北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

Hyperbolic Learning:
Theory and Applications

Presenters: Pengxiang Li[1], Peilin Yu[1], Yangkai Xue[1], Yuwei Wu[1] , Zhi Gao[1]

[1]Beijing Key Laboratory of Intelligent Information Technology, School of Computer Science & Technology, Beijing Institute of Technology, China

德以明理　学以精工

1 Hyperbolic geometry

2 Hyperbolic learning

Hyperbolic geometry

- Why hyperbolic?
- Riemannian manifold
- Hyperbolic Models
- Basic Operation
- δ-Hyperbolicity

Hyperbolic geometry

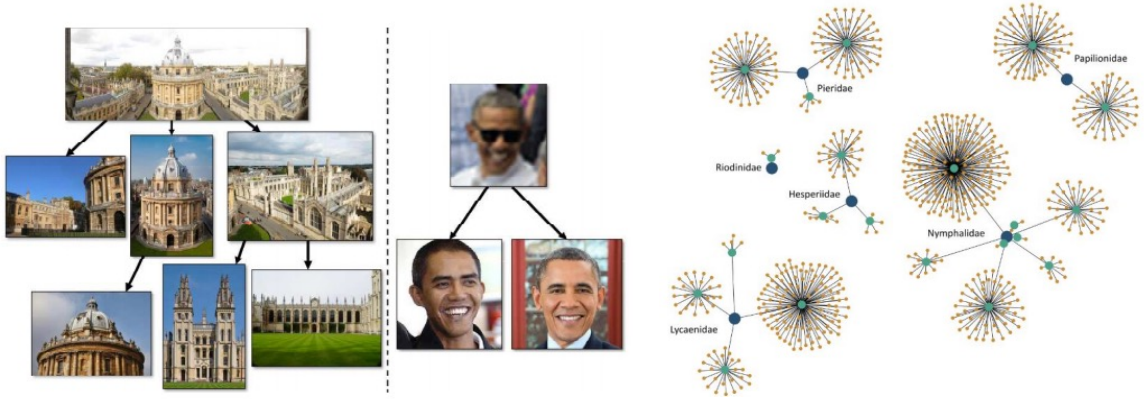
- Why hyperbolic?
- Riemannian manifold
- Hyperbolic Models
- Basic Operation
- δ-Hyperbolicity

Why hyperbolic?

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

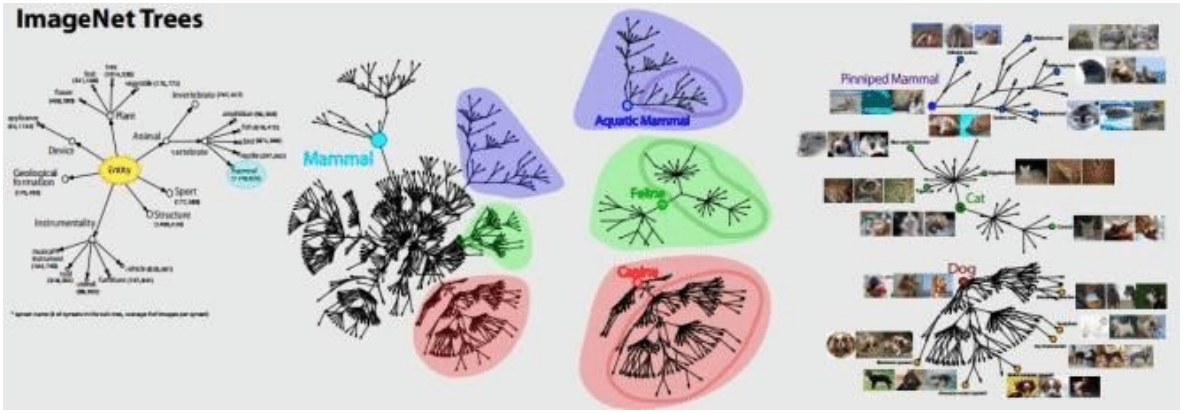
1. High modeling capacity
   - Hierarchical/tree-like structures

Hierarchical data



Visual hierarchies [1] Semantic hierarchies [2]

Hierarchical knowledge [3]

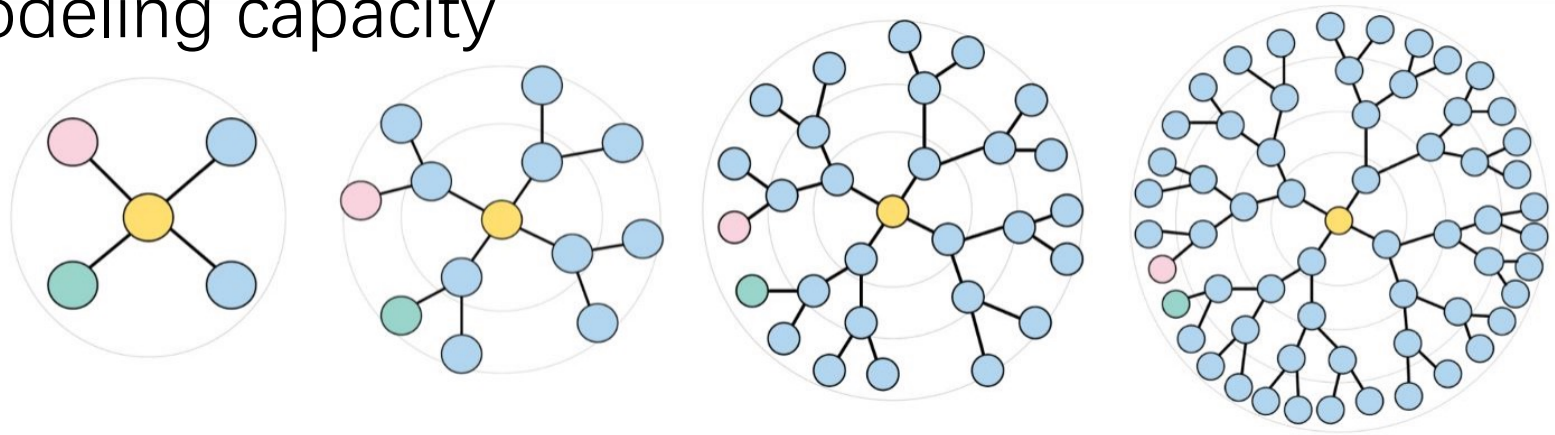


[1] Hyperbolic Image Embeddings, Mirvakhabova et al. CVPR 2020.
[2] Hierarchical Image Classification using Entailment Cone Embeddings. Dhall et al. CVPRw 2020.
[3] Hyperbolic Representation Learning for Computer Vision.Mettes et al. ECCV 2022

Why hyperbolic?

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

1. High modeling capacity



[1]

Hierarchies grow **exponentially** in depth.

Euclidean space grows linearly with norm.

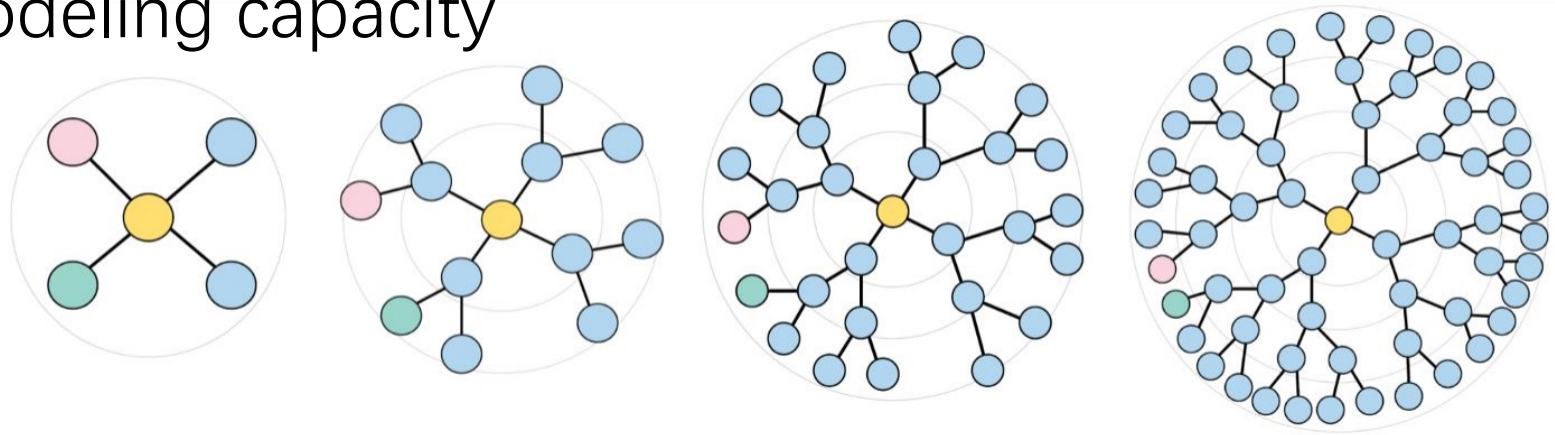
Distances in hyperbolic space grow **exponentially**.

[1] Hyperbolic Representation Learning for Computer Vision.Mettes et al. ECCV 2022

Why hyperbolic?

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

1. High modeling capacity



[1]

Hierarchies grow **exponentially** in depth.

Euclidean space grows linearly with norm.

Distances in hyperbolic space grow **exponentially**.

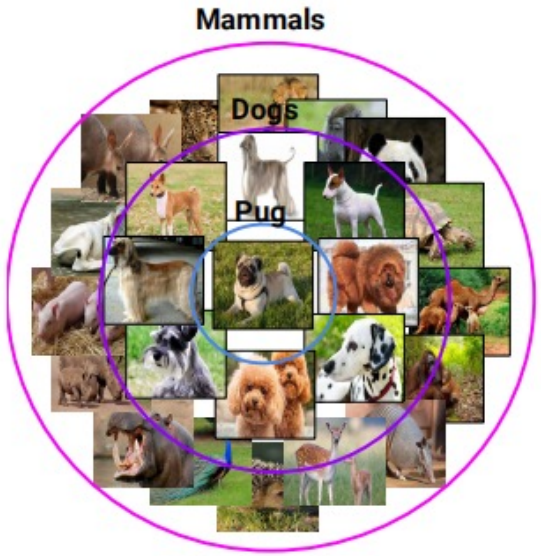
Hyperbolic spaces **more closely resemble hierarchical structures** than Euclidean space.

[1] Hyperbolic Representation Learning for Computer Vision.Mettes et al. ECCV 2022

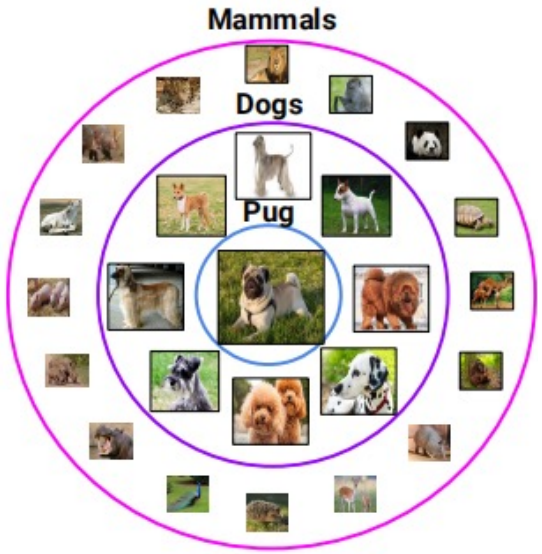
Why hyperbolic?

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

2. More space than Euclidean geometry

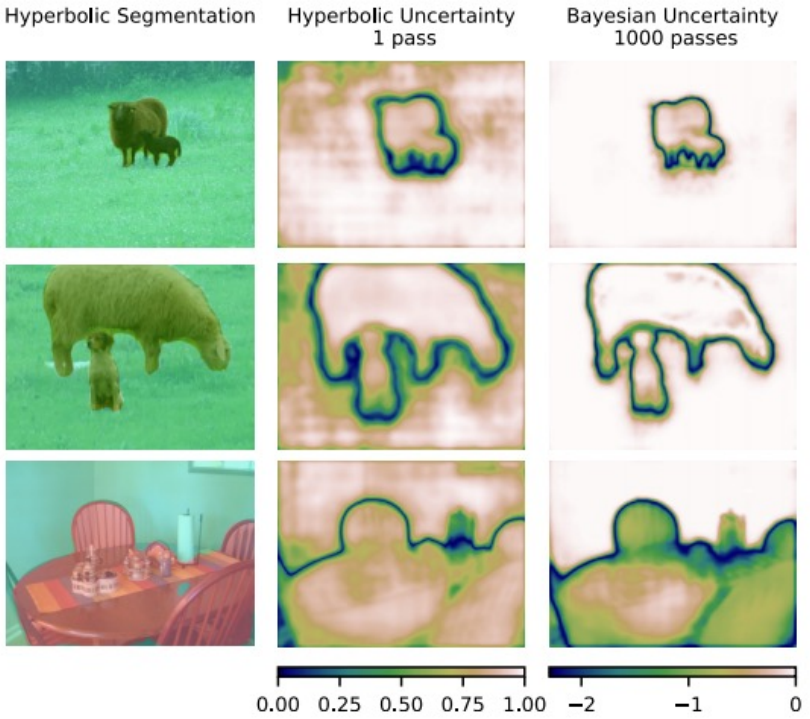


Euclidean Space [1]



Hyperbolic Space [1]

3. Natural measure of uncertainty



Hyperbolic vs Euclidean uncertainty [2]

[1] Hyperbolic Attention Network. Gulcehre et al. ICLR 2019.
[2] Hyperbolic Image Segmentation GhadimiAtigh et al. CVPR 2022.

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

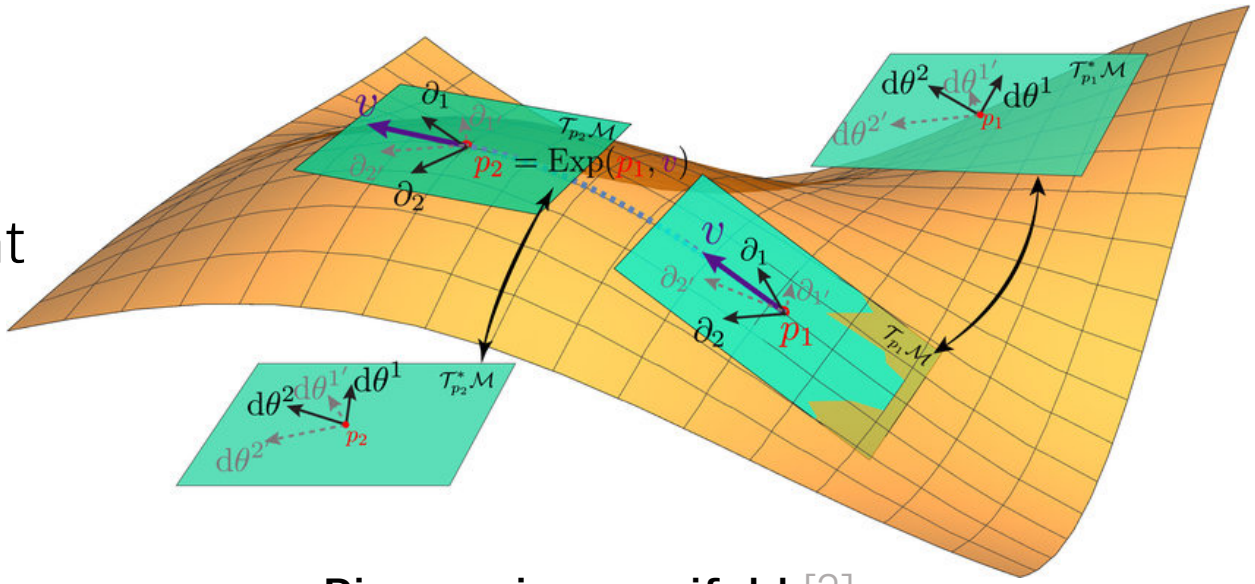
Hyperbolic geometry

- Why hyperbolic?
- Riemannian manifold
- Hyperbolic Models
- Basic Operation
- δ-Hyperbolicity

Riemannian manifold

In differential geometry, a Riemannian manifold or Riemannian space (M, g), is a real, **smooth** manifold M equipped with a positive-definite inner product g_p on the tangent space T_pM at each point p.[1]

The tangent bundle of a smooth manifold M assigns to each point p of M a vector space T_pM called the tangent space of M at p. [1]



Riemannian manifold [2]

[1] https://en.wikipedia.org/wiki/Riemannian_manifold

[2] Accelerating Natural Gradient with Higher-Order Invariance. Song, Yang & Ermon, Stefano. (2018).

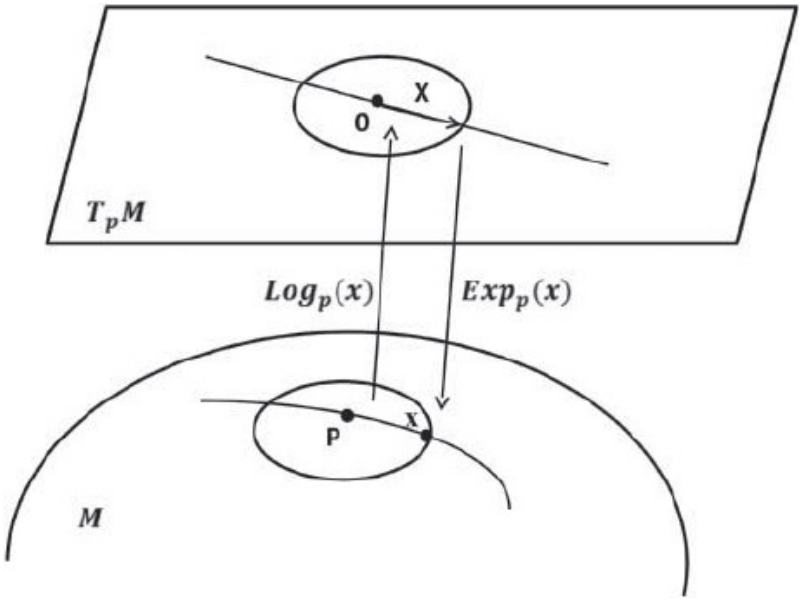
Riemannian manifold

A Riemannian metric (by its definition) assigns to each p a positive-definite inner product $g_p: T_pM \times T_pM \to \mathbb{R}$. The smooth manifold M endowed with this metric g is a Riemannian manifold, denoted (M, g). [1]

$$Exp_p : T_pM \to M$$

$$Log_p : M \to T_pM$$

As a Riemannian manifold, a **hyperbolic space** is a complete contractible **smooth manifold** of constant **negative Riemann curvature** (equivalently, Gaussian curvature).

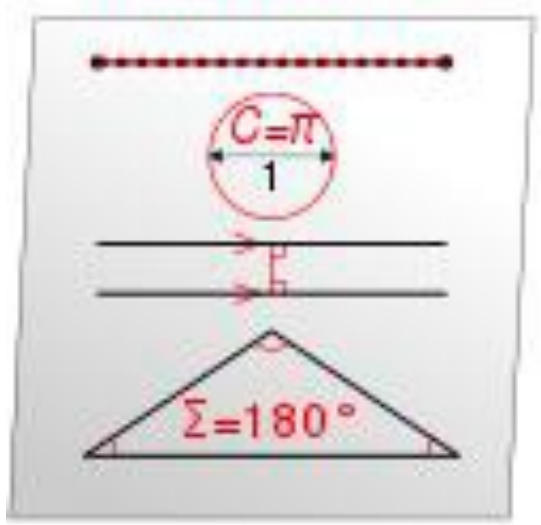


Log and exp mapping [2]

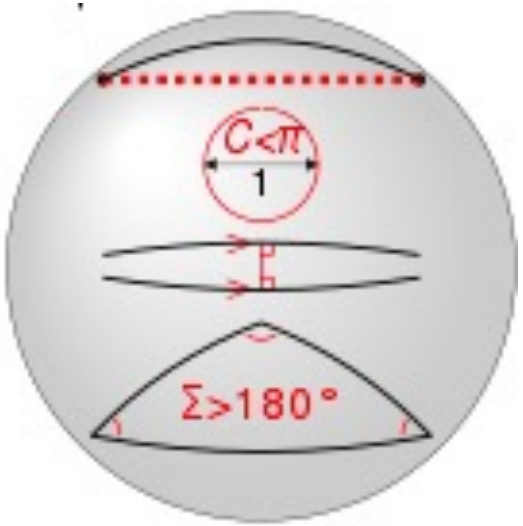
[1] https://en.wikipedia.org/wiki/Riemannian_manifold

[2] Nodehi, Anahita & Golalizadeh, Mousa & Heydari, Abbas. (2015). Dihedral angles principal geodesic analysis using nonlinear statistics. Journal of Applied Statistics.

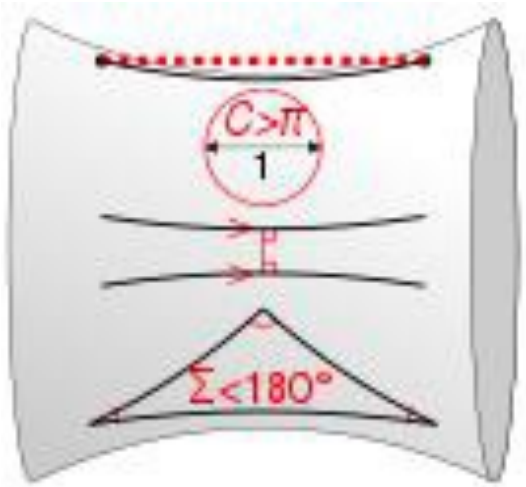
Euclidean Space vs Non-Euclidean Space



Euclidean



Spherical



Hyperbolic

Curvature = 0
Circumference/Diameter = π
One parallel line
Sum of angles in a Δ = 180°

Curvature > 0
Circumference/Diameter > π
No parallel lines
Sum of angles in a Δ > 180°

Curvature < 0
Circumference/Diameter < π
Infinitely many parallel lines
Sum of angles in a Δ < 180°

https://thereaderwiki.com/en/Hyperbolic_geometry

Geodesic in Different Geometries

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Geodesic: Some kind of shortest distance between two points on the manifold. Applies to graphs too.

Hyperbolic geodesic

$$\lambda(t) = \cosh\left(t\sqrt{\kappa}\right)x + \sinh\left(t\sqrt{\kappa}\right)v \qquad \lambda''(t) - \kappa\lambda(t) = 0$$

Euclidean geodesic

$$\lambda(t) = x + tv \qquad \lambda''(t) = 0$$

德以明理 学以精工

[1] Hyperbolic Representation Learning for Computer Vision.Mettes et al. ECCV 2022

Euclidean Space vs Hyperbolic Space

Five postulates for plane geometry.

Euclidean

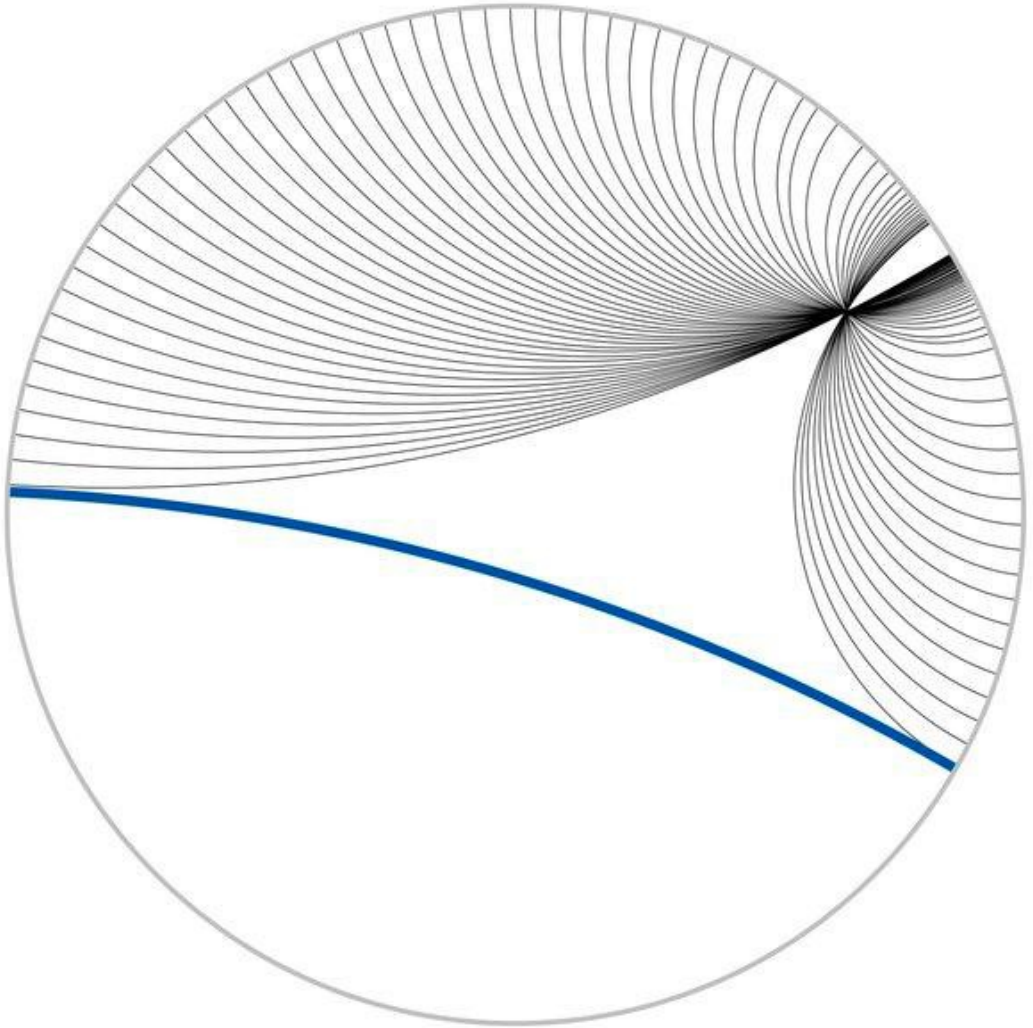
1. Each pair of points can be joined by one and only one straight line segment.
2. Any straight line segment can be indefinitely extended in either direction.
3. There is exactly one circle of any given radius with any given center.
4. All right angles are congruent to one another.
5. **(Parallel Postulate)** In a plane, through a point not on a given straight line, at most one line can be drawn that never meets the given line.

Hyperbolic

1. Each pair of points can be joined by one and only one straight line segment.
2. Any straight line segment can be indefinitely extended in either direction.
3. There is exactly one circle of any given radius with any given center.
4. All right angles are congruent to one another.
5. **(Hyperbolic Alternative)** Given a line and a point not on it, there is **more than** one line going through the given point that does not intersect the given line.

Euclidean Space vs Hyperbolic Space

Five postulates for plane geometry.



德以明理　学以精工

Hyperbolic

1. Each pair of points can be joined by one and only one straight line segment.
2. Any straight line segment can be indefinitely extended in either direction.
3. There is exactly one circle of any given radius with any given center.
4. All right angles are congruent to one another.
5. **(Hyperbolic Alternative)** Given a line and a point not on it, there is **more than** one line going through the given point that does not intersect the given line.

Hyperbolic geometry

- Why hyperbolic?
- Riemannian manifold
- Hyperbolic Models
- Basic Operation
- δ-Hyperbolicity

Hyperbolic function

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

$$\operatorname{arsinh} x = \ln(x + \sqrt{x^2 + 1})$$

$$\operatorname{arcosh} x = \ln(x \pm \sqrt{x^2 - 1})$$

$$\operatorname{artanh} x = \frac{1}{2} \ln\left(\frac{1 + x}{1 - x}\right)$$

The five isometric models

- $\mathbb{L}$: **Lorentz Model**
- $\mathbb{B}$: **Poincaré Model**
- $\mathbb{K}$: Klein Model
- $\mathbb{H}$: Poincaré Half Model
- $\mathbb{J}$: Hemisphere model

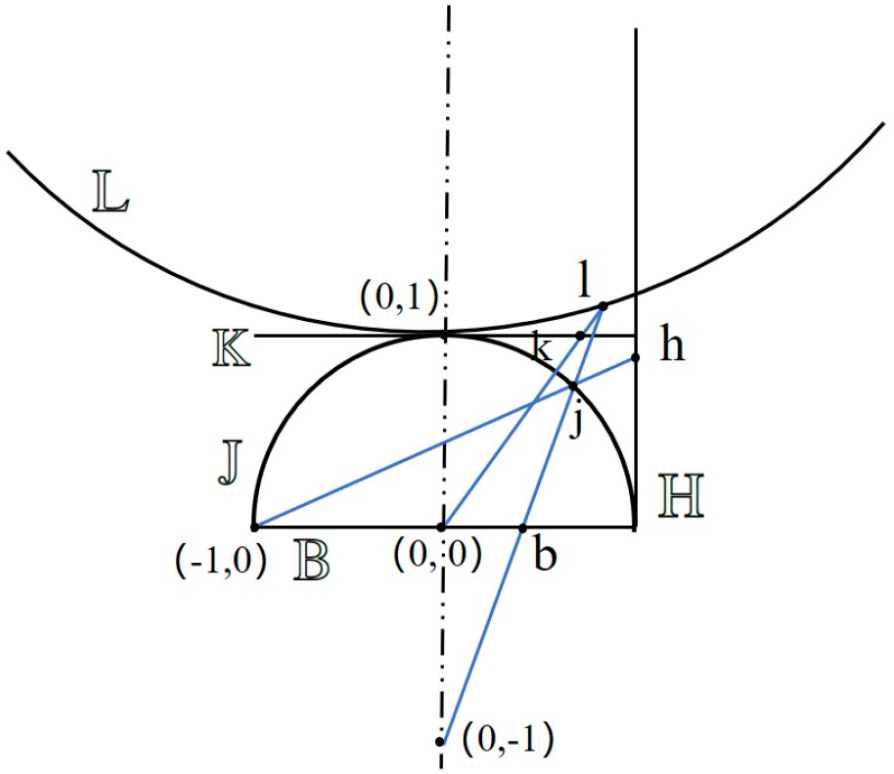


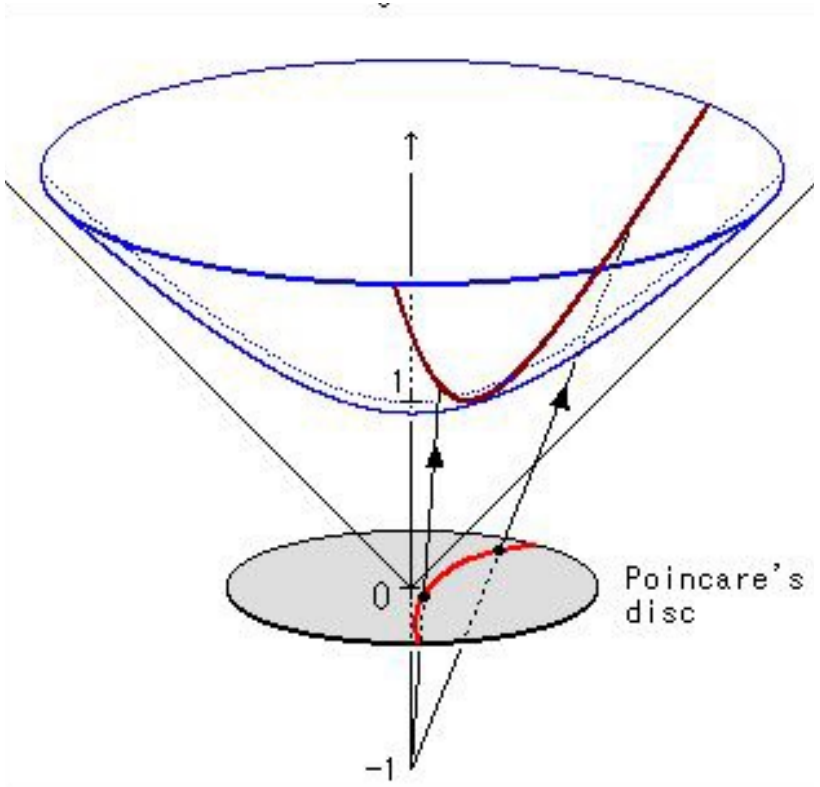
Illustration of five isometric model[1]

[1] Hyperbolic Deep Neural Networks: A Survey. Wei Peng et al. TPAMI 2022.

Isometric Models

Lorentz Model⇔Poincaré Model

$$x = (x_0, \ldots, x_n) \in \mathbb{L}^n \Leftrightarrow \left(\frac{x_1}{1 + x_0}, \ldots, \frac{x_n}{1 + x_0} \right) \in \mathbb{B}^n$$



Isometric Models

Poincaré Model⇔Poincaré Half Model

$$x = (x_0, \ldots, x_{n-1}) \in \mathbb{B}^n \Leftrightarrow \left(\frac{1 - ||x||^2, 2x_1, \ldots, 2x_{n-1}}{1 + 2x_0 + ||x||^2} \right) \in \mathbb{H}^n$$

Lorentz Model ⇔ Klein Model

$$x = (x_0, \ldots, x_n) \in \mathbb{L}^n \Leftrightarrow \left(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0} \right) \in \mathbb{K}^n$$

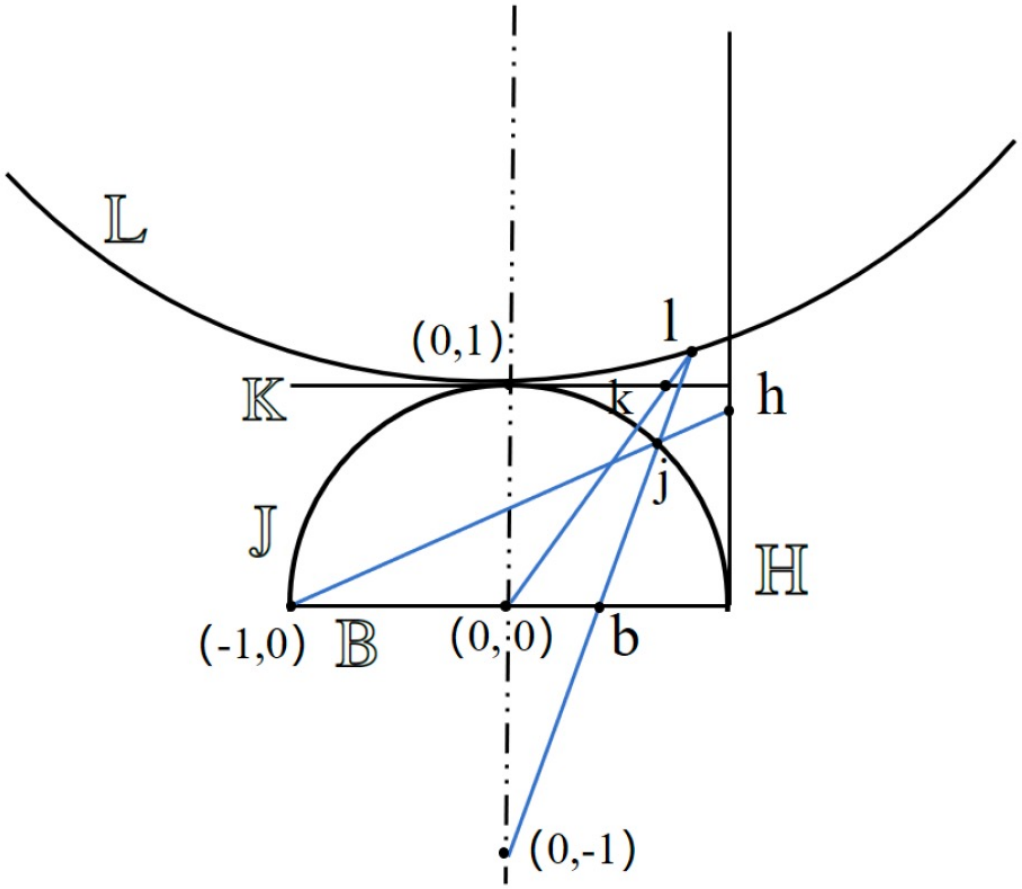


Illustration of five isometric model[1]

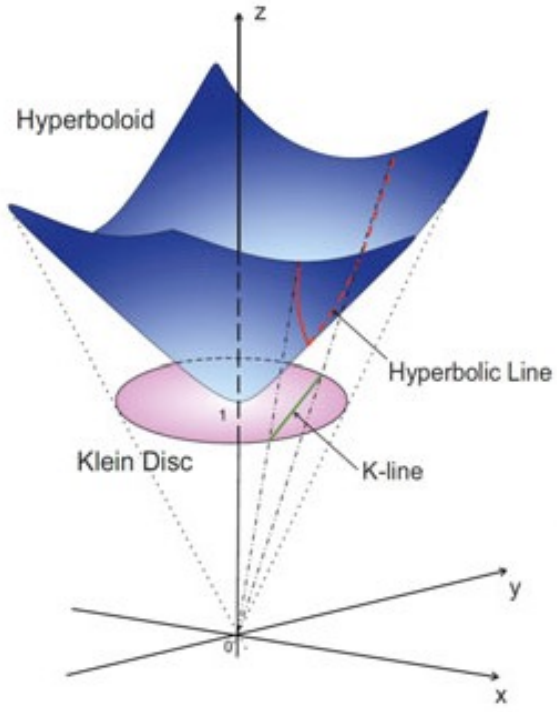
[1] Hyperbolic Deep Neural Networks: A Survey. Wei Peng et al. TPAMI 2022.

Isometric Models

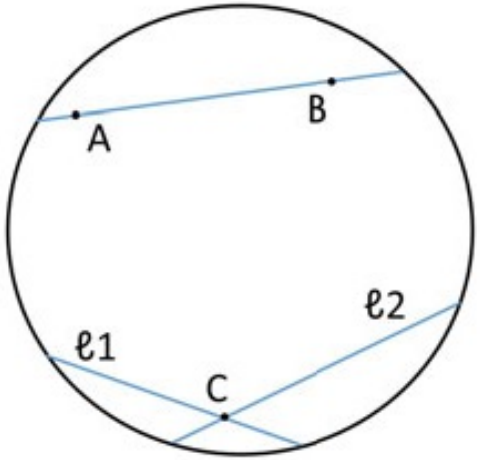
北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Poincaré Model ⇔ Klein Model

Lorentz Model

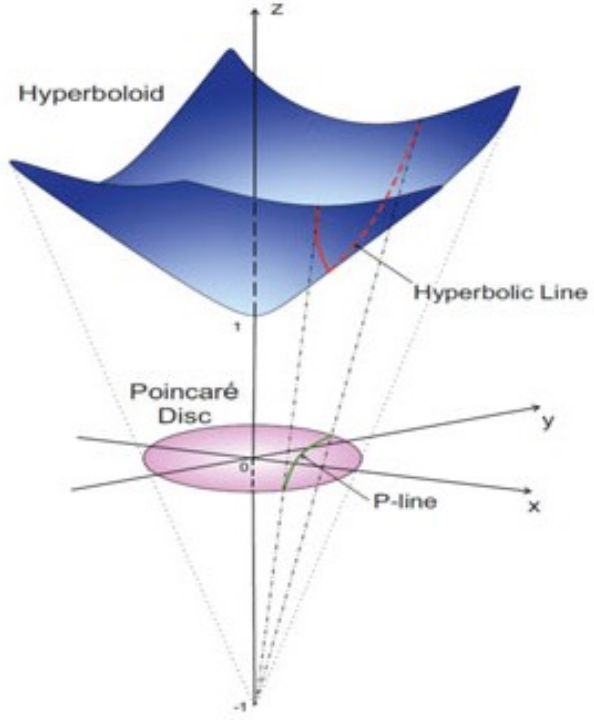


[1]

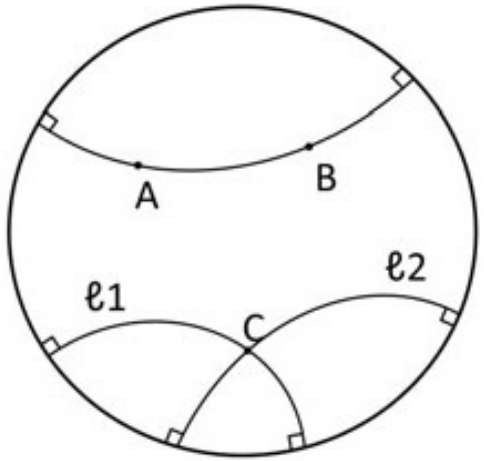


Klein Model

Lorentz Model



[1]



Poincaré Model

德以明理 学以精工

[1] Hyperbolic Deep Neural Networks: A Survey. Wei Peng et al. TPAMI 2022.

The five isometric models

- $\mathbb{L}$: **Lorentz Model**
- $\mathbb{B}$: **Poincaré Model**
- $\mathbb{K}$: Klein Model
- $\mathbb{H}$: Poincaré Half Model
- $\mathbb{J}$: Hemisphere model

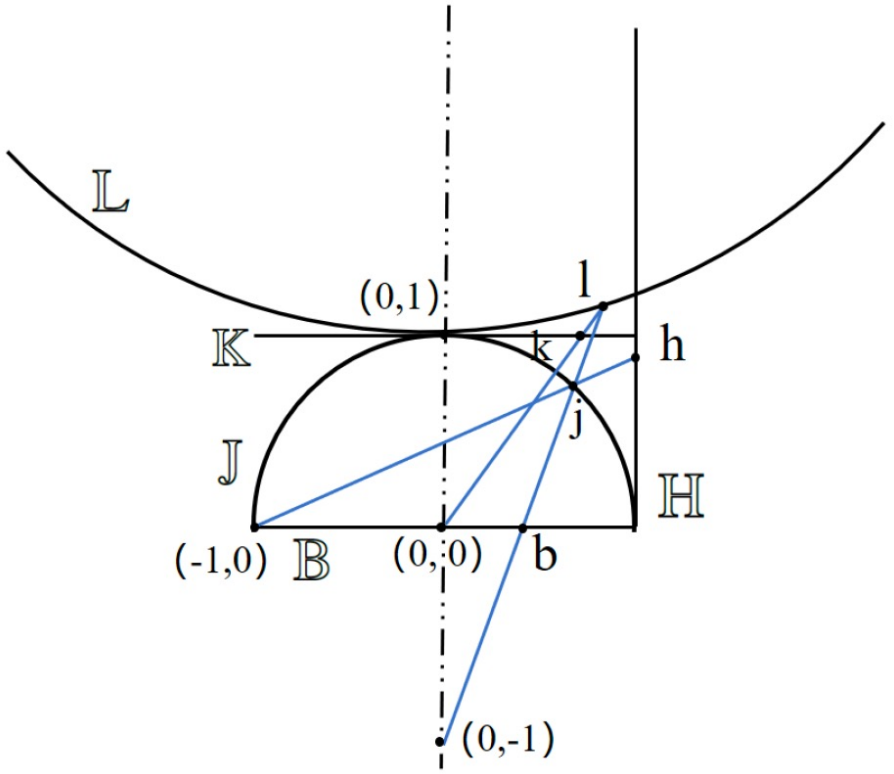


Illustration of five isometric model[1]

[1] Hyperbolic Deep Neural Networks: A Survey. Wei Peng et al. TPAMI 2022.

Lorentz Model

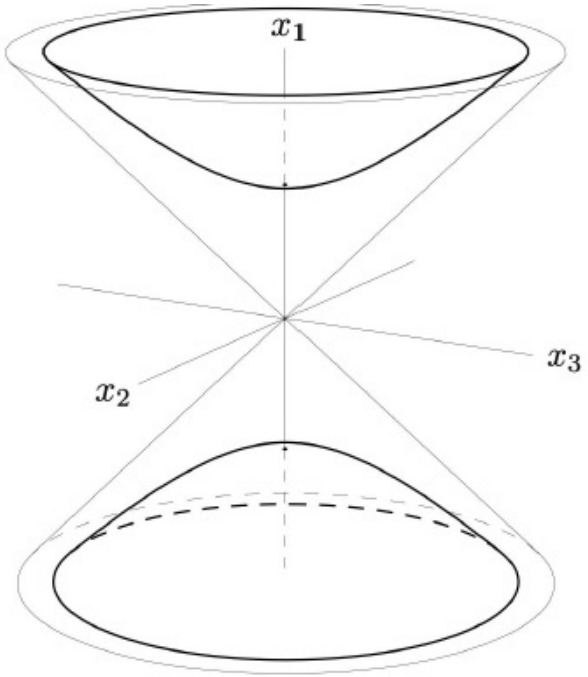
The Lorentz model $\mathbb{L}^n$ of an n dimensional hyperbolic space is a manifold embedded in the $n+1$ dimensional Minkowski space.

$$\mathbb{L}^n = \{x = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} : < x, x >_{\mathbb{L}} = -1, x_0 > 0\},$$

in which $< x, x >_{\mathbb{L}}$ is the Lorentzian inner product

$$\langle x, y \rangle_{\mathbb{L}} = x^T \mathfrak{g}^{\mathbb{L}} y = -x^0 y^0 + \sum_{i=1}^{n} x^i y^i, x \text{ and } y \in \mathbb{R}^{n+1},$$

where $\mathfrak{g}^{\mathbb{L}}$ is a diagonal matrix with entries of 1s, except for the first element being -1.



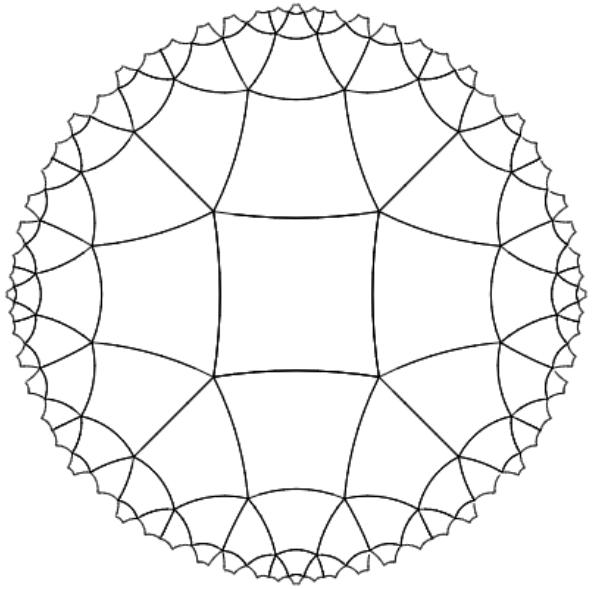
$$d(x, y) = \operatorname{arcosh}(- < x, y >_{\mathbb{L}}).$$

Poincaré Model

$$\mathbb{B}^n = \{x \in \mathbb{R}^n : ||x|| < 1\},$$

A Poincaré model is a manifold equipped with a Riemannian metric $\mathbf{g}^B = \boldsymbol{\lambda}_x^2 \, \mathbf{g}^B$, where $\boldsymbol{\lambda}_x = \frac{2}{1-||x||^2}$.

The Poincaré model $\mathbb{B}^n$ is obtained by mapping $x \in \mathbb{L}^{n+1}$ to the hyperplane $x_0 = 0$, using rays emanating from the origin.

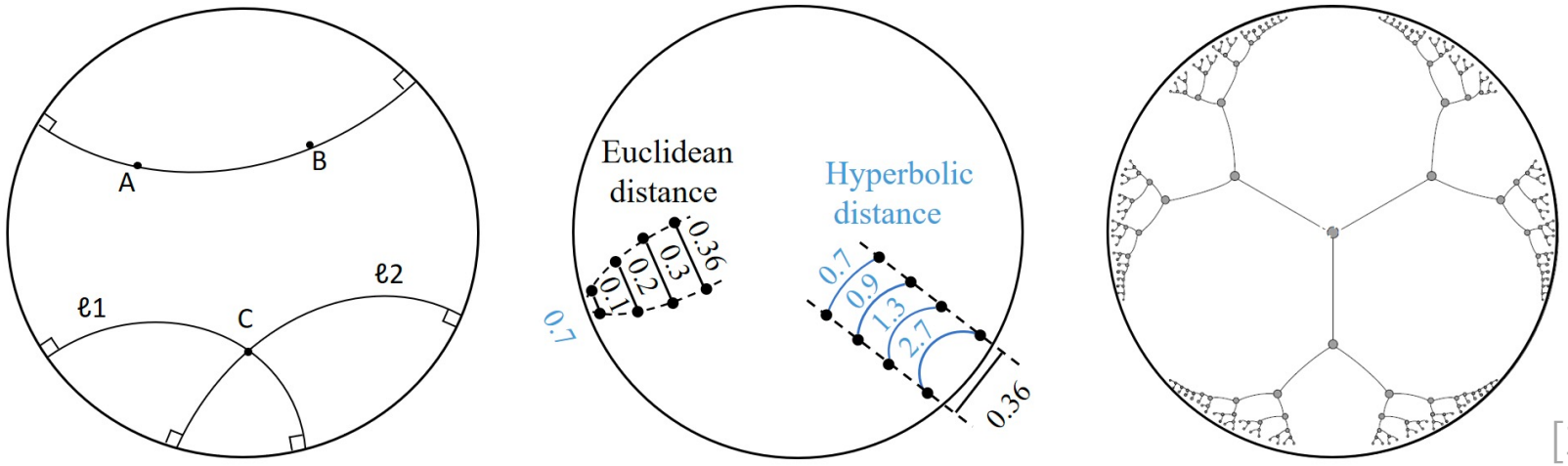


https://wxyhly.github.io/three/HyperbolicSpace.html

Poincaré Model

The distance between $p, q \in \mathbb{B}^n$ is defined as

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



[1]

[1] Hyperbolic Deep Neural Networks: A Survey. Wei Peng et al. TPAMI 2022.

Most commonly used two models.

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Lorentz Model

+ has linear structure (Translations and other isometries are linear maps)
+ simpler formulas for geodesics, distance, etc.

- needs one redundant dimension
- not as easy to visualize

Poincaré Model

+ visually attractive
+ conformal model (Euclidean angle = hyperbolic angle)
+ geodesic lines are Euclidean circles or line segments

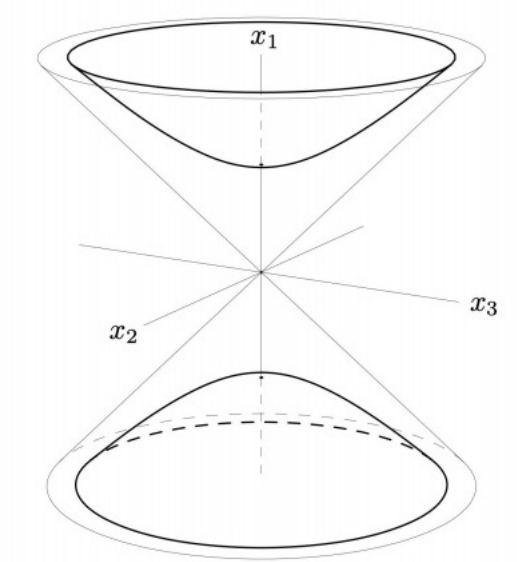
- no linear structure
- formulas for geodesics, distance, etc. not as nice, but: gyrovectorspace calculus

[1] Hyperbolic Representation Learning for Computer Vision.Mettes et al. ECCV 2022

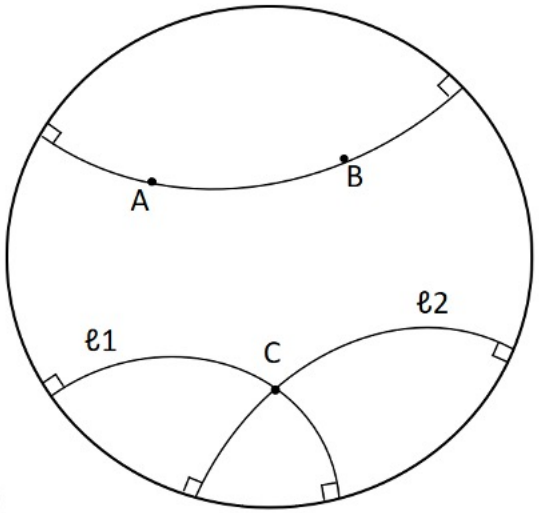
Most commonly used two models.

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Lorentz Model



Poincaré Model



Klein Model

The Klein model $\mathbb{K}^n$ is obtained by mapping $x \in \mathbb{L}^{n+1}$ to the hyperplane $x_0 = 1$, using rays emanating from the origin.

$$\mathbb{K}^n = \{x \in \mathbb{R}^n : ||x|| < 1\},$$

The distance betwen $x, y \in \mathbb{K}^n$ is

$$d(x, y) = \operatorname{arcosh}\left(1 + \frac{(y_0 - x_0)^2 + (y_1 - x_1)^2}{2x_1 y_1}\right).$$

Poincaré Half Model

A Poincaré half model is a manifold equipped with a Riemannian metric $(\mathbb{H}^n, \mathfrak{g}^H)$, where

$$\mathbb{H}^n = \{x \in \mathbb{R}^n : x_n > 0\} \qquad \mathfrak{g}^H = \frac{\mathfrak{g}^E}{x_n^2}.$$

The Poincaré half model can be obtained by taking the inverse of the Poincaré model and the discance for $x, y \in \mathbb{H}^n$ is

$$d(x, y) = \operatorname{arcosh}\left(1 + \frac{||x - y||^2}{2x_n y_n}\right).$$

Hemisphere Model

The hemisphere model is generally used to visualize the transformation between various models.

$$\mathbb{J}^n = \{x = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} : ||x|| = 1, x_0 > 0\},$$

Hyperbolic geometry

- Why hyperbolic?
- Riemannian manifold
- Hyperbolic Models
- Basic Operation
- δ-Hyperbolicity

Basic Operation in Hyperbolic Space

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- Exponential map
- Logarithmic map
- Möbius addition
- Möbius scalar multiplication
- Möbius vector multiplication
- Einstein midpoint

To keep the hyperbolicity of the output.

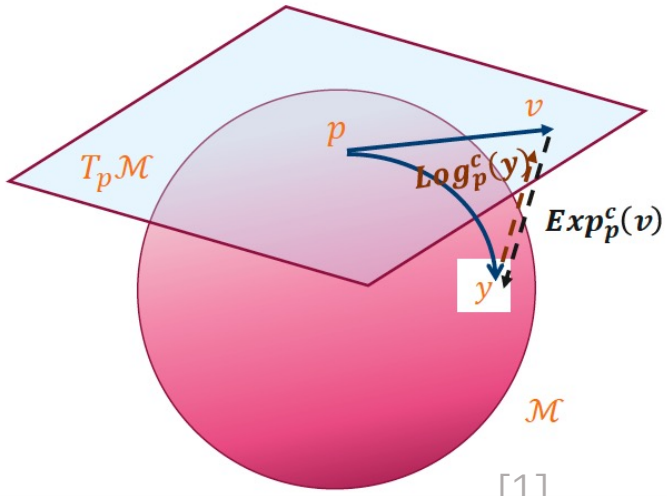
Exponential map

The exponential map takes a vector $\mathbf{v} \in \boldsymbol{\mathcal{T}_x\mathcal{M}}$ of a point $x \in \mathcal{M}$ to a point on the manifold $\boldsymbol{\mathcal{M}}$.

$$\mathrm{Exp}_x : \mathcal{T}_x\mathcal{M} \rightarrow \mathcal{M}$$

For a vector $\mathbf{v} \in \boldsymbol{\mathcal{T}_x\mathcal{M}}$ in the tangent space, the exponential r is defined as

$$\mathrm{Exp}_x(v) = x \oplus \left(\tanh\left(\frac{\lambda_x ||v||}{2}\right) \frac{v}{||v||}\right),$$



[1]

[1] Hyperbolic Networks: Theory, Architectures and Applications. Department of Computer Science, Virginia Tech, Arlington, VA.

Logarithmic map

The logarithmic map is the inverse of the aforementioned exponential map, which projects a point $y \in \mathcal{M}$ on the manifold to the tangent space of another point x $\in \mathcal{M}$.

$$\mathrm{Log}_x : \mathcal{M} \rightarrow \mathcal{T}_x\mathcal{M}$$

As the inverse operation of the exponential map, for a point $y \in \mathbb{B}^n$ on the poincaré model, the logarithmic map is defined as

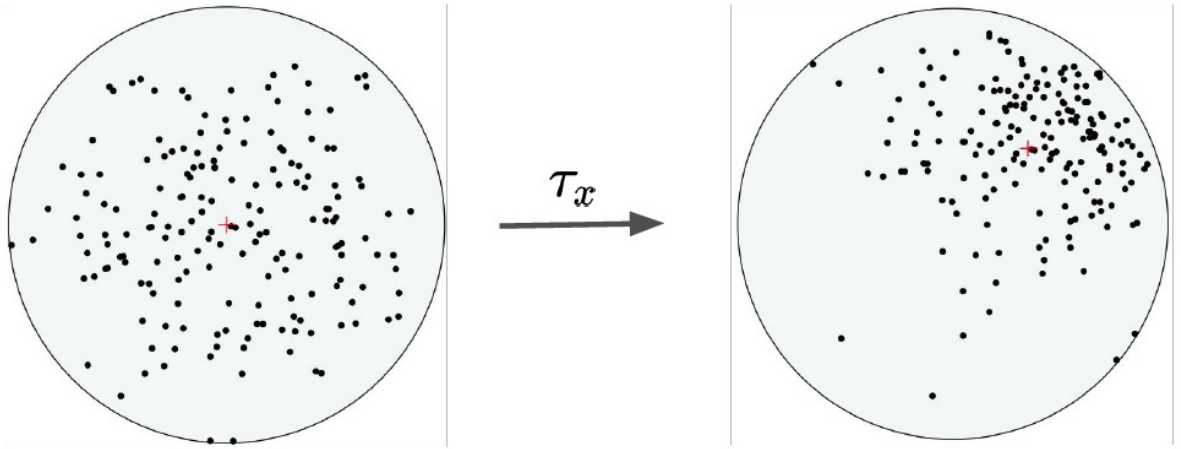
$$\mathrm{Log}_x(y) = \frac{2}{\lambda_x} \operatorname{artanh}\left(||-x \oplus y||\right) \frac{-x \oplus y}{||-x \oplus y||}.$$

Möbius addition

In the Gyrovector space, the Möbius addition $\oplus$ for x and y in model $\mathbb{B}$ is defined as

$$\tau_x(y) = x \oplus y = \frac{(1 + 2\langle x, y\rangle + ||y||^2)x + (1 - ||x||^2)y}{1 + 2\langle x, y\rangle + ||x||^2||y||^2}.$$

$x \oplus y$ will recover to $x + y$ when the curvature goes to zero. the Möbius subtraction $\ominus$ is simply defined as: $x \ominus y = x \oplus (-y)$.



[1]

[1] Hyperbolic Representation Learning for Computer Vision.Mettes et al. ECCV 2022

Möbius scalar multiplication

In the Gyrovector space, the Möbius scalar multiplication $\otimes$ for x in model $\mathbb{B}^n$ is defined as

$$r \otimes x = \begin{cases} \tanh\left(r \operatorname{artanh}\left(||x||\right)\right)\frac{x}{||x||}, & x \in \mathbb{B}^n \\ 0, & x = 0, \end{cases}$$

where r is a scalar factor.

In addition, the Möbius scalar multiplication can also be conducted in the tangent space by using the exponential and logarithmic maps.

$$r \otimes x = \mathrm{Exp}_0(r\mathrm{Log}_0(x)).$$

[1] Hyperbolic Deep Neural Networks: A Survey. Wei Peng et al. TPAMI 2022.

Möbius vector multiplication

In the Gyrovector space, the Möbius vector multiplication $M^{\otimes}(x)$ for metric M and x in model $\mathbb{B}^n$ is defined as

$$M^{\otimes}(x) = \tanh\left(\frac{||Mx||}{||x||}\,\operatorname{actanh}\,(||x||)\right)\frac{Mx}{||Mx||}$$

Mean in hyperbolic space

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Tangential aggregation. Map x to tangent space, calculate the mean and back to hyperbolic space.

$$\mu = \mathrm{Exp}_x\Big(\sum_{j\in\mathcal{N}(i)} w_{ij}\mathrm{Log}_x(x_j)\Big).$$

Lou et al. NIPS 2019

Einstein midpoint

Einstein midpoint is an extension of the mean operation to hyperbolic spaces, which has the most concise form in the Klein coordinates.

$$\mu = \frac{\sum_{i=1}^{N} \gamma_i x^{(i)}}{\sum_{i=1}^{N} \gamma_i}, \qquad \gamma_i = \frac{1}{||x^{(i)}||^2}$$

in which the x_i is the i-th sample represented using coordinates in Klein model. The mean in Poincaré model can be conducted from the results in Klein model by

$$\mu_{\mathbb{B}} = \frac{\mu}{1 + \sqrt{1 - ||\mu||^2}}.$$

Gulcehre et al. 2018

Einstein midpoint

A closed-form expression for Poincaré model to compute the average (midpoint) in the Gyrovector spaces.

$$m(x^{(1)}, \ldots, x^{(N)}; \alpha) = \frac{1}{2} \oplus \left(\sum_{i=1}^{N} \frac{\alpha_i \gamma_i}{\sum_{j=1}^{N} \alpha_j (\gamma_j - 1)} x^{(i)} \right)$$

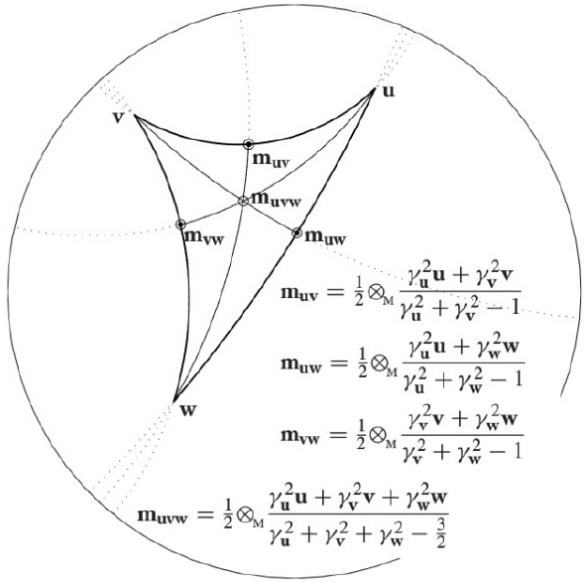


Illustration of the Möbius gyromidpoint. [1]

with $\alpha = (\alpha_1, \ldots, \alpha_N)$ as the weights, and Lorzent factor as $\gamma_i = \frac{2}{||x^{(i)}||^2}$

[1] A gyrovector space approach to hyperbolic geometry. A. A. Ungar. Synthesis Lectures on Mathematics and Statistics, vol. 1, no. 1, pp. 1194, 2008.

Hyperbolic geometry

- Why hyperbolic?
- Riemannian manifold
- Hyperbolic Models
- Basic Operation
- δ-Hyperbolicity

δ-Hyperbolicity

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Gromov δ-hyperbolicity is used to evaluate the hyperbolicity of a dataset/space X.

Gromov product

$$(y,z)_x = \frac{1}{2}(d(x,y) + d(x,z) - d(y,z))$$

δ is defined as the minimal value such that the following four-point condition holds for all points $x, y, z, w \in X$:

$$(x,z)_w \geq \min((x,y)_w, (y,z)_w) - \delta$$

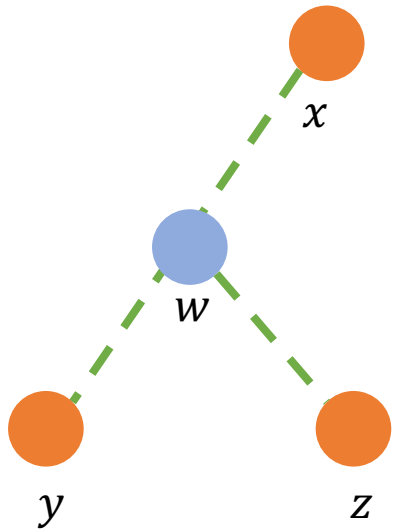
δ-Hyperbolicity

Gromov δ-hyperbolicity is used to evaluate the **hyperbolicity** of a dataset/space X.

Gromov product $(y,z)_x = \frac{1}{2}(d(x,y) + d(x,z) - d(y,z))$

δ is defined as the minimal value such that the following four-point condition holds for all points $x, y, z, w \in X$:

$$(x,z)_w \geq \min((x,y)_w, (y,z)_w) - \delta$$



δ-Hyperbolicity

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

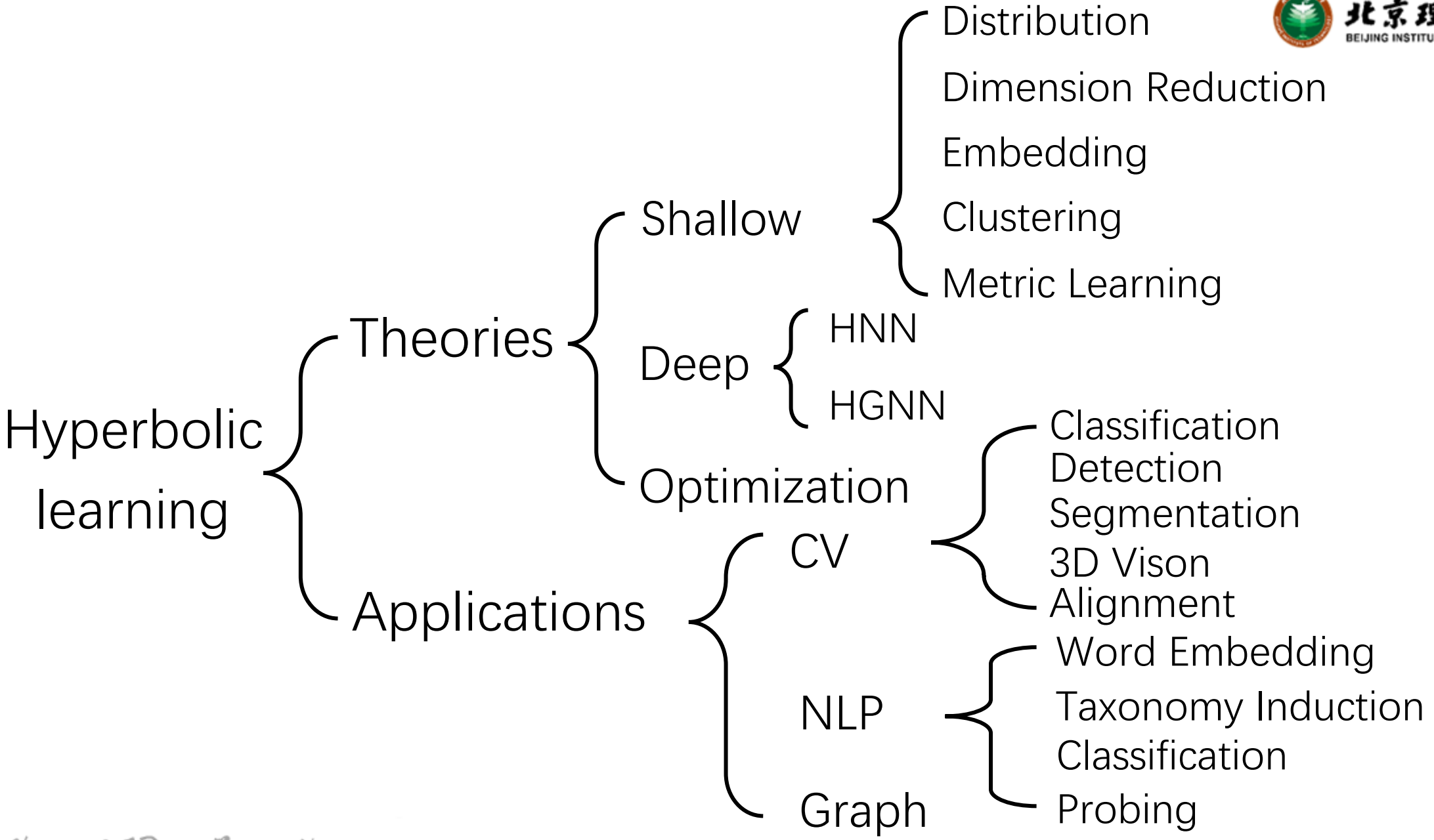
Encoder	Dataset			
	CIFAR10	CIFAR100	CUB	*Mini*ImageNet
Inception v3 [49]	0.25	0.23	0.23	0.21
ResNet34 [14]	0.26	0.25	0.25	0.21
VGG19 [42]	0.23	0.22	0.23	0.17

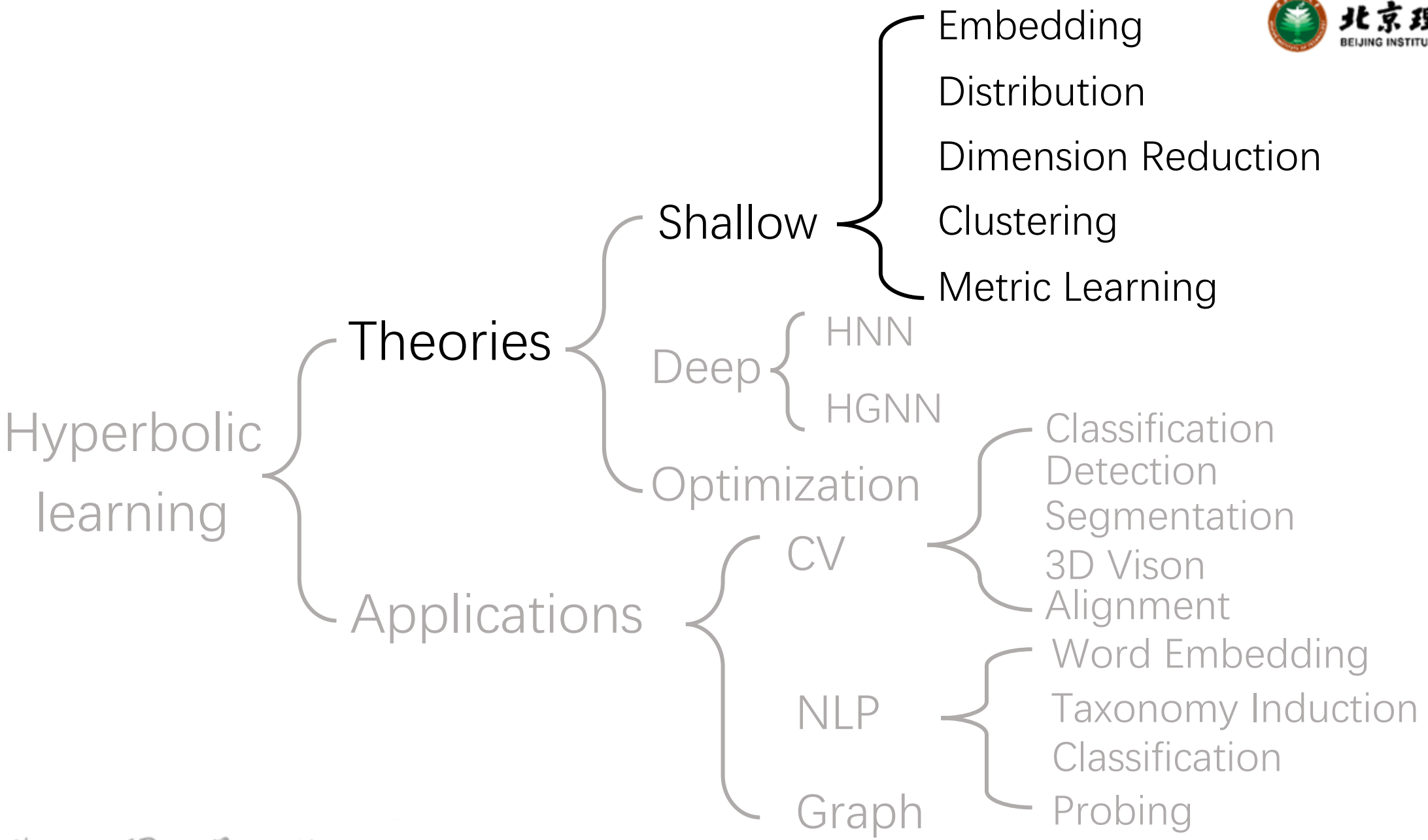
The relative delta δ_{rel} values calculated for different datasets. [1]

德以明理 学以精工

[1] Hyperbolic Image Embeddings, Mirvakhabova et al. CVPR 2020

Hyperbolic learning


Hyperbolic learning
Theories
Shallow
Distribution
Dimension Reduction
Embedding
Clustering
Metric Learning
Deep
HNN
HGNN
Optimization
Applications
CV
Classification
Detection
Segmentation
3D Vison
Alignment
NLP
Word Embedding
Taxonomy Induction
Classification
Graph
Probing


Hyperbolic learning
Theories
Shallow
Embedding
Distribution
Dimension Reduction
Clustering
Metric Learning
Deep
HNN
HGNN
Optimization
Applications
CV
Classification
Detection
Segmentation
3D Vison
Alignment
NLP
Word Embedding
Taxonomy Induction
Classification
Graph
Probing

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

Theories

Shallow Learning

德以明理　学以精工

Embedding

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

Why do we model embedding hierarchies in hyperbolic space?

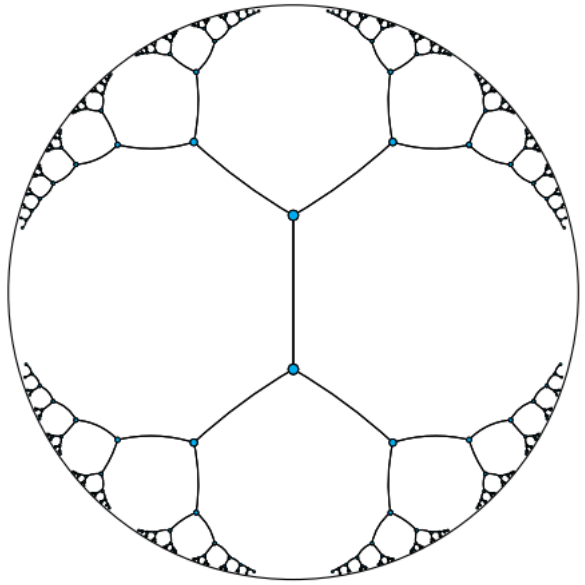
- Hierarchies grow exponentially in depth, Euclidean space grows linearly with norm.
- The exponential growth of distances in hyperbolic space is a direct fit with hierarchies.

Embedding

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Poincaré embedding[1]

The earliest paper that studied hyperbolic embeddings.



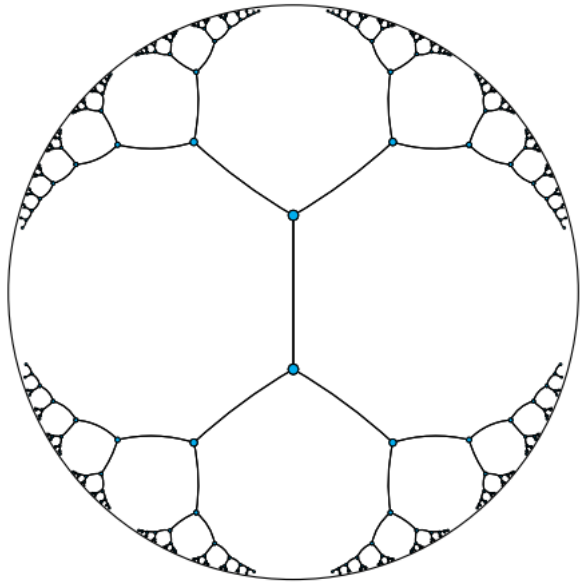
Embedding of a tree

Learning hierarchical representations of data in a low-dimensional space while preserving the inherent geometric structure of the data.

[1] Poincaré Embeddings for Learning Hierarchical Representations, Nickel et al. NIPS2017

Embedding

Poincaré embedding[1]



Embedding of a tree

INPUT

Nodes $\mathcal{S} = \{x_i\}_{i=1}^{n}$

Positive Samples $\mathcal{D} = \{(u, v)\}$

Negative Samples $\mathcal{N}(u) = \{v' \mid (u, v') \notin \mathcal{D}\} \cup \{v\}$

OPTIMIZATION

$$\Theta' \leftarrow \arg\min_{\Theta} \mathcal{L}(\Theta) \qquad \text{s.t. } \forall \boldsymbol{\theta}_i \in \Theta : \|\boldsymbol{\theta}_i\| < 1$$

$$\mathcal{L}(\Theta) = \sum_{(u,v)\in\mathcal{D}} \log \frac{e^{-d(\boldsymbol{u},\boldsymbol{v})}}{\sum_{\boldsymbol{v}'\in\mathcal{N}(u)} e^{-d(\boldsymbol{u},\boldsymbol{v}')}}$$

OUTPUT

Embeddings in poincaré model. $\Theta = \{\boldsymbol{\theta}_i\}_{i=1}^{n}$

[1] Poincaré Embeddings for Learning Hierarchical Representations, Nickel et al. NIPS2017

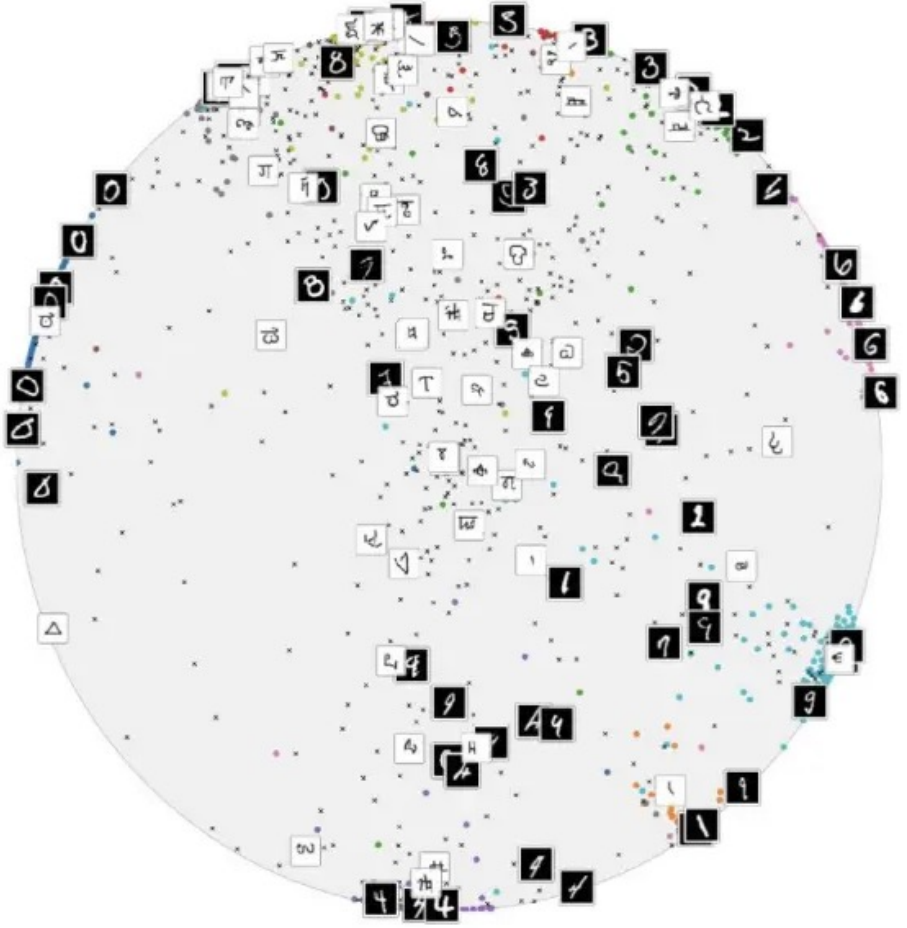
Embedding

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Hyperbolic embeddings for computer vision.

Hyperbolic Image embeddings[1]

- Develop a new method for embedding high-dimensional data, such as images, in a **low-dimensional** space that **retains the important relationships** between the data points.
- The motivation for using hyperbolic space is that it better captures the hierarchical structure present in many real-world data sets, such as image categories.



德以明理 学以精工

[1] Hyperbolic Image Embeddings, Mirvakhabova et al. CVPR 2020

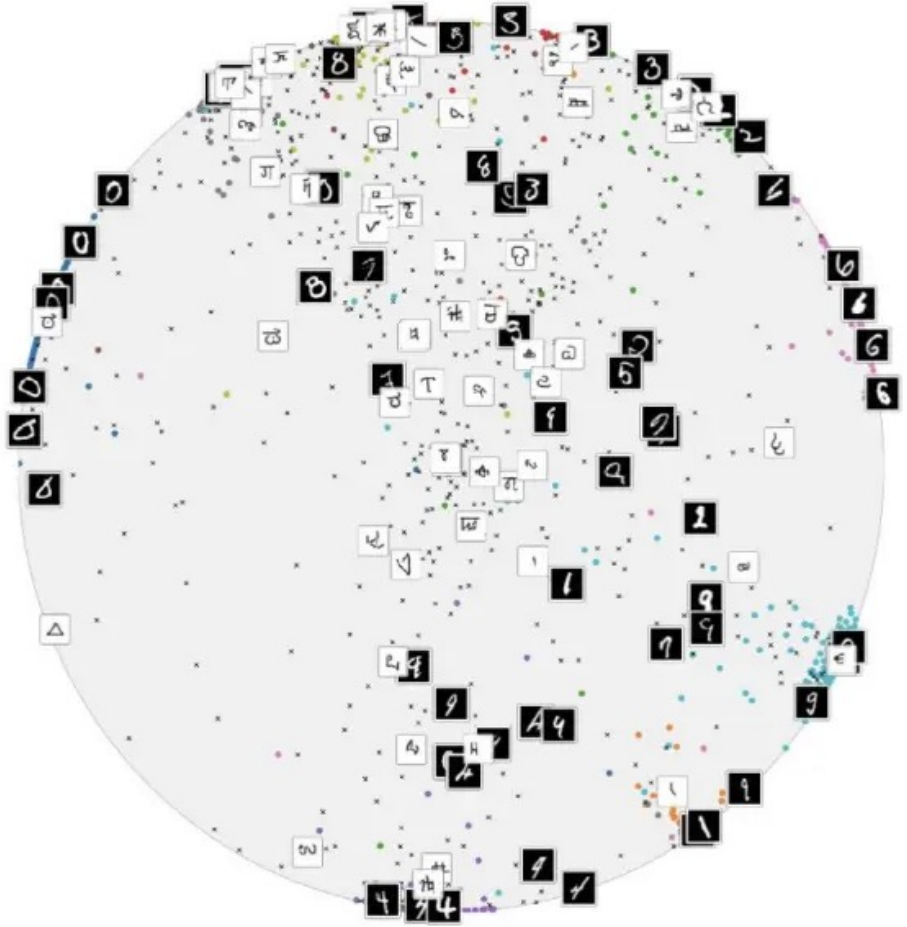
Embedding

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Hyperbolic embeddings for computer vision.

Hyperbolic Image embeddings[1]

- Input the images to CNNs.
- Map the features to hyperbolic space.
- Generate prototypes in hyperbolic spaces.



$$\mathrm{HypAve}(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \sum_{i=1}^{N} \gamma_i \mathbf{x}_i \Big/ \sum_{i=1}^{N} \gamma_i,$$

$$\gamma_i = \frac{1}{\sqrt{1-c\|\mathbf{x}_i\|^2}}$$

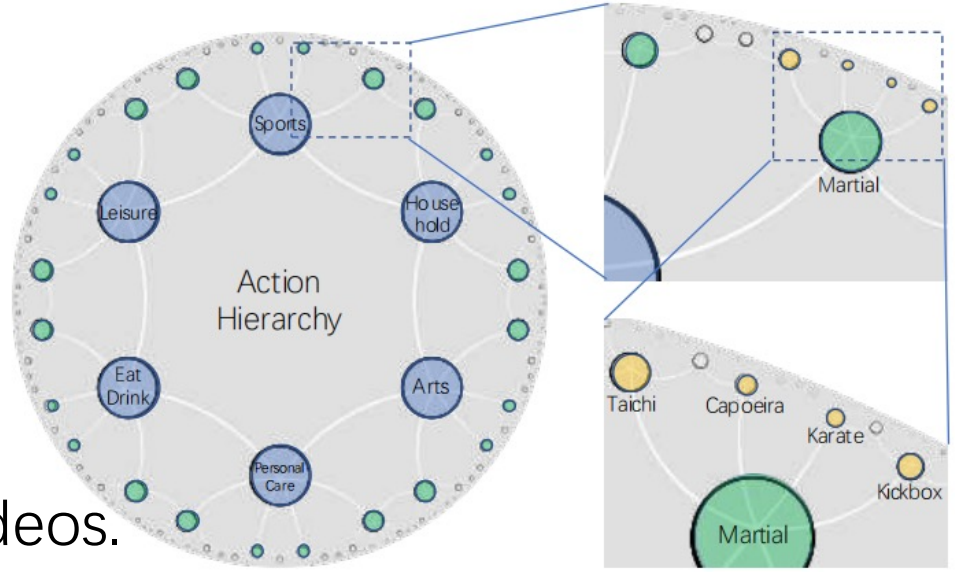
德以明理 学以精工

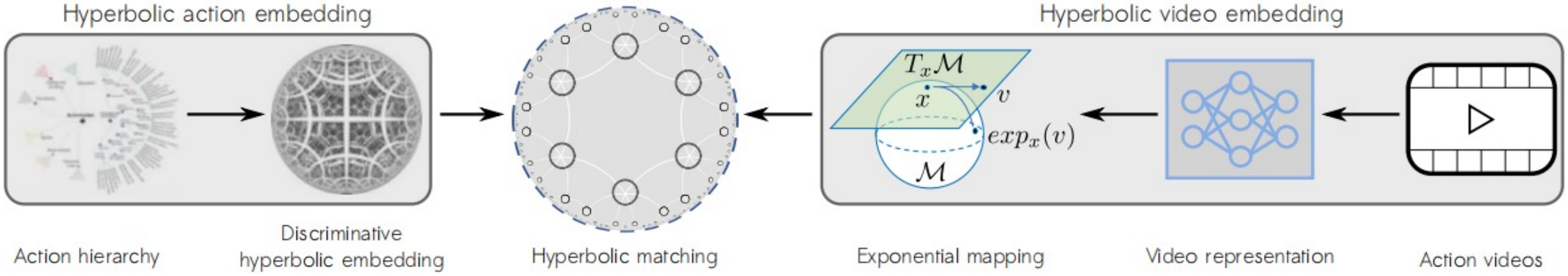
[1] Hyperbolic Image Embeddings, Mirvakhabova et al. CVPR 2020

Embedding

Searching for Actions on the Hyperbole[1]

1. Extract poincaré embedding of actions.
2. Map the representation of videos in Euclidean space to hyperbolic space.
3. Minimize the distance between actions and videos.





[1] Searching for Actions on the Hyperbole, Long et al. CVPR 2020.

Distribution

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Wrapped normal distribution[1]

A normal distribution in hyperbolic space

$$\mathcal{N}^{\mathrm{W}}_{\mathbb{B}^d_c}(\boldsymbol{z}|\boldsymbol{\mu},\Sigma) = \mathcal{N}\left(\lambda^c_{\boldsymbol{\mu}}\log_{\boldsymbol{\mu}}(\boldsymbol{z})\middle|\mathbf{0},\Sigma\right)\left(\frac{\sqrt{c}\,d^c_p(\boldsymbol{\mu},\boldsymbol{z})}{\sinh(\sqrt{c}\,d^c_p(\boldsymbol{\mu},\boldsymbol{z}))}\right)^{d-1}$$

Riemannian (max-entropy) normal distribution[2]

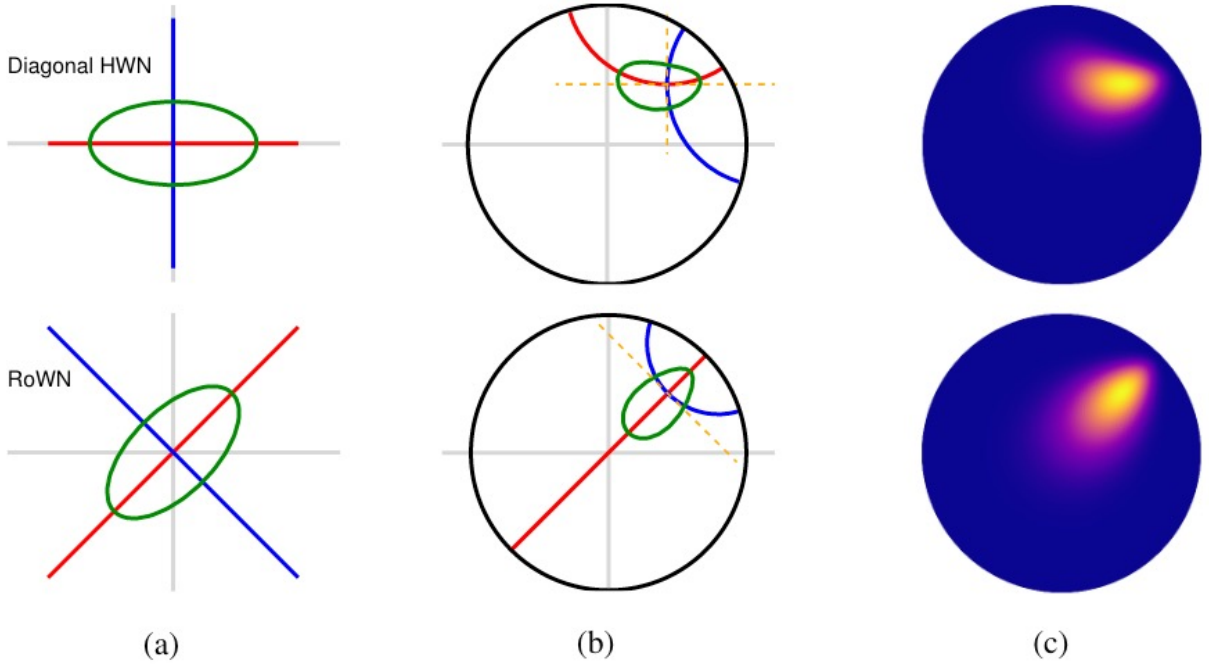
$$\mathcal{N}^{\mathrm{R}}_{\mathbb{B}^d_c}(\boldsymbol{z}|\boldsymbol{\mu},\sigma^2) = \frac{d\nu^{\mathrm{R}}(\boldsymbol{z}|\boldsymbol{\mu},\sigma^2)}{d\mathcal{M}(\boldsymbol{z})} = \frac{1}{Z^{\mathrm{R}}}\exp\left(-\frac{d^c_p(\boldsymbol{\mu},\boldsymbol{z})^2}{2\sigma^2}\right)$$

[1] A Wrapped Normal Distribution on Hyperbolic Space for Gradient-Based Learning. Yoshihiro Nagano et al. ICML 2019.
[2] Continuous Hierarchical Representations with Poincaré Variational Auto-Encoders. Mathieu et al. NIPS 2019.

Distribution

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

RoHWN (Rotated Hyperbolic Wrapped Normal Distribution)



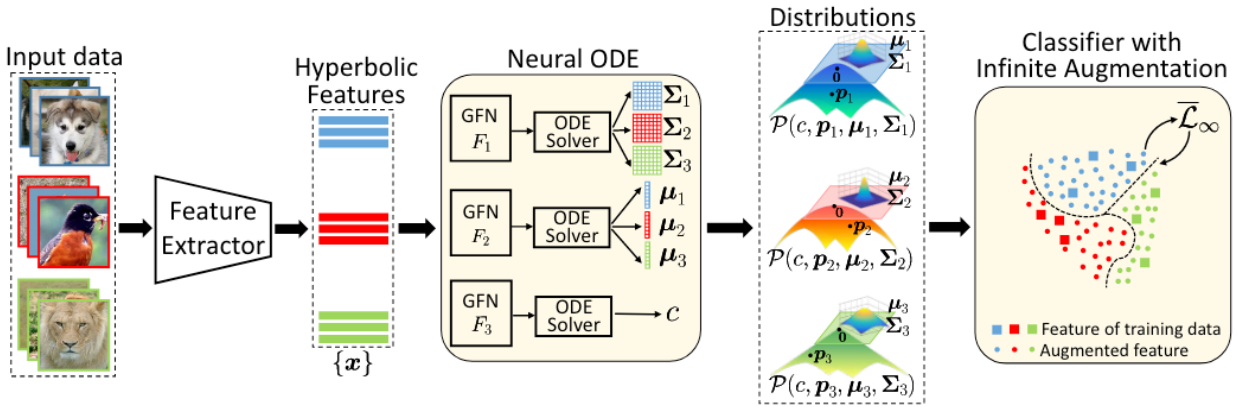
Analyze HWN's geometry and limits in representation learning.

[1] A Rotated Hyperbolic Wrapped Normal Distribution for Hierarchical Representation Learning. Seunghyuk Cho et al. CSED POSTECH

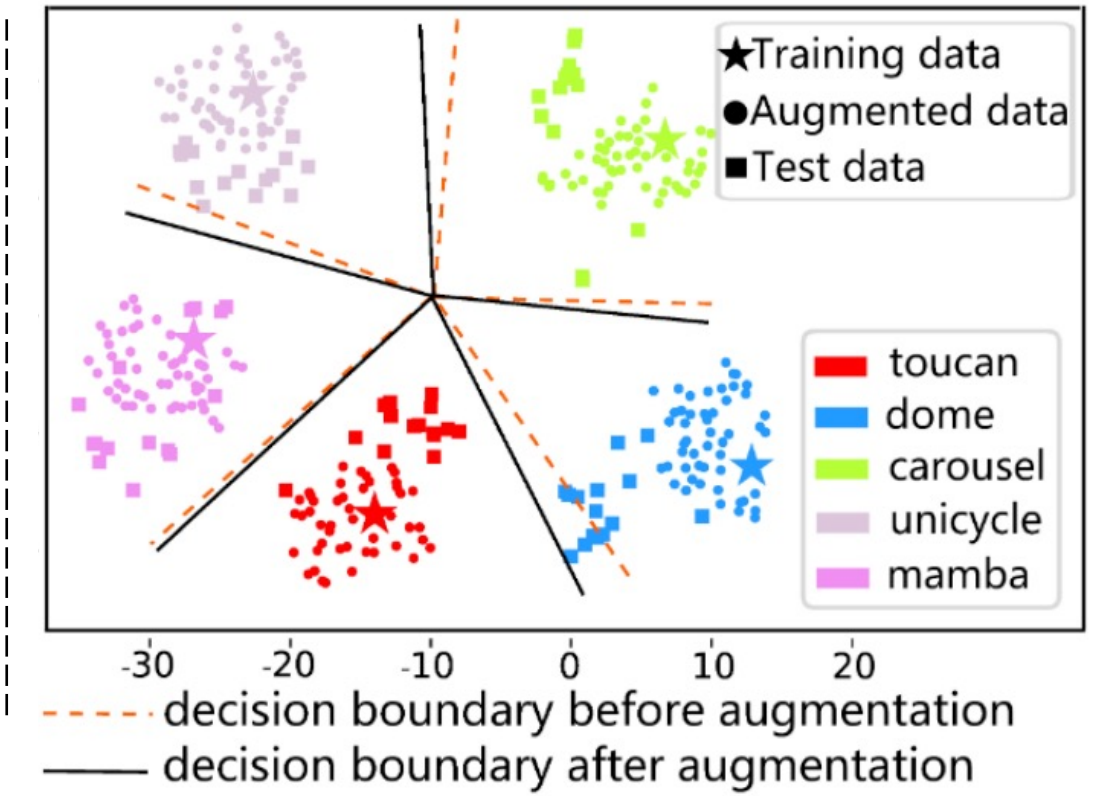
Distribution

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Hyperbolic Feature Augmentation via Distribution Estimation[1]



Sample an infinite number of data augmentation points from HWN.

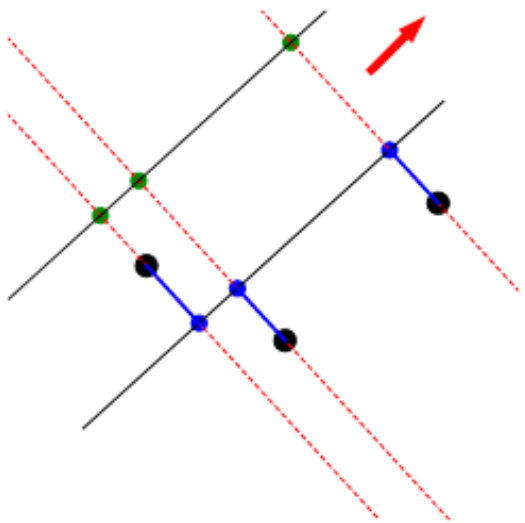


[1] Hyperbolic Feature Augmentation via Distribution Estimation and Infinite Sampling on Manifolds. Gao Zhi et al. NIPS 2022

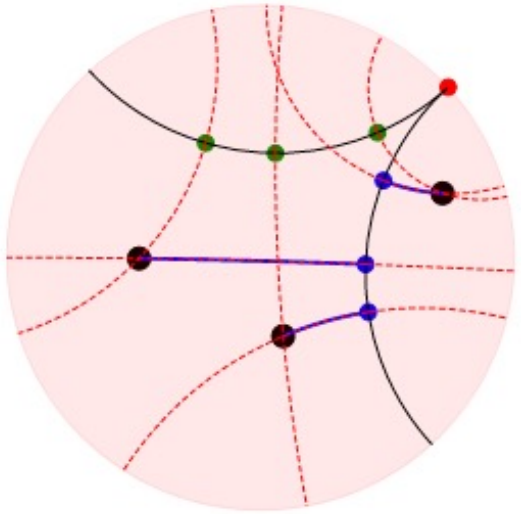
Dimension Reduction

HoroPCA

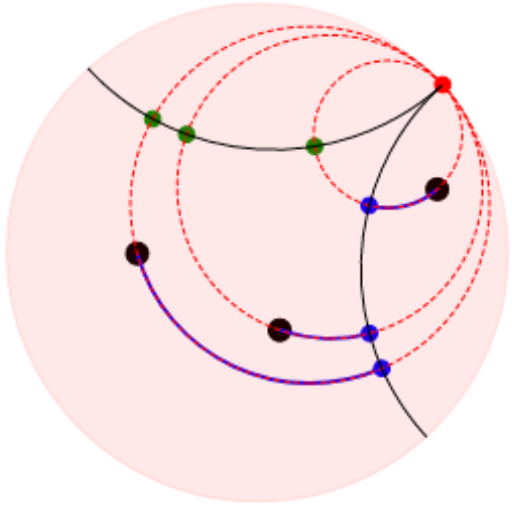
Look for directions that explain the data in hyperbolic space.



(a) Euclidean projections.



(b) Hyperbolic geodesic projections.

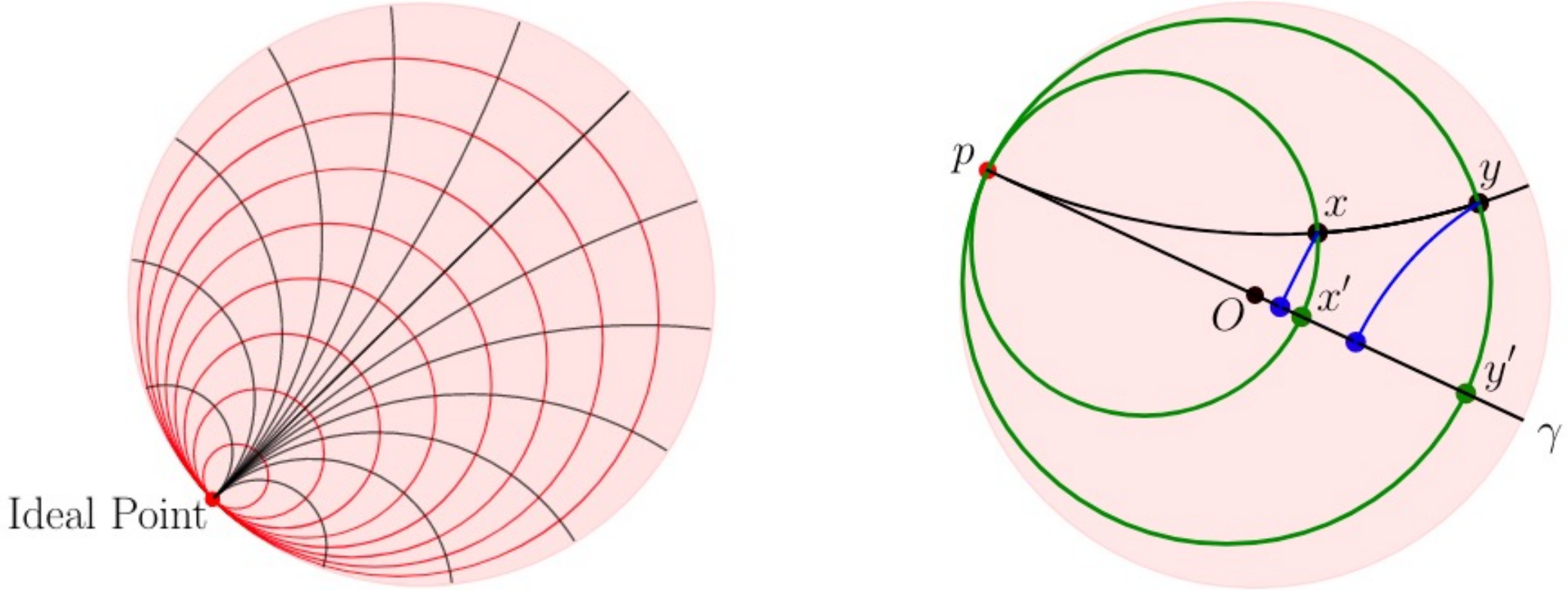


(c) Hyperbolic horospherical projections.

[1]

[1] HoroPCA: Hyperbolic Dimensionality Reduction via Horospherical Projections. Ines Chami et al. ICML 2021.

Dimension Reduction

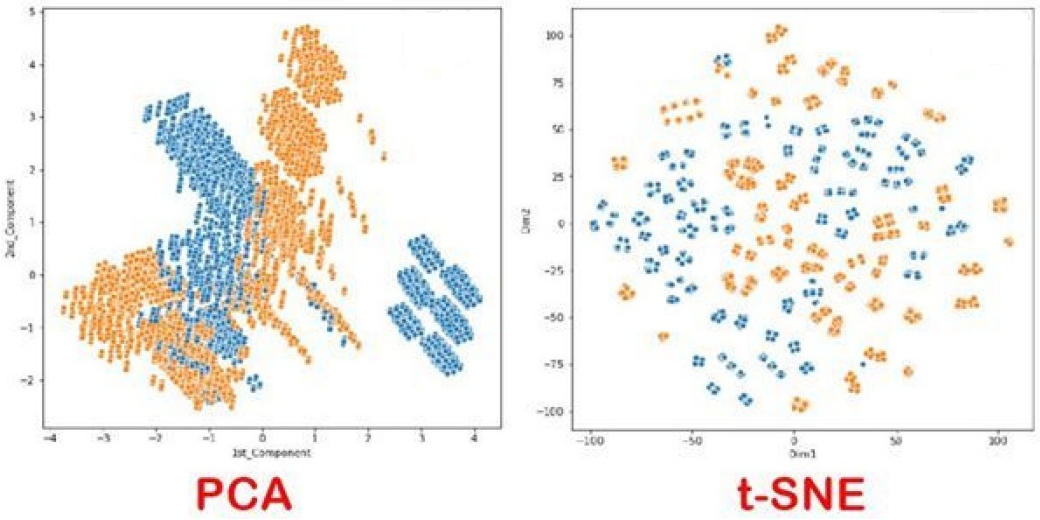


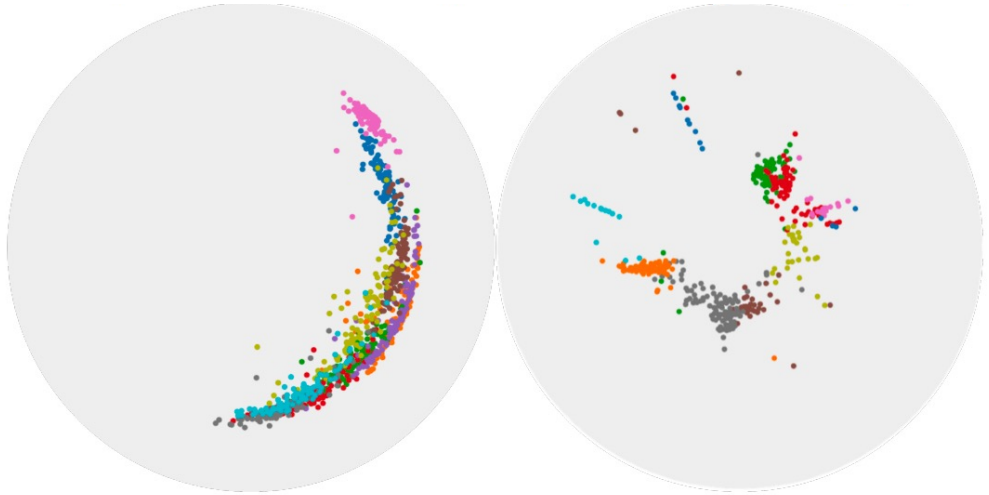
PGA vs HoroPCA [1]

[1] HoroPCA: Hyperbolic Dimensionality Reduction via Horospherical Projections. Ines Chami et al. ICML 2021.

Dimension Reduction

CO-SNE [1]





[1]

HoroPCA cannot preserve the **local** similarity,T-SNE cannot preserve the **global** hierarchy.

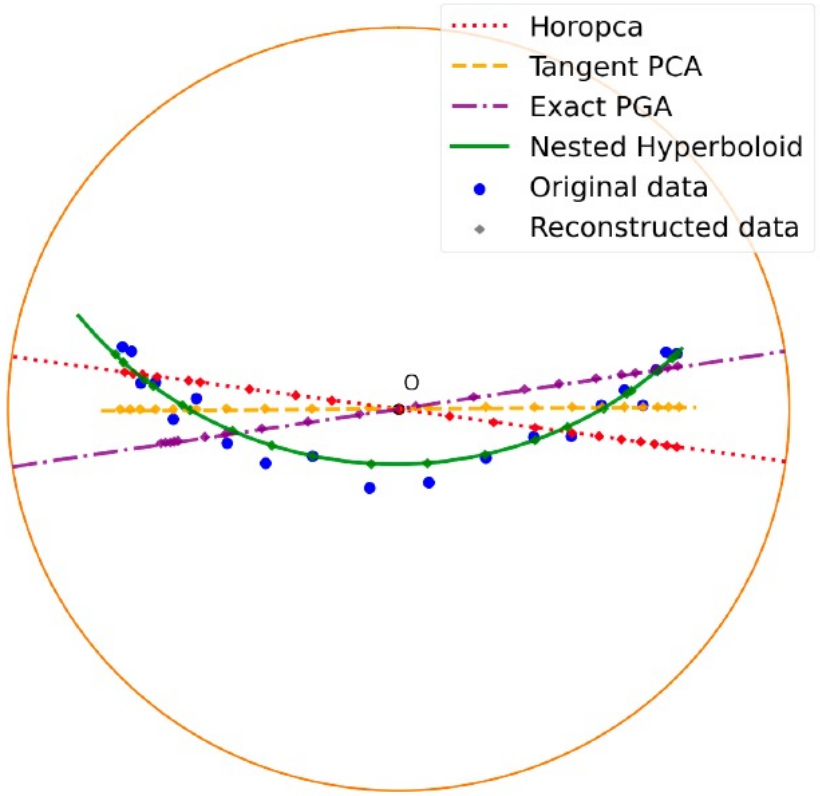
CO-SNE can preserve **both the global hierarchy and local similarity** of high-dimensional hyperbolic embeddings in a low-dimensional hyperbolic space.

[1] CO-SNE: Dimensionality Reduction and Visualization for Hyperbolic Data. Yunhui Guo et al. CVPR 2022

Dimension Reduction

Nested Hyperboloid [1]

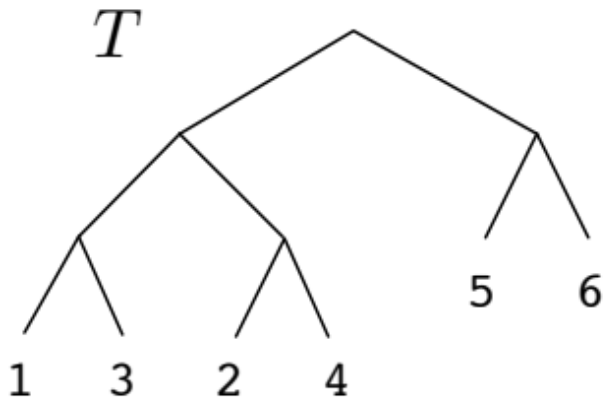
Capture the main trend of the data.
Proposed curved principal component axes, better suited to the data

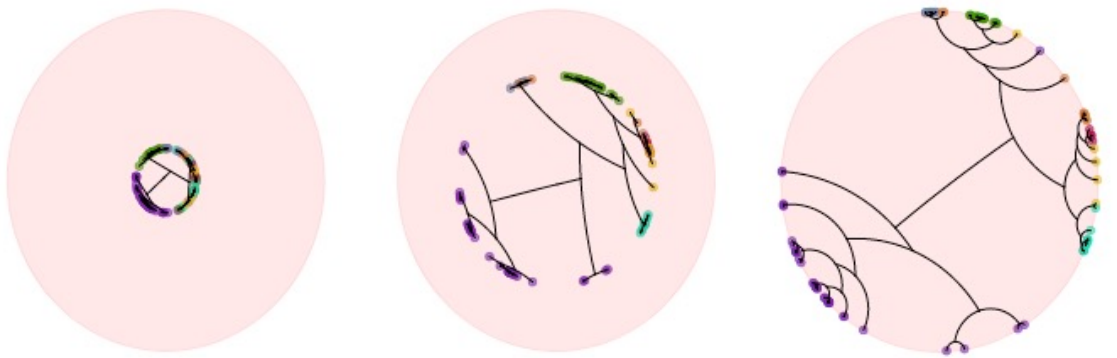


[1] Nested Hyperbolic Spaces for Dimensionality Reduction and Hyperbolic NN Design. Xiran Fan et al. CVPR 2022.

Clustering

HypHC (Hyperbolic Hierarchical Clustering)[1]





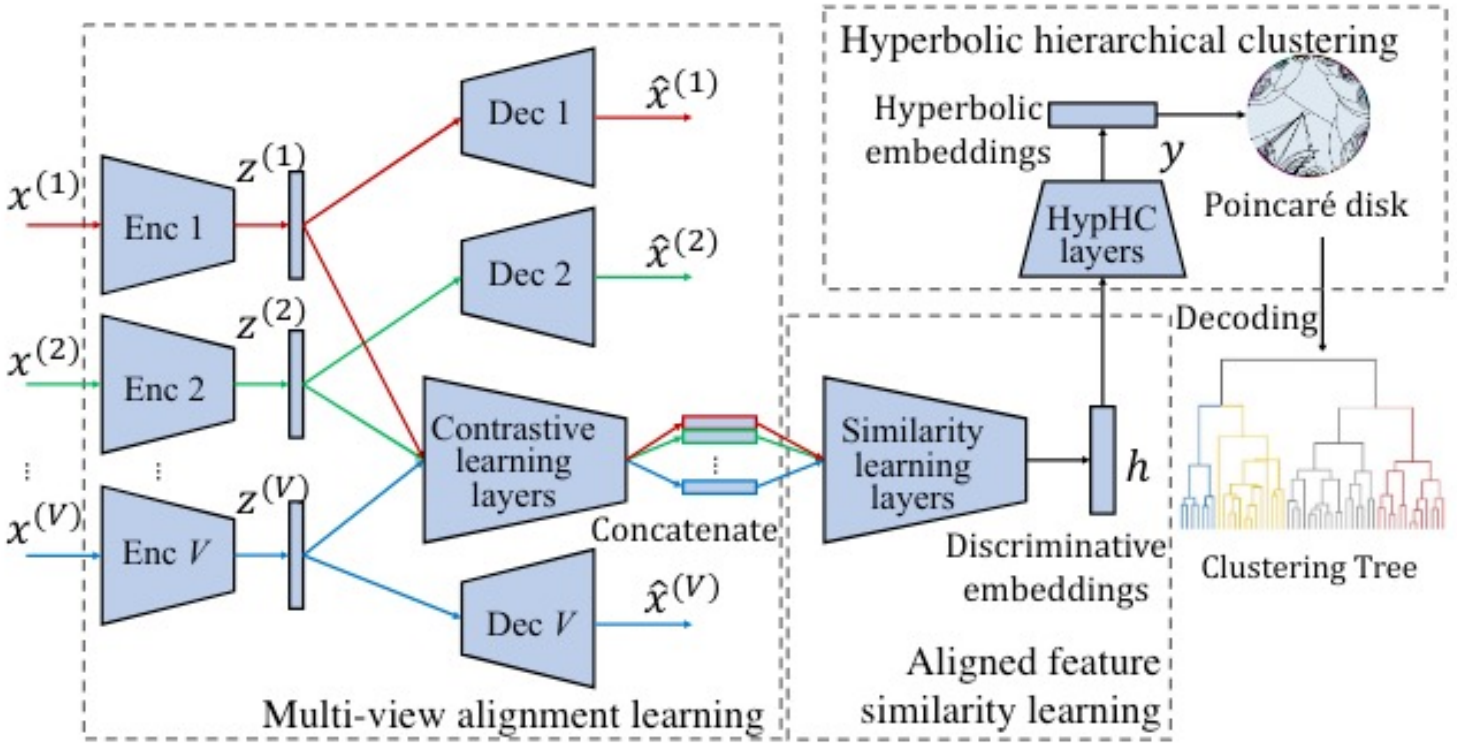
[1]

Show a direct correspondence from discrete trees to **continuous** representations and back, allowing us to search the space of discrete binary trees with continuous optimization.

[1] From Trees to Continuous Embeddings and Back: Hyperbolic Hierarchical Clustering. Ines Chami et al. NIPS 2020.
[2] A cost function for similarity-based hierarchical clustering. Sanjoy Dasgupta.

Clustering

CMHHC (Contrastive Multi-view Hyperbolic Hierarchical Clustering)[1]



Align data of different modalities in feature space to obtain hierarchical clustering in hyperbolic space

[1] Contrastive Multi-view Hyperbolic Hierarchical Clustering. Fangfei Lin et al. IJCAI 2022.

Metric Learning

Kernel Methods

Table 1. Summary of the proposed positive definite kernels in hyperbolic spaces and their properties.

Kernel	Formulation: $k(z_i, z_j)$	Condition	Properties
$f_{\mathbb{D}}(z) = \tanh^{-1}(\sqrt{c}\|z\|)\frac{z}{\sqrt{c}\|z\|},\ c > 0$ and $z \in \mathbb{D}_c^n$			
Hyperbolic tangent kernel	$k^{\tan}(z_i, z_j) = \langle f_{\mathbb{D}}(z_i), f_{\mathbb{D}}(z_j)\rangle$	-	pd
Hyperbolic RBF kernel	$k^{\mathrm{rbf}}(z_i, z_j) = \exp\left(-\xi\|f_{\mathbb{D}}(z_i), f_{\mathbb{D}}(z_j)\|^2\right)$	$\xi > 0$	pd, universal
Hyperbolic Laplace kernel	$k^{\mathrm{lap}}(z_i, z_j) = \exp\left(-\xi\|f_{\mathbb{D}}(z_i), f_{\mathbb{D}}(z_j)\|\right)$	$\xi > 0$	pd, universal
Generalized Hyperbolic Laplace kernel	$k^{\mathrm{glap}}(z_i, z_j) = \exp\left(-\xi\|f_{\mathbb{D}}(z_i), f_{\mathbb{D}}(z_j)\|^{2\alpha}\right)$	$\xi > 0, 0 < \alpha < 1$	pd, universal
Hyperbolic binomial kernel	$k^{\mathrm{bin}}(z_i, z_j) = \left(1 - \langle f_{\mathbb{D}}(z_i), f_{\mathbb{D}}(z_j)\rangle\right)^{-\alpha}$	$\alpha > 0$	pd, universal

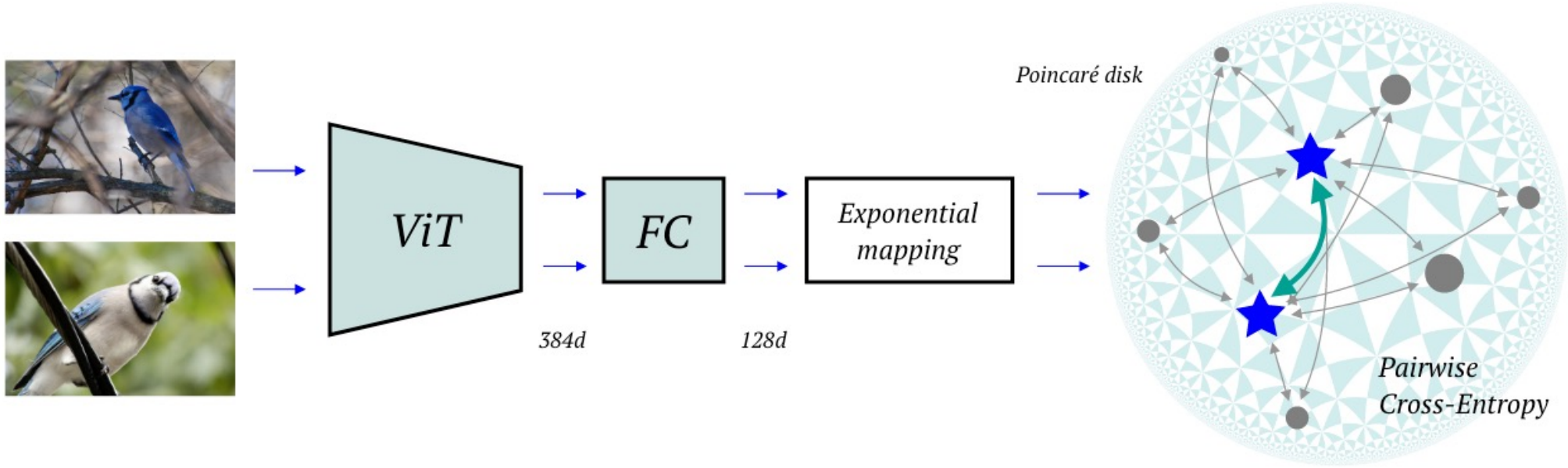
- Benefit from kernel machines and hyperbolic embeddings. [1]
- Simplify various operations involving hyperbolic data because of the rich structure of the Hilbert spaces.

德以明理 学以精工

[1] Kernel Methods in Hyperbolic Spaces. Pengfei Fang et al. ICCV 2021.

Metric Learning

Hyperbolic metric works well on vision transformer.



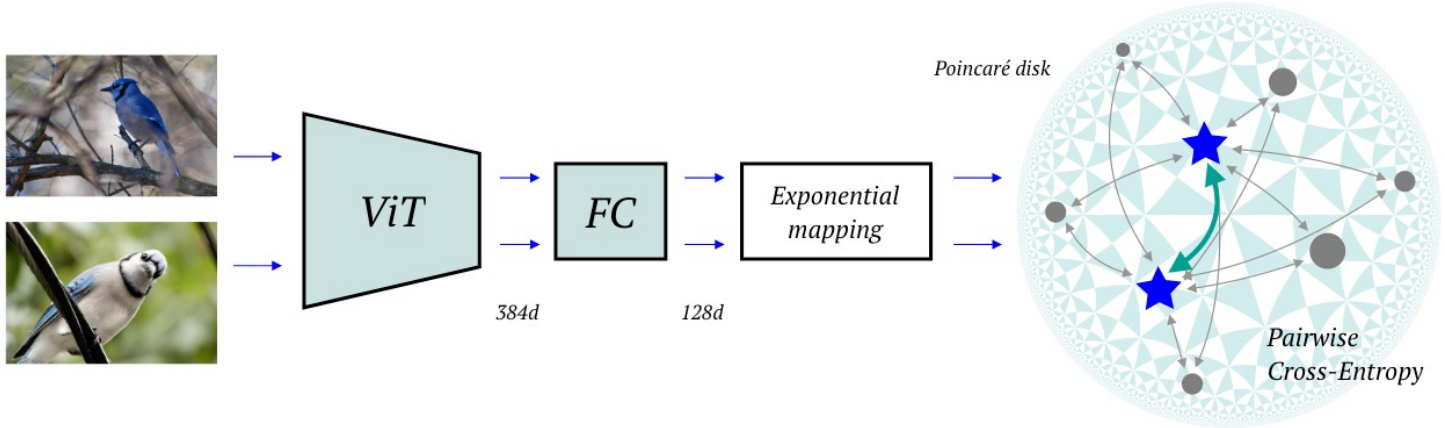
New state-of-the-art on well-known fine-grained datasets.

[1]

[1] Hyperbolic Vision Transformers: Combining Improvements in Metric Learning. Aleksandr Ermolov et al. CVPR2022.

Metric Learning

Hyperbolic metric works well on vision transformer.



$$D_{cos}(\mathbf{z}_i, \mathbf{z}_j) = \left\| \frac{\mathbf{z}_i}{\|\mathbf{z}_i\|_2} - \frac{\mathbf{z}_j}{\|\mathbf{z}_j\|_2} \right\|_2^2 = 2 - 2\frac{\langle \mathbf{z}_i, \mathbf{z}_j \rangle}{\|\mathbf{z}_i\|_2 \cdot \|\mathbf{z}_j\|_2}$$

$$l_{i,j} = -\log \frac{\exp\left(-D(\mathbf{z}_i, \mathbf{z}_j)/\tau\right)}{\sum_{k=1,k\neq i}^{K} \exp\left(-D(\mathbf{z}_i, \mathbf{z}_k)/\tau\right)},$$

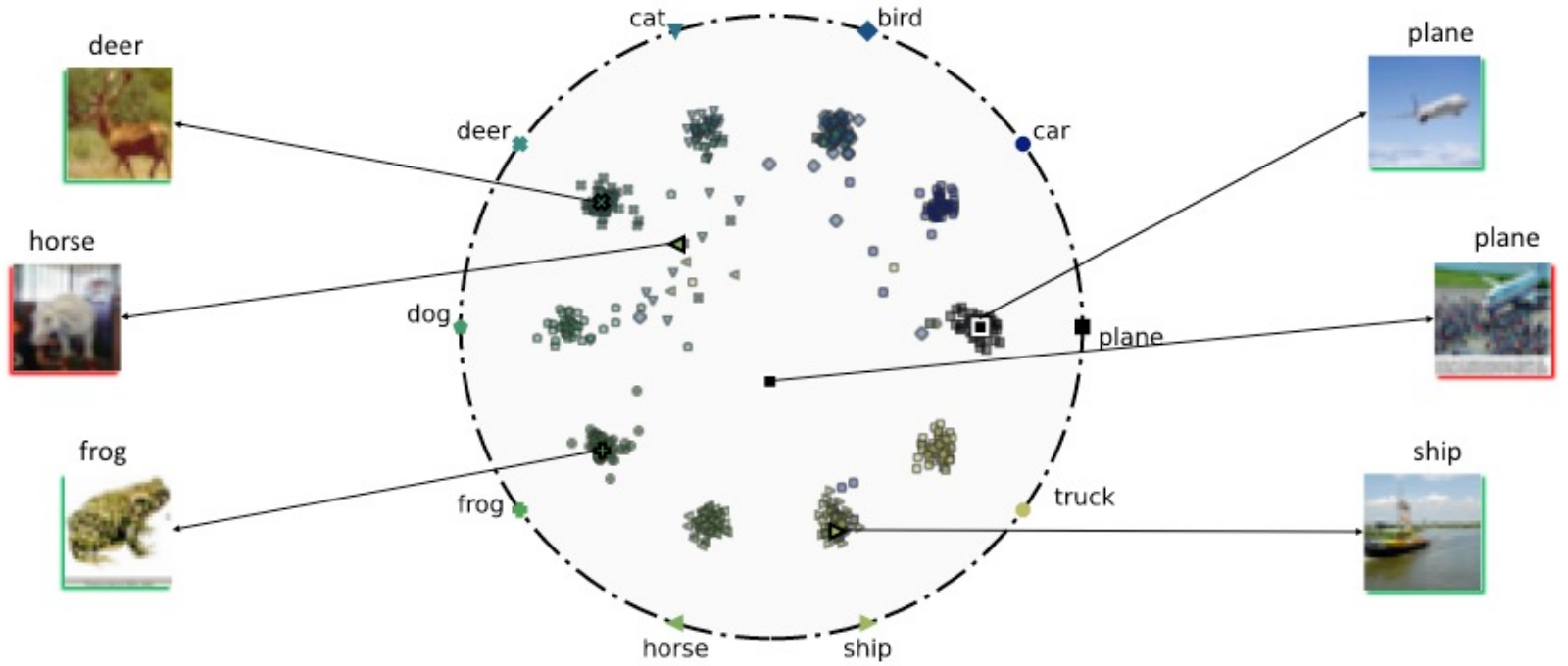
[1]

[1] Hyperbolic Vision Transformers: Combining Improvements in Metric Learning. Aleksandr Ermolov et al. CVPR2022.

Metric Learning

Hyperbolic Busemann Learning[1]

Samples closer to the boundary are more robust.



[1]

Embedding hyperbolic prototypes on ideal boundaries of Poincaré without prior label knowledge.

[1] Hyperbolic Busemann Learning with Ideal Prototypes. Mina Ghadimi Atigh et al. NIPS 2021.

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Hyperbolic Busemann Learning[1]

Ideal points are at infinite geodesic distance from all other points in $\mathbb{B}_d$.

The Busemann function with respect to $\mathbf{p}$ is defined for $\mathbf{z} \in \mathbb{B}_d$ as

$$b_{\mathbf{p}}(\mathbf{z}) = \lim_{t \to \infty} (d_{\mathbb{B}}(\gamma_{\mathbf{p}}(t), \mathbf{z}) - t).$$

p: Ideal point.
γ_p: Geodesic ray, parametrized by arc length, tending to p.

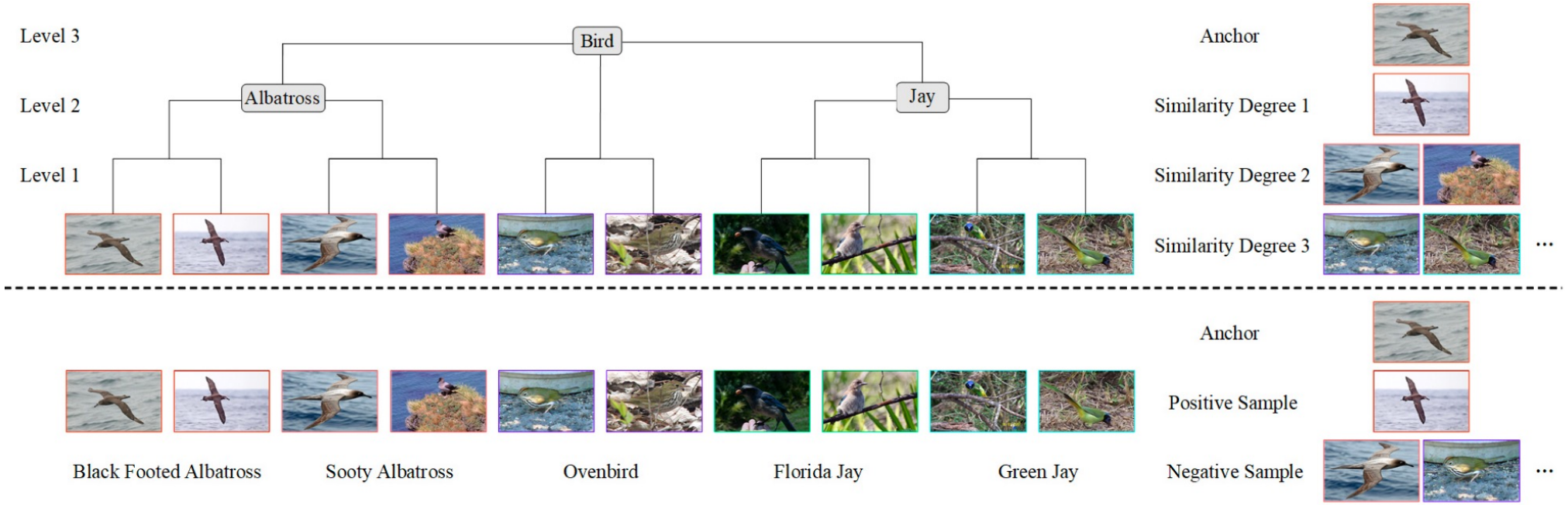
Busemann function in Poincaré Model

$$b_{\mathbf{p}}(\mathbf{z}) = \log \frac{||\mathbf{p} - \mathbf{z}||^2}{(1 - ||\mathbf{z}||^2)}.$$

[1] Hyperbolic Busemann Learning with Ideal Prototypes. Mina Ghadimi Atigh et al. NIPS 2021.

Metric Learning

Unsupervised Hyperbolic Metric Learning [1]



[1]

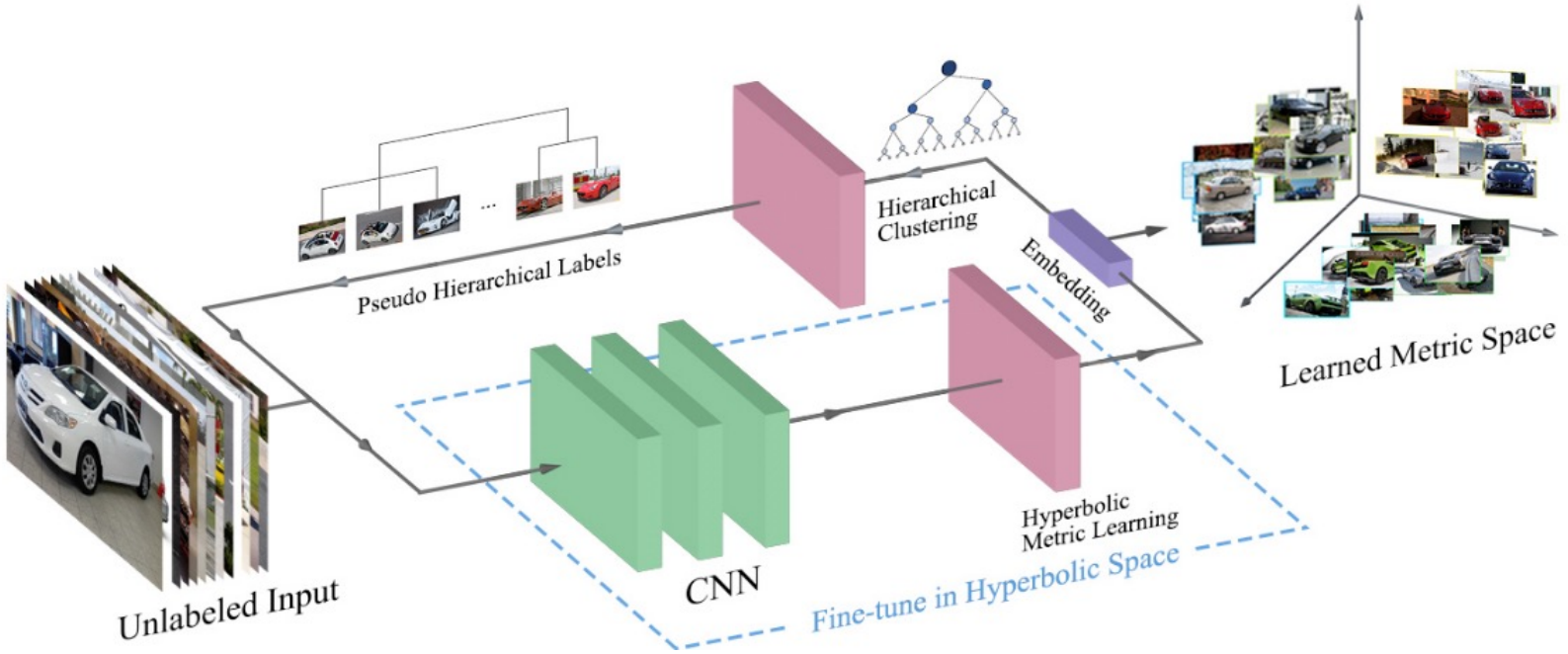
First hyperbolic unsupervised deep metric learning framework.

德以明理 学以精工

[1] Unsupervised Hyperbolic Metric Learning. Jiexi Yan et al. CVPR 2021.

Metric Learning

Unsupervised Hyperbolic Metric Learning [1]



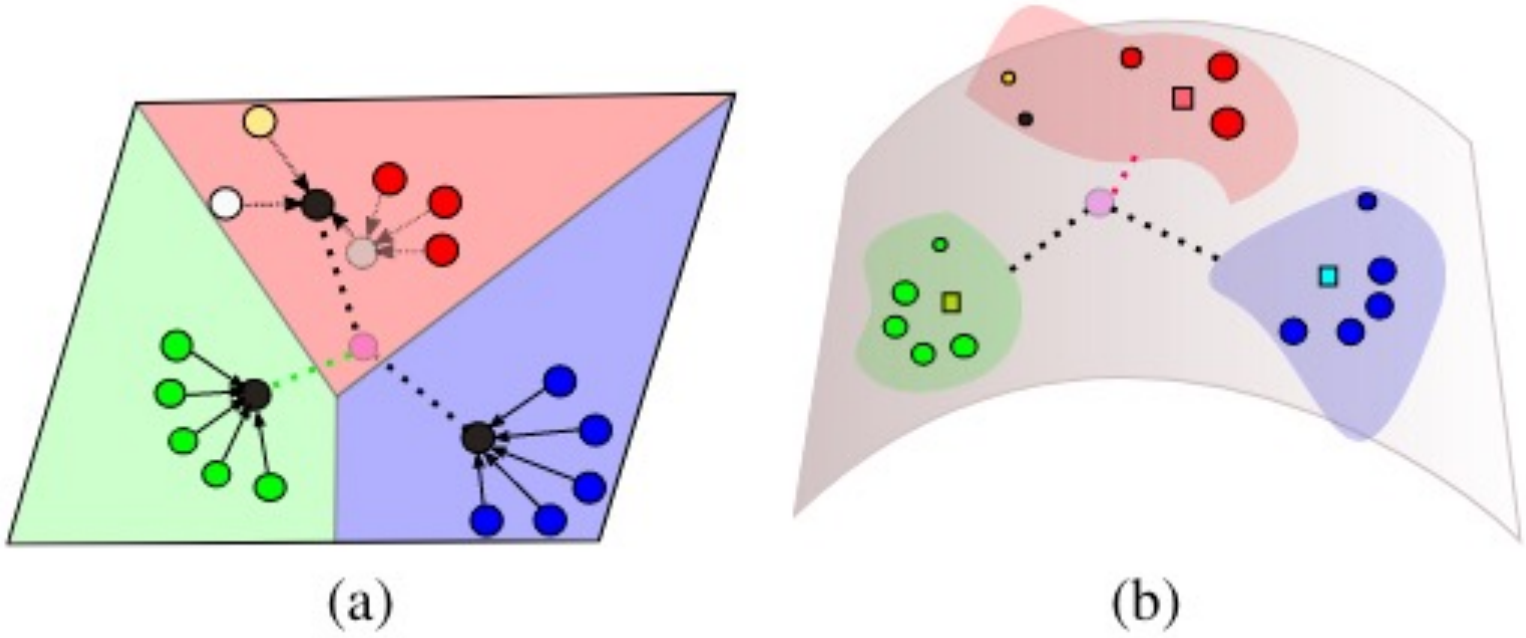
[1]

First hyperbolic unsupervised deep metric learning framework.

[1] Unsupervised Hyperbolic Metric Learning. Jiexi Yan et al. CVPR 2021.

Metric Learning

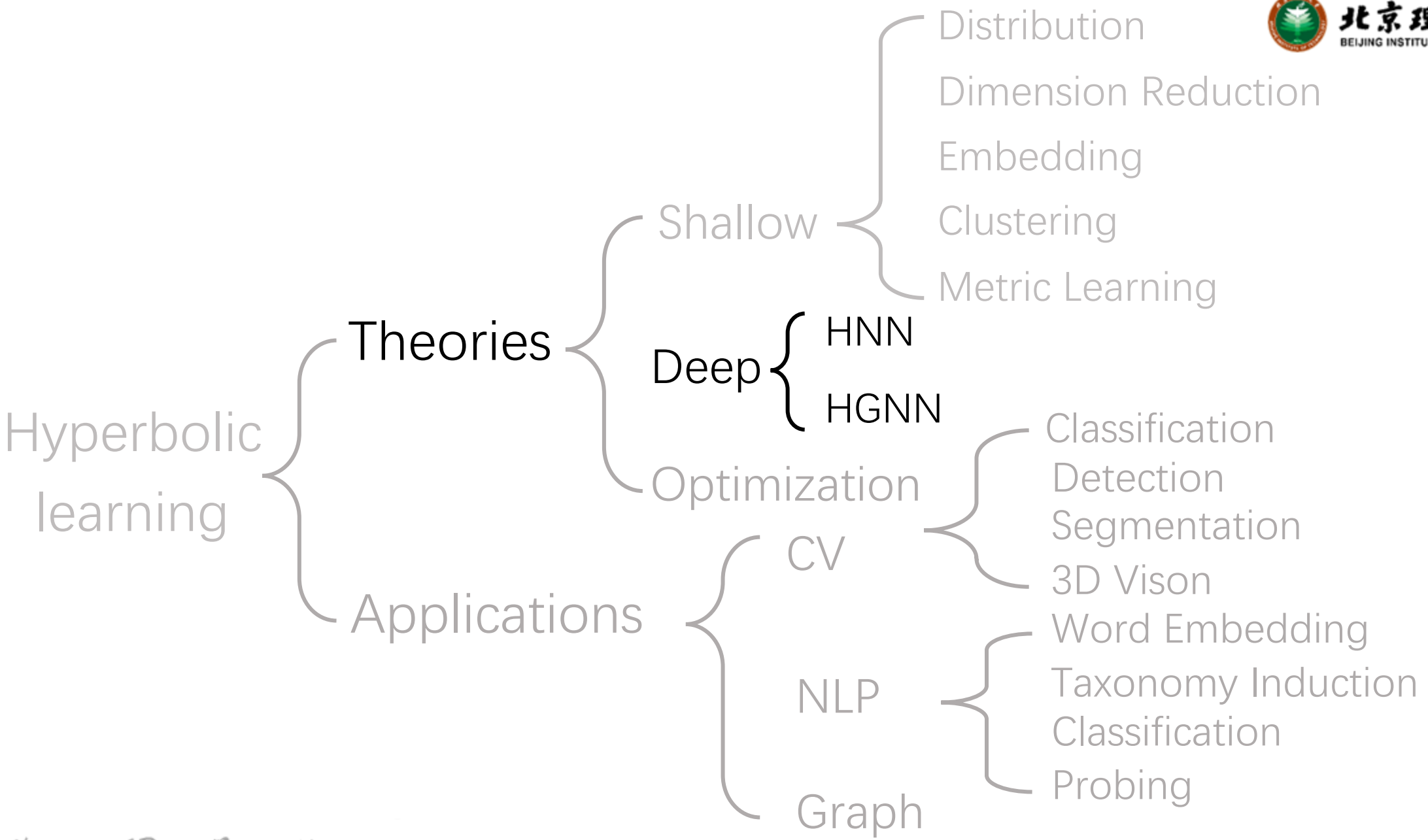
Adaptive Poincaré Point to Set Distance[1]



[1]

Robust metric against **the noises and outliers** in few-show learning.

[1] Adaptive Poincaré Point to Set Distance for Few-Shot Classification. Rongkai Ma et al. AAAI 2022.


Hyperbolic learning
Theories
Shallow
Distribution
Dimension Reduction
Embedding
Clustering
Metric Learning
Deep
HNN
HGNN
Optimization
Applications
CV
Classification
Detection
Segmentation
3D Vison
NLP
Word Embedding
Taxonomy Induction
Classification
Probing
Graph

Theories

Deep Learning

HNN

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Hyperbolic Neural Networks[1]

Multinomial logistic regression, Feed-forward, Recurrent layers in hyperbolic space.[1]

Hyperbolic Neural Networks++[2]

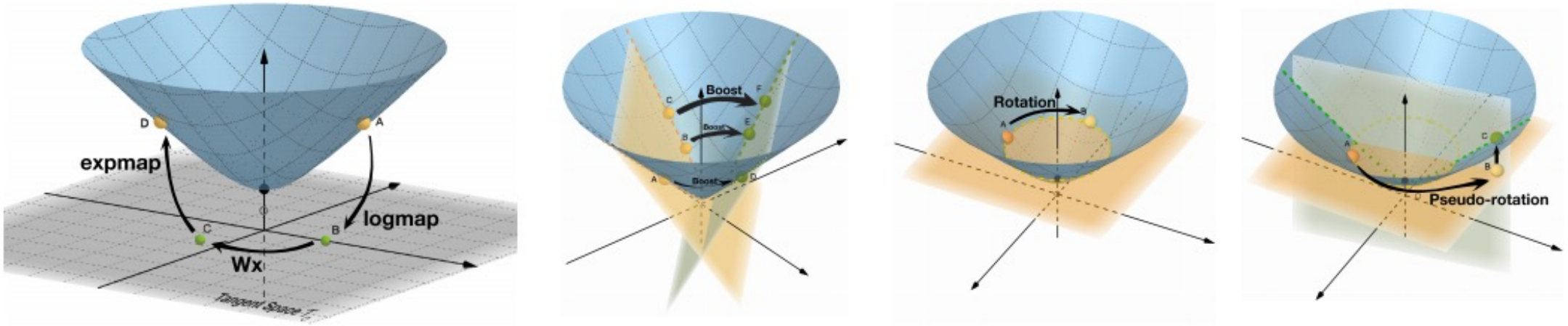
Multinomial logistic regression, Fully-connected layers, Convolutional layers, Attention mechanisms.[2]

[1] Hyperbolic Neural Networks. Octavian-Eugen Ganea et al. NIPS 2018.
[2] Hyperbolic Neural Networks++. Ryohei Shimizu et al. ICLR 2021.

HNN

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Fully Hyperbolic Neural Networks[1]

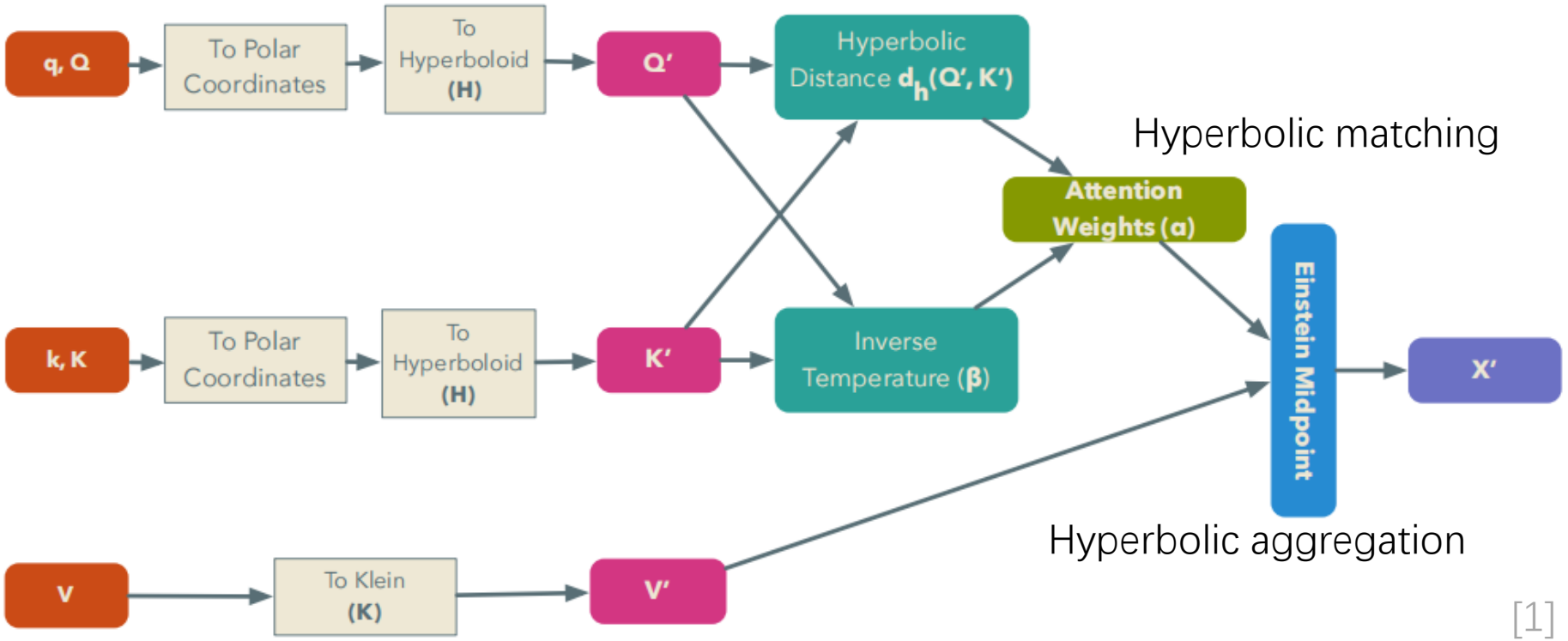


(a) Linear layer formalized in tangent space (b) Lorentz boost (c) Lorentz rotation (d) Pseudo-rotation

[1]

[1] Fully Hyperbolic Neural Networks. Chen et al. ACL 2022.

Hyperbolic Attention Network[1]

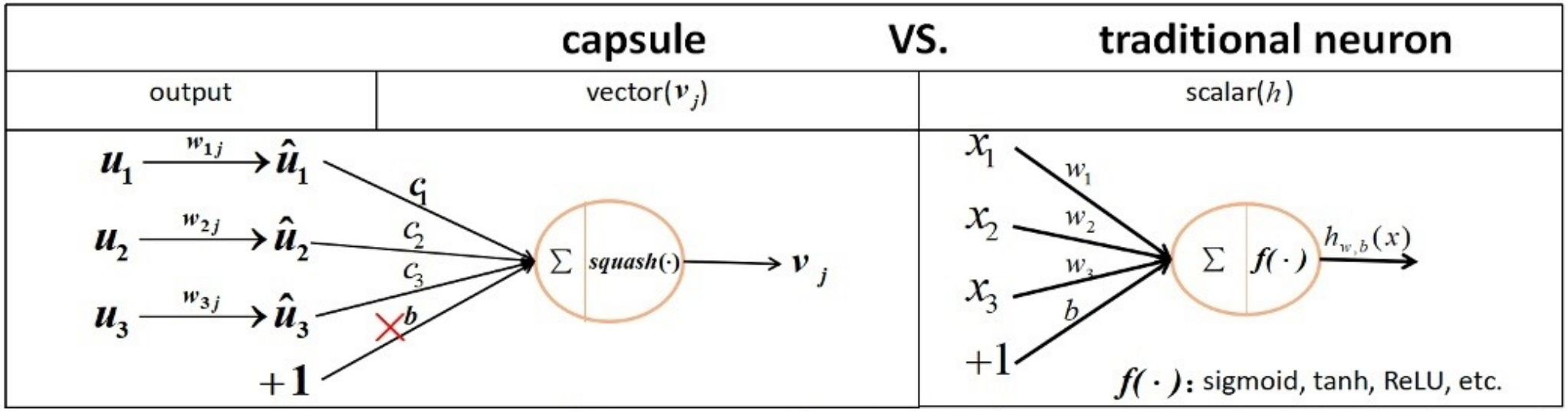


Make the algorithm match more complicated data.

[1] Hyperbolic Attention Network. Gulcehre et al. ICLR 2019

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

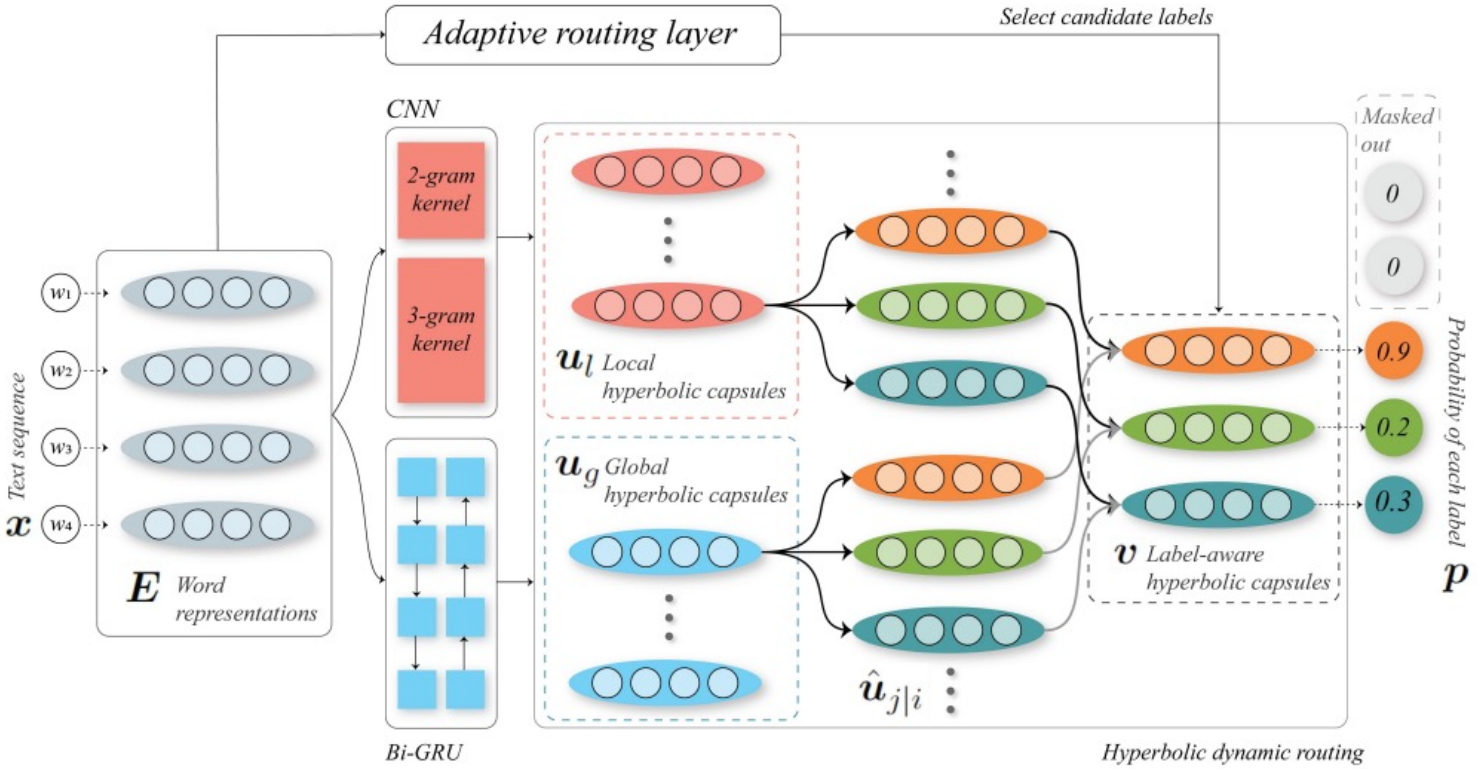
Hyperbolic Capsule Networks[1]



[1] Hyperbolic Capsule Networks for Multi-Label Classification. Chen et al. ACL 2020

HNN

Hyperbolic Capsule Networks[1]



Capture complicated structures among labels and documents

[1]

[1] Hyperbolic Capsule Networks for Multi-Label Classification. Chen et al. ACL 2020

GNN

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

HGNN[1]

Propagation rule in GNN:

$$\mathbf{h}_u^{k+1} = \sigma\left(\sum_{v\in\mathcal{I}(u)} \tilde{\mathbf{A}}_{uv}\mathbf{W}^k\mathbf{h}_v^k\right)$$

Propagation rule in HGNN:

$$\mathbf{h}_u^{k+1} = \sigma\left(\exp_{\mathbf{x}'}\left(\sum_{v\in\mathcal{I}(u)} \tilde{\mathbf{A}}_{uv}\mathbf{W}^k \log_{\mathbf{x}'}(\mathbf{h}_v^k)\right)\right)$$

[1] Hyperbolic Graph Neural Networks. Liu et al. NIPS 2019..

GNN

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

HGCN[1]

GCN

$$\mathbf{h}_i^{\ell,E} = W^\ell \mathbf{x}_i^{\ell-1,E} + \mathbf{b}^\ell$$
$$\mathbf{x}_i^{\ell,E} = \sigma(\mathbf{h}_i^{\ell,E} + \sum_{j\in\mathcal{N}(i)} w_{ij}\mathbf{h}_j^{\ell,E})$$

HGCN

$$\mathbf{h}_i^{\ell,H} = (W^\ell \otimes^{K_{\ell-1}} \mathbf{x}_i^{\ell-1,H}) \oplus^{K_{\ell-1}} \mathbf{b}^\ell$$
$$\mathbf{y}_i^{\ell,H} = \mathrm{AGG}^{K_{\ell-1}}(\mathbf{h}^{\ell,H})_i$$
$$\mathbf{x}_i^{\ell,H} = \sigma^{\otimes^{K_{\ell-1},K_\ell}}(\mathbf{y}_i^{\ell,H})$$

Attention

$$w_{ij} = \mathrm{SOFTMAX}_{j\in\mathcal{N}(i)}(\mathrm{MLP}(\log_{\mathbf{o}}^K(\mathbf{x}_i^H)||\log_{\mathbf{o}}^K(\mathbf{x}_j^H)))$$
$$\mathrm{AGG}^K(\mathbf{x}^H)_i = \exp_{\mathbf{x}_i^H}^K\left(\sum_{j\in\mathcal{N}(i)} w_{ij}\log_{\mathbf{x}_i^H}^K(\mathbf{x}_j^H)\right).$$

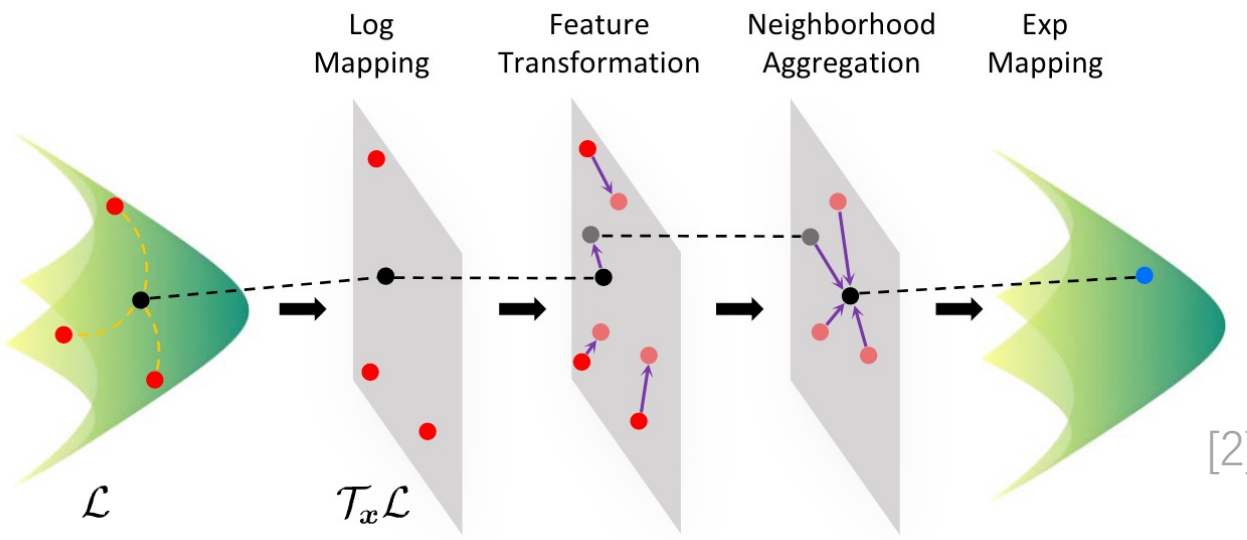
[1] Hyperbolic Graph Convolutional Neural Networks. Chami et al. NIPS 2019.

GNN

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

HGCN [1] /HGNN [2]

Map the nodes to tangent space for aggregation.



[2]

Transformations – Tangent space
Neighborhood Aggregation - Tangent space
Non-linear Activations - Tangent space

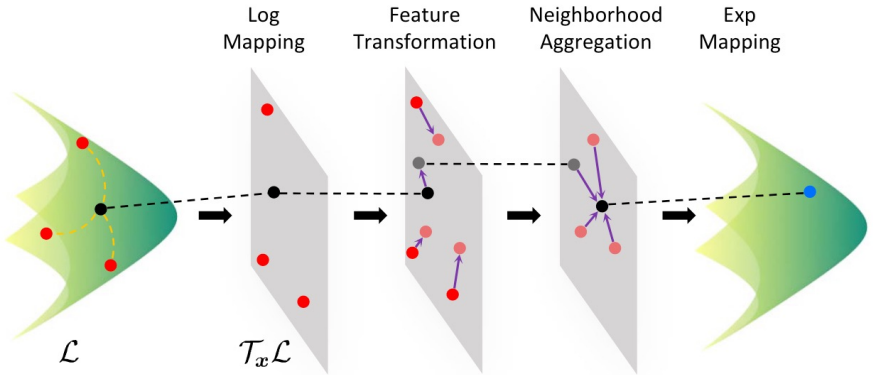
[1] Hyperbolic Graph Convolutional Neural Networks. Chami et al. NIPS 2019.
[2] Hyperbolic Graph Neural Networks. Liu et al. NIPS 2019..
[3] A Hyperbolic-to-Hyperbolic Graph Convolutional Network. Jindou Dai. CVPR 2021.

GNN

H2H-GCN[1]

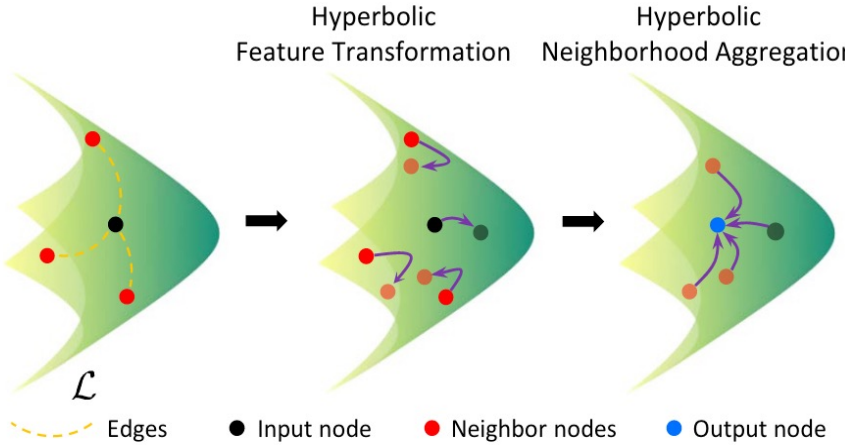
Fully hyperbolic GNN, avoiding approximations.

HGCN



Transformations – Tangent space
Neighborhood Aggregation - Tangent space
Non-linear Activations - Tangent space

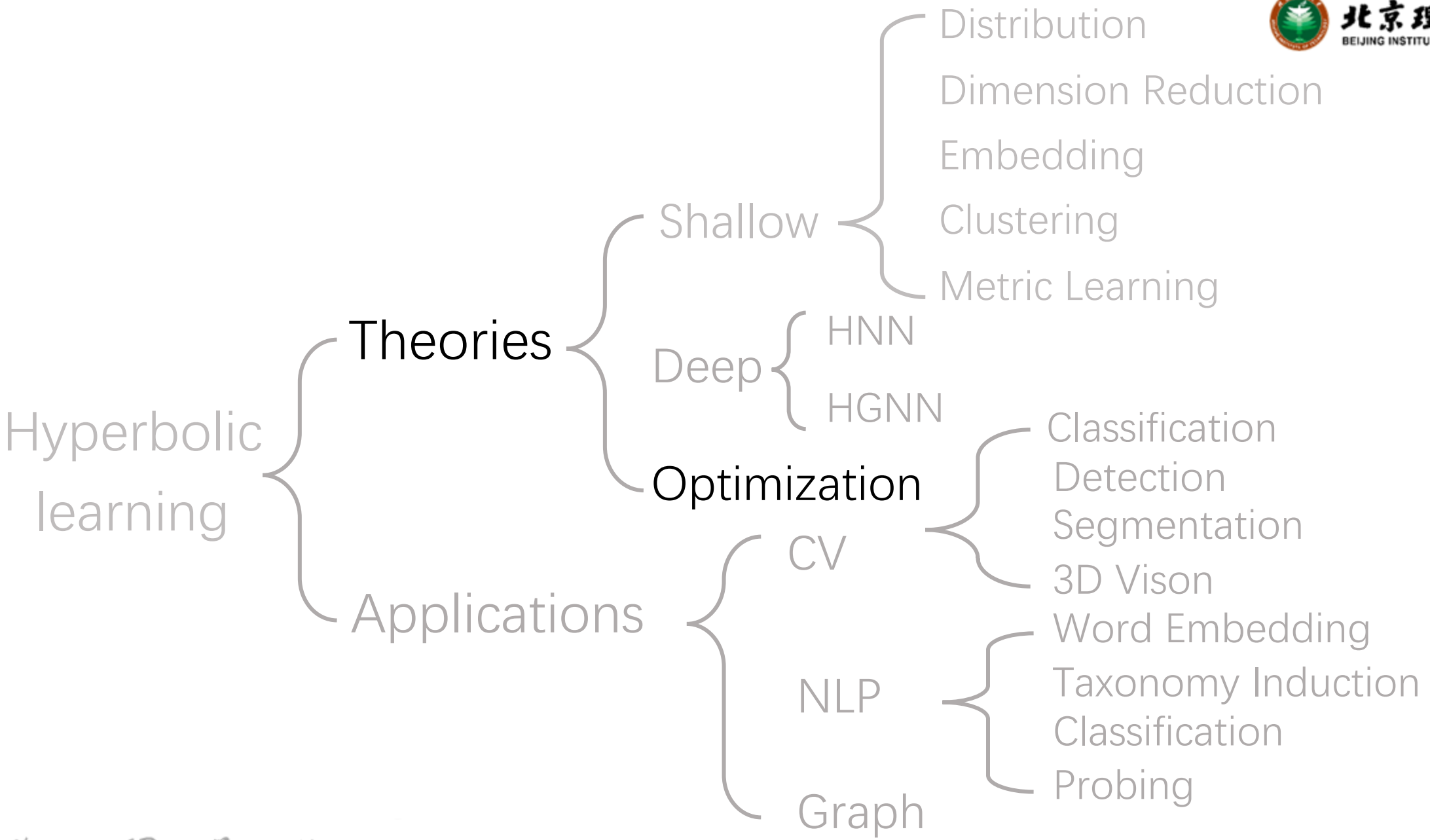
H2H-GCN



[1]

Transformations - Lorentz model
Neighborhood Aggregation - Klein model
Non-linear Activations - Poincaré model

[1] A Hyperbolic-to-Hyperbolic Graph Convolutional Network. Jindou Dai. CVPR 2021.


Hyperbolic learning
Theories
Shallow
Distribution
Dimension Reduction
Embedding
Clustering
Metric Learning
Deep
HNN
HGNN
Optimization
Applications
CV
Classification
Detection
Segmentation
3D Vison
NLP
Word Embedding
Taxonomy Induction
Classification
Probing
Graph

Theories

Optimization

Optimization

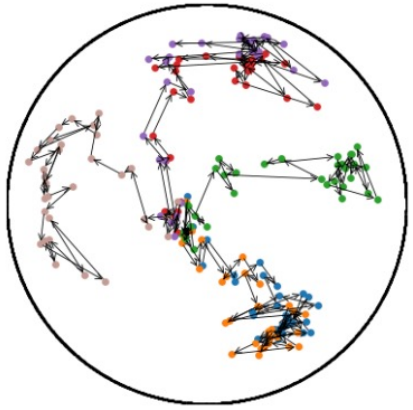
北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- Non-convex problem in the Euclidean view becomes geodesically convex with the right geometry.
- Hyperbolic form of Nesterov-like accelerated gradient
- No analogue of accelerated gradient descent for geodesically convex functions
- The hyperbolic plane needs to search over a much larger area than any fixed dimension of Euclidean space within a ball of radius r.

[1] A No-go Theorem for Robust Acceleration in the Hyperbolic Plane, Hamilton et al. NIPS2021

Optimization

Clipped Hyperbolic Classifiers



[1]

Vanishing gradient problem during backpropagation

Vanishing Gradient Problem

$$p(y=k|\mathbf{x}) \propto \exp(\langle -\mathbf{p}_k \oplus_c \mathbf{x}, \mathbf{a}_k \rangle) \sqrt{\mathfrak{g}^c_{\mathbf{p_k}}(\mathbf{a}_k, \mathbf{a}_k)} d_c(\mathbf{x}, \tilde{H}^c_{\mathbf{a}_k, \mathbf{p}})$$

$\|\mathbf{x}^H\|^2$ approaches 1

$$\frac{\partial \ell}{\partial \mathbf{x}^H} = \frac{(1-\|\mathbf{x}^H\|^2)^2}{4} \nabla \ell(\mathbf{x}^H)$$

$$\frac{\partial \ell}{\partial \mathbf{w}^E} = (\frac{\partial \mathbf{x}^H}{\partial \mathbf{w}^E})^T \frac{\partial \ell}{\partial \mathbf{x}^H}$$

[1] Clipped Hyperbolic Classifiers Are Super-Hyperbolic Classifiers Yunhui Guo. CVPR 2022.

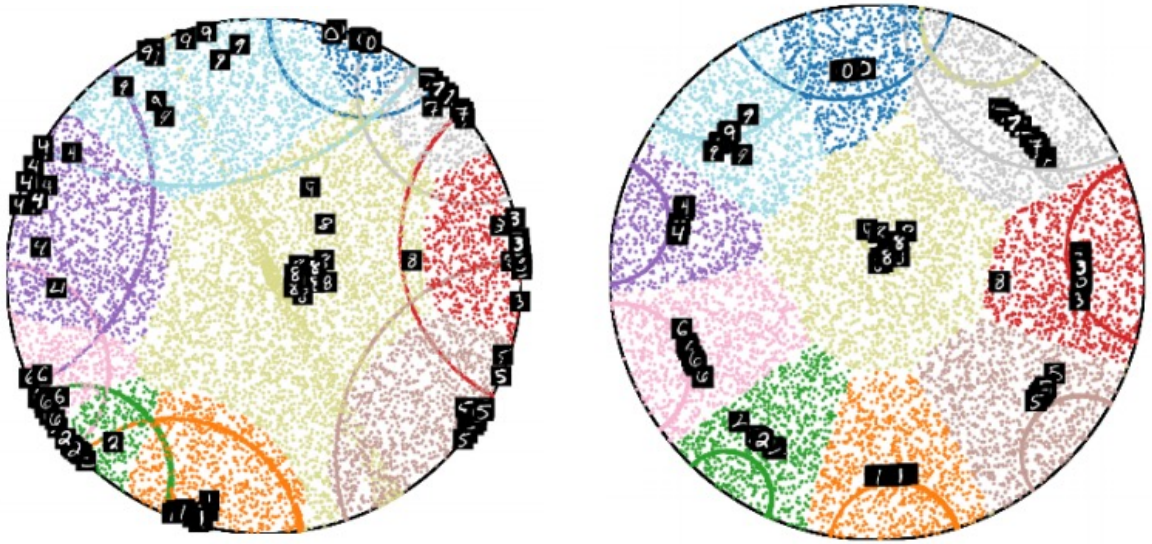
Optimization

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Clipped Hyperbolic Classifiers

Euclidean Feature Clipping

$$\mathrm{CLIP}(\mathbf{x}^E; r) = \min\{1, \frac{r}{\|\mathbf{x}^E\|}\} \cdot \mathbf{x}^E$$



b) Baseline HNNs　　c) Clipped HNNs

[1]

Alleviating the vanishing gradient problem without compromising accuracy

[1] Clipped Hyperbolic Classifiers Are Super-Hyperbolic Classifiers Yunhui Guo. CVPR 2022.

Optimization

- Linear graph embedding (LGE): graph embedding using linear space equipped with an inner product function.
- Hyperbolic graph embedding (HGE): graph embedding in hyperbolic space.

- HGE's performance is limited to ideal noiseless. LGE and HGE's performance with noise has not been discussed.

- LGE and HGE cause a polynomial and exponential error with respect to the radius, and imbalanced data distribution can worsen the error.

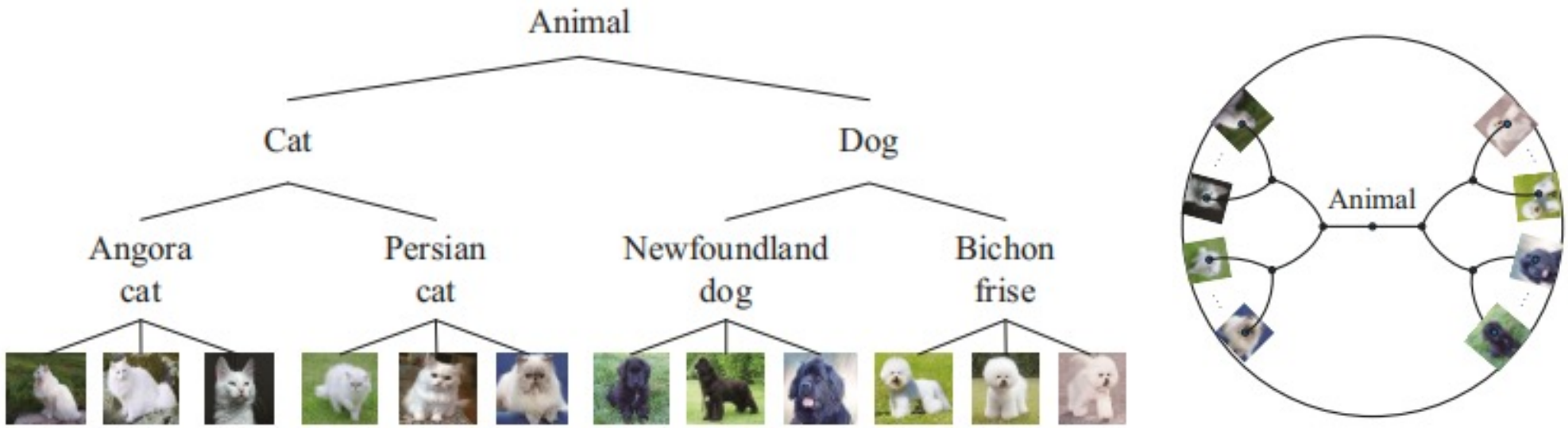
[1] Generalization Error Bounds for Graph Embedding Using Negative Sampling: Linear vs Hyperbolic, Suzuki et al. NIPS2021
[2] Generalization Error Bound for Hyperbolic Ordinal Embedding, Suzuki et al. ICML2021

- Hyperbolic learning
  - Theories
    - Optimization
      - Shallow
        - Distribution
        - Dimension Reduction
        - Embedding
        - Clustering
        - Metric Learning
      - Deep
        - HNN
        - HGNN
  - Applications
    - CV
      - Classification
      - Detection
      - Segmentation
      - 3D Vison
      - Alignment
    - NLP
      - Word Embedding
      - Taxonomy Induction
      - Classification
      - Probing
    - Graph

Application

Computer Vision

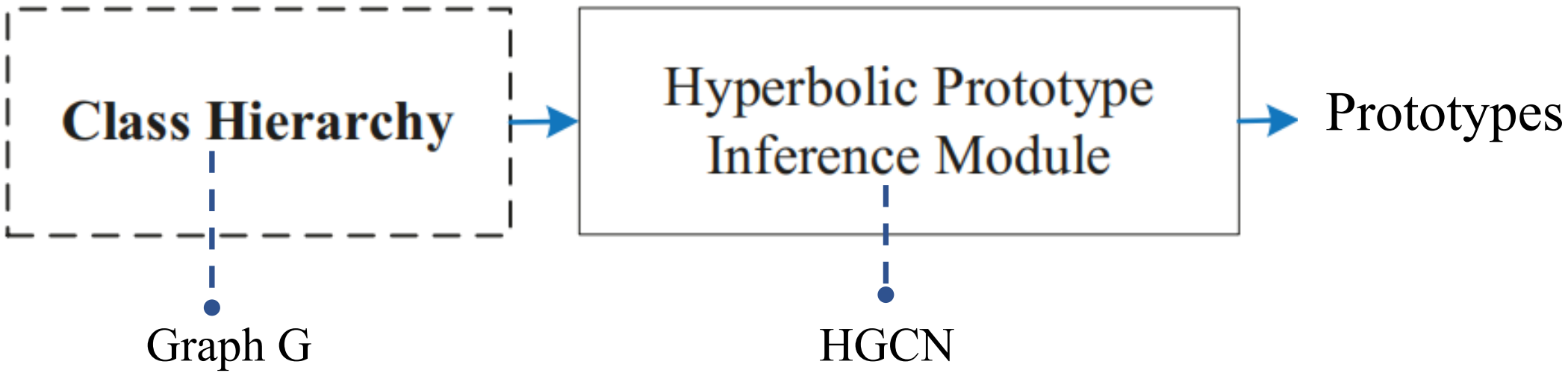
Classification



Class hierarchy [1]

[1] Hyperbolic Knowledge Transfer with Class Hierarchy for Few-Shot Learning, Zhang et al. IJCAI2022

Classification

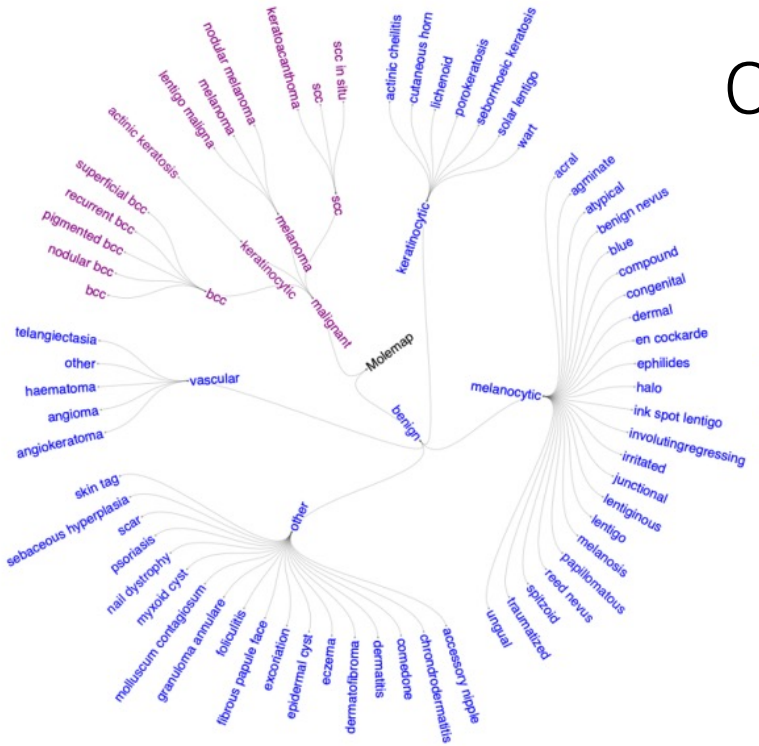


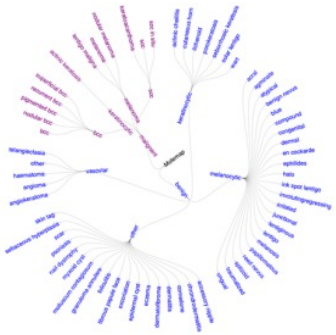
Use class hierarchy in hyperbolic space as prior for Few-shot Learning.

[1] Hyperbolic Knowledge Transfer with Class Hierarchy for Few-Shot Learning, Zhang et al. IJCAI2022

Classification

Class distance guided hyperbolic prototype learning. [1]





Class hierarchy

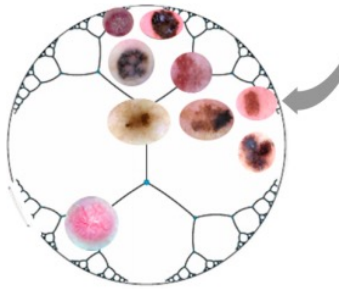
Relation pairs

$\mathcal{P} = \{(u, v)|u = \mathrm{h}(v)\}$

$\mathcal{N}(u) = \{v|(u, v) \notin \mathcal{P}\} \cup \{u\}$

Class relation encoding

Class distance matrix: $\mathbf{D} \in \mathbb{R}_{+}^{K \times K}$



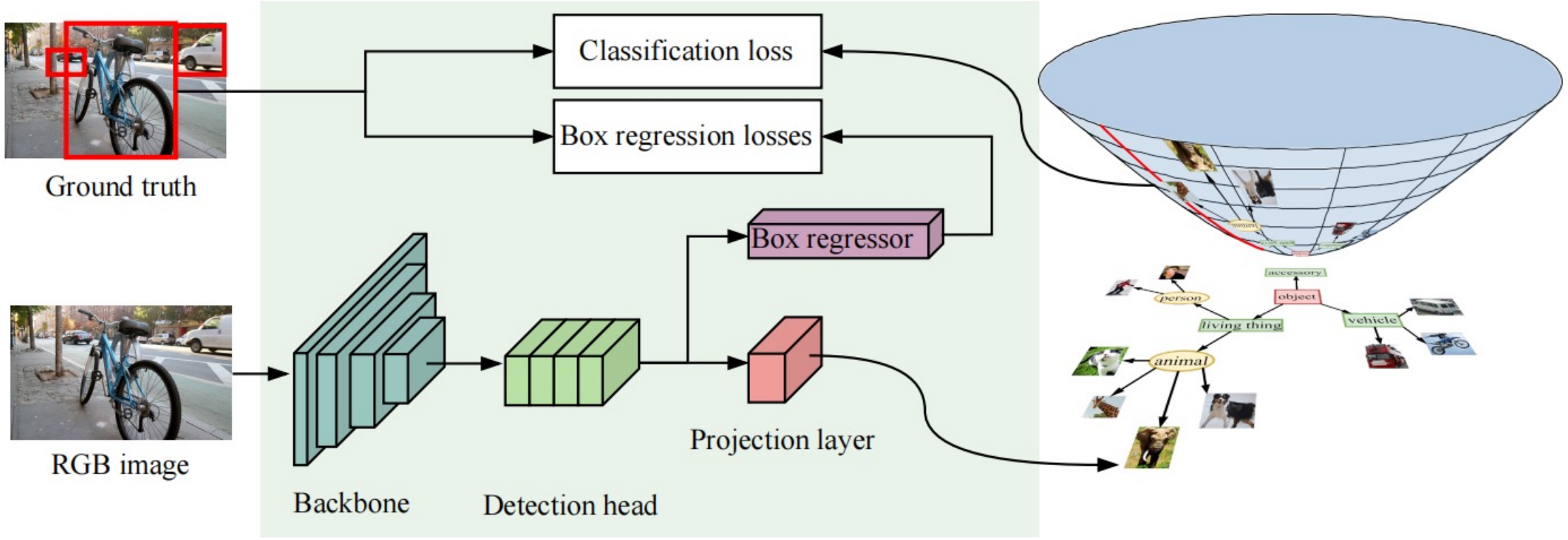
Hyperbolic label representations

德以明理 学以精工

[1]Skin Lesion Recognition with Class-Hierarchy Regularized Hyperbolic Embeddings. Yu et al. MICCAI 2022

Detection

Use hyperbolic embeddings to learn class prototypes for object detector.



[1]On Hyperbolic Embeddings in Object Detection. Lang et al. arXiv:2203.08049

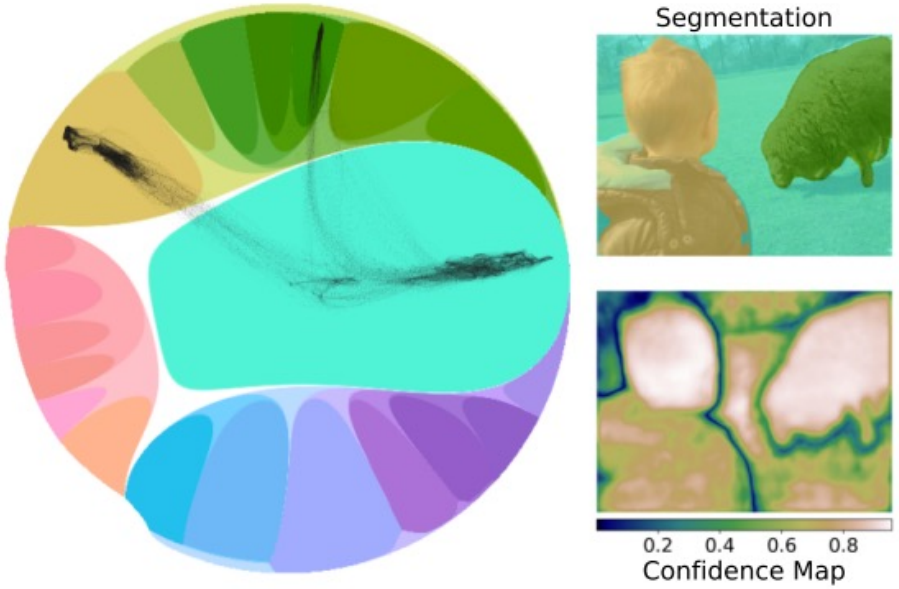
Segmentation

Per-pixel classification in hyperbolic space.

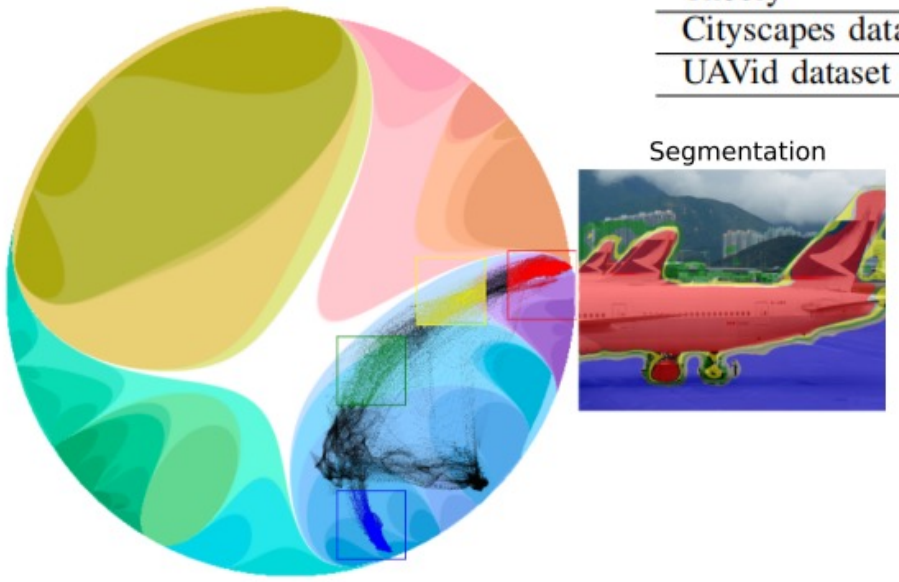
TABLE I [1]

THEORETICAL δ-HYPERBOLICITY AND REAL δ-HYPERBOLICITY VALUES IN CITYSCAPES AND UAVID DATASETS. $\mathbb{B}^n$ IS THE POINCARÉ BALL MODEL OF HYPERBOLIC SPACE.

Spaces	$\mathbb{B}^n$
Theory	0
Cityscapes dataset (δ)	0.16
UAVid dataset (δ)	0.16



(a) Prediction uncertainty for free



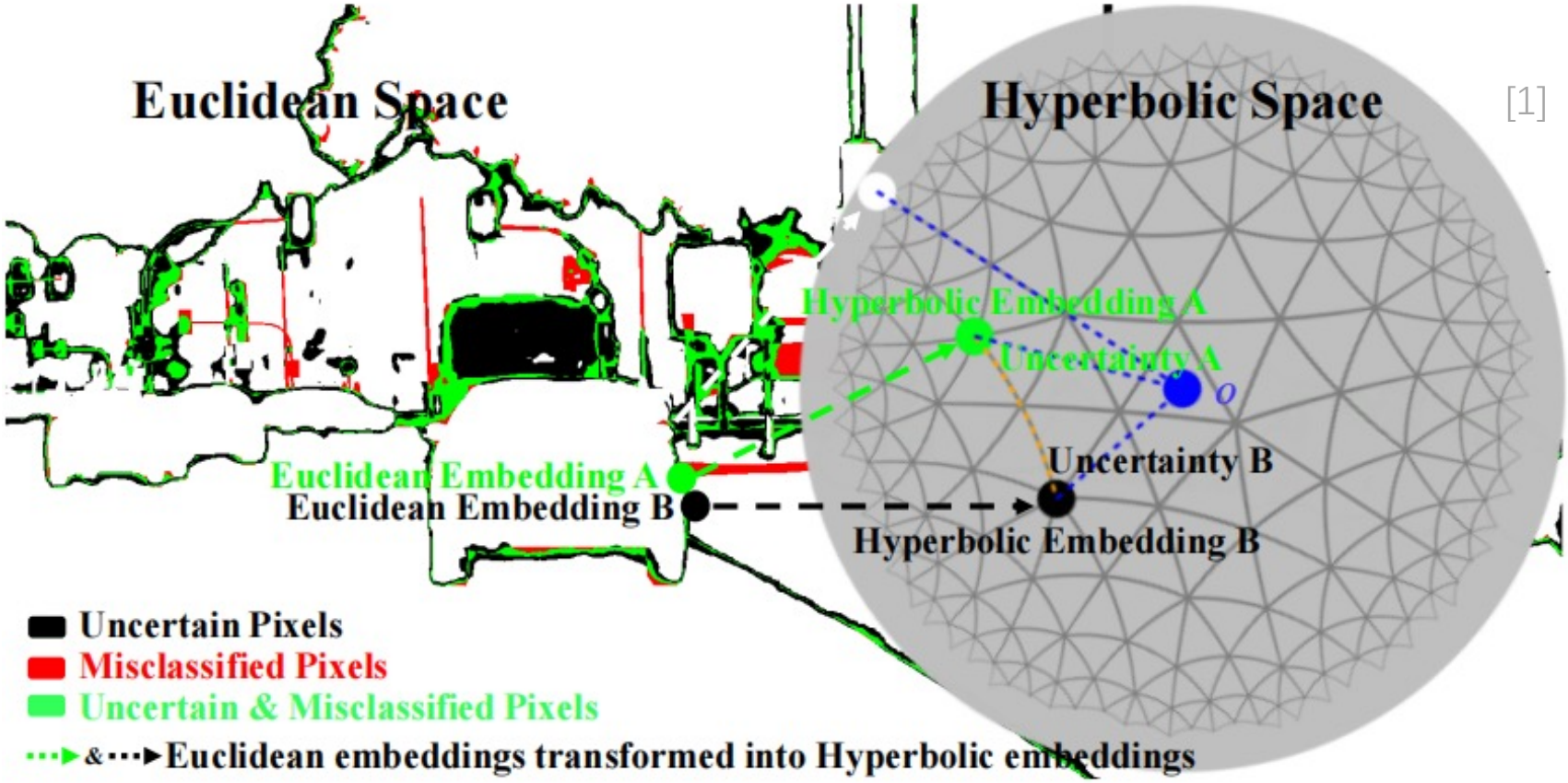
(b) Boundary information for free [2]

德以明理 学以精工

[1] Hyperbolic Uncertainty Aware Semantic Segmentation. Chen et al. arXiv:2203.08881
[2] Hyperbolic Image Segmentation GhadimiAtigh et al. CVPR 2022

Segmentation

Improve segmentation through hyperbolic embeddings with uncertainty.



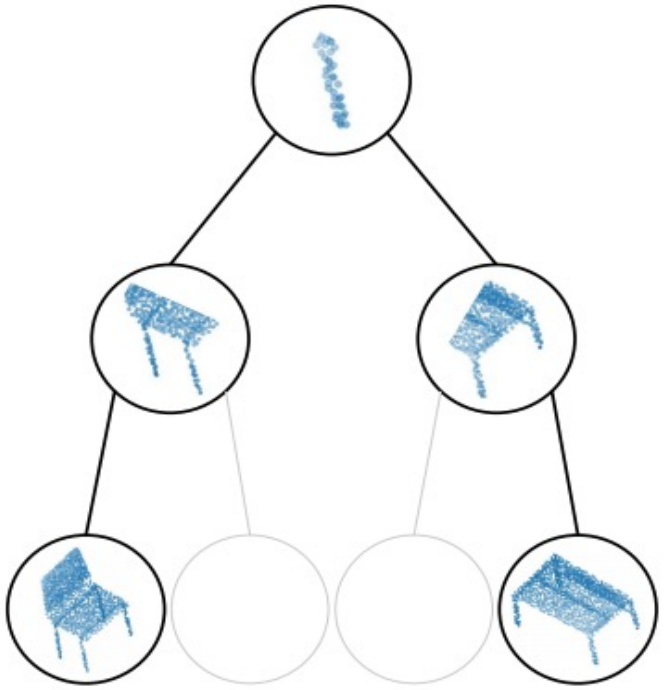
[1]

德以明理 学以精工

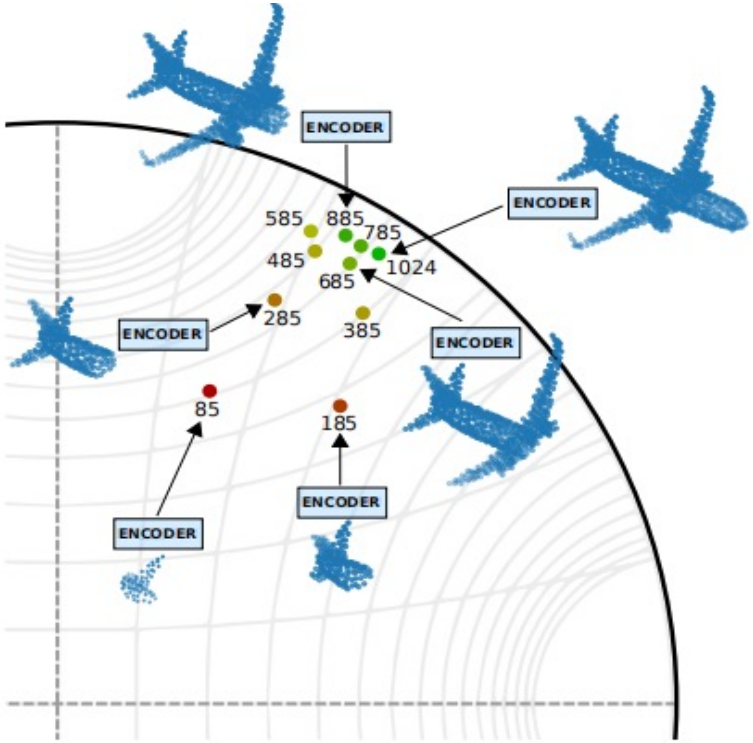
[1] Hyperbolic Uncertainty Aware Semantic Segmentation. Chen et al. arXiv:2203.08881

3D Vision

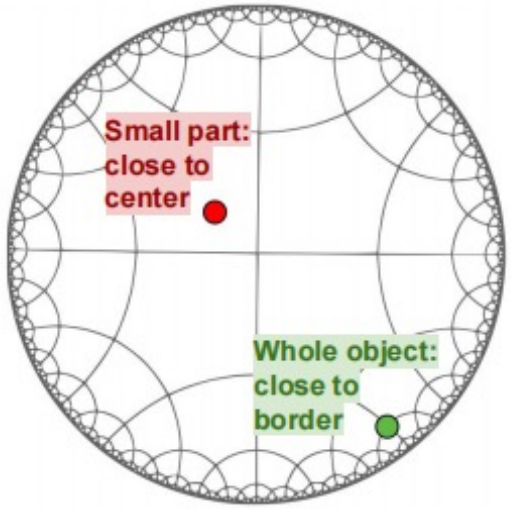
Compositional part-whole hierarchy in 3D objects. [1]



(a)

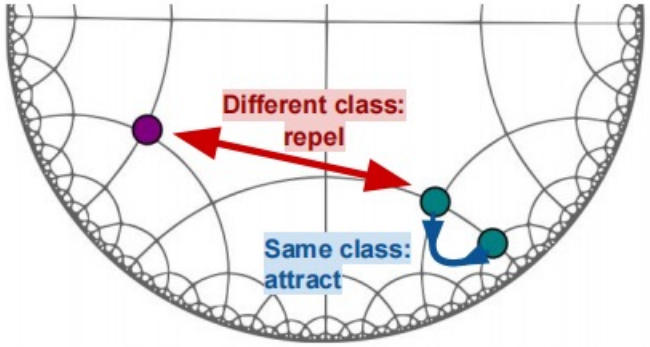


(b)



Hierarchy Regularizer R_{Hier}

(c) [1]



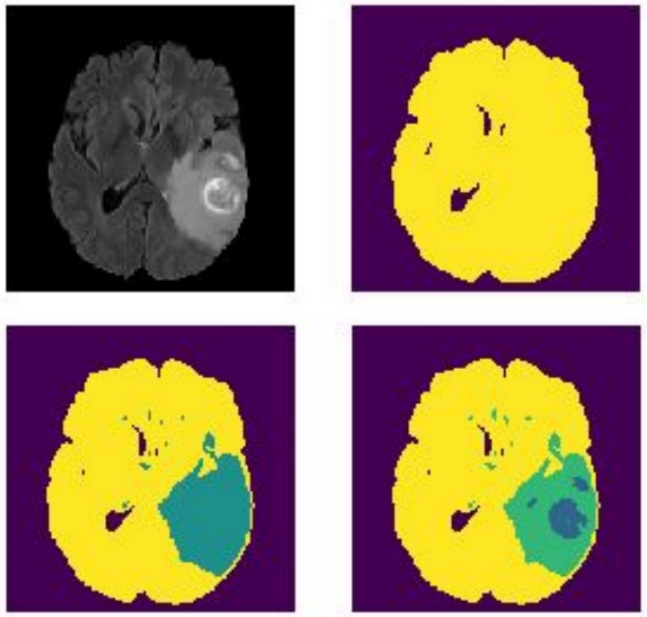
Contrastive Regularizer R_{Contr}

(d) [1]

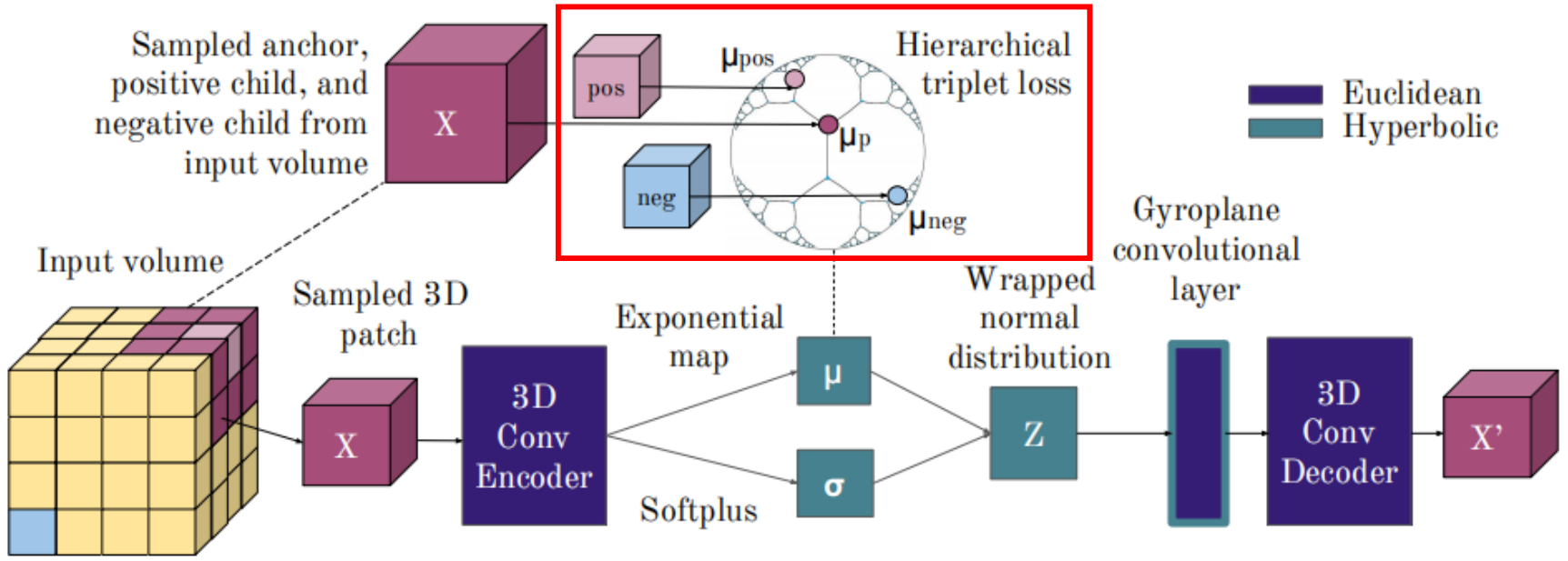
[1] Rethinking the compositionality of point clouds through regularization in the hyperbolic space, Montanaro et al. NIPS2022

3D Vision

Capturing implicit hierarchical structure in 3D biomedical images.



(a) BraTS 2019 [1]



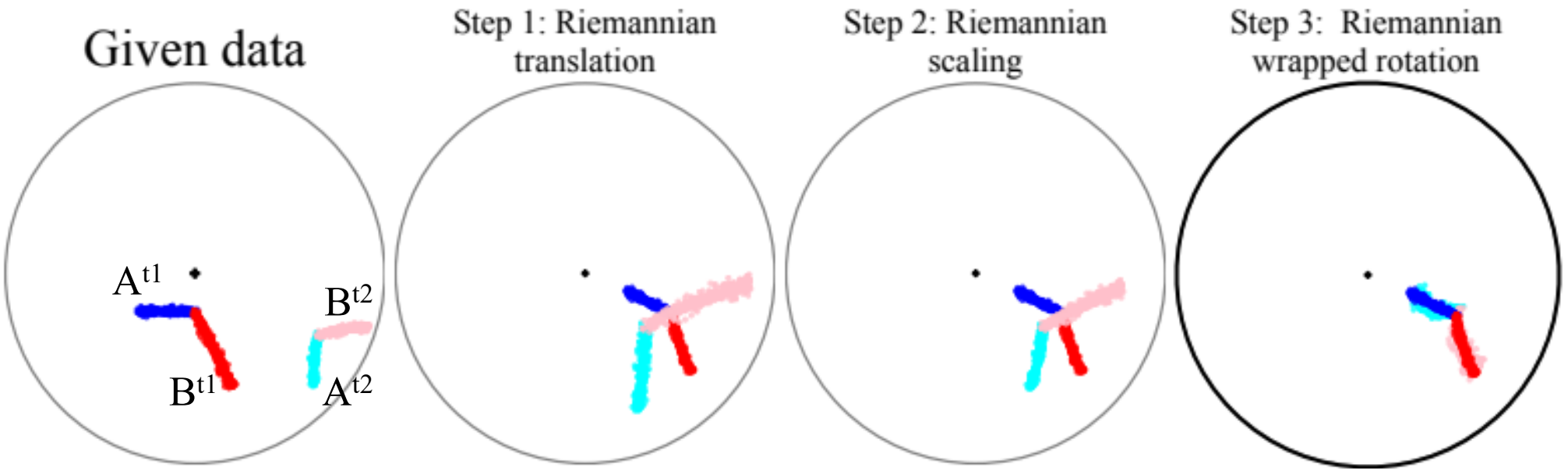
(b) 3D hyperbolic VAE [1]

德以明理 学以精工

[1] Capturing implicit hierarchical structure in 3D biomedical images with self-supervised hyperbolic representations, Hsu et al. NIPS2021

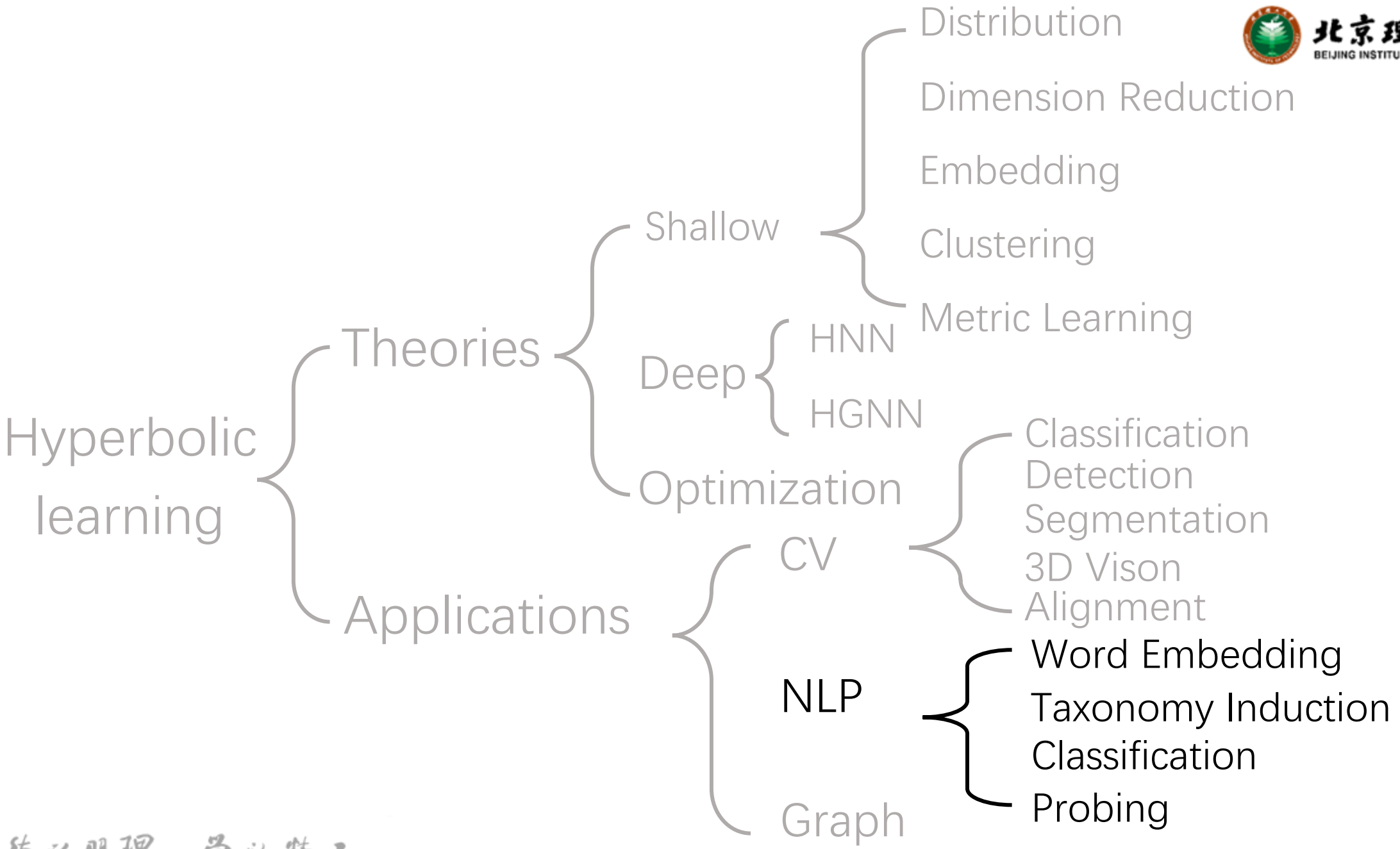
Alignment

Hyperbolic Procrustes analysis (HPA) for label-free alignment of hierarchical datasets.



(a) (b) HPA [1]

[1] Hyperbolic Procrustes Analysis Using Riemannian Geometry, Lin et al. NIPS2021

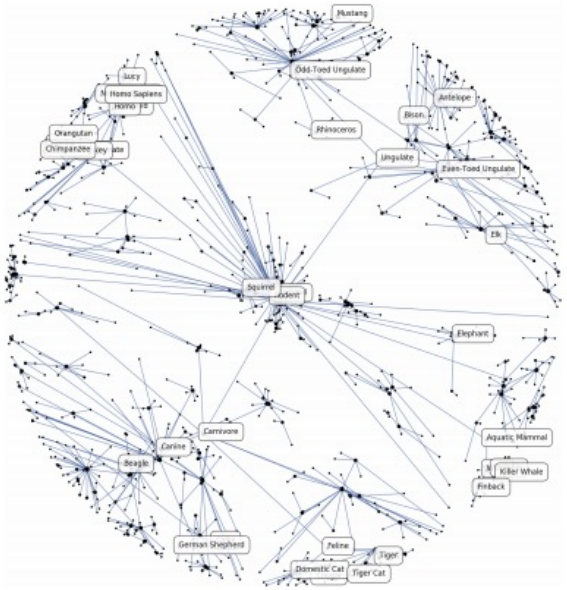


Application

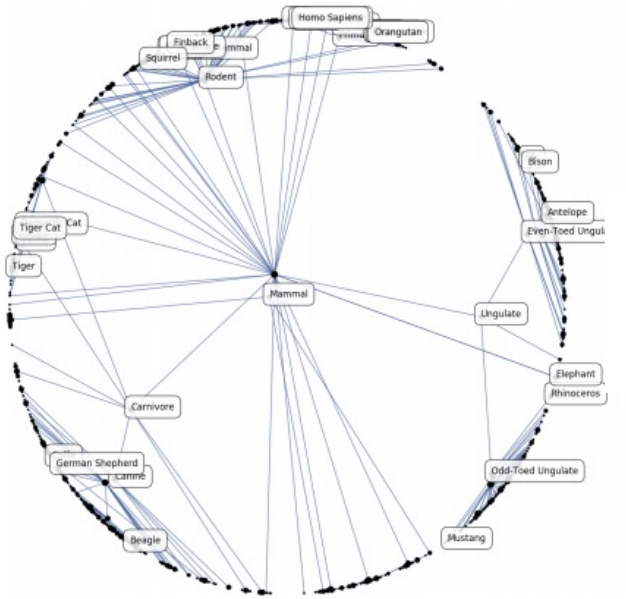
NLP

Word Embedding

Embed the WordNet Noun hierarchy into the Poincaré ball.[1]



(a) Intermediate embedding after 20 epochs



(b) Embedding after convergence

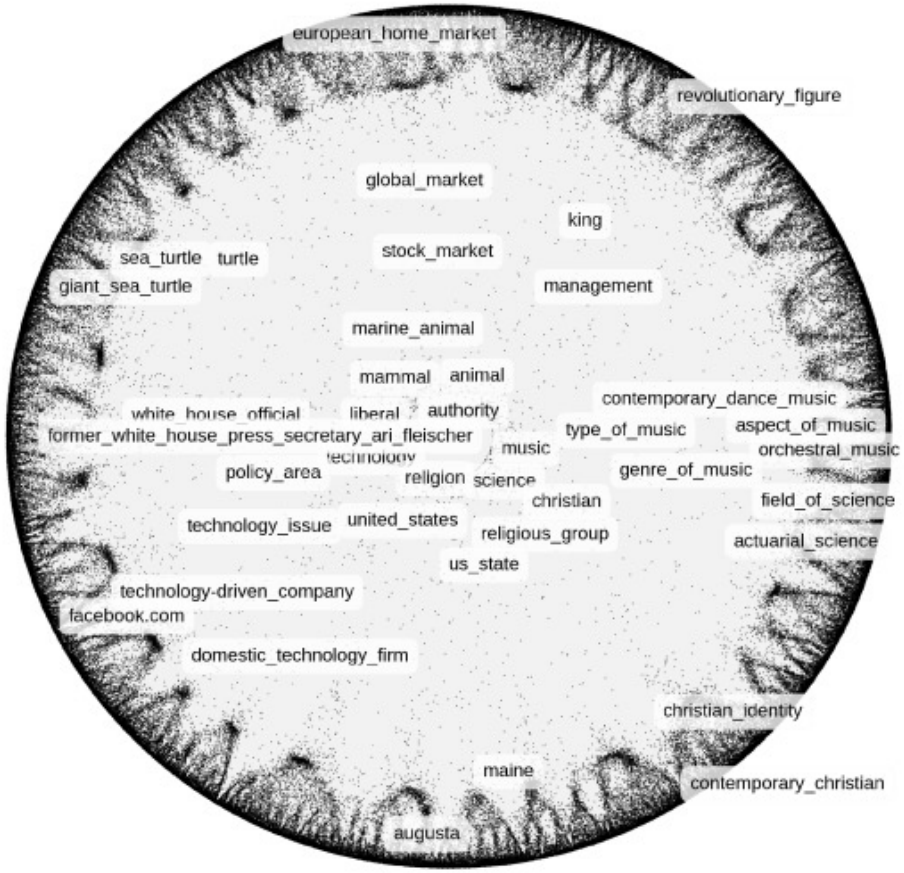
$$\mathcal{L}(\Theta) = \sum_{(u,v)\in\mathcal{D}} \log \frac{e^{-d(u,v)}}{\sum_{v'\in\mathcal{N}(u)} e^{-d(u,v')}}$$

德以明理 学以精工

[1] Poincaré Embeddings for Learning Hierarchical Representations. Nickel et al. NIPS 2017

Word Embedding

Hyperbolic embedding of Hearst Graph.[1]



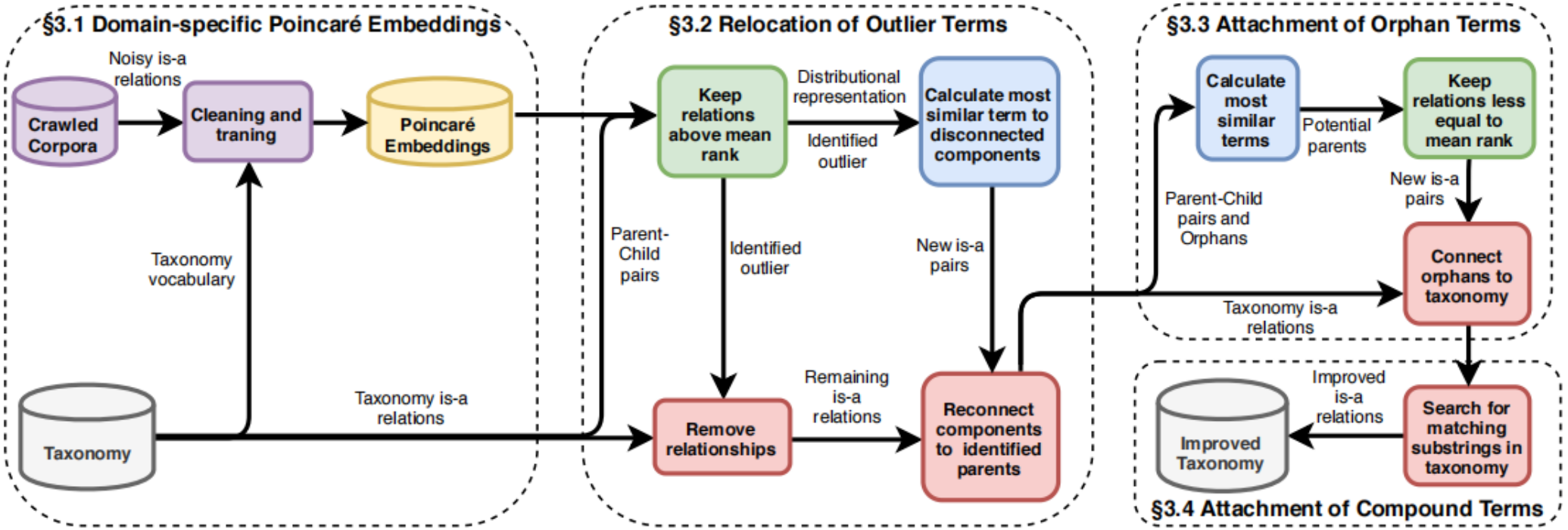
- extract potential is-a relationships using Hearst patterns
- build Hearst Graph
- embed Hearst Graph in hyperbolic space

[1] Inferring Concept Hierarchies from Text Corpora via Hyperbolic Embeddings. Le et al. ACL 2019

Taxonomy Induction

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Taxonomy Refinement using Hyperbolic Word Embeddings.[1]

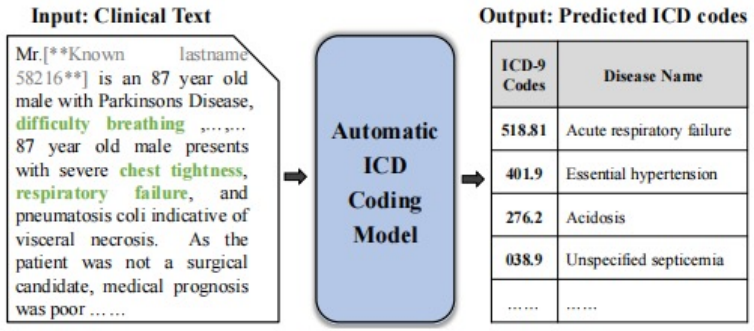


Better capture hierarchical lexical-semantic relationships

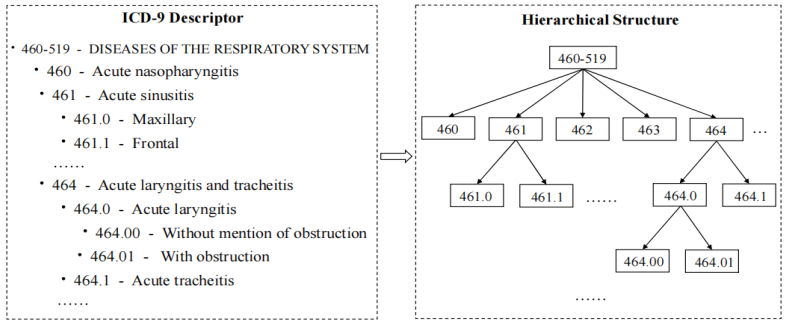
[1] Every child should have parents: a taxonomy refinement algorithm based on hyperbolic term embeddings. Aly et al. ACL 2019

Classification

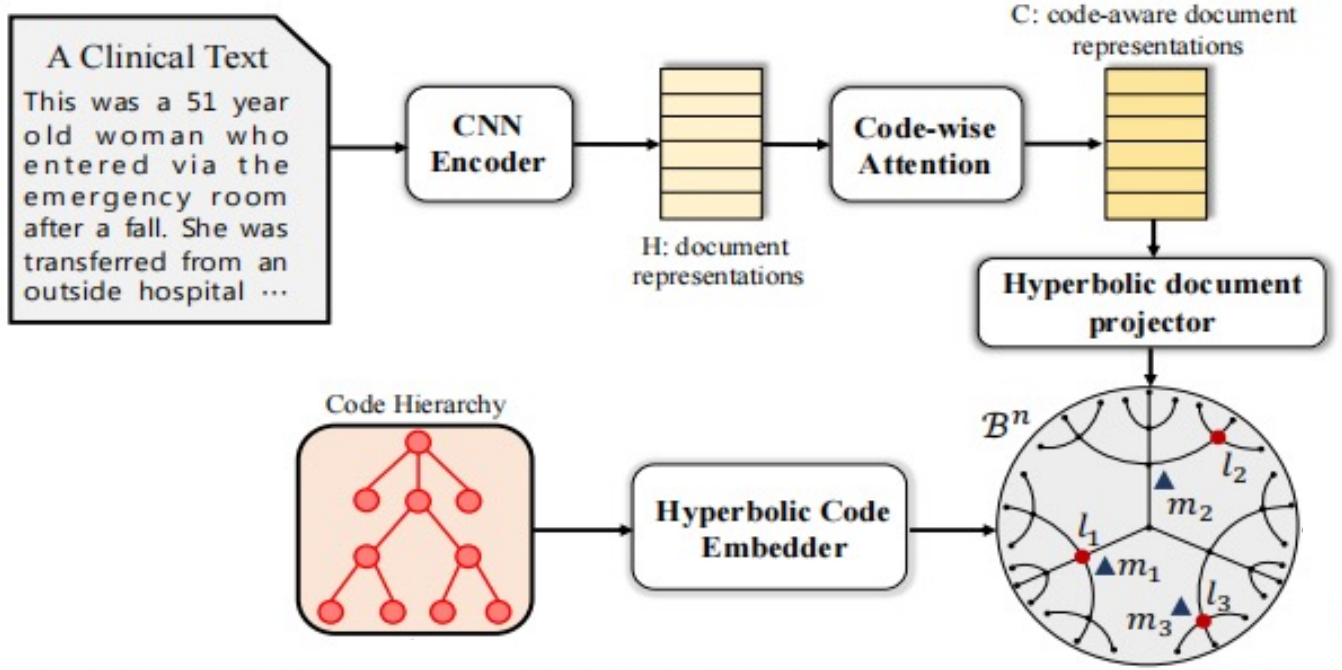
Automatic ICD Coding——a multi-label classification task.



(a) ICD coding[1]



(b) Code Hierarchy[1]

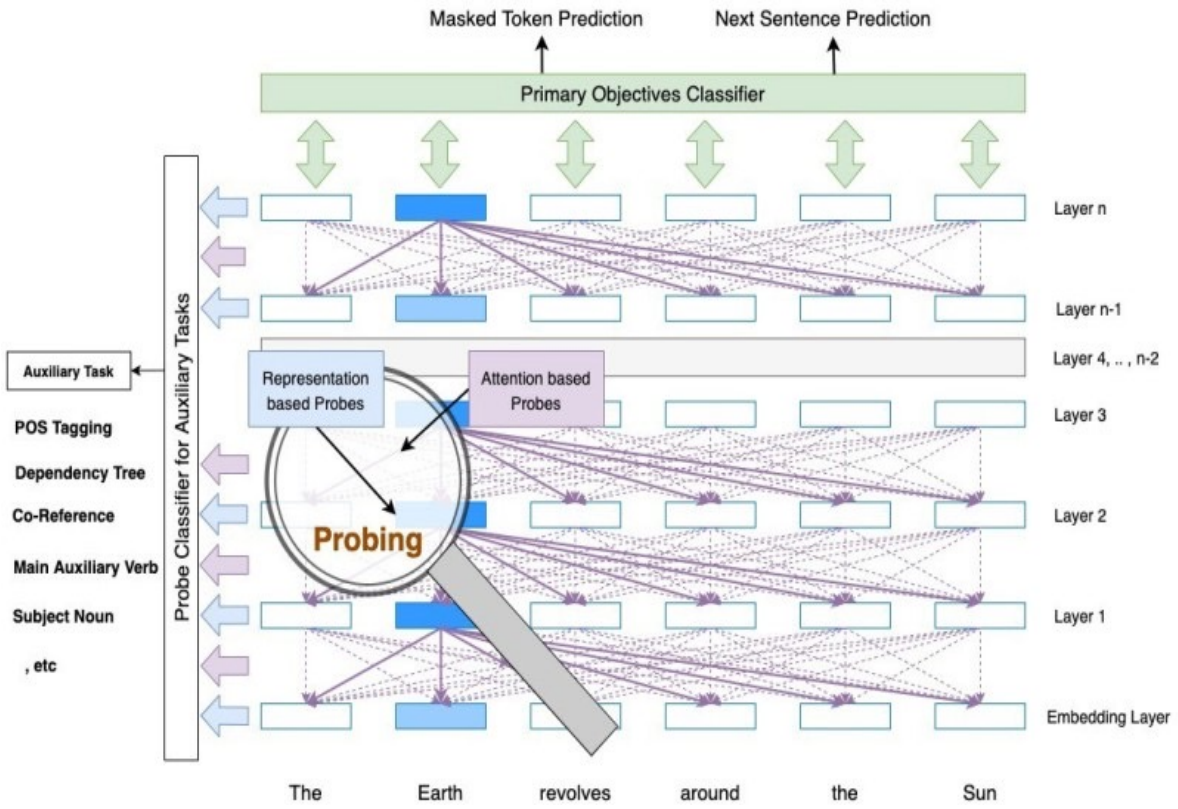


(c) Hyperbolic module [1]

德以明理 学以精工

[1] HyperCore: Hyperbolic and Co-graph Representation for Automatic ICD Coding. Nickel et al. NIPS 2017

Probing

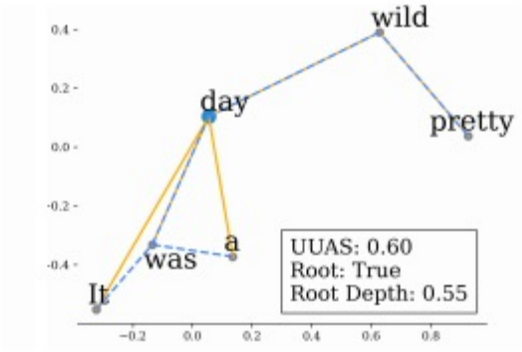


(a) Probing BERT[1]

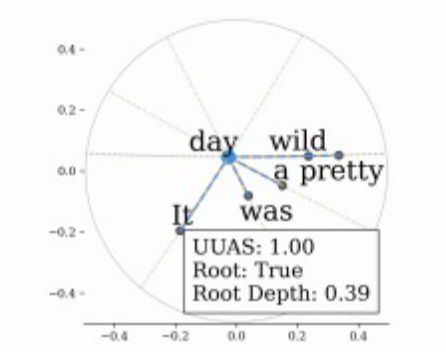
Poincaré probe discovers more syntactic information.

$$\boldsymbol{p}_i = \exp_{\mathbf{0}}(\boldsymbol{P}\boldsymbol{h}_i)$$

$$\boldsymbol{q}_i = \boldsymbol{Q} \otimes_c \boldsymbol{p}_i$$



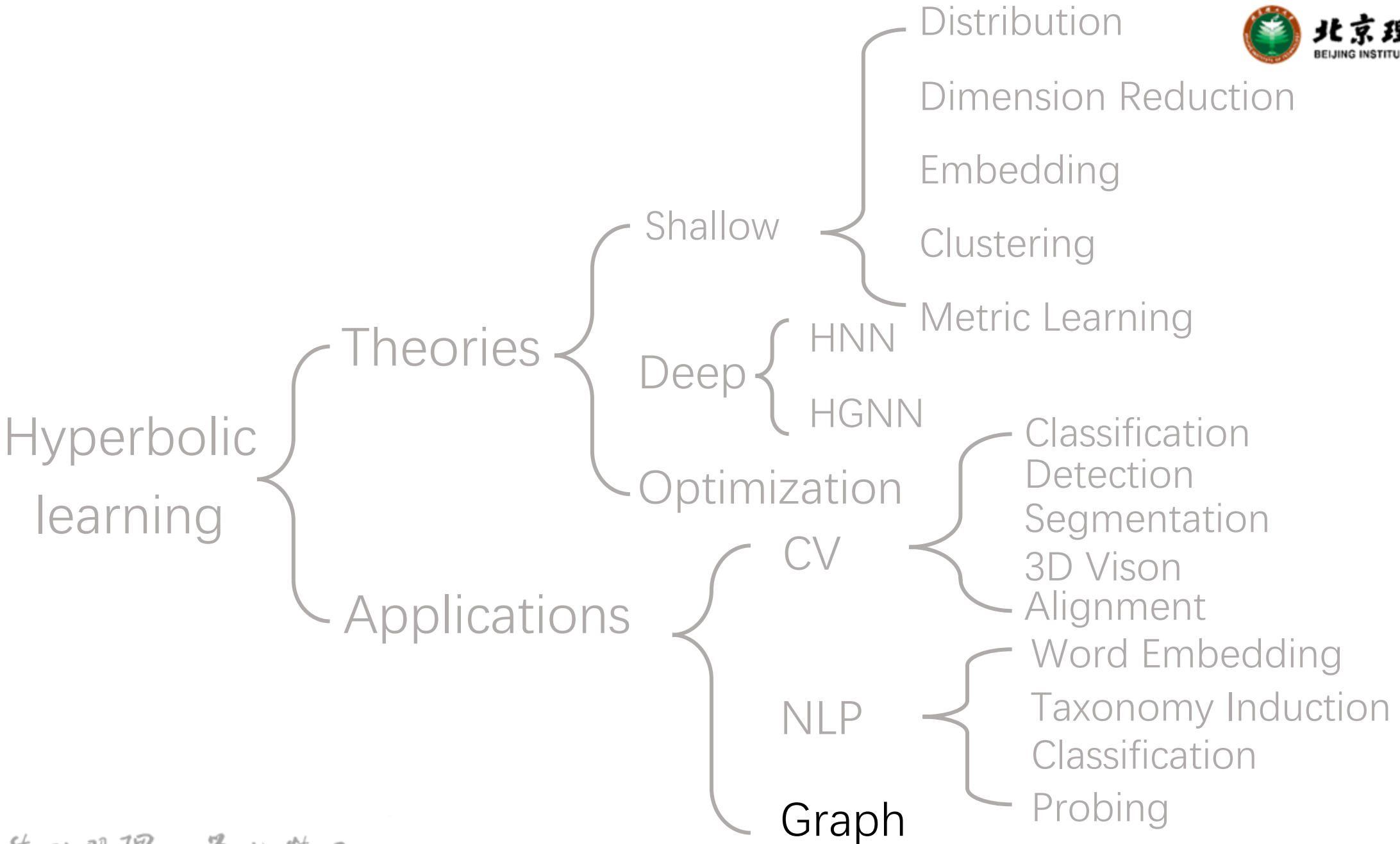
(C) BERTBASE7



(D) BERTBASE7 [2]

德以明理 学以精工

[1] https://www.topbots.com/linguistics-wisdom-of-nlp-models/
[2]Probing BERT in Hyperbolic Spaces. Chen et al. ICLR 2021

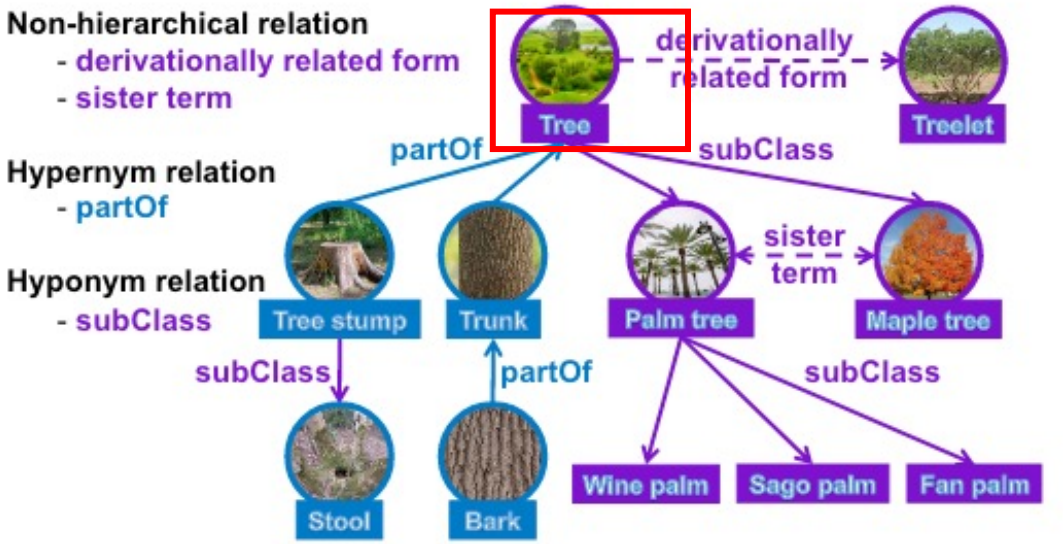


Application

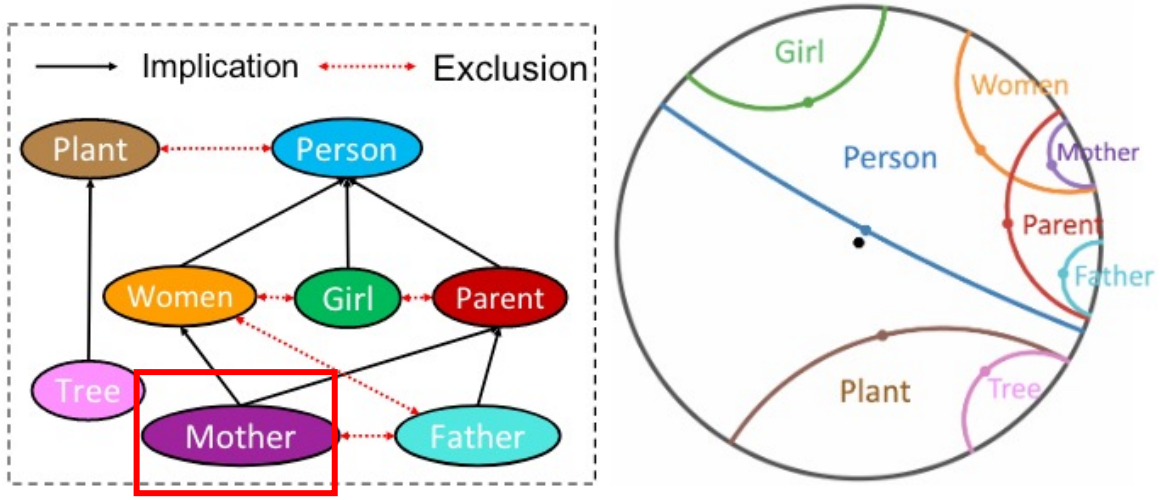
Graph

Knowledge Graph Embedding

Multi-relation or Multi-label



Modeling Heterogeneous Hierarchies[1]



Hyperbolic structured multi-label prediction[1]

[1] Modeling Heterogeneous Hierarchies with Relation-specific Hyperbolic Cones, Yushi Bai et al. NIPS 2021.
[2] Hyperbolic Embedding Inference for Structured Multi-Label Prediction. Bo Xiong et al. NIPS 2022.

Knowledge Graph Embedding

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

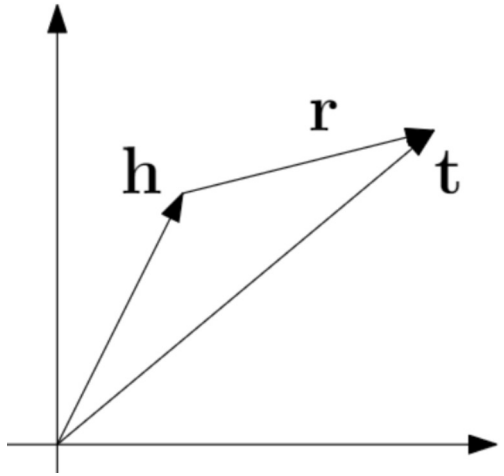
Solve multi-relation problem in hyperbolic KGE.

Bilinear models： regard relations as linear transformations.

For a triple (h, r, t), $hW_r \approx t$

Translational models： regard relations as translations.

$h + r \approx t$

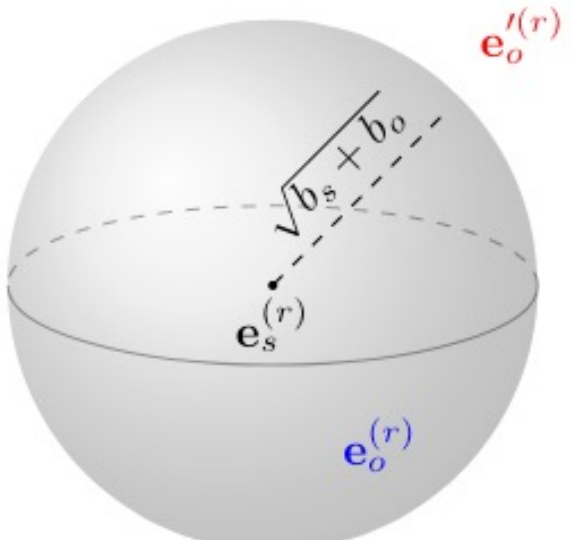


Knowledge Graph Embedding

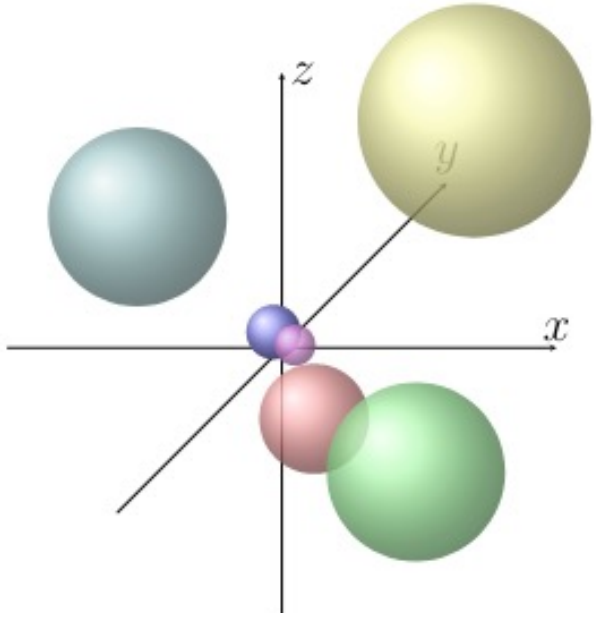
Different “balls” for different hierarchies.

Basic score function

$$\begin{aligned}\phi(e_s, r, e_o) &= -d(\mathbf{e}_s^{(r)}, \mathbf{e}_o^{(r)})^2 + b_s + b_o \\ &= -d(\mathbf{R}\mathbf{e}_s, \mathbf{e}_o + \mathbf{r})^2 + b_s + b_o \end{aligned} \quad (1)$$



(a) [1]



(b) [1]

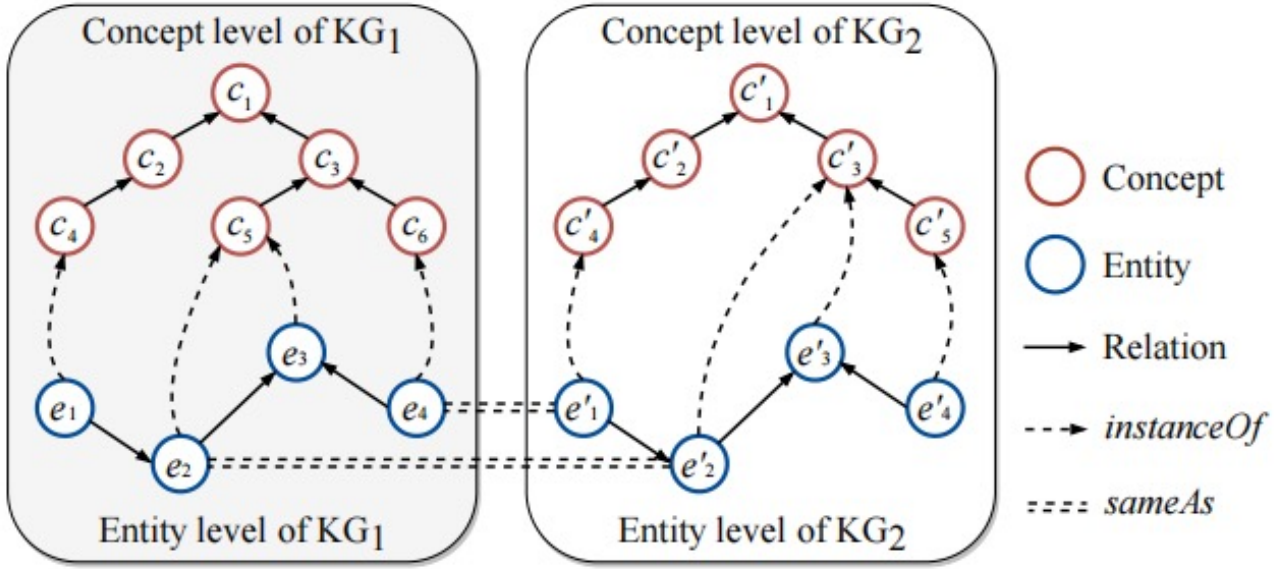
Hyperbolic score function

$$\begin{aligned}\phi_{\text{MuRP}}(e_s, r, e_o) &= -d_{\mathbb{B}}(\mathbf{h}_s^{(r)}, \mathbf{h}_o^{(r)})^2 + b_s + b_o \\ &= -d_{\mathbb{B}}(\exp_{\mathbf{0}}^c(\mathbf{R}\log_{\mathbf{0}}^c(\mathbf{h}_s)), \mathbf{h}_o \oplus_c \mathbf{r}_h)^2 + b_s + b_o \end{aligned} \quad (2)$$

[1] Multi-relational Poincaré Graph Embeddings, Balaževic et al. NIPS 2019.

Knowledge Graph Embedding

Hyperbolic knowledge association with translation in the hyperbolic space.



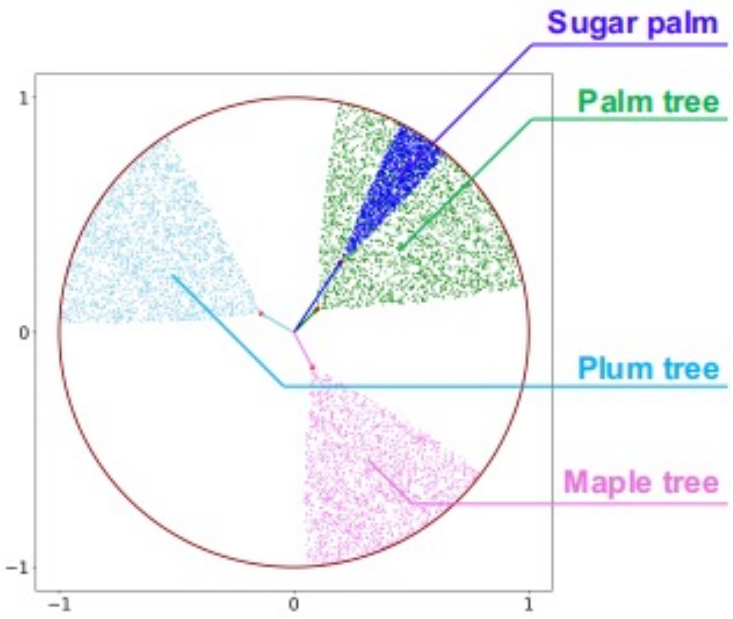
$$f(\tau) = d_{\mathbb{D}}(\mathbf{u}_h^{(0)} \oplus \mathbf{u}_r^{(0)}, \mathbf{u}_t^{(0)}) \quad (1)$$

Figure 1: Illustration of two kinds of knowledge associations (i.e., *sameAs* and *instanceOf*) in KGs. [1]

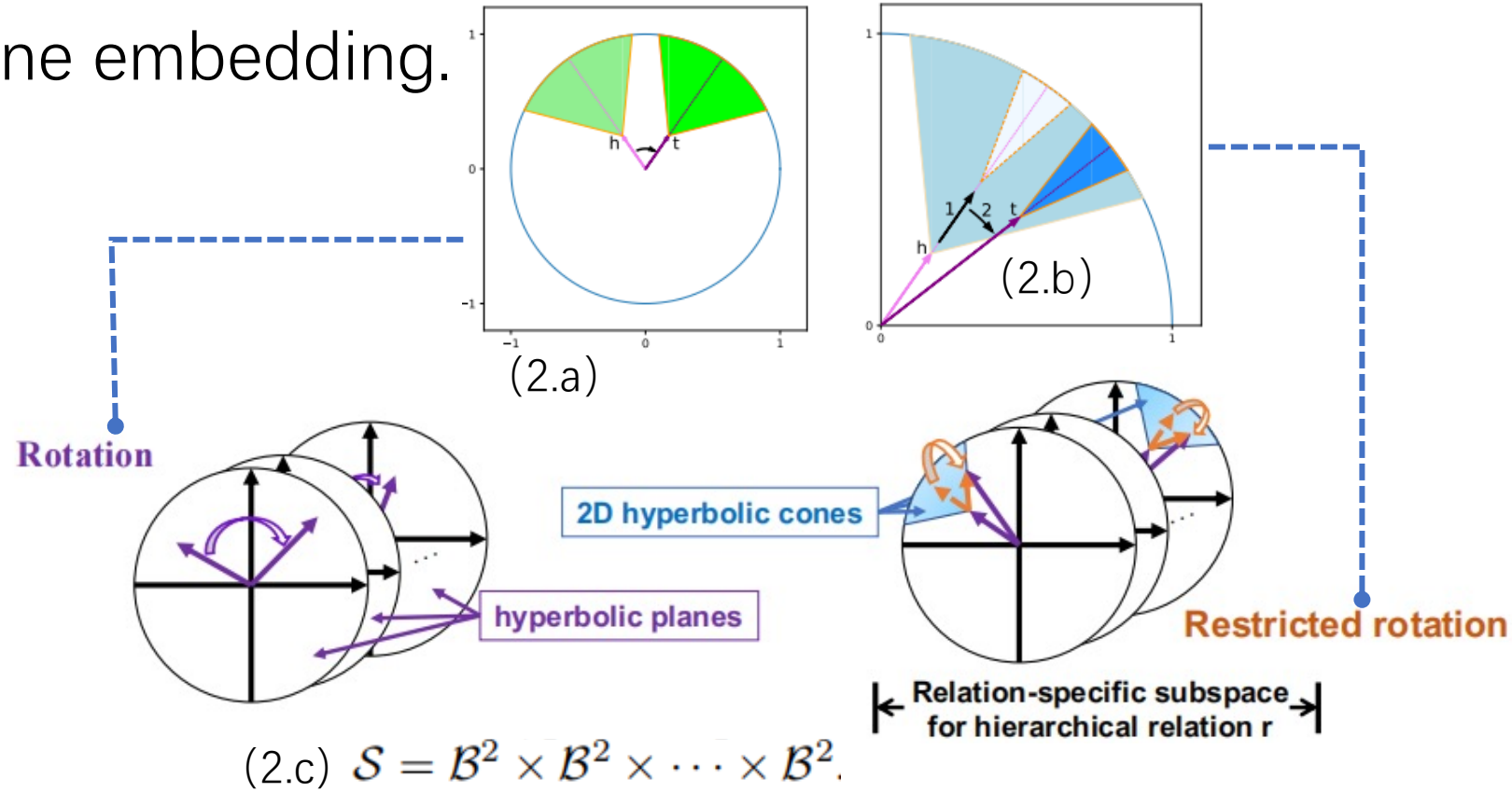
德以明理 学以精工 [1] Knowledge Association with Hyperbolic Knowledge Graph Embeddings, Sun et al. EMNLP 2020.

Knowledge Graph Embedding

Model hierarchies by cone embedding.



(1) Hyperbolic entailment cones [1]

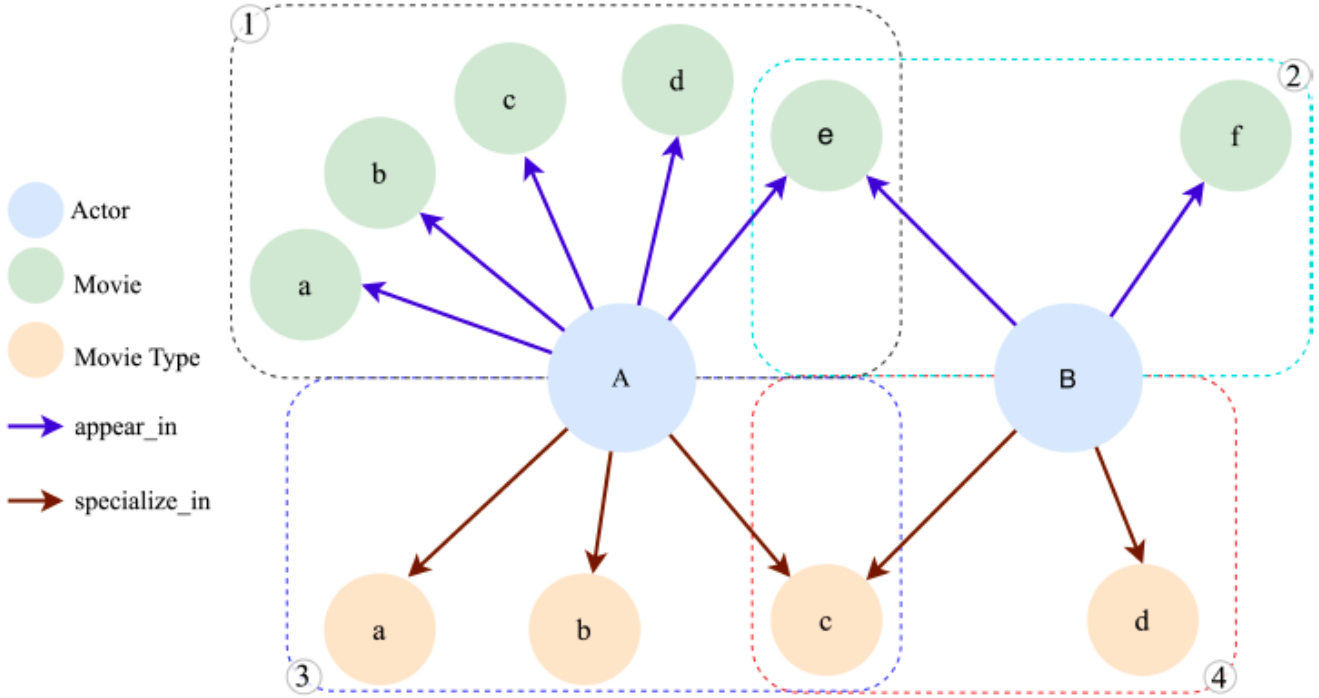


(2) ConE Model[1]

[1] Modeling Heterogeneous Hierarchies with Relation-specific Hyperbolic Cones, Bai et al. NIPS2021

Knowledge Graph Embedding

Choose suitable curvatures for various hierarchies. [1]



$$s(h,r,t) = -d_{\mathbb{B}}^{c_{h,r}}(\mathbf{h}_{e,rot}^{\mathbf{H}} \oplus^{c_{h,r}} \boldsymbol{\varepsilon}_r^{\mathbf{H}}, \mathbf{t}^{\mathbf{H}})^2 + b_h + b_t$$

[1] Hyperbolic Hierarchical Knowledge Graph Embeddings for Link Prediction in Low Dimensions. Wenjie Zheng, et al. EMNLP 2021.

Challenges

Challenges in Hyperbolic learning

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- Currently only limited to representation learning.
- Objective functions remains limited to hyperbolic distance, which is equivalent to L1 norm. There is scope for development of more complex classification and regression objectives.
- **Unstable** training of hyperbolic networks..
- Learn to **adjust curvature** to data and problem.
- **Large-scale** hyperbolic learning.

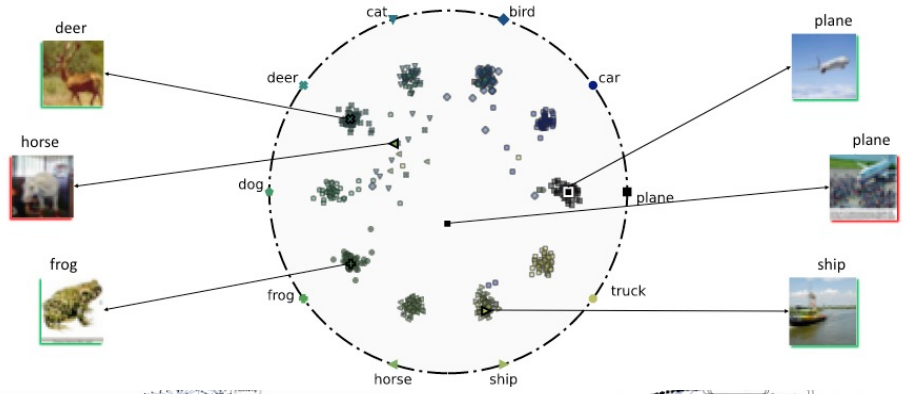
德以明理 学以精工

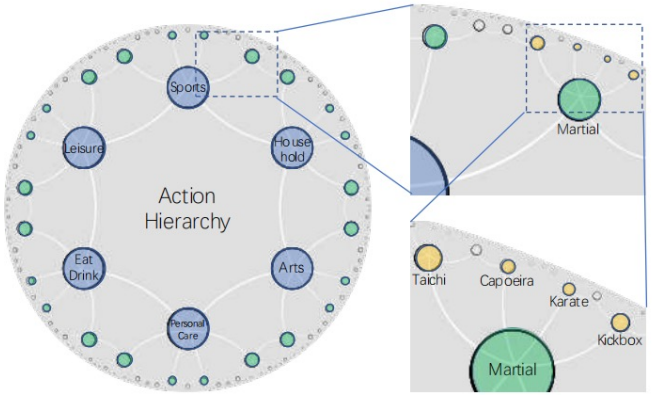
[1] Hyperbolic Networks: Theory, Architectures and Applications. Department of Computer Science, Virginia Tech, Arlington, VA.

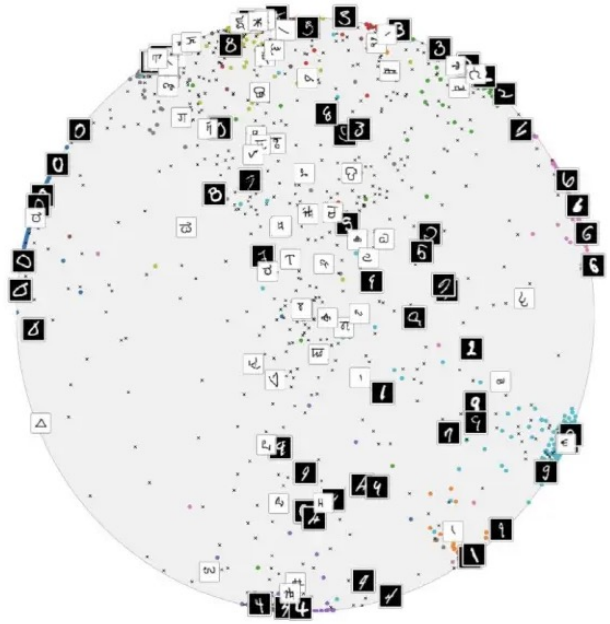
Challenges in Hyperbolic learning

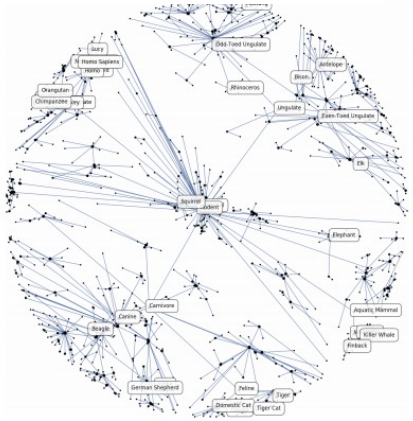
北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

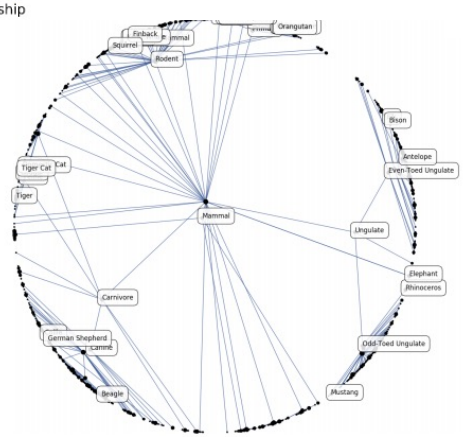
- Currently only **limited to** representation learning.





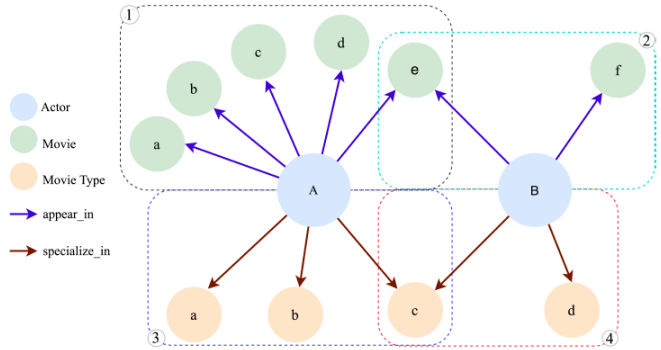






(a) Intermediate embedding after 20 epochs

(b) Embedding after convergence



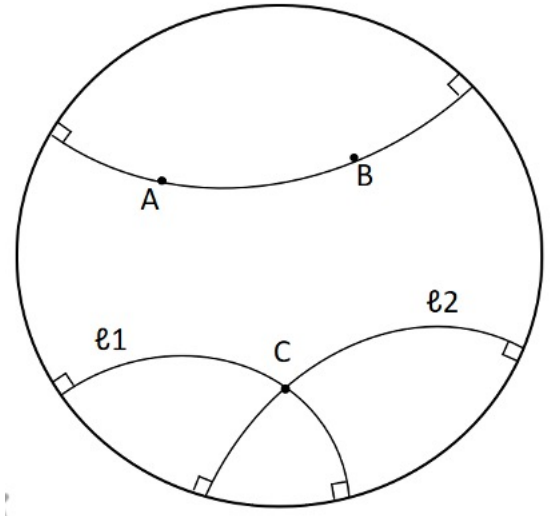
德以明理 学以精工

[1] Hyperbolic Networks: Theory, Architectures and Applications. Department of Computer Science, Virginia Tech, Arlington, VA.

Challenges in Hyperbolic learning

- Currently only **limited to** representation learning
- Objective functions remains limited to hyperbolic distance, which is equivalent to L1 norm. There is scope for development of **more suitable** classification and regression objectives.

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



德以明理 学以精工

[1] Hyperbolic Networks: Theory, Architectures and Applications. Department of Computer Science, Virginia Tech, Arlington, VA.

Challenges in Hyperbolic learning

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- Currently only **limited to** representation learning
- Objective functions remains limited to hyperbolic distance, which is equivalent to L1 norm. There is scope for development of **more suitable** classification and regression objectives.
- **Unstable** training of hyperbolic networks.

德以明理 学以精工

[1] Hyperbolic Networks: Theory, Architectures and Applications. Department of Computer Science, Virginia Tech, Arlington, VA.

Challenges in Hyperbolic learning

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- Currently only **limited to** representation learning
- Objective functions remains limited to hyperbolic distance, which is equivalent to L1 norm. There is scope for development of **more suitable** classification and regression objectives.
- Further study into the **hyperbolic gradient descent** is needed to provide new techniques for stable training.
- Learn to **adjust curvature** to data and problem.

德以明理 学以精工

[1] Hyperbolic Networks: Theory, Architectures and Applications. Department of Computer Science, Virginia Tech, Arlington, VA.

Challenges in Hyperbolic learning

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

- Currently only **limited to** representation learning
- Objective functions remains limited to hyperbolic distance, which is equivalent to L1 norm. There is scope for development of **more suitable** classification and regression objectives.
- Further study into the **hyperbolic gradient descent** is needed to provide new techniques for stable training.
- Learn to **adjust curvature** to data and problem.
- **Large-scale** hyperbolic learning.

德以明理 学以精工

[1] Hyperbolic Networks: Theory, Architectures and Applications. Department of Computer Science, Virginia Tech, Arlington, VA.

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

Q & A

德以明理　学以精工

Thanks for your listening!