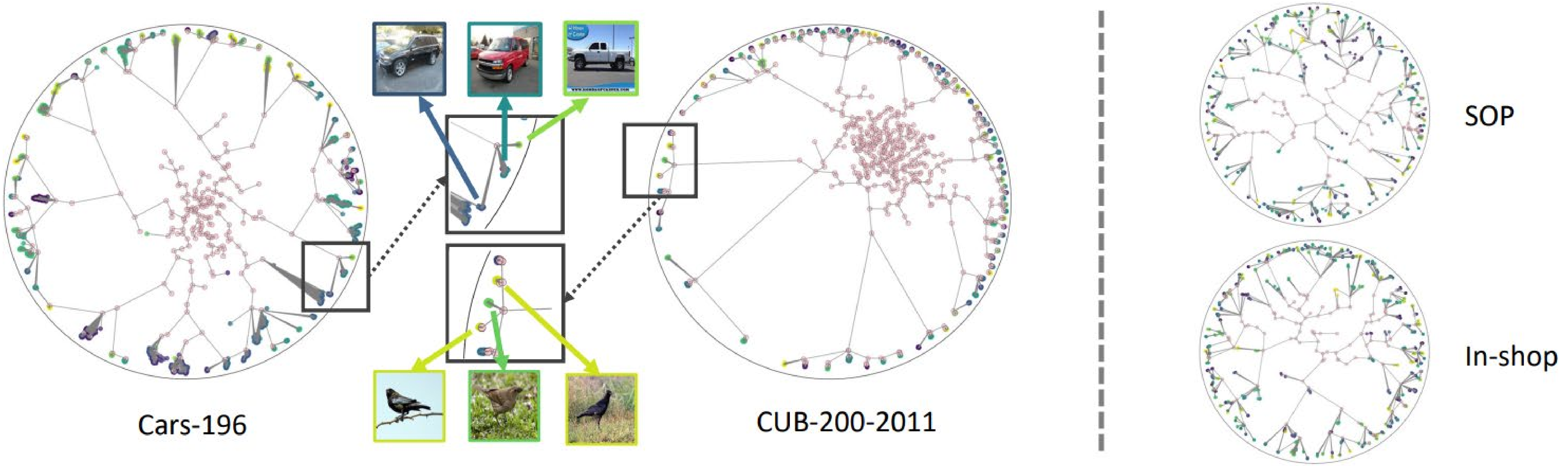
Experiments

- Quantitative results on the image retrieval benchmarks

Methods	Arch.	CUB			Cars			SOP			In-Shop		
		R@1	R@2	R@4	R@1	R@2	R@4	R@1	R@10	R@100	R@1	R@10	R@20
Backbone architecture: CNN													
MS	B^{512}	65.7	77.0	86.3	84.1	90.4	94.0	78.2	90.5	96.0	89.7	97.9	98.5
SoftTriple	B^{512}	65.4	76.4	84.5	84.5	90.7	94.5	78.6	86.6	91.8	-	-	-
PA	B^{512}	68.4	79.2	86.8	86.1	91.7	95.0	79.1	90.8	96.2	91.5	98.1	98.8
NSoftmax	R^{512}	61.3	73.9	83.5	84.2	90.4	94.4	78.2	90.6	96.2	86.6	97.5	98.4
†ProxyNCA++	R^{512}	69.0	79.8	87.3	86.5	92.5	95.7	80.7	92.0	96.7	90.4	98.1	98.8
Hyp	R^{512}	65.5	76.2	84.9	81.9	88.8	93.1	79.9	91.5	96.5	90.1	98.0	98.7
HIER (ours)	R^{512}	70.1	79.4	86.9	88.2	93.0	95.6	80.2	91.5	96.6	92.4	98.2	98.8
Backbone architecture: ViT													
DeiT-S	De^{384}	70.6	81.3	88.7	52.8	65.1	76.2	58.3	73.9	85.9	37.9	64.7	72.1
Hyp	De^{384}	77.8	86.6	91.9	86.4	92.2	95.5	83.3	93.5	97.4	90.5	97.8	98.5
HIER (ours)	De^{384}	78.7	86.8	92.0	88.9	93.9	96.6	83.0	93.1	97.2	90.6	98.1	98.6
DINO	DN^{384}	70.8	81.1	88.8	42.9	53.9	64.2	63.4	78.1	88.3	46.1	71.1	77.5
Hyp	DN^{384}	80.9	87.6	92.4	89.2	94.1	96.7	85.1	94.4	97.8	92.4	98.4	98.9
HIER (ours)	DN^{384}	81.1	88.2	93.3	91.3	95.2	97.1	85.7	94.6	97.8	92.5	98.6	99.0
ViT-S	V^{384}	83.1	90.4	94.4	47.8	60.2	72.2	62.1	77.7	89.0	43.2	70.2	76.7
Hyp	V^{384}	85.6	91.4	94.8	86.5	92.1	95.3	85.9	94.9	98.1	92.5	98.3	98.8
HIER (ours)	V^{384}	85.7	91.3	94.4	88.3	93.2	96.1	86.1	95.0	98.0	92.8	98.4	99.0

Experiments

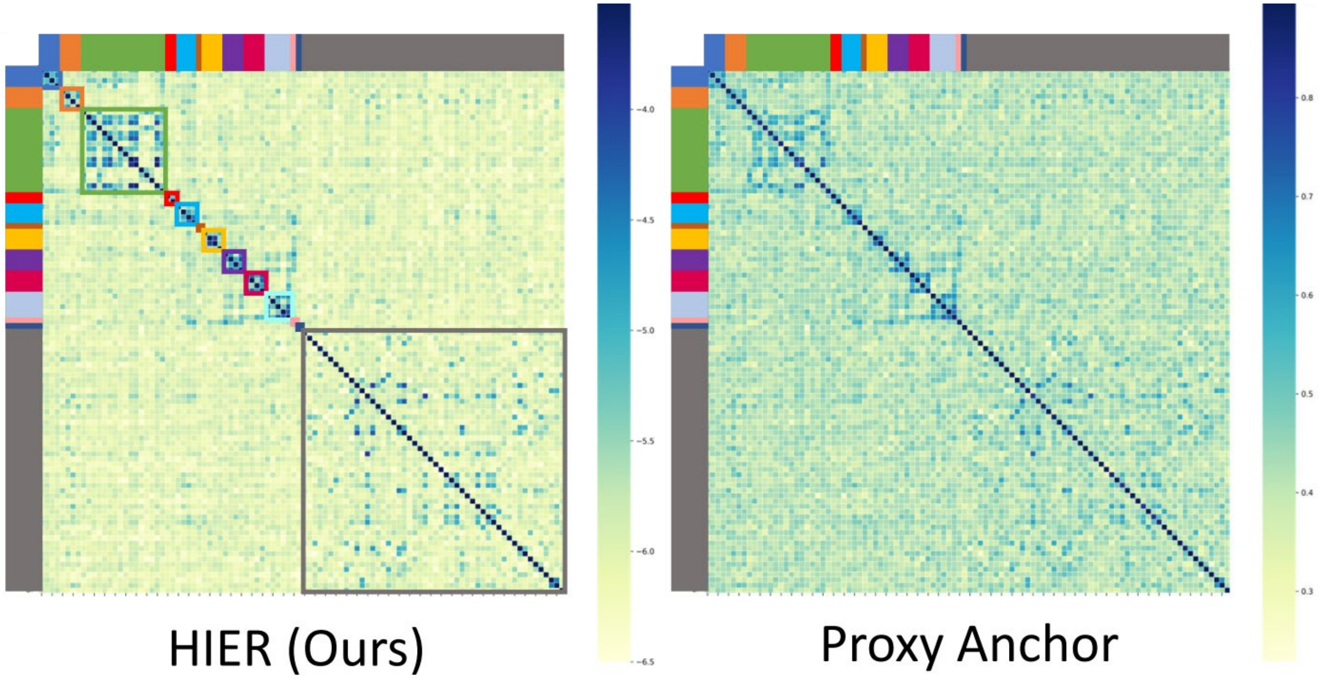
- UMAP visualization of our embedding space



Hierarchical proxies   Samples of distinct classes   Ancestor-descendant relations

Experiments

- Analysis on semantic hierarchy



Class-to-class affinity matrices
[Super-class]

Neighbors of leaf hierarchical proxies
[Sub-class]

Conclusion

- Introducing HIER: a novel hierarchical regularization technique.
- Capturing the semantic hierarchy in a self-supervised learning fashion.
- Consistent performance improvements and state-of-the-art achievements.
- Seamless integration with existing metric learning losses.

Thank you for your attention!

Github Link