CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

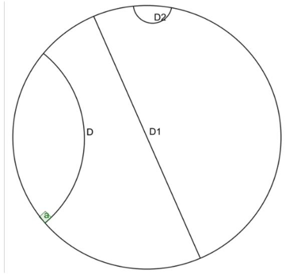
CO-SNE: Dimensionality Reduction & Visualization for Hyperbolic Data



Yunhui Guo   Haoran Guo   Stella X. Yu

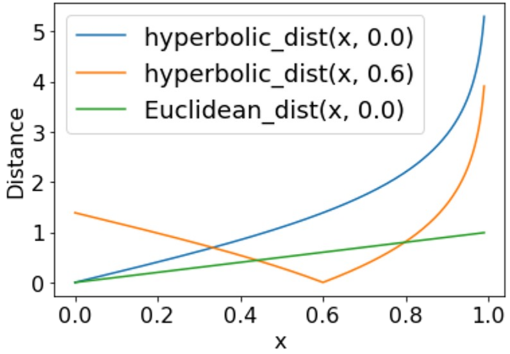
Poincaré Ball of Hyperbolic Space

- Points in hyperbolic space: $\mathbb{B}^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1\}$
- Lines in hyperbolic space:



- Hyperbolic distance:

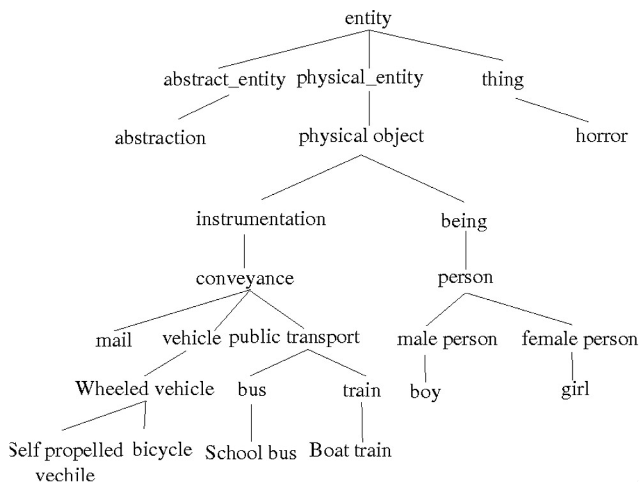
$$d_{\mathbb{B}^n}(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arcosh}\left(1 + 2\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{(1 - \|\boldsymbol{u}\|^2)(1 - \|\boldsymbol{v}\|^2)}\right)$$



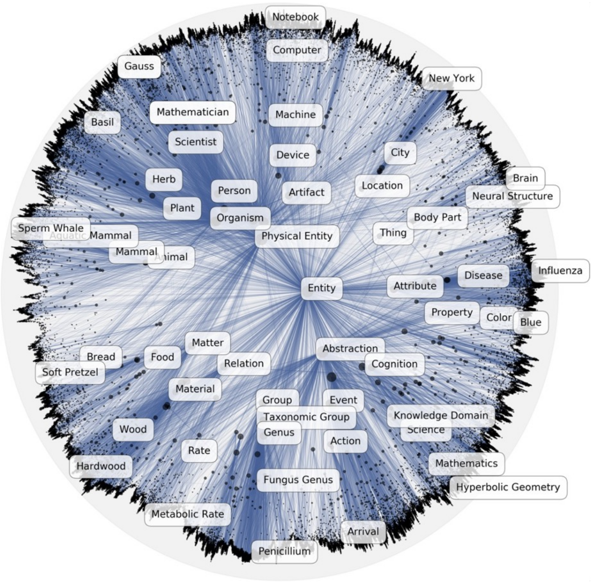
Visualizing 2D Hyperbolic Space is Easy

- Embedding WordNet hierarchy in hyperbolic space

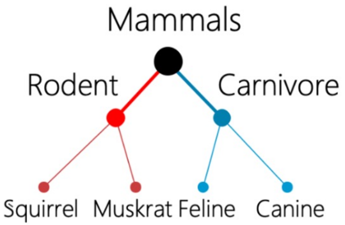
WordNet



Hyperbolic embeddings

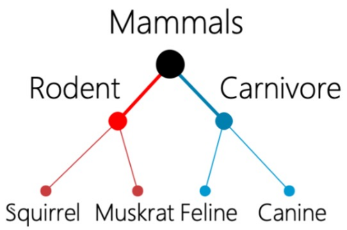


Interpreting High-dimensional Hyperbolic Data



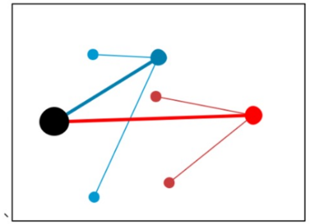
High-Dimensional Poincaré Embeddings

Interpreting High-dimensional Hyperbolic Data

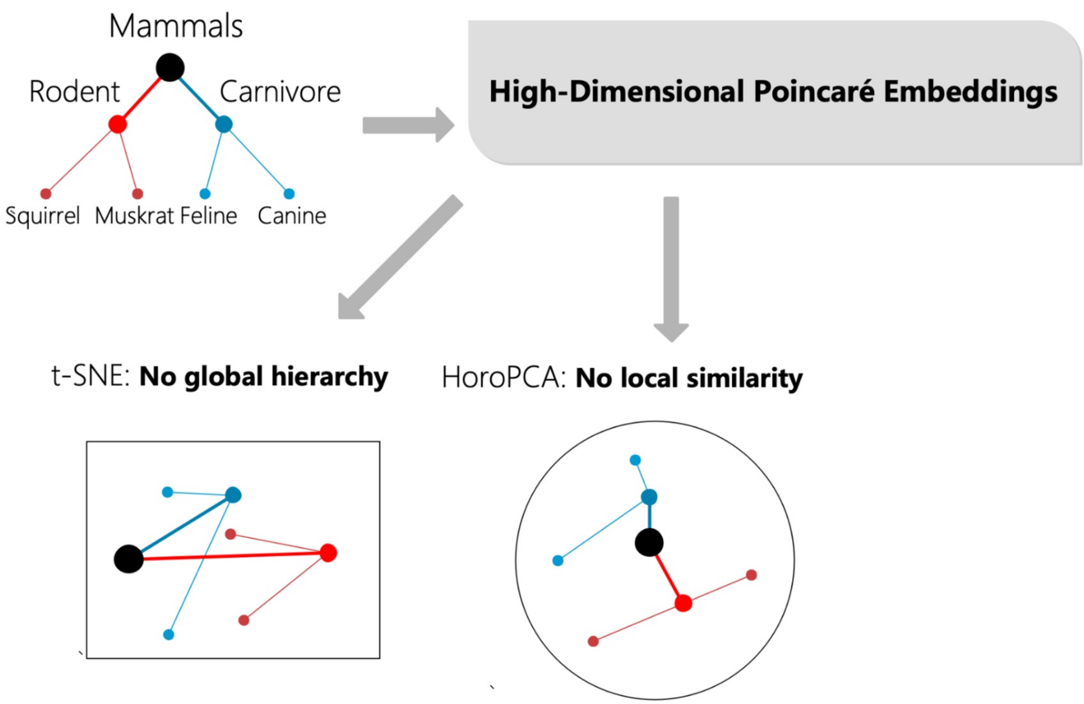


High-Dimensional Poincaré Embeddings

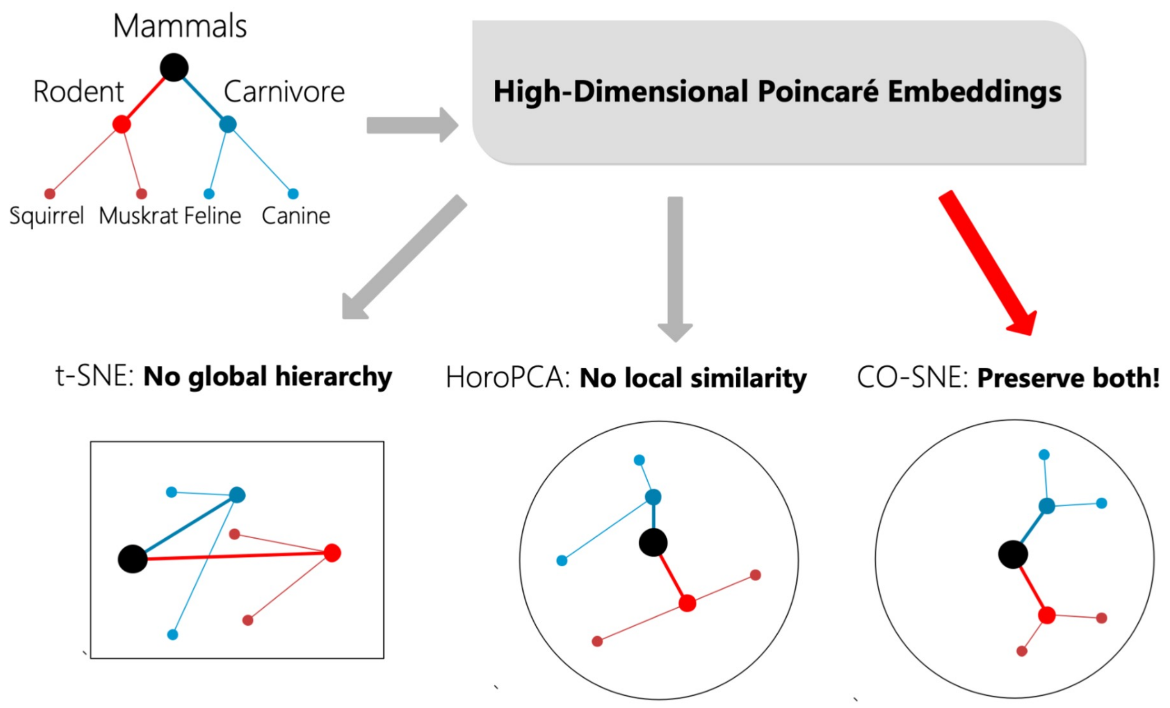
t-SNE: **No global hierarchy**



Interpreting High-dimensional Hyperbolic Data



Interpreting High-dimensional Hyperbolic Data



t-SNE for Visualizing High-Dimensional Euclidean Data

- Given high-dimensional data **x**, seek low-dimensional data **y** such that their distance-based joint probabilities *P* and *Q* match:

$$\mathcal{C} = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

- A normal distribution is used for *P* :

$$p_{j|i} = \frac{\exp(-d(\mathbf{x}_i, \mathbf{x}_j)^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-d(\mathbf{x}_i, \mathbf{x}_k)^2 / 2\sigma_i^2)}$$

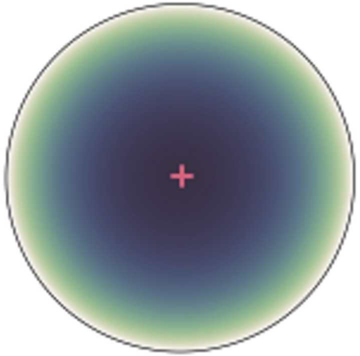
- A heavier tailed *t*-distribution is used for *Q* to make up for smaller volumes in low-dimensional space at the same distance:

$$q_{ij} = \frac{(1 + d(\mathbf{y}_i, \mathbf{y}_j)^2)^{-1}}{\sum_{k \neq l} (1 + d(\mathbf{y}_k, \mathbf{y}_l)^2)^{-1}}$$

Extending t-SNE to Hyperbolic Space

Hyperbolic Normal Distribution

$$\mathcal{N}_{\mathbb{B}^n}(\boldsymbol{x}|\boldsymbol{\mu}, \sigma^2) = \frac{1}{\boldsymbol{Z}} \exp(-\frac{d_{\mathbb{B}^n}(\boldsymbol{\mu}, \boldsymbol{x})^2}{2\sigma^2})$$

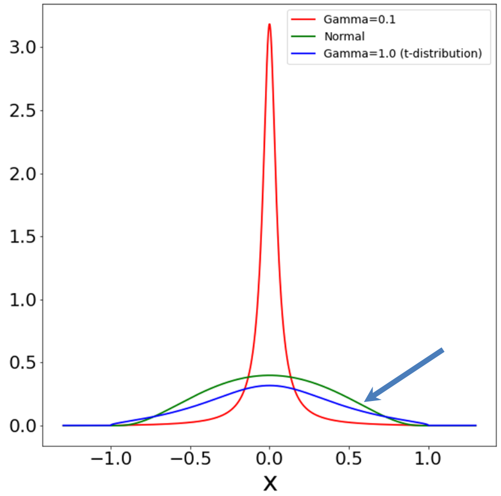


Hyperbolic Student's *t*-distribution

$$f_{\mathbb{B}^n}(t; t_0) = \frac{1}{\pi(1 + d_{\mathbb{B}^n}(t, t_0)^2)}$$

Extension of t-SNE: The Wrong Way

	Metric	Low-dimensional Distribution	Losses
t-SNE	Euclidean	t-distribution	KL-divergence
HT-SNE	Hyperbolic	t-distribution	KL-divergence

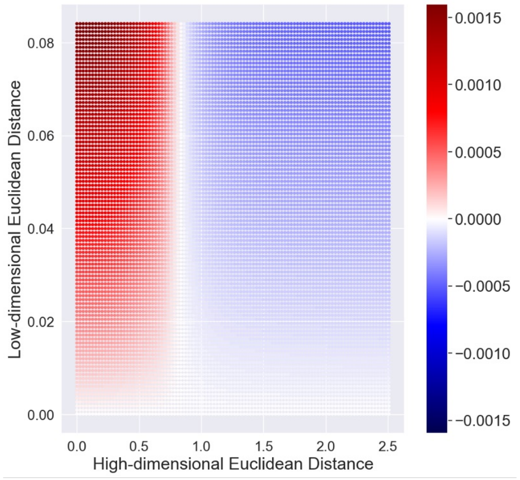


Hyperbolic Student's t-distribution is not heavy-tailed

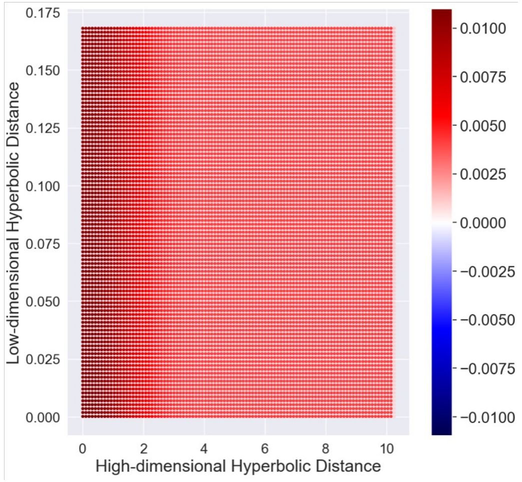
Extension of t-SNE: The Wrong Way

	Metric	Low-dimensional Distribution	Losses
t-SNE	Euclidean	t-distribution	KL-divergence
HT-SNE	Hyperbolic	t-distribution	KL-divergence

t-SNE



HT-SNE

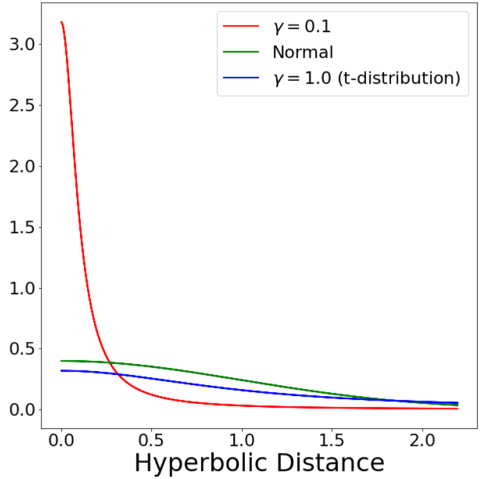


No repulsion between dissimilar high-dimensional points

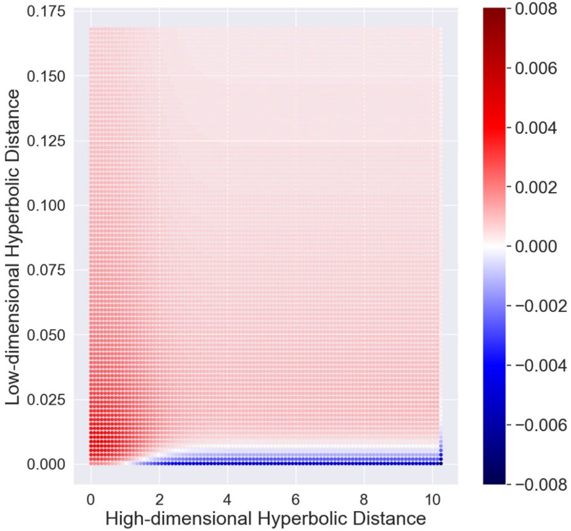
From t-SNE to CO-SNE

	Metric	Low-dimensional Distribution	Losses
t-SNE	Euclidean	t-distribution	KL-divergence
HT-SNE	Hyperbolic	t-distribution	KL-divergence
CO-SNE	Hyperbolic	Cauchy	KL-divergence + Distance

CO-SNE uses a small λ



CO-SNE



Strong repulsion between dissimilar high-dimensional points

Losses in CO-SNE

- Similarity Loss

$$\mathcal{C} = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

→ Preserve local similarity!

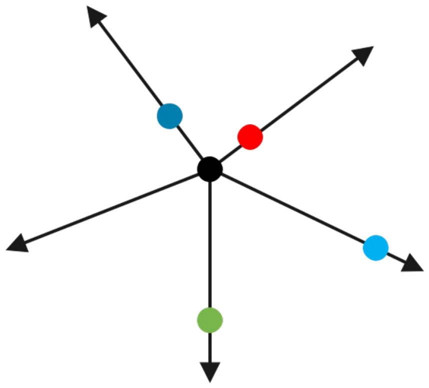
- Distance Loss

$$\mathcal{H} = \frac{1}{m} \sum_{i=1}^{m} (\|\mathbf{x}_i\|^2 - \|\mathbf{y}_i\|^2)^2$$

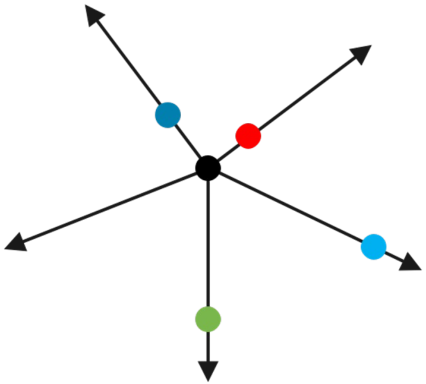
→ Preserve global hierarchy!

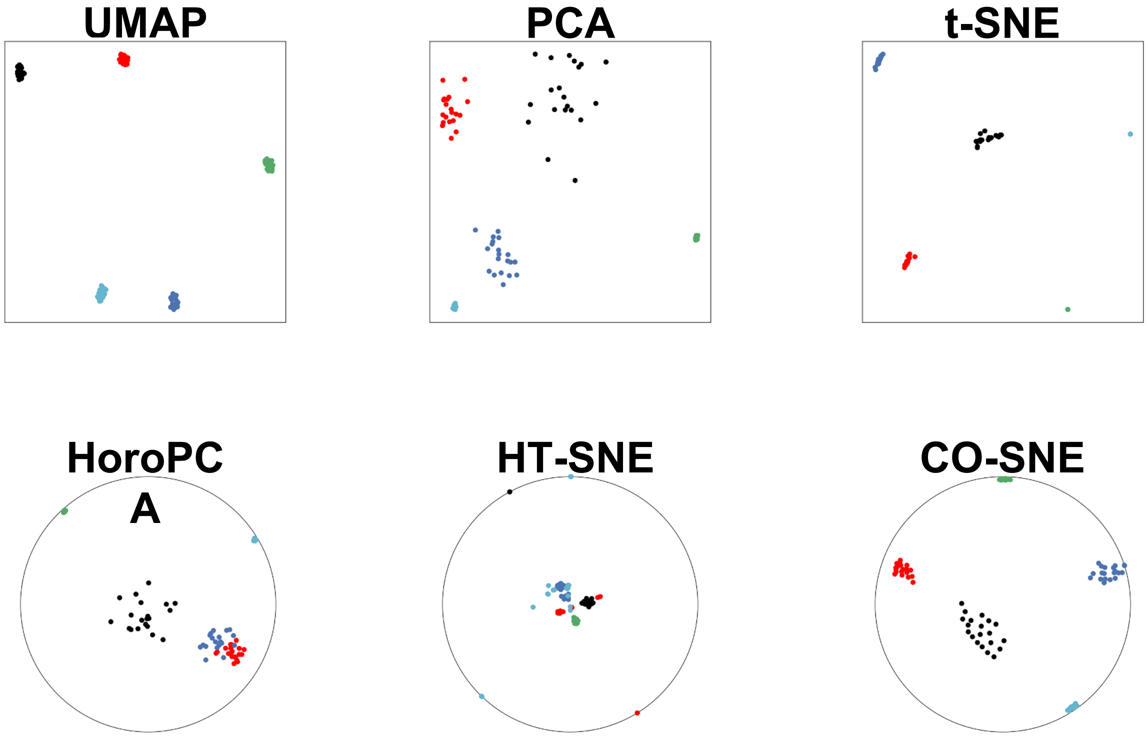
Visualizing Synthetic Point Clouds

- Visualize data in a five-dimensional hyperbolic space
  - Sample data from a mixture of hyperbolic Gaussians with means as,
    - [0, 0, 0, 0, 0] --> origin
    - [0.1, 0, 0, 0, 0]
    - [0, -0.2, 0, 0, 0]
    - [0, 0, 0.9, 0, 0]
    - [0, 0, 0,-0.9, 0]



Visualizing Synthetic Point Clouds





Why Not Existing Methods?

The standard t-SNE and UMAP:

- Underestimates the distance between points that are close to the boundary of the Poincaré ball.

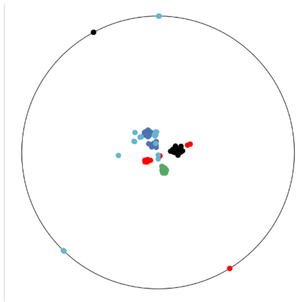
PCA and HoroPCA:

- As a linear dimensionality method, PCA cannot reduce high-dimensional data to two dimensions in a meaningful way for visualization

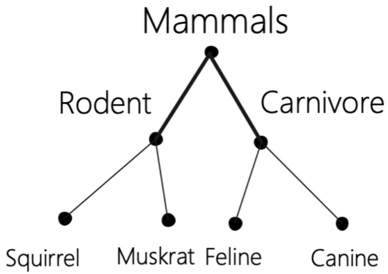
HT-SNE:

- No repulsion between dissimilar high-dimensional points

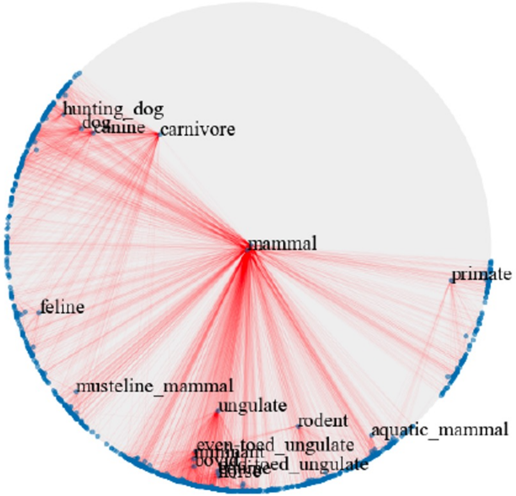
HT-SNE



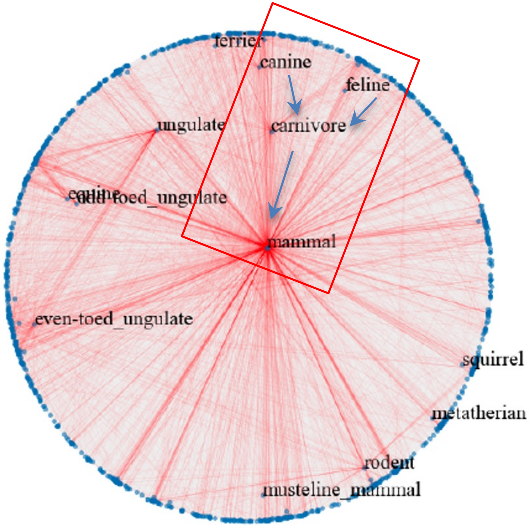
Visualizing High-dimensional Poincaré Word Vectors



HoroPCA

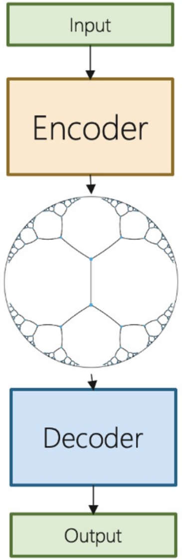


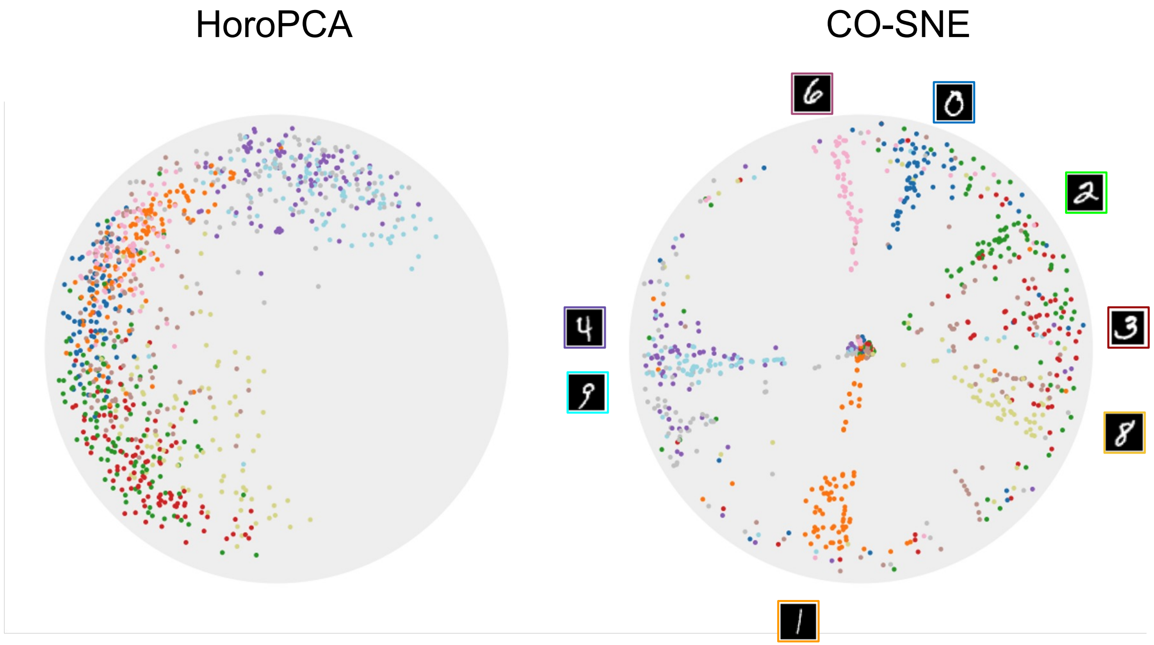
CO-SNE



Nickel, Maximillian, et al. "Poincaré embeddings for learning hierarchical representations." NeurIPS 2017.

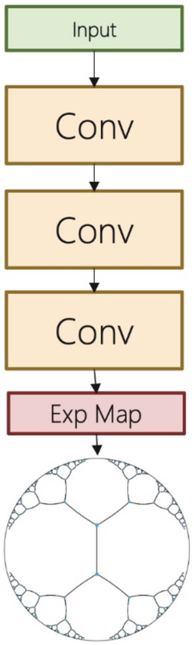
Visualizing Pioncaré Variational Auto-Encoder Features

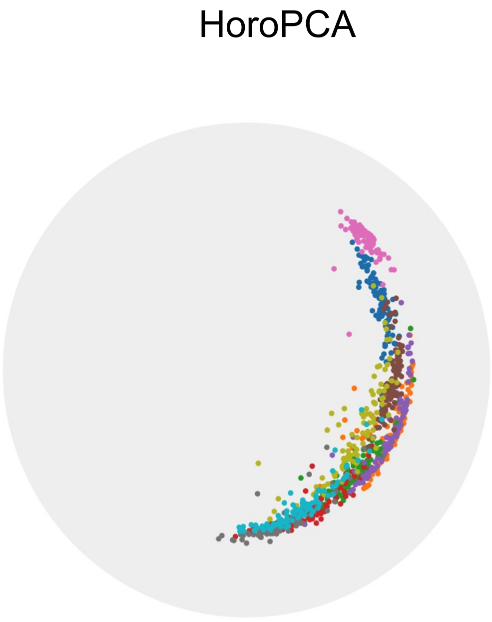


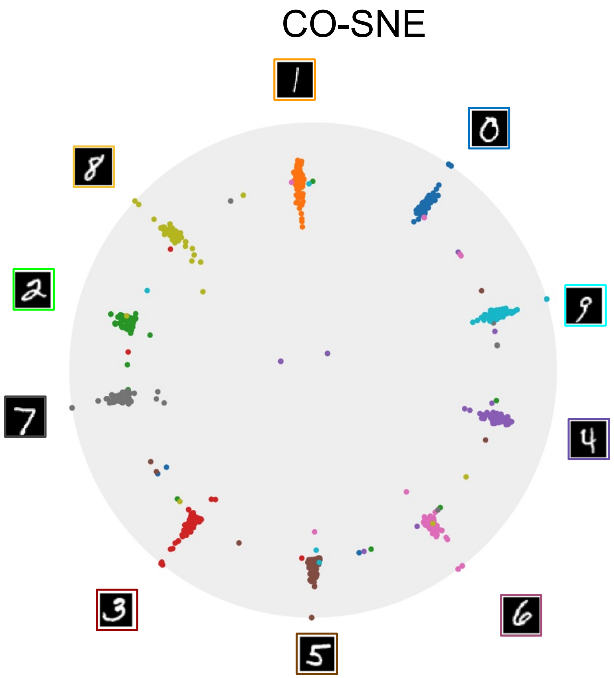


Mathieu, Emile, et al. "Continuous Hierarchical Representations with Poincaré Variational Auto-Encoders." NeurIPS 2019

Visualizing Hyperbolic Neural Net Features







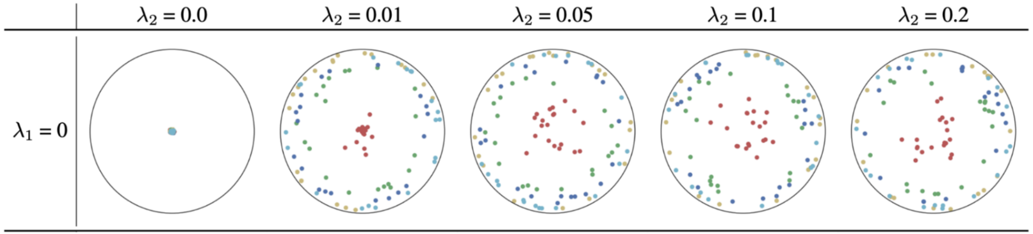
Ganea, Octavian,et al. "Hyperbolic neural networks." NeurIPS 2018.

Impact of t-SNE Loss and Distance Loss

Objective function of CO-SNE: $\mathcal{L} = \lambda_1 \cdot \text{t-SNE Loss} + \lambda_2 \cdot \text{Distance Loss}$

Maintaining global hierarchy

Maintaining local similarity

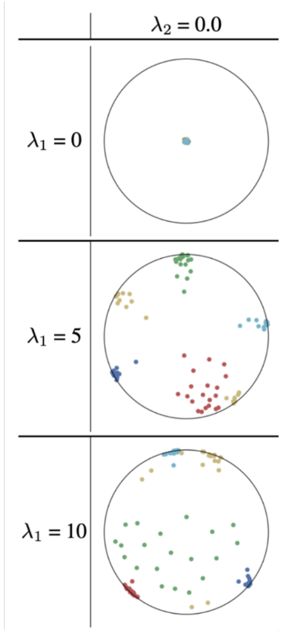


Impact of t-SNE Loss and Distance Loss

Objective function of CO-SNE: $\mathcal{L} = \lambda_1 \cdot \text{t-SNE Loss} + \lambda_2 \cdot \text{Distance Loss}$

Maintaining global hierarchy

Maintaining local similarity

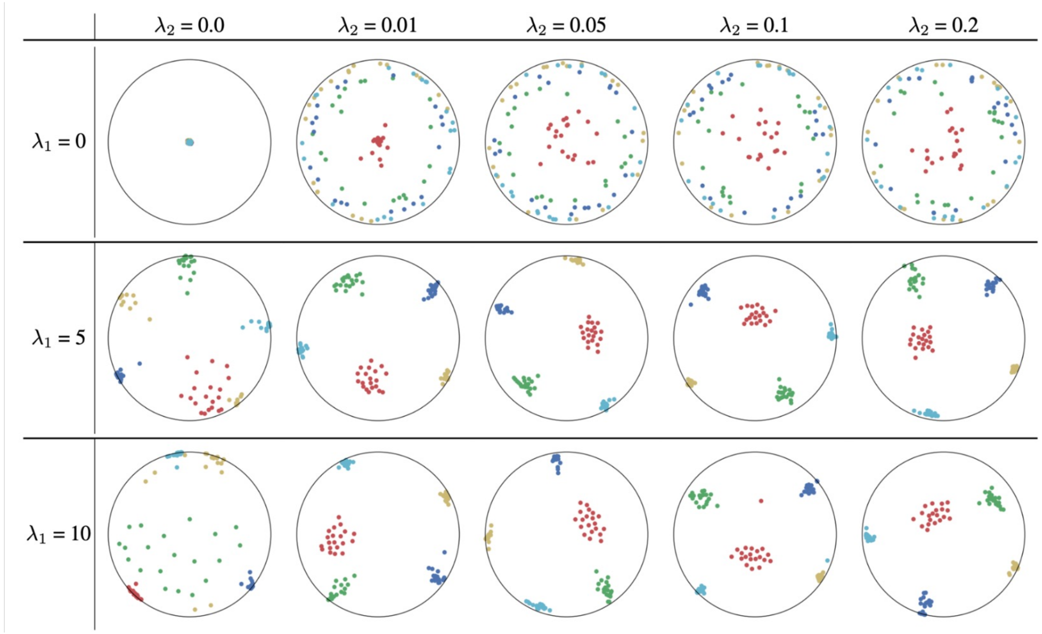


Impact of t-SNE Loss and Distance Loss

Objective function of CO-SNE: $\mathcal{L} = \lambda_1 \cdot \text{t-SNE Loss} + \lambda_2 \cdot \text{Distance Loss}$

Maintaining global hierarchy

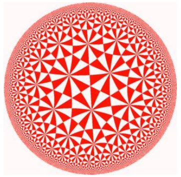
Maintaining local similarity

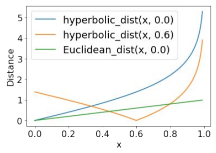


CO-SNE: Dimensionality Reduction & Visualization for Hyperbolic Data

Yunhui Guo   Haoran Guo   Stella X. Yu

Hyperbolic Space





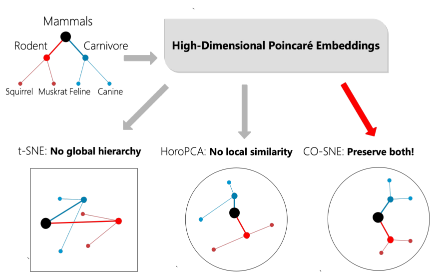
Hyperbolic Distance

$$d_{\mathbb{B}^n}(\boldsymbol{u},\boldsymbol{v}) = \operatorname{arcosh}\left(1+2\frac{\|\boldsymbol{u}-\boldsymbol{v}\|^2}{(1-\|\boldsymbol{u}\|^2)(1-\|\boldsymbol{v}\|^2)}\right)$$

- **Non-Euclidean space with constant negative curvature**
- **Can embed tree-like data continuously with low distortion**

Contributions

CO-SNE: A novel visualization method designed specifically for high-dimensional hyperbolic data



Losses in CO-SNE

Total Loss: $\mathcal{L} = \lambda_1\mathcal{C} + \lambda_2\mathcal{H}$

t-SNE Loss:

$$\mathcal{C} = KL(P||Q) = \sum_i\sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}}$$

Maintaining local similarities

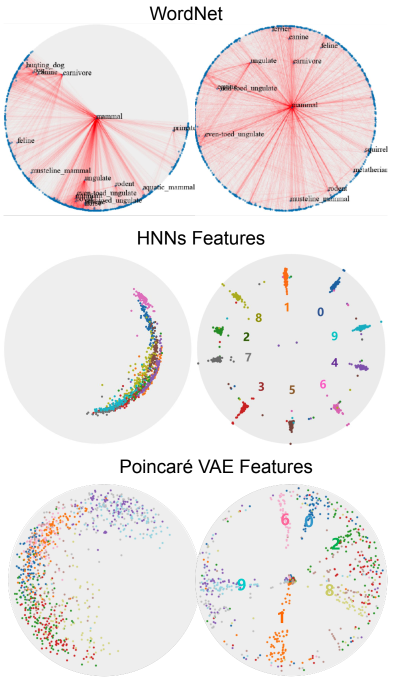
Distance Loss:

$$\mathcal{H} = \frac{1}{m}\sum_{i=1}^{m}(||\mathbf{x}_i||^2 - ||\mathbf{y}_i||^2)^2$$

Maintaining global hierarchy

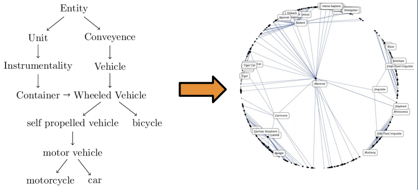
Visualizing Hyperbolic Features

HoroPCA **CO-SNE**



CO-SNE produces much better visualization than HoroPCA

Visualizing Two-Dimensional Hyperbolic Space is Easy

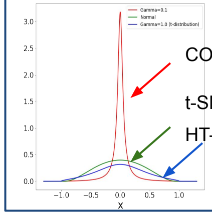


Embedding WordNet in a Two-dimensional Poincaré Ball

CO-SNE Uses Hyperbolic Cauchy Distribution

	Metric	low-dimensional Dist.	Losses
t-SNE	Euclidean	t-distribution	KL-div
HT-SNE	Hyperbolic	t-distribution	KL-div
CO-SNE	Hyperbolic	Cauchy	KL-div + Distance

CO-SNE uses a small γ



Hyperbolic Normal

$$\mathcal{N}_{\mathbb{B}^n}(x|\mu,\sigma^2) = \frac{1}{Z}\exp\left(-\frac{d_{\mathbb{B}^n}(\mu,x)^2}{2\sigma^2}\right)$$

Hyperbolic Cauchy

$$f(t;t_0,\gamma) = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{d_{\mathbb{B}^n}(t,t_0)^2+\gamma^2}\right]$$

CO-SNE Produces Stronger Repulsion Force

The gradients as a function of high-dimensional and low-dimensional hyperbolic distance