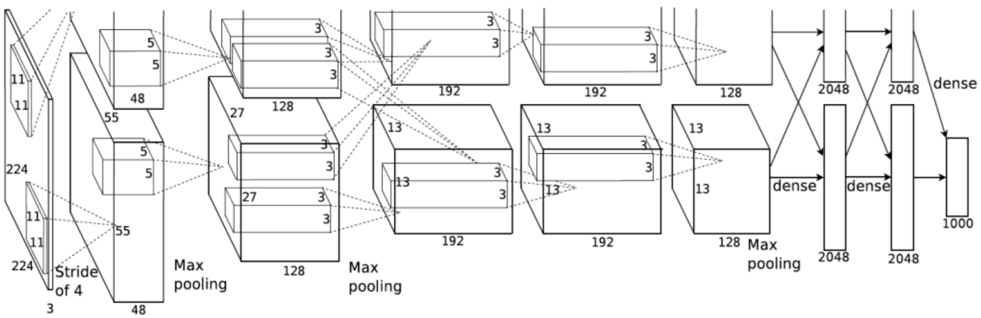
Hyperbolic Deep Learning

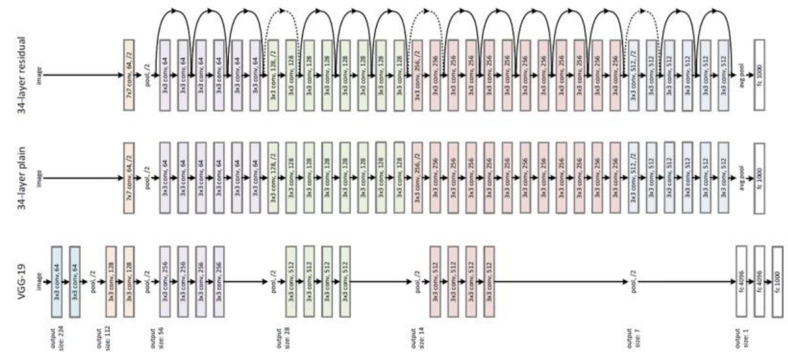
Fabio Galasso | 4 September 2023

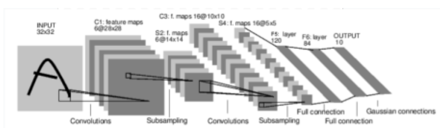
PINLAB Perception and Intelligence

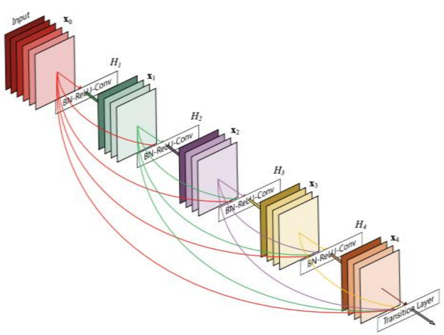
SAPIENZA UNIVERSITÀ DI ROMA

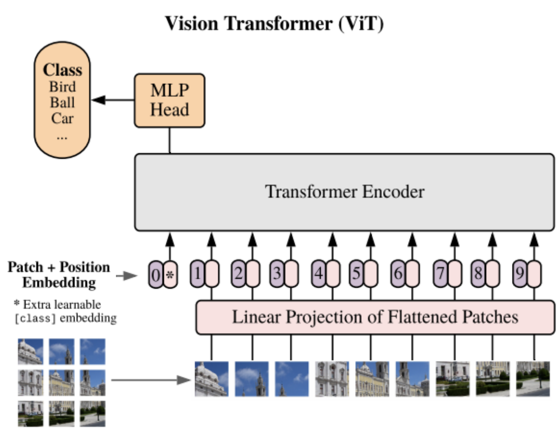
The long cycle of new deep network architectures

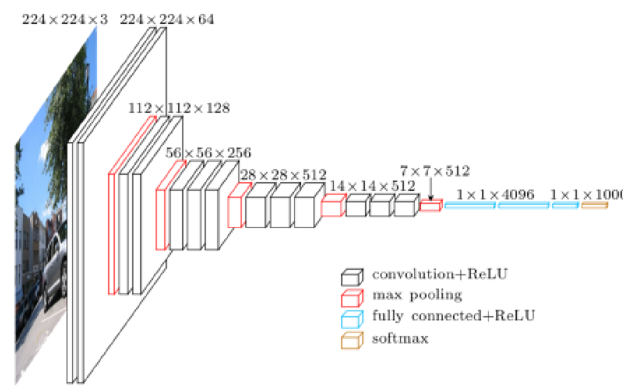








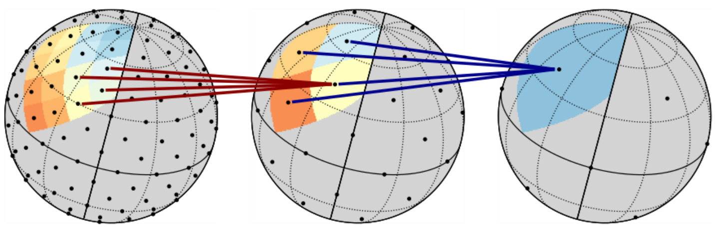




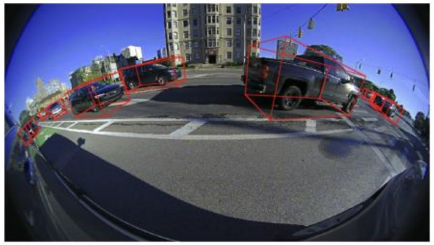
PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

But what if data is not on regular grids?

- Often, our data is not best represented in Euclidean space

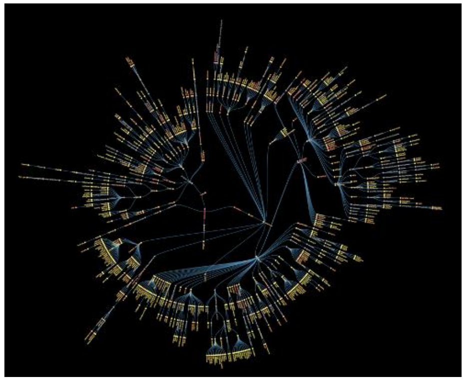


Perraudin et al. Astronomy and Computing 2019

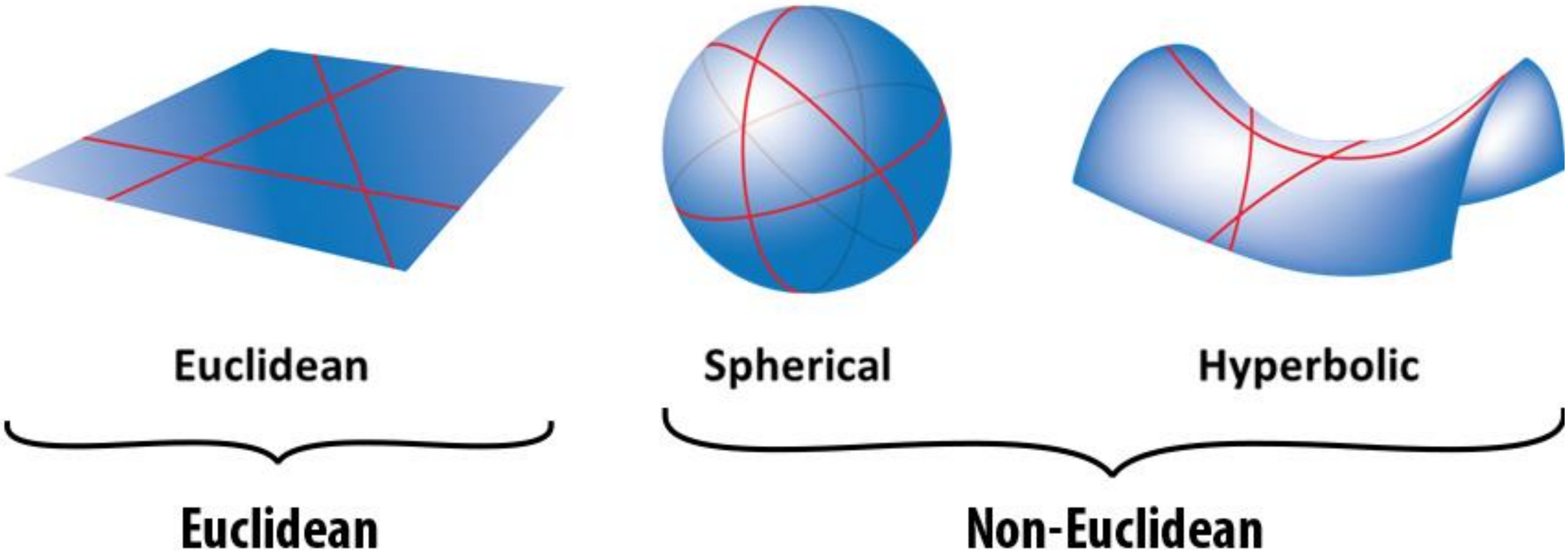


Plaut et al. CVPRw 2021

- How to do deep learning in such settings?

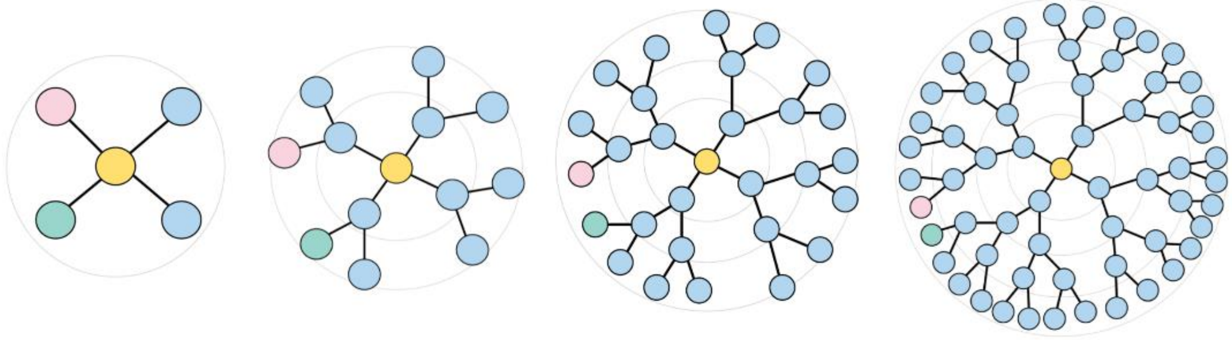


Curved and non-curved spaces



Hyperbolic geometry: the natural geometry of hierarchies

- Hierarchies and Euclidean space are a mismatch
  - exponential vs. linear growth

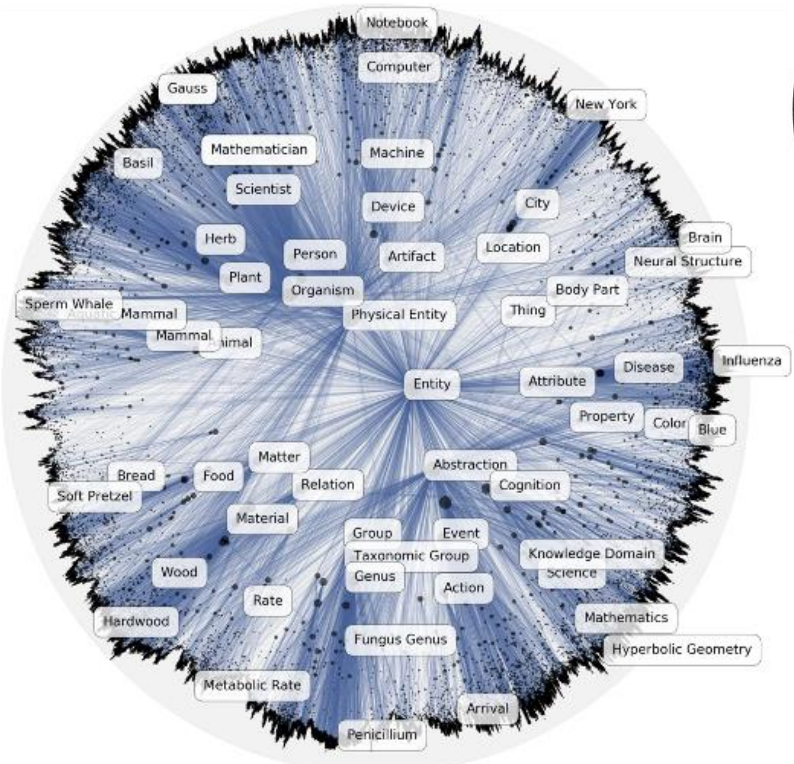


Bachmann et al. ICML 2020

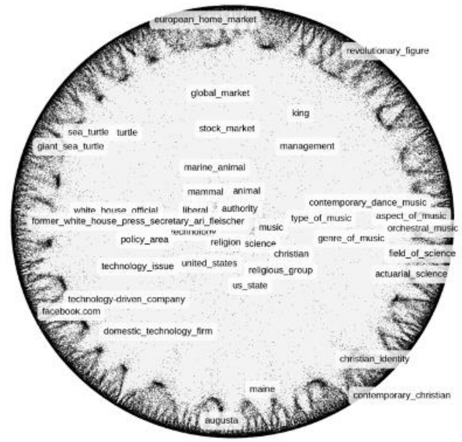
M. C. Escher (1958)

- In hyperbolic space, distances also grow exponentially
- This allows us to embed hierarchical data with minimal distortion

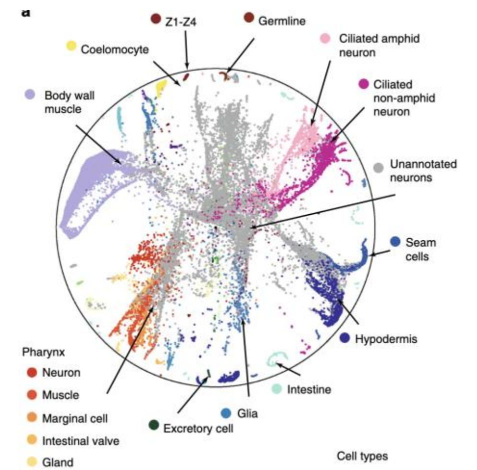
Early success:
embedding hierarchies, graphs, and text



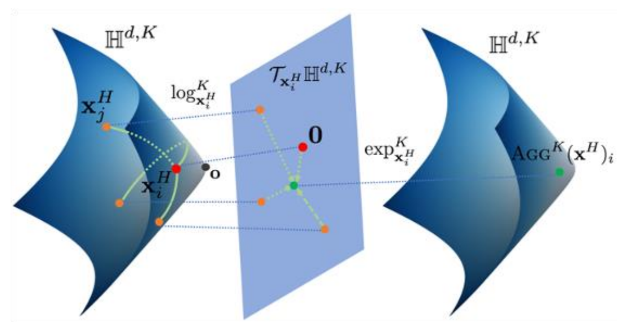
Hierarchical text embeddings
Nickel and Kiela. NeurIPS 2017



Hierarchical text embeddings
Tifrea et al. ICLR 2019



Molecular regrepresentation learning
Klimovskaia et al. Nat Comm 2020

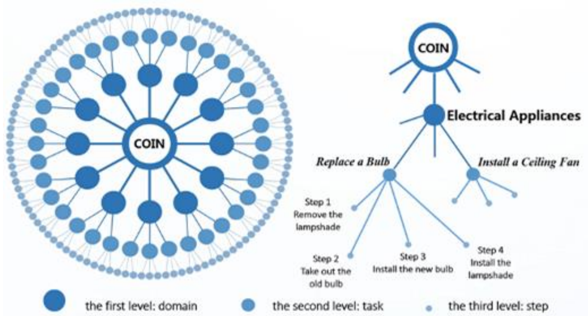


GNN on Riemannian manifold
Liu et al. NeurIPS 2019

PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

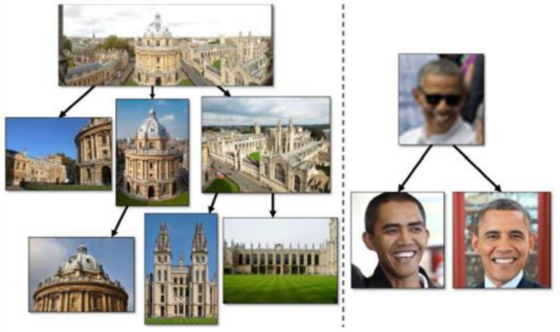
What makes hyperbolic learning interesting for vision?

Tang et al. CVPR 2019



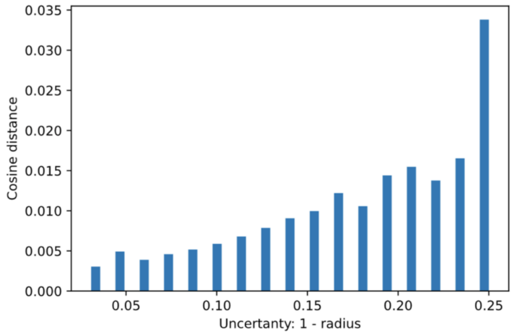
Semantics are commonly hierarchical

Khrulkov et al. CVPR 2020



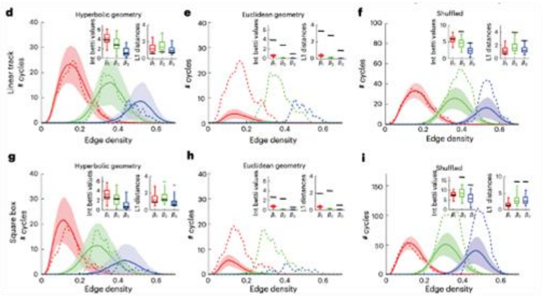
Visual collections are commonly hierarchical

Franco et al. ICLR 2023



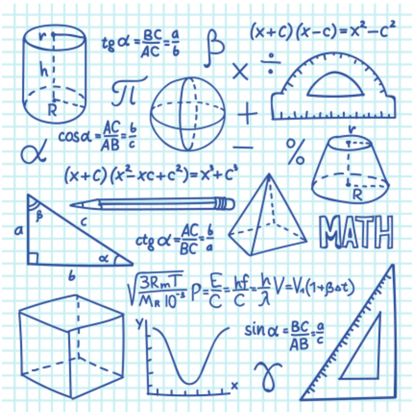
End-to-end estimate of uncertainty

Zhang et al. Nat. Neur. 2022



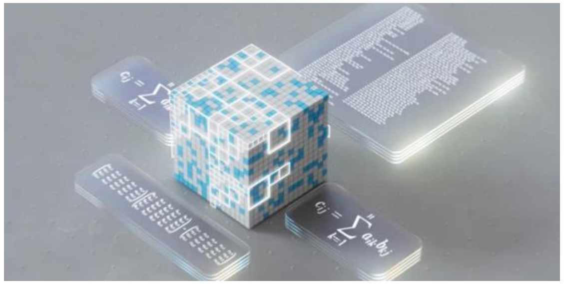
The brain is hyperbolic?

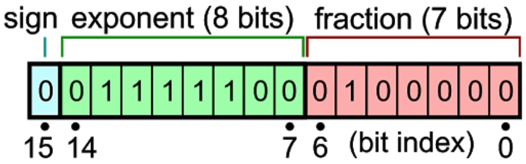
Why is hyperbolic learning not the standard (yet)?



PyTorch

TensorFlow





Our school curricula are Euclidean

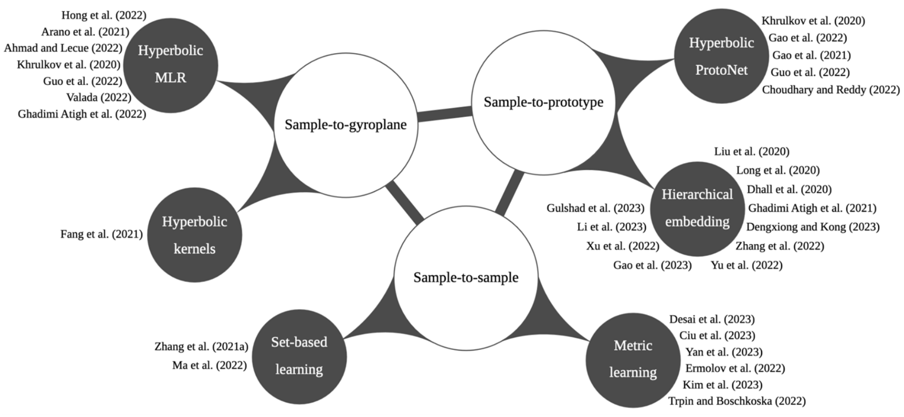
Our deep learning tools are Euclidean

Our computers work best with Euclidean data

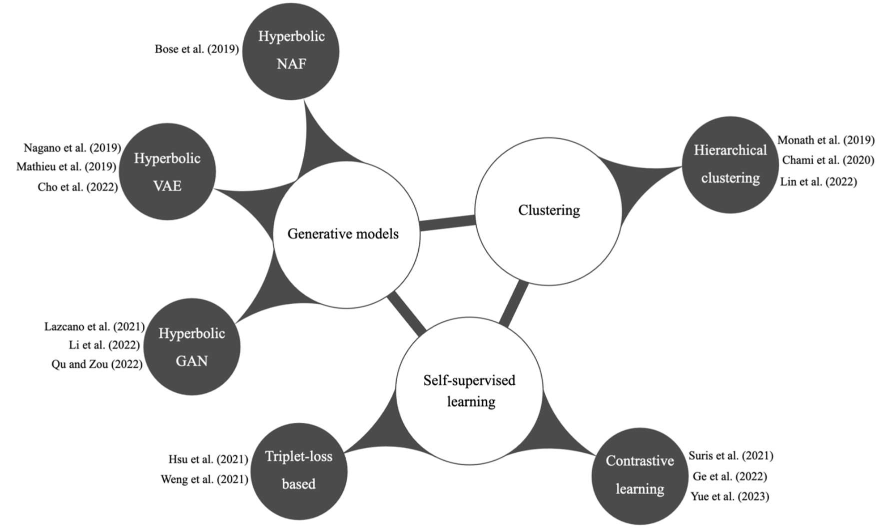
PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

How far are we with hyperbolic learning in vision?

Supervised learning



Unsupervised learning



Mettes et al. Hyperbolic Deep Learning in Computer Vision: A Survey. *arXiv:2305.06611*. 2023

Overview

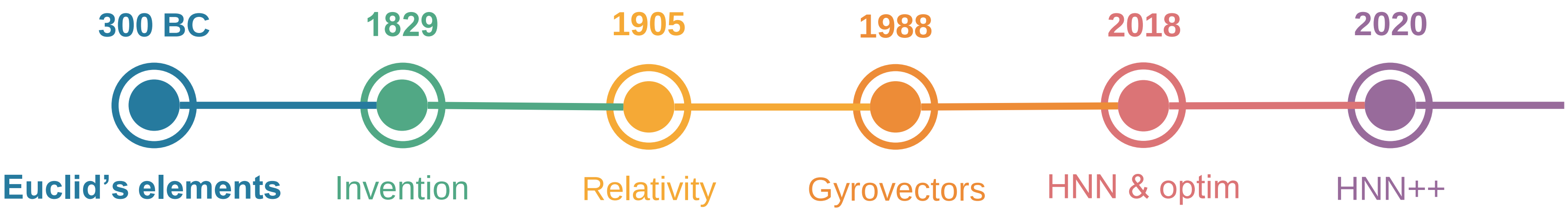
- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

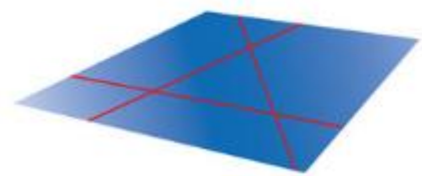
Overview

- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

What is hyperbolic geometry?

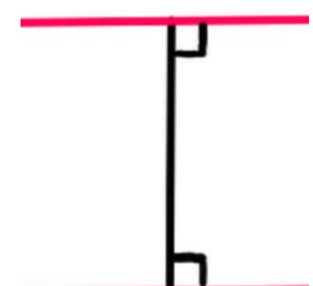
Euclid's Axioms



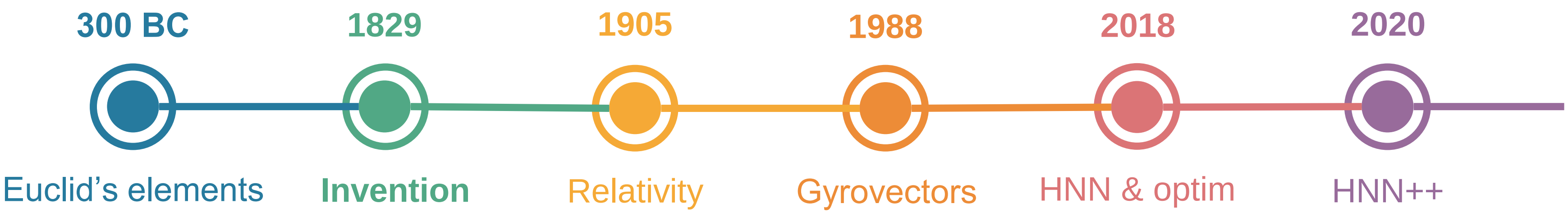


Euclidean geometry and the Euclid's axioms

1. A line can be drawn from any two points
2. Any straight line can be indefinitely extended
3. There is one circle for any center and radius
4. All right angles are congruent
5. Given a line and a point not on it, there is exactly one line going through the given point that does not intersect the given line (parallel postulate)

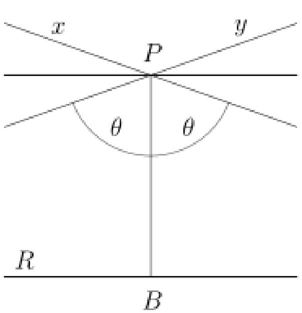


Replace the Euclid's Fifth Axiom



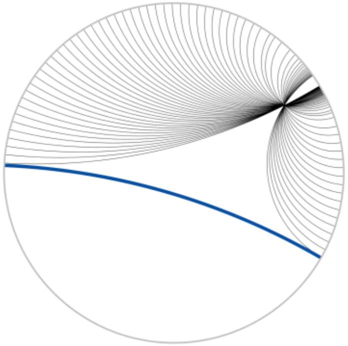
- In the early 19th century, Gauss (1824), Taurinus (1826), Lobachevsky (1829) and Bolyai (1832) discovered hyperbolic geometry by replacing the parallel postulate

5. Given a point P and a line l not passing through P, there is more than one line through P, which do not meet l
(The parallel postulate in hyperbolic geometry)



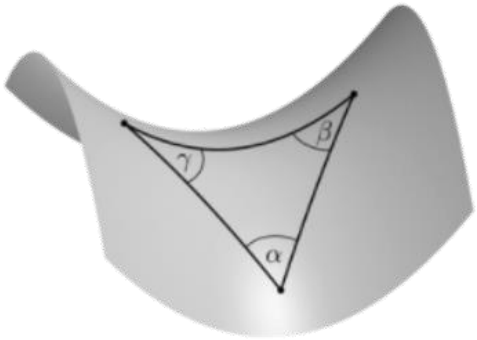
Facts about Hyperbolic Geometry

- For a line l and a point P, there are infinitely many lines through P parallel to l



- The sum of angles of a triangle is less than 180°

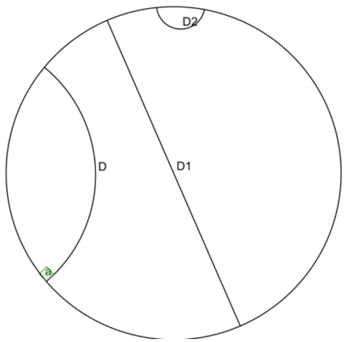
$\alpha+\beta+\gamma < 180°$



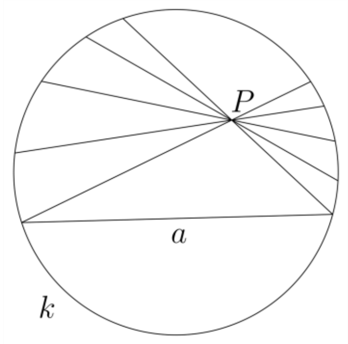
- In hyperbolic geometry, the circumference of a circle of radius r is greater than $2\pi r$

Four Models for Hyperbolic Geometry

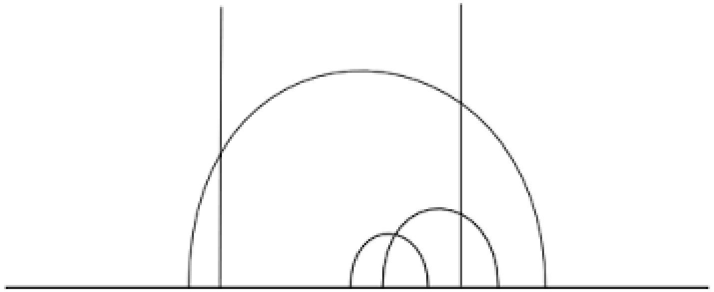
Poincaré ball model



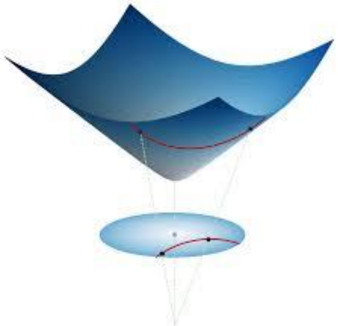
The Beltrami–Klein model



The Poincaré half-plane model



The hyperboloid model

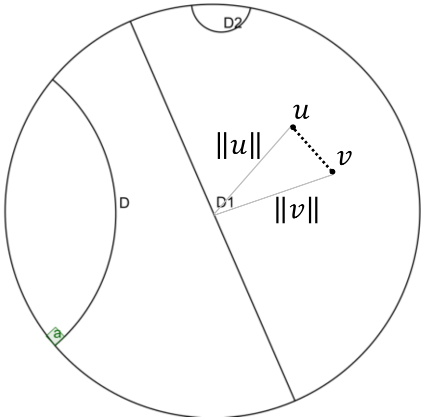


Four Models for Hyperbolic Geometry

	Pros	Cons
Poincaré ball model	Conformal	Numerically unstable
The Beltrami–Klein model	Lines are straight	Angles are not preserved
The Poincaré half-plane model	Conformal	Hard to visualize
The hyperboloid model	Numerically stable	Hard to interpret

Poincaré Ball Model

- Points lie within a disc of radius 1 $\mathbb{B}^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1\}$



- Straight lines (geodesics) are the diameters and the circular arcs perpendicular to the boundary

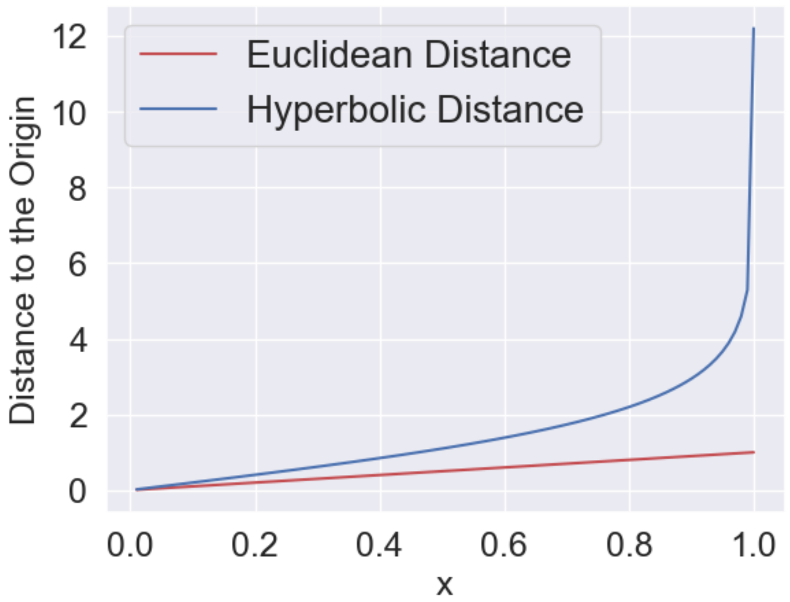
- The Poincaré distance between u and v depends *also* on the hyperbolic radii of $\|u\|$ and $\|v\|$

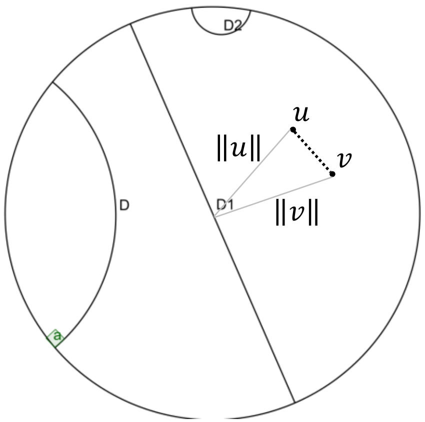
$$d_{\mathbb{B}^n}(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arcosh}\left(1 + 2\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{(1 - \|\boldsymbol{u}\|^2)(1 - \|\boldsymbol{v}\|^2)}\right)$$

Distance in the Poincaré Ball Model

- Sample distance plot (assuming v is in the origin)

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

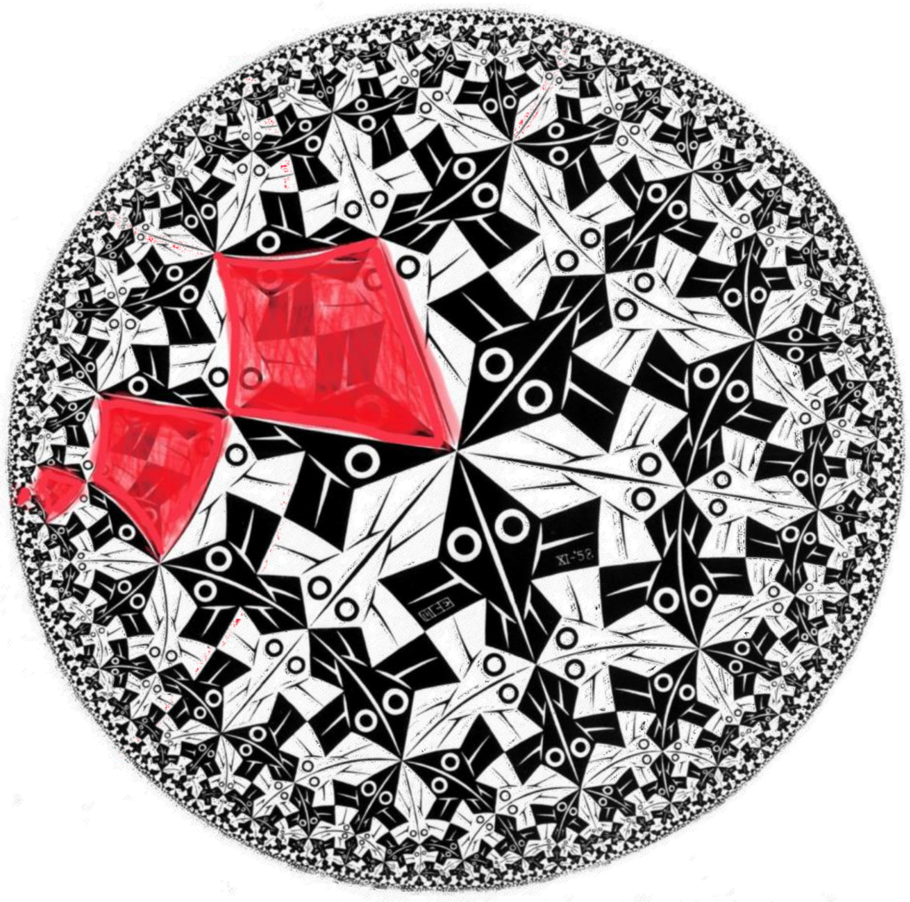




Visual Illustration of the Poincaré Ball

- Geodesic lines
  - the diameters and the circular arcs perpendicular to the boundary

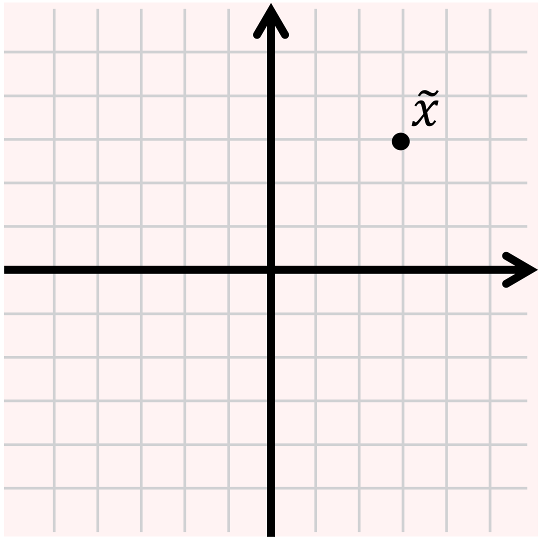
- Regions of equal area
  - Volume grows exponentially when moving towards the ball edge



M. C. Escher (1958)

In a Nutshell

- From Euclidean to the Poincaré ball



PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

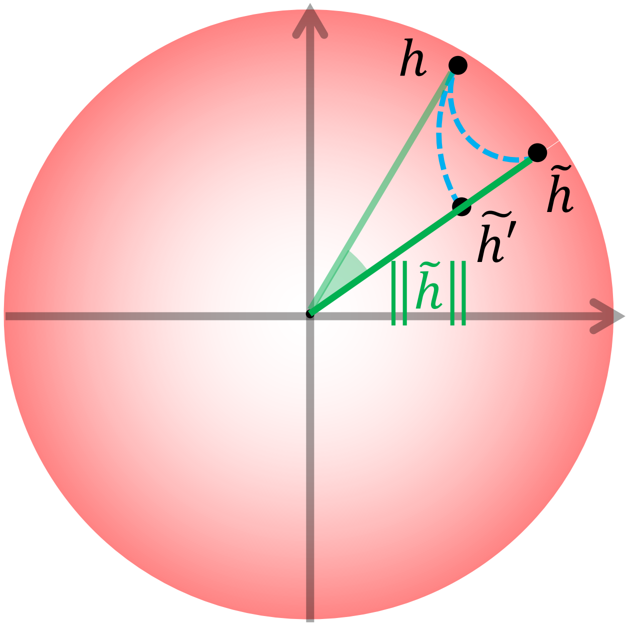
In a Nutshell

- From Euclidean to the Poincaré ball
  - One-to-one mapping:

$$\tilde{h} = Exp_0^c(\tilde{x}) = tanh(\sqrt{c}\,\|\tilde{x}\|)\frac{\tilde{x}}{\sqrt{c}\,\|\tilde{x}\|}$$ (curvature c)

  - Conformal: angles are preserved
  - Poincaré distance:

$$d_{\mathbb{D}}(h,\widetilde{h}) = \cosh^{-1}\left(1 + 2\frac{\|h-\widetilde{h}\|^2}{(1-\|h\|^2)(1-\|\widetilde{h}\|^2)}\right)$$



- Relevant to the application of the Poincaré Ball
  - Exponentially growing volume and distances allow to embed hierarchies with low distortion
  - The exponential growth of the Poincaré distance with the hyperbolic radius makes the hyperbolic radius $\|\tilde{h}\|$ a proxy for uncertainty

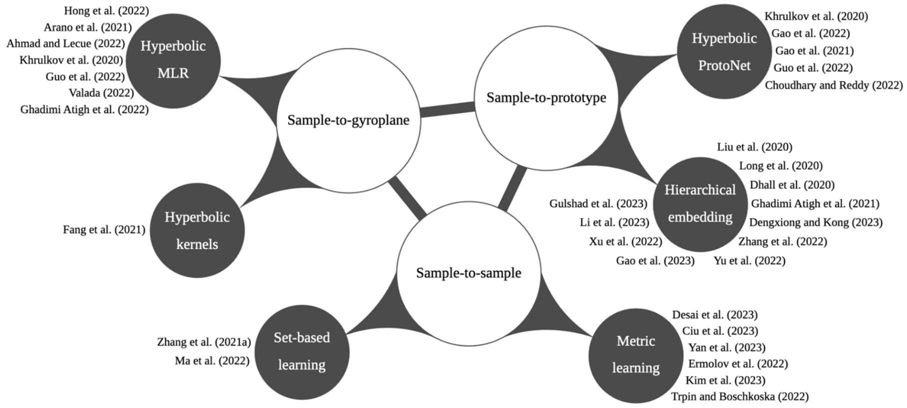
Overview

- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

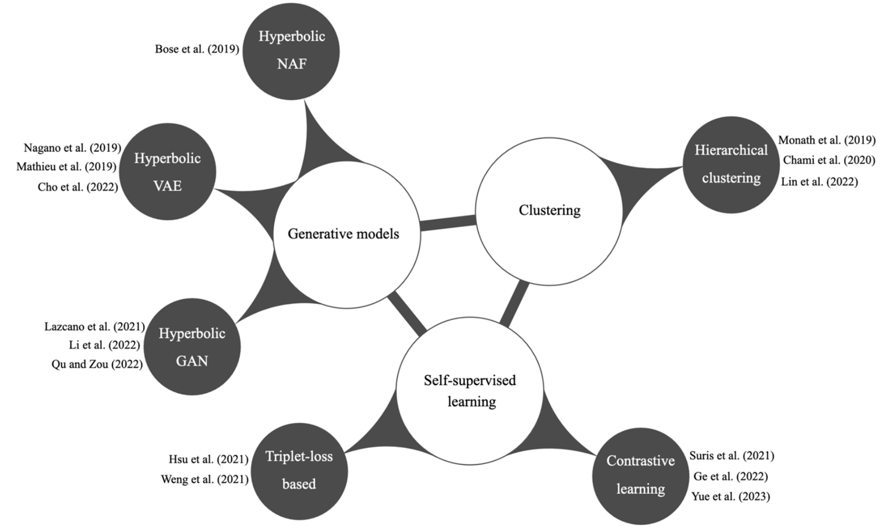
Overview of the field

How far are we with hyperbolic learning in vision?

Supervised learning

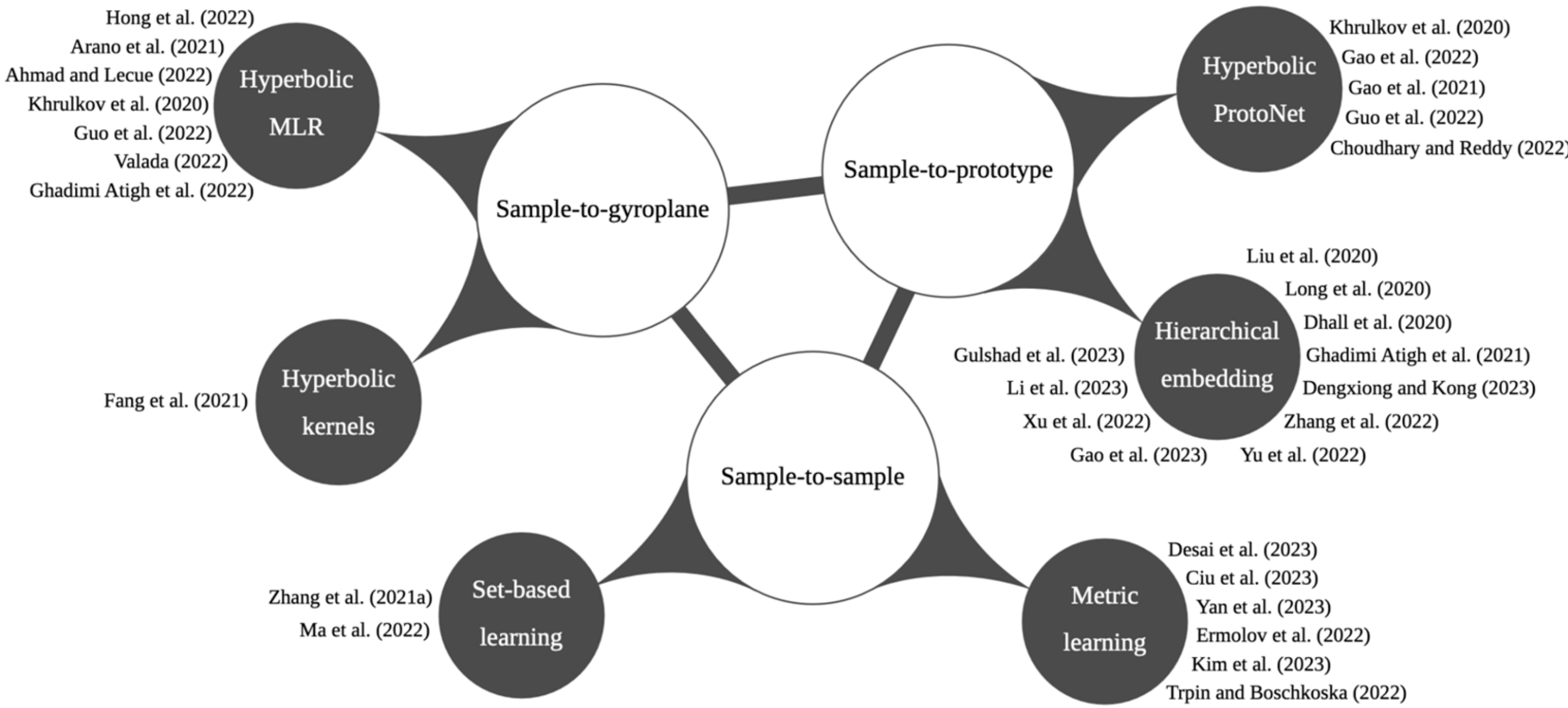


Unsupervised learning



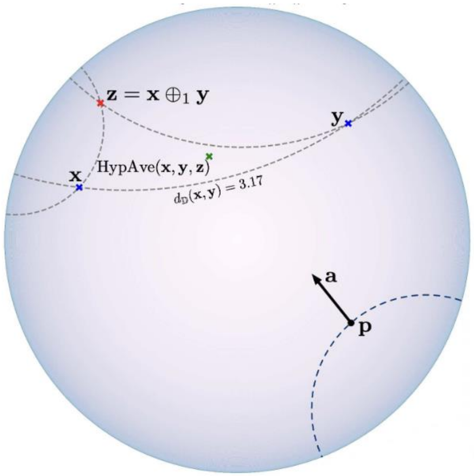
Mettes et al. Hyperbolic Deep Learning in Computer Vision: A Survey. *arXiv:2305.06611*. 2023

Supervised learning in hyperbolic space

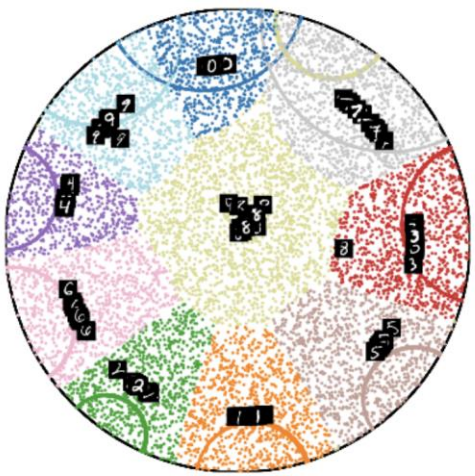


Sample-to-gyroplane learning

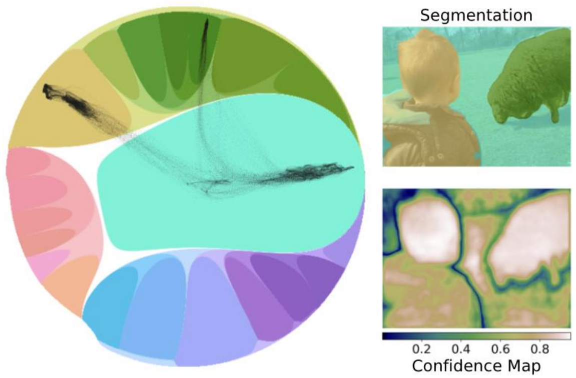
Khrulkov et al. (2020)



Guo et al. (2022)



Ghadimi Atigh et al. (2022)



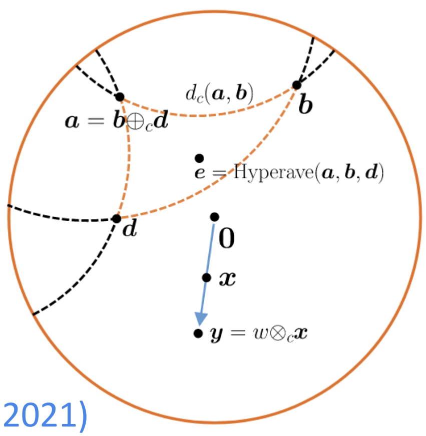
- Learn normal representations with hyperbolic logistic regression on top
- *Improves hierarchical classification, robustness, structured prediction, and more*

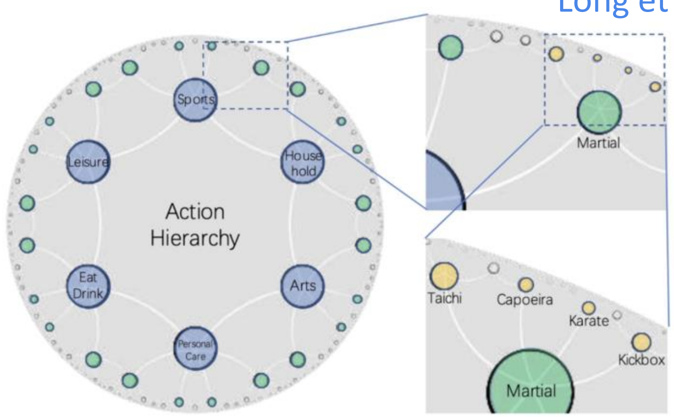
Sample-to-prototype learning

Perspective 1: prototypes from sample mean

Mostly used in few-shot learning, outperforming Euclidean conventions

Long et al. (2020)





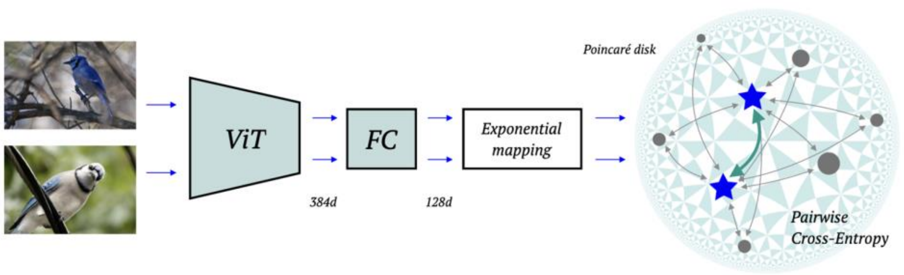
Gao et al. (2021)

Perspective 2: prototypes from hierarchical embeddings

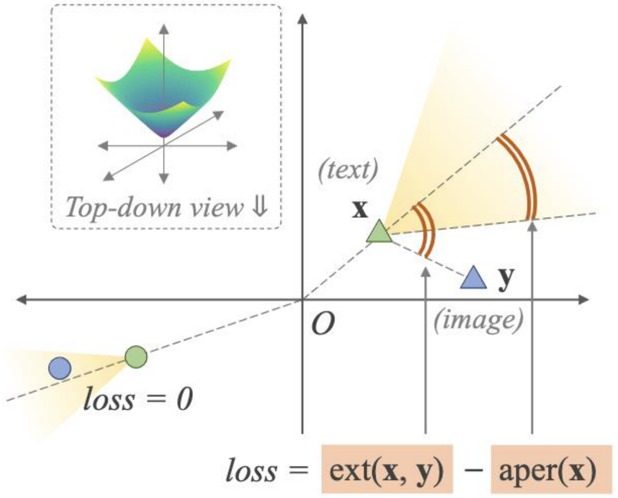
Ideal for learning with prior hierarchical knowledge and zero-shot generalization

Sample-to-sample learning

Hyperbolic Vision Transformers
Ermolov et al. (2022)



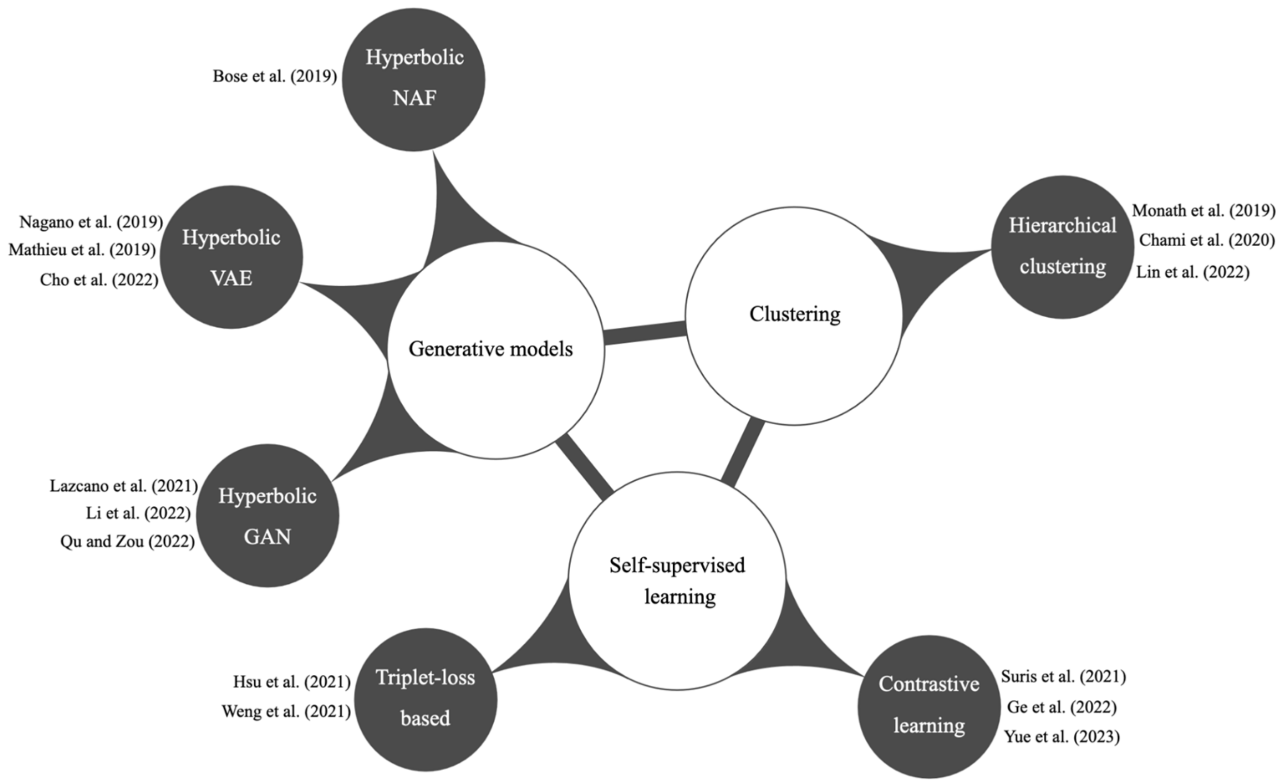
Hyperbolic CLIP
Desai et al. (2023)



- Pull and push sample pairs akin to contrastive and metric learning in Euclidean space
- *Great for fine-grained and multi-modal tasks, even at scale*

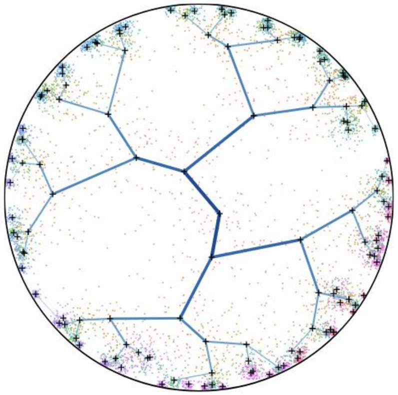
PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

Unsupervised learning in hyperbolic space

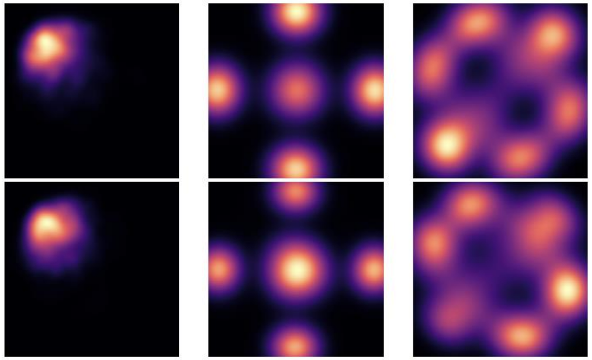


Generative approaches

Poincaré VAE
Mathieu et al. (2019)



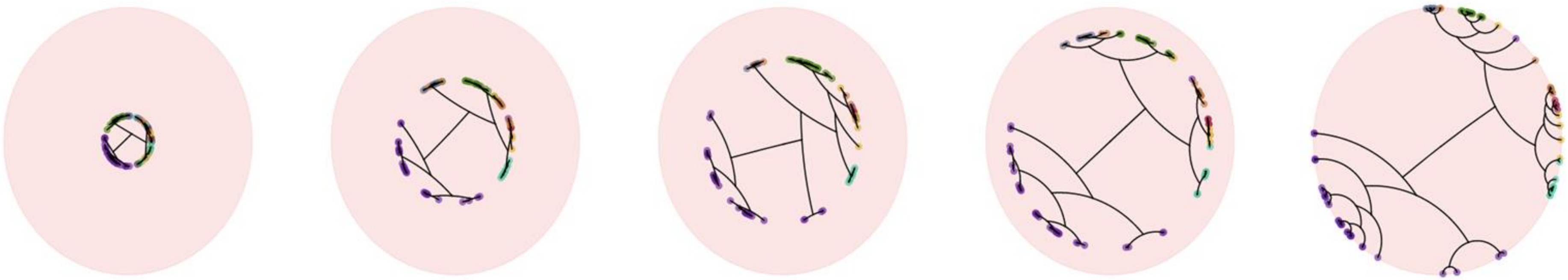
Modelling densities with
Riemannian diffusion models
Huang et al. (2022)



- For all well-known generative algorithms, there is a hyperbolic alternative
- *From hyperbolic VAE to diffusion, best results when space is latently hierarchical*

Clustering

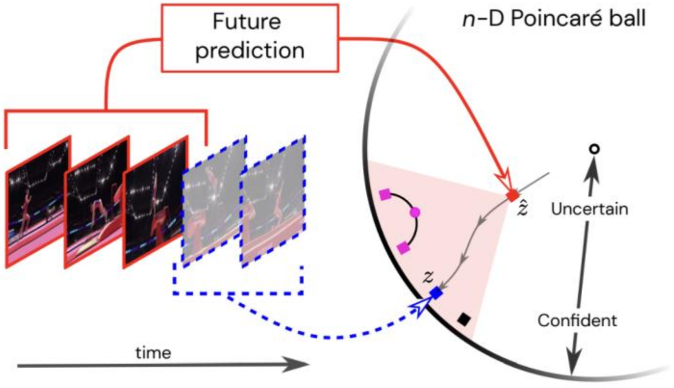
Chami et al. (2020)



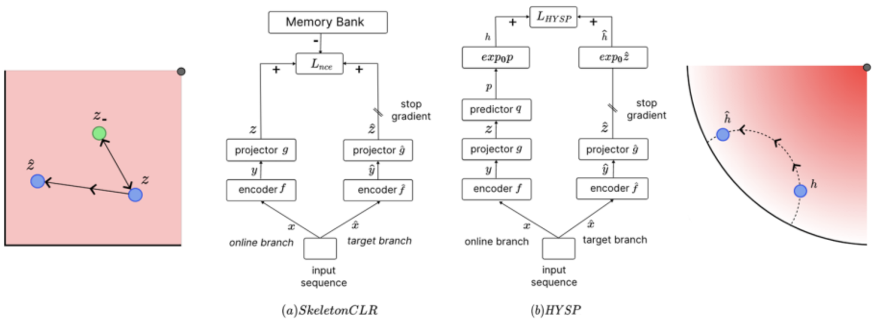
- Similarity-based hierarchical clustering is a classical machine learning task
- *Hyperbolic space is natural for this task, enabling us to cluster and discover hierarchies*

Self-supervised learning

Surís et al. (2021)



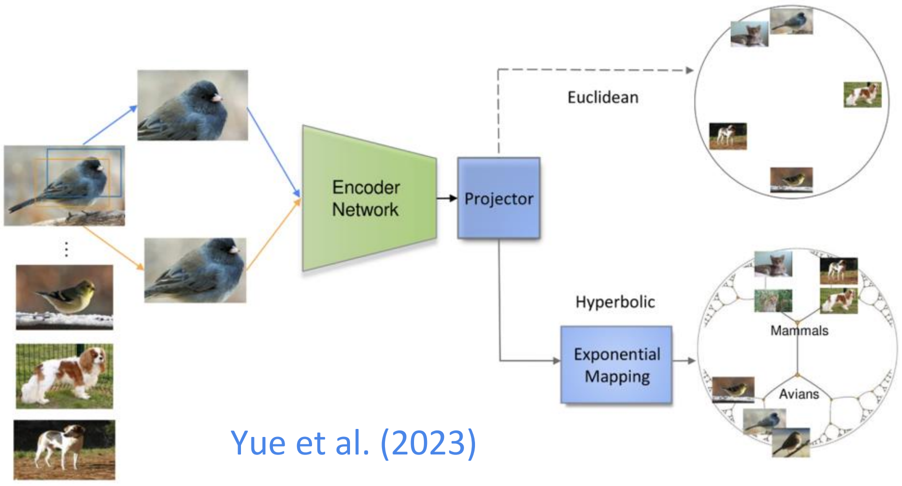
Franco et al. (2023)



- Self-supervised objective transfer naturally to hyperbolic space
- *Hierarchical representations without labels and inherent uncertainty quantification*

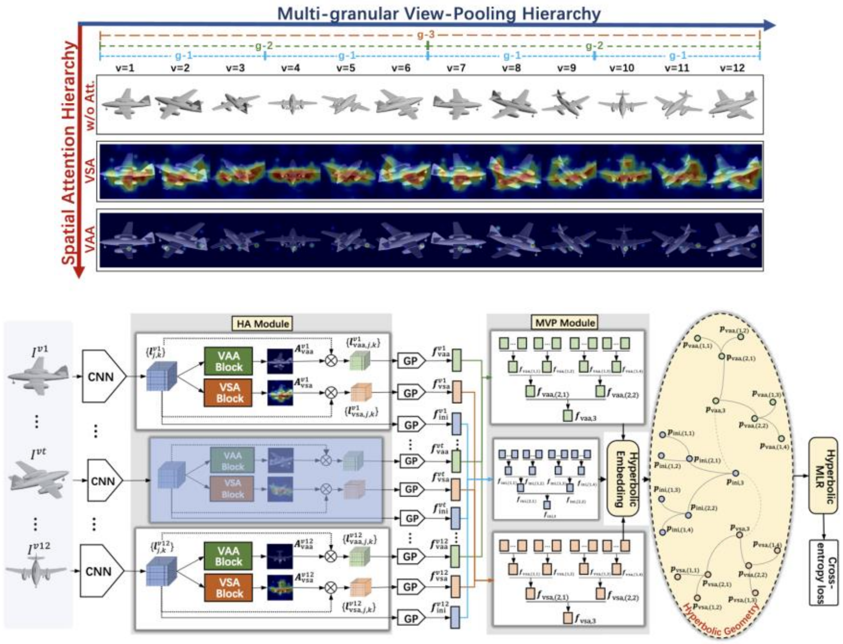
PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

Hyperbolic 3D vision



Yue et al. (2023)

Chen et al. (2020)



- Scene-object, point clouds, and LiDAR also deal with hierarchies and uncertainty
- *Hyperbolic embeddings help with supervised and self-supervised learning on 3D data*

Pros and cons

Most effective for:

- hierarchical learning
- end-to-end uncertainty estimation
- few-sample learning
- robust learning
- low-dimensional learning

Open challenges include:

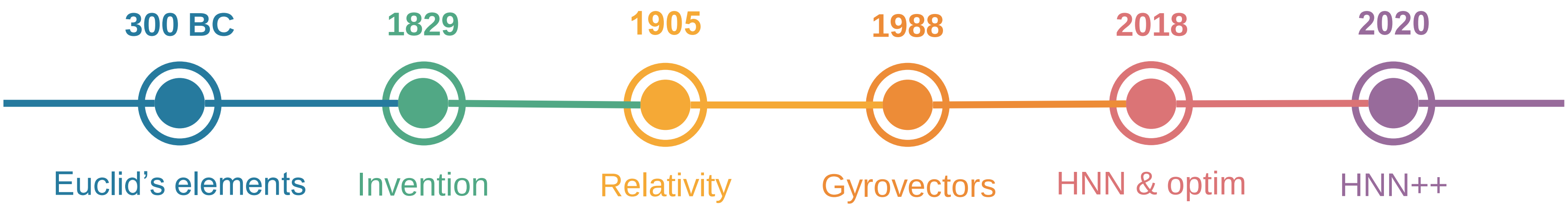
- fully-hyperbolic learning
- computational hurdles
- open-source development
- large-scale learning

Overview

- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

From Euclid to Hyperbolic Deep Learning

Some history of hyperbolic geometry



Some history of hyperbolic geometry

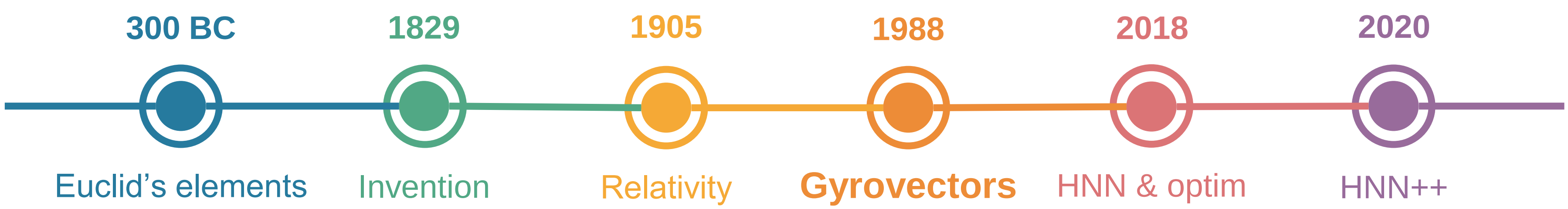
300 BC	1829	1905	1988	2018	2020
Euclid's elements	**Invention**	**Relativity**	Gyrovectors	HNN & optim	HNN++

After Euclid, geometers tried to prove the parallel postulate from the others for over 2,000 years

In the 19th century non-Euclidean and hyperbolic geometry were discovered

Einstein used hyperbolic geometry for his theory of special relativity

Some history of hyperbolic geometry



Ungar proposed gyrovector spaces for studying hyperbolic geometry and special relativity

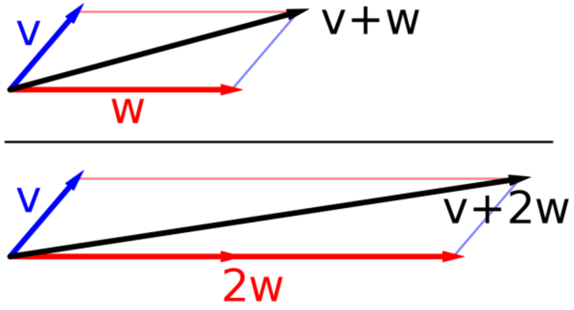
Gyrovector spaces allow for an algebraic approach to hyperbolic geometry, similar to linear algebra in Euclidean geometry

These unlock hyperbolic geometry in ML

Why not just linear algebra?

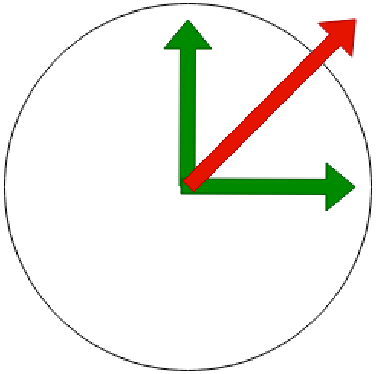
A vector space has two important operations:

- Vector addition v + w
- Scalar multiplication a * v



Doesn’t work in hyperbolic space

Need something similar, but weaker

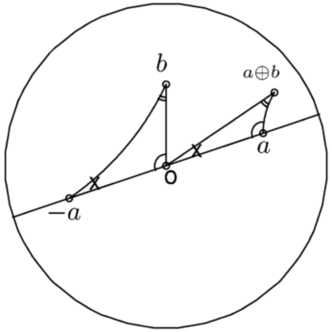


Gyrovector space

Ungar replaced vector addition and scalar multiplication by two new operations:

- Möbius addition:

$$x \oplus_c y = \frac{(1 + 2c\langle x, y\rangle + c||y||^2)x + (1 - c||x||^2)y}{1 + 2c\langle x, y\rangle + c^2||x||^2||y||^2}$$



- Möbius scalar multiplication:

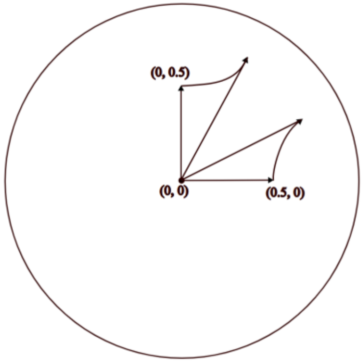
$$r \otimes_c x = \frac{1}{\sqrt{c}} \tanh\left(r \tanh^{-1}(\sqrt{c}||x||)\right) \frac{x}{||x||}$$

A nonlinear scaling with some nice vector space-like properties

Gyrovector space

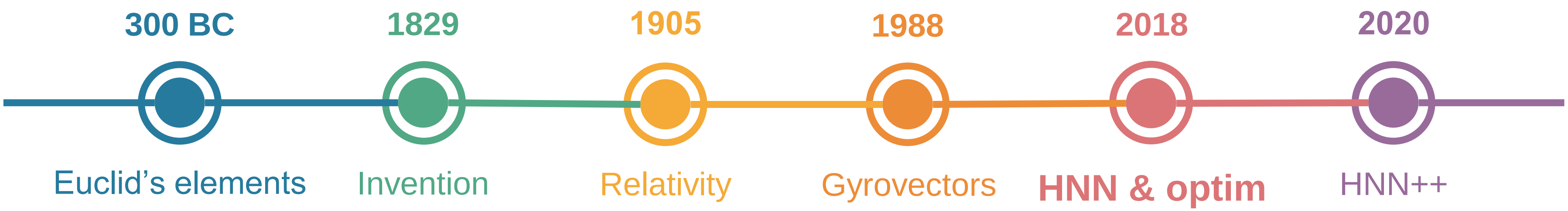
So why is it weaker? We lose commutativity, so: $x \oplus_c y \neq y \oplus_c x$

Example:



It turns out that we can still do a lot without this commutativity

Some history of hyperbolic geometry



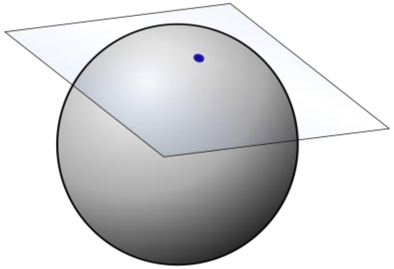
Ganea et al. derive additional important computational tools for the Poincaré ball

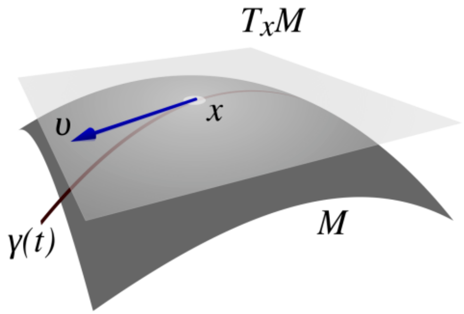
They propose a first fully hyperbolic formulation of multinomial logistic regression (MLR)

Tangent space

The tangent space at a point of a manifold can be seen as the collection of lines that are tangent to the manifold at this point

The tangent space at a point on a sphere is a plane





Useful for defining directions and velocities

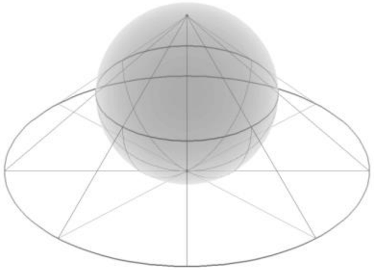
It has Euclidean geometry

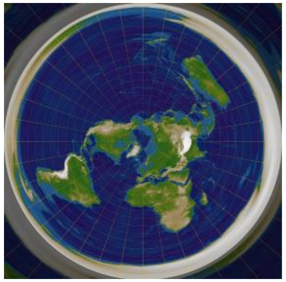
Exponential map and logarithmic map

The expmap maps from the tangent space at a point x to the manifold itself

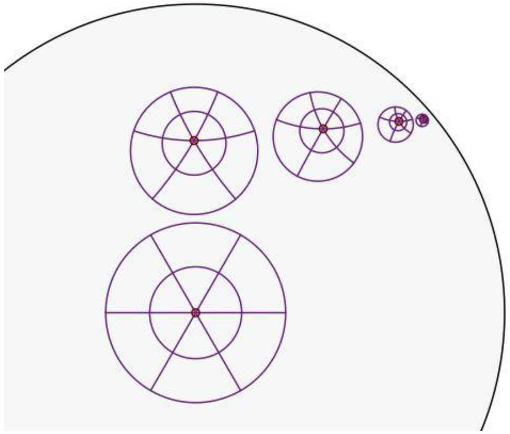
The logarithmic map at a point x does the exact opposite

An example is stereographic projection:





Ganea et al. derived these maps for the Poincaré ball



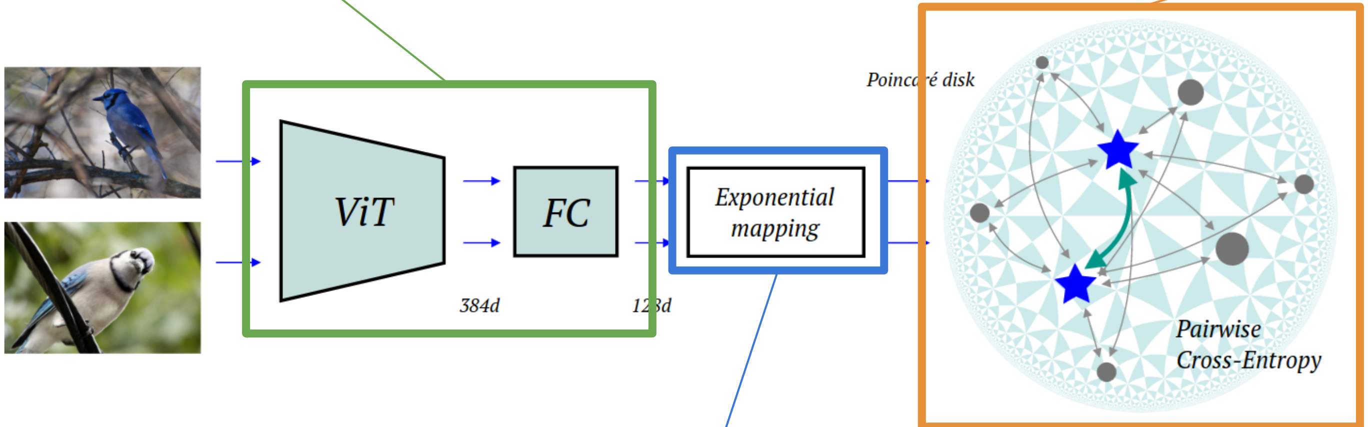
Important for many operations, such as

- Gradient descent
- Moving along geodesics (straight lines)

Hyperbolic image embeddings

1. Compute image embeddings with a backbone

3. Perform downstream task



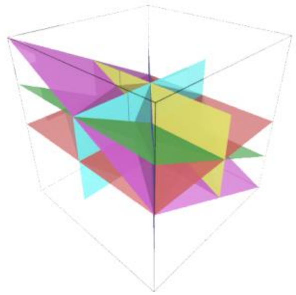
Ermolov et al. (2022)

2. Map the embeddings to the Poincaré ball

Geometric Interpretation of MLR

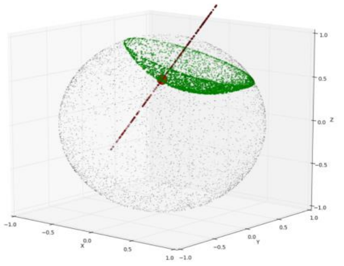
Euclidean logits in MLR: $$Ax - b = (\langle a_k, x \rangle - b_k)_k$$

Signed distance to a hyperplane: $$\frac{\langle a, x \rangle - b}{||a||}$$



Logits in MLR are equivalent to scaled signed distances of the input to hyperplanes

Ganea et al. 2018



We can do the same in other geometries

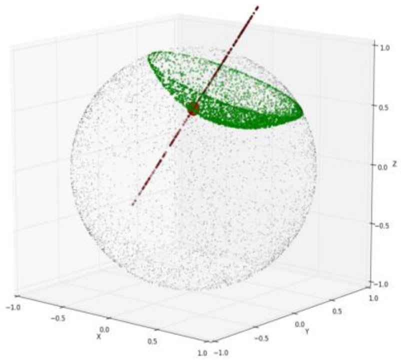
But we need a way to compute these distances

Hyperbolic MLR formulation

For each of *n* classes we have parameters:

$$p_k \in \mathbb{D}^n_c,\ a_k \in T_{p_k}\mathbb{D}^n_c \setminus \{\mathbf{0}\}:$$

Ganea et al. 2018



With which we can compute the output probabilities as:

$$p(y=k|x) \propto \exp\left(\underbrace{\frac{\lambda^c_{p_k}\|a_k\|}{\sqrt{c}}\sinh^{-1}\left(\frac{2\sqrt{c}\langle -p_k \oplus_c x, a_k\rangle}{(1-c\|-p_k \oplus_c x\|^2)\|a_k\|}\right)}_{\text{Scaled signed distances to hyperplanes}}\right)$$

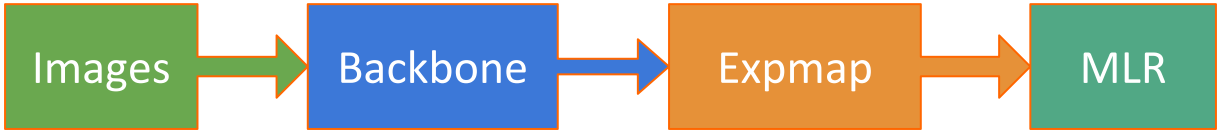
Scaled signed distances to hyperplanes

Application of MLR

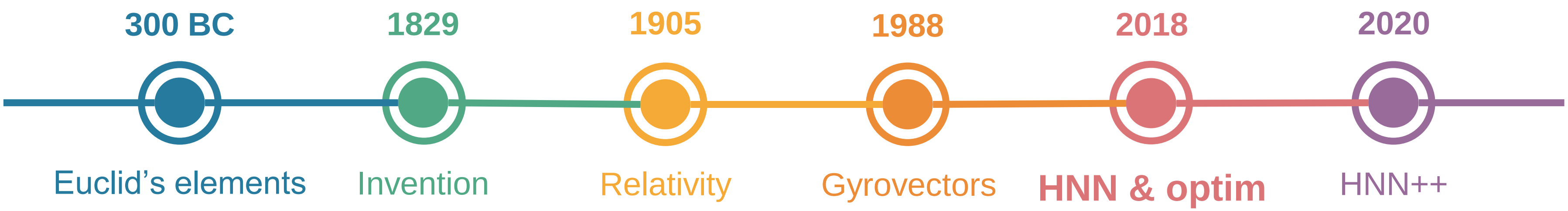
Obtain hyperbolic image embeddings (upon Euclidean backbone + expmap)

Then apply hyperbolic MLR to these hyperbolic embeddings

An example of this approach can be found in Khrulkov et al. (2020)



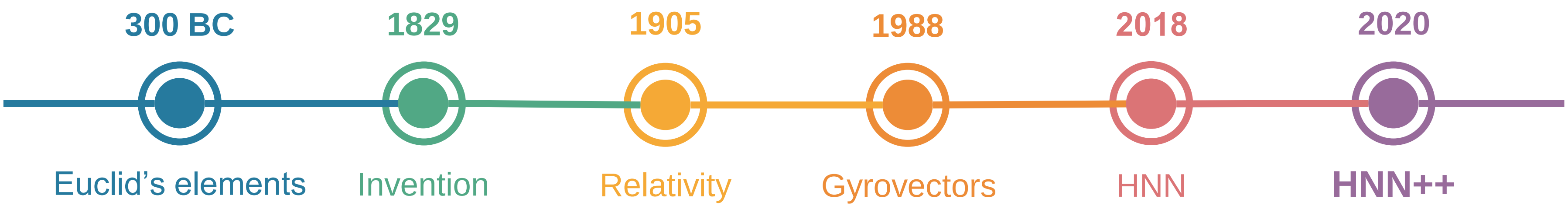
Some history of hyperbolic geometry



Bécigneul et al. also introduce Riemannian versions of several optimizers, based on the Riemannian SGD by Bonnabel (2013)

This connects the tools from HNN with the Riemannian optimizers and it yields hyperbolic optimizers

Some history of hyperbolic geometry



Shimizu et al. (2020) propose fully hyperbolic layers

They propose a new way to concatenate and split hyperbolic gyrovectors, which enables hyperbolic ConvNets

Van Spengler et al. (2023) build the first fully-hyperbolic ResNet, which includes hyperbolic convolutions, skip connections and batch-norm

HypLL

New Python hyperbolic learning library can be installed via pip

pip install hypll

Or by installing from our GitHub repository:

https://github.com/maxvanspengler/hyperbolic_learning_library

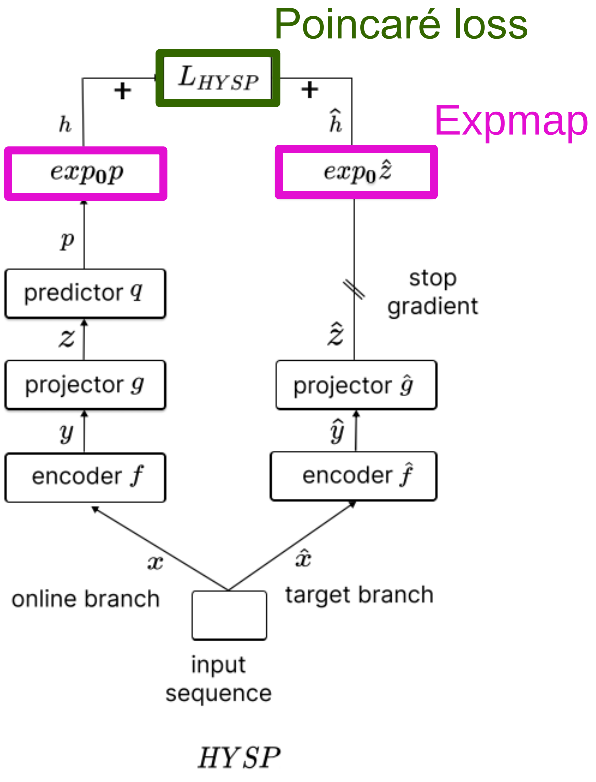
Overview

- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

Leading Interpretations of the Hyperbolic Radius

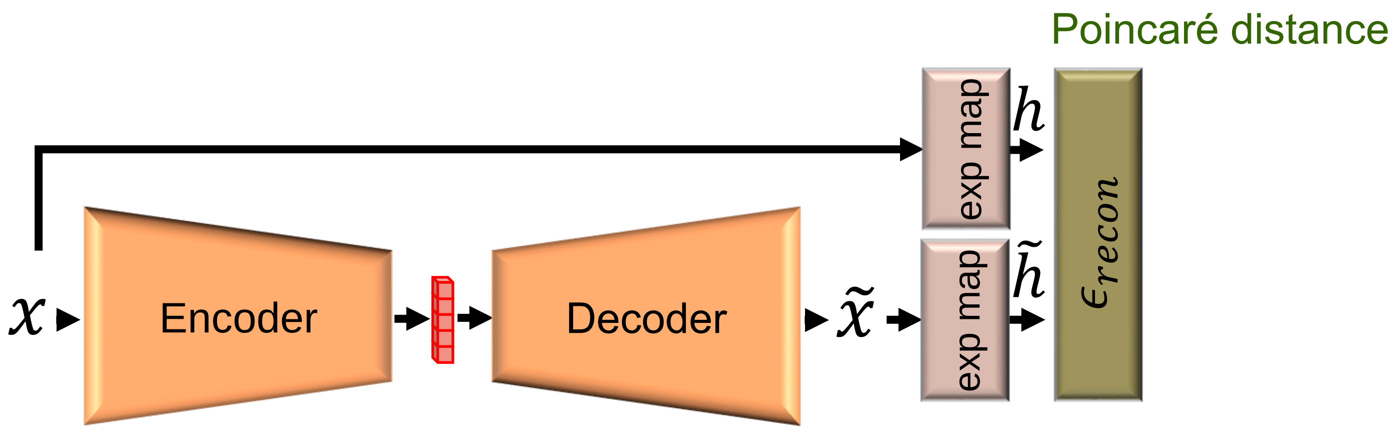
Hyperbolic Radius in Computer Vision

- Radius → certainty
  - Weight samples in SSL via the hyperbolic radius
    - **Franco et al. ICLR’23**



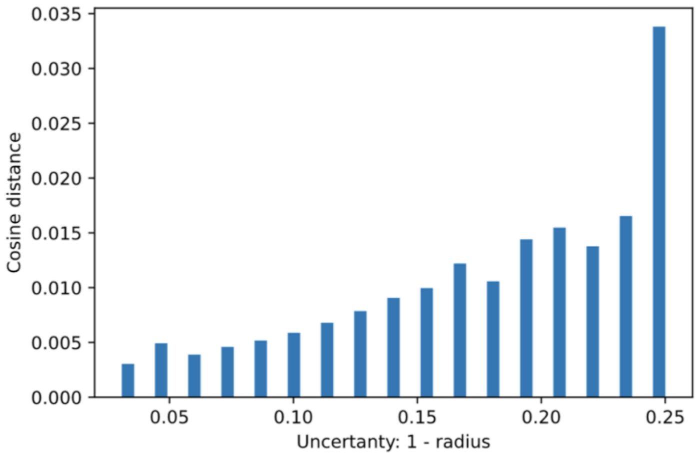
Hyperbolic Radius in Computer Vision

- Radius → certainty
  - Weight samples in SSL via the hyperbolic radius
    - **Franco et al. ICLR'23**
  - Detect anomalies and abstain if uncertain
    - **Flaborea et al. CVPR'23 Wks**

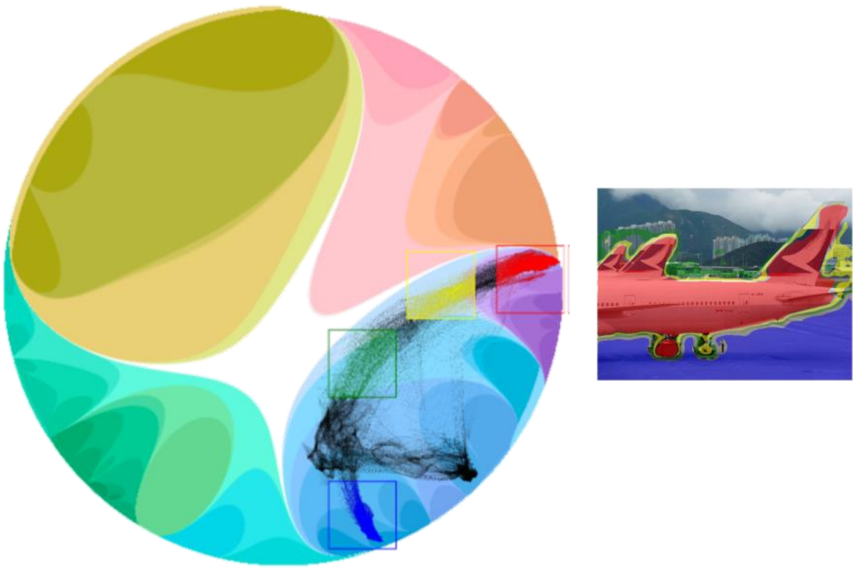


Hyperbolic Radius in Computer Vision

- Radius → certainty
  - Weight samples in SSL via the hyperbolic radius
    - **Franco et al. ICLR’23**
  - Detect anomalies and abstain if uncertain
    - **Flaborea et al. CVPR’23 Wks**
  - Larger error when more uncertain

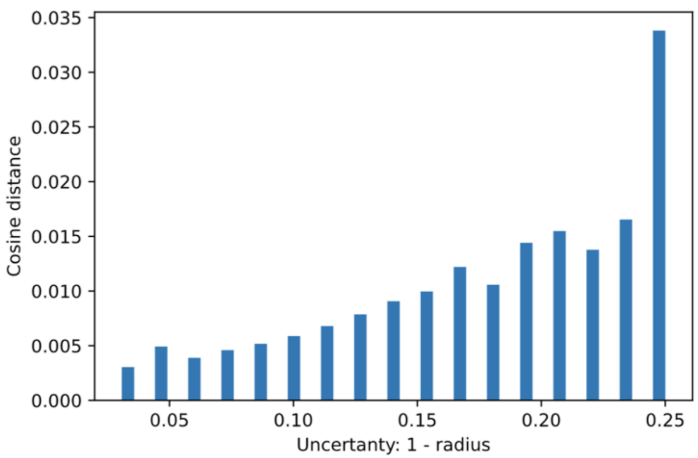


- Radius → hierarchy (parent-to-child)
  - In hierarchical classification, parent classes have lower radii
    - Ghadimi Atigh et al. CVPR’22

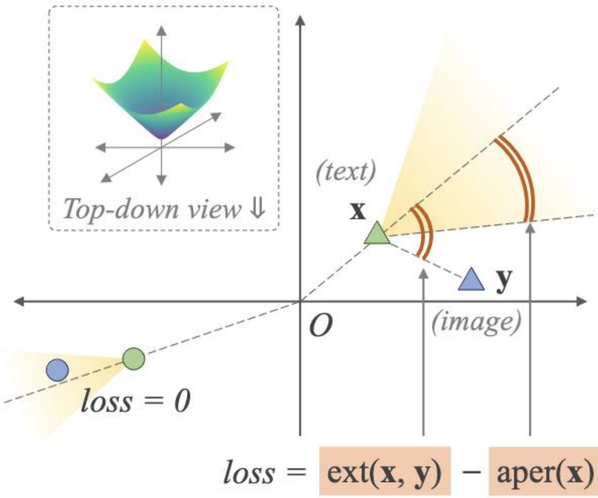


Hyperbolic Radius in Computer Vision

- Radius → certainty
  - Weight samples in SSL via the hyperbolic radius
    - **Franco et al. ICLR'23**
  - Detect anomalies and abstain if uncertain
    - **Flaborea et al. CVPR'23 Wks**
  - Larger error when more uncertain



- Radius → hierarchy (parent-to-child)
  - In hierarchical classification, parent classes have lower radii
    - Ghadimi Atigh et al. CVPR'22
  - Enforce image-text hierarchies by an entailment loss
    - Desai et al. PMLR'23



Overview

- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

Hyperbolic Uncertainty for Anomaly Detection

- Prenkaj, Aragona, Flaborea, Galasso, Gravina, Podo, Reda, Velardi (2023). A self-supervised algorithm to detect signs of social isolation in the elderly from daily activity sequences. In *Artificial Intelligence in Medicine*
- Flaborea, Prenkaj, Munjal, Sterpa, Aragona, Podo, Galasso (2023). Are we certain it's anomalous? In Proc. *CVPR wks*

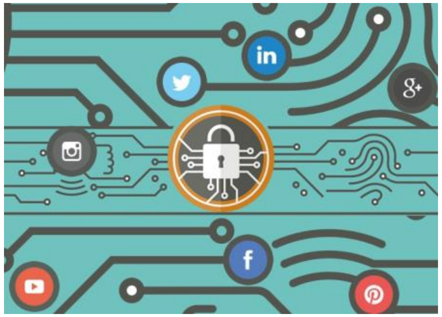
PINLAB Perception and Intelligence

SAPIENZA UNIVERSITÀ DI ROMA

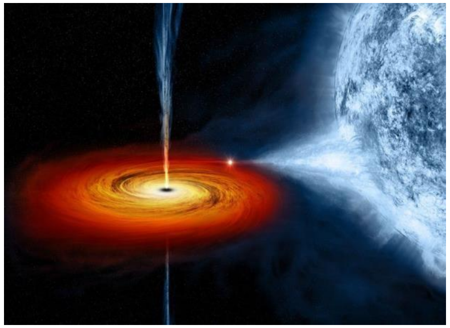
Anomaly Detection Applications

Cybersecurity:
attacks, malware, malicious apps/URLs, biometric spoofing

Social Network and Web Security:
false/malicious accounts, false/hate/toxic information

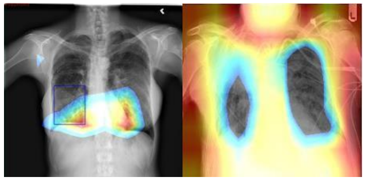


Astronomy:
Anomalous events

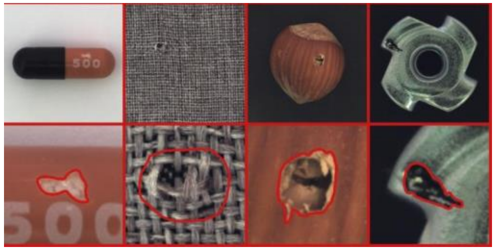


Finance:
credit card/insurance frauds, market manipulation, money laundering, etc.

Healthcare:
lesions, tumours, events in IoT/ICU monitoring, etc.



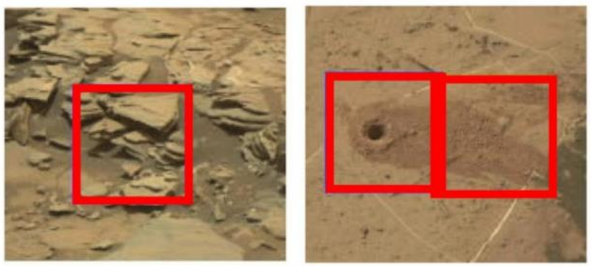
Industrial Inspection:
Defects, micro-cracks



Slide credit: Guansong Pang, Longbing Cao, Charu Aggarwal

Anomaly Detection Applications

Rover-Based Space Exploration:
unknown textures



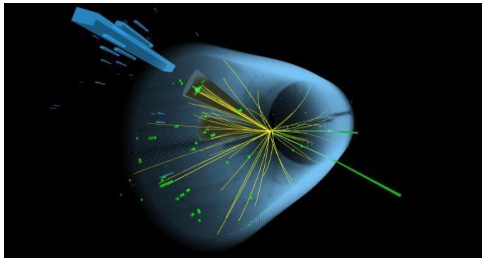
Bedrock
(Sol 1032)

Drill hole and tailings
(Sol 1496)

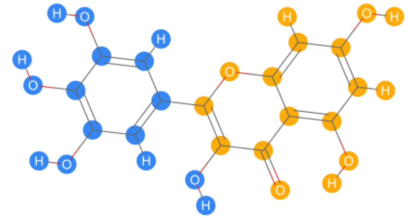
Video surveillance:
anomalous behavior, accidents, fights..

Normal Frame **Anomalous Frame**

High-Energy Physics:
Higgs boson particles



Material Science:
exceptional molecule graphs



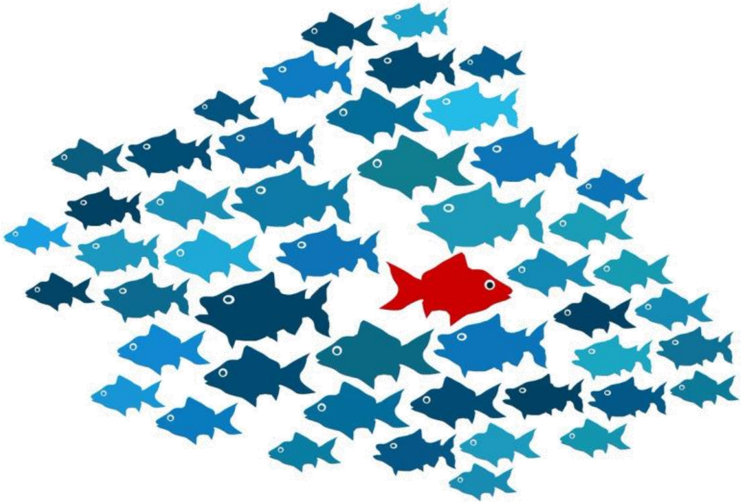
Drug Discovery:
rare active substances

Slide credit: Guansong Pang, Longbing Cao, Charu Aggarwal

Anomaly Detection
AIM'23, CVPR-wks'23, Pattern Recognition'23 (u. rev.)

- Target data
  - Financial series (NAB)
  - IT systems (YAHOO)
  - Mars aerospace measurements (NASA)
  - Medical data on elderly from sensor data (CASA)
  - Industrial water treatment (SWaT)
  - Anomalous human behavior (UBnormal)

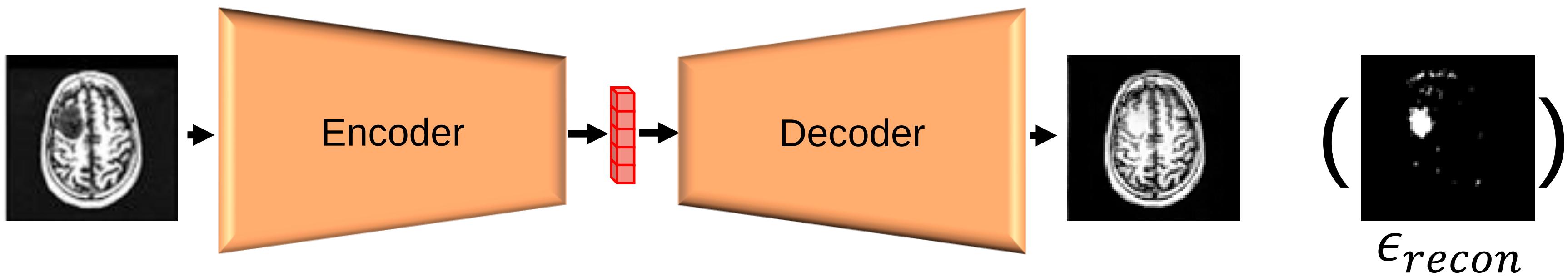
- **Real-world problem formulation**
  - Train on normalcy just (*aka* **OCC**)
  - Novel classes of *test* anomaly (**open set**)



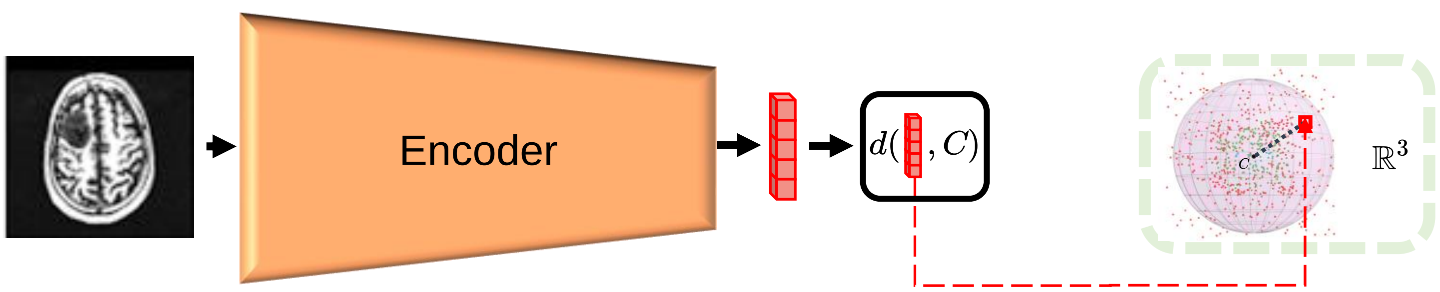
https://static.tildacdn.com/tild3131-3237-4364-b662-663731666262/anomaly_detection.png

Anomaly Detection

- Learn to reconstruct normalcy, compare input Vs. reconstructed

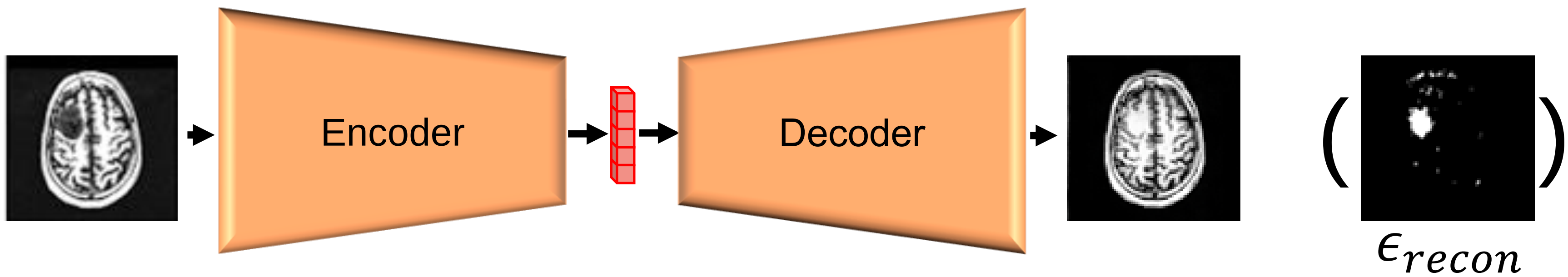


- Constrain normalcy into a hypersphere, measure dist. from center



Anomaly Detection

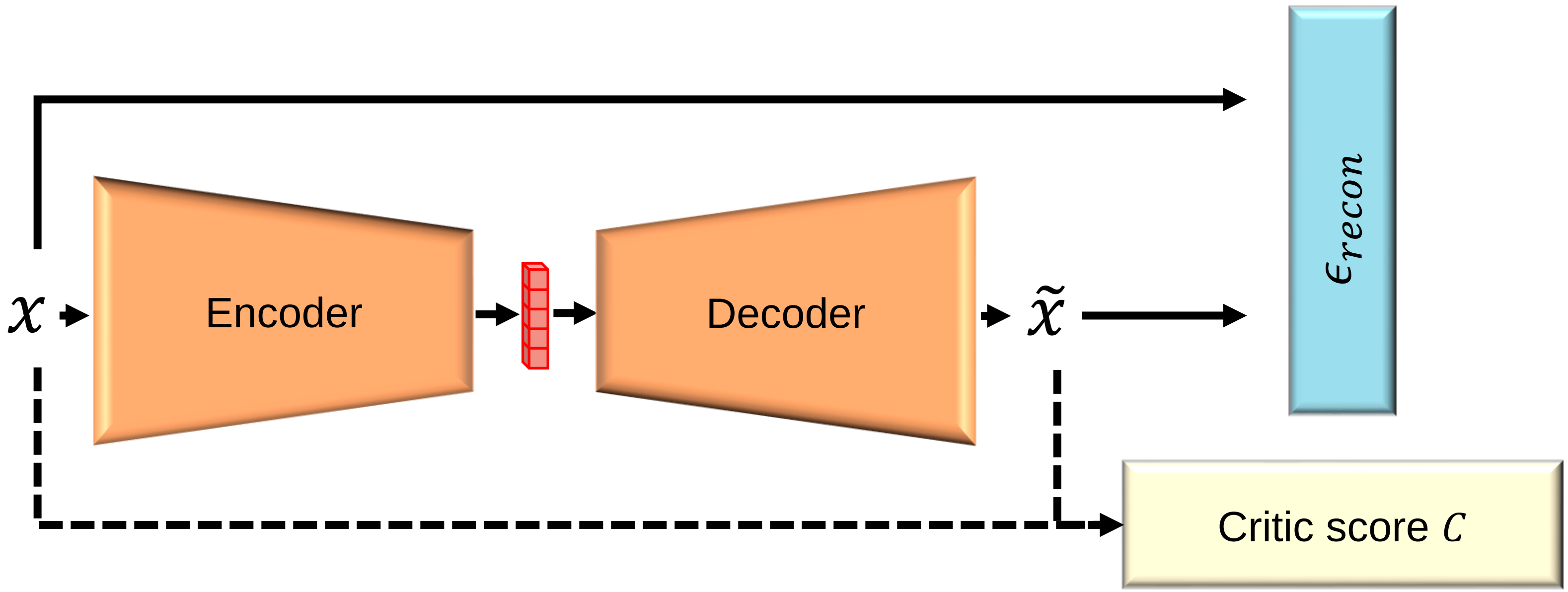
- Learn to reconstruct normalcy, compare input Vs. reconstructed



- Constrain normalcy into a hypersphere, measure dist. from center

Anomaly Detection by Reconstruction Error

- Build on the current best (TADGAN)
  - $\epsilon_{recon} = \mathrm{Dist}(x, \tilde{x})$
  - Critic score C for adversarial training
  - anomaly $= \epsilon_{recon} * C$

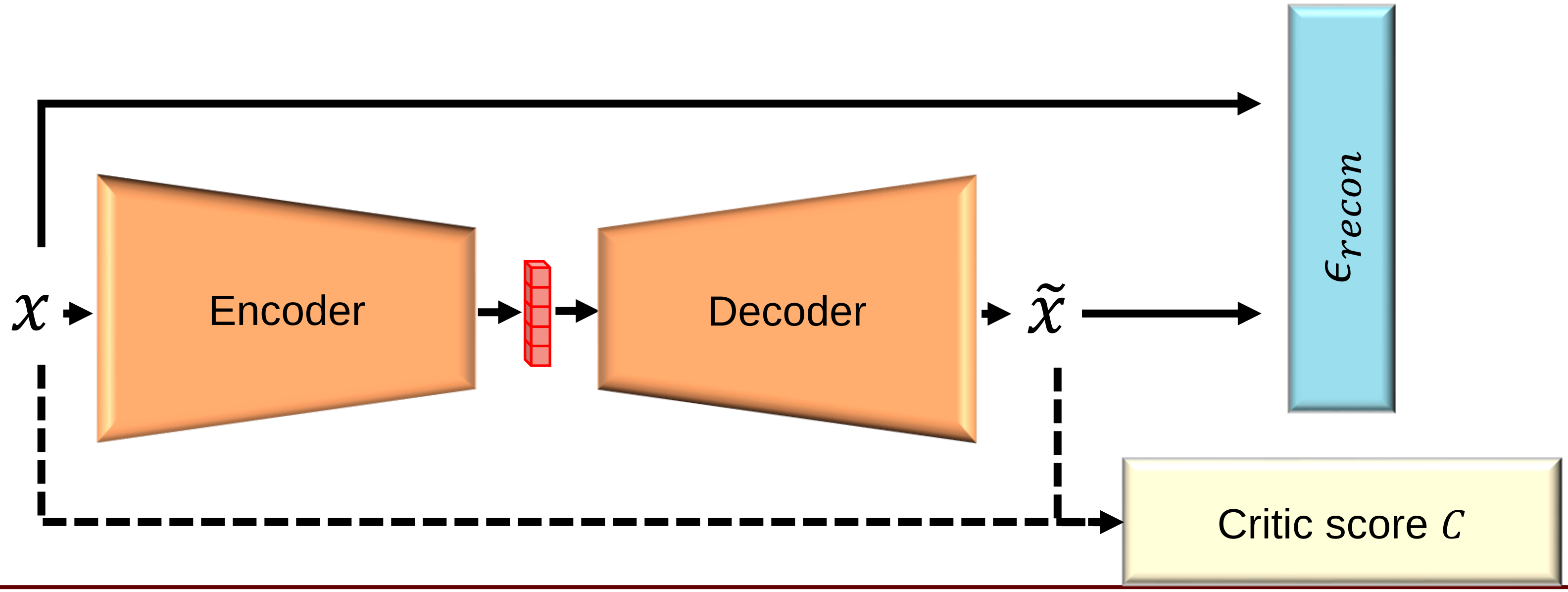


Geiger et al. (2020). "TADGAN: Time series anomaly detection using generative adversarial networks". In IEEE Int. Conf. on Big Data'20

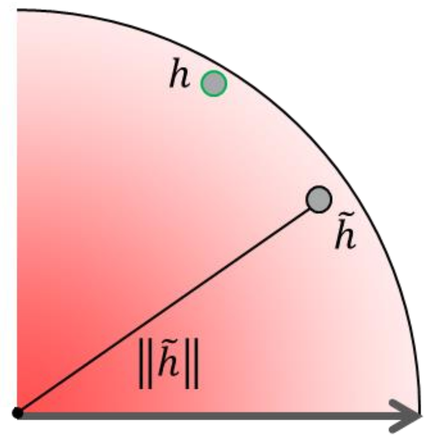
Proposed: HypAD
Hyperbolic Uncertainty for Anomaly Detection

- Proposed: *trust* reconstruction errors if *certain* about the sample

→ estimate the uncertainty of the samples



Proposed: HypAD
Hyperbolic Uncertainty for Anomaly Detection

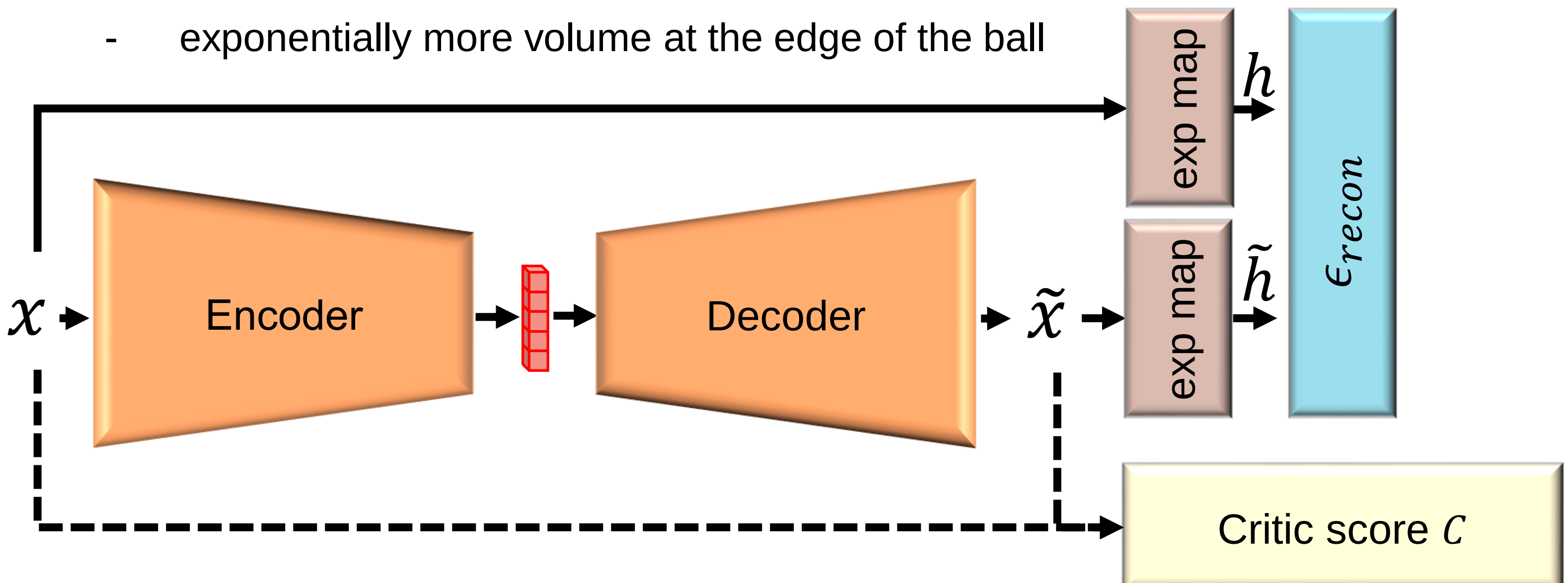


- Minimize reconstruction errors in Hyperbolic space
  - Map x and the reconstructed $\tilde{x}$ into the Poincaré ball via exp map [Ganea et al. NeurIPS'18]

$$\tilde{h} = Exp_0^c(\tilde{x}) = tanh(\sqrt{c}\,\|\tilde{x}\|)\frac{\tilde{x}}{\sqrt{c}\,\|\tilde{x}\|}$$

where c is the curvature

- exponentially more volume at the edge of the ball



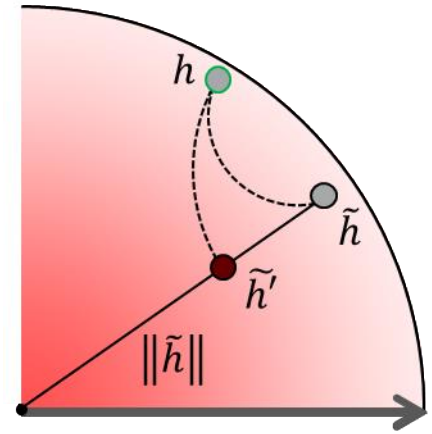
Fabio Galasso

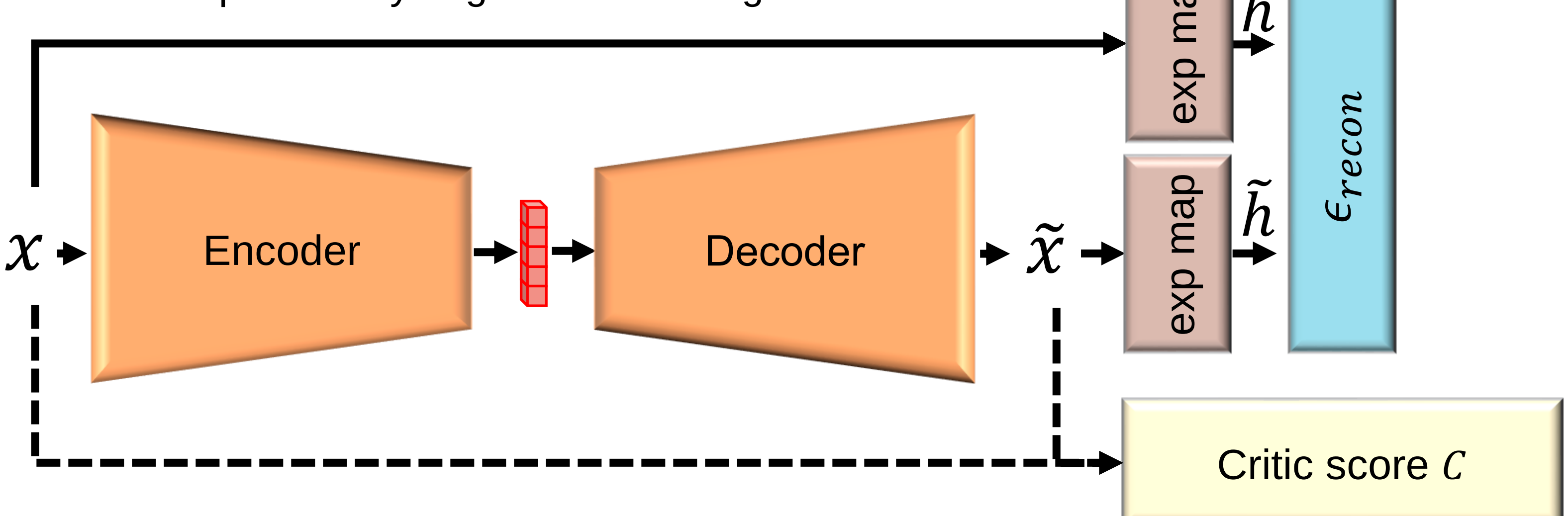
Proposed: HypAD
Hyperbolic Uncertainty for Anomaly Detection

- Minimize reconstruction errors in Hyperbolic space
  - Map x and the reconstructed $\tilde{x}$ into the Poincaré ball via exp map [Ganea et al. NeurIPS'18]
  - Compare the embeddings via the Poincaré distance:

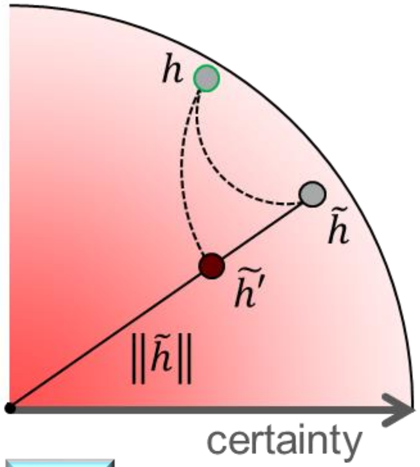
$$Z_{RE}(x) = \cosh^{-1}\left(1 + 2\frac{\|h - \tilde{h}\|^2}{\left(1 - \|h\|^2\right)\left(1 - \|\tilde{h}\|^2\right)}\right)$$

    - exponentially larger loss with larger radii





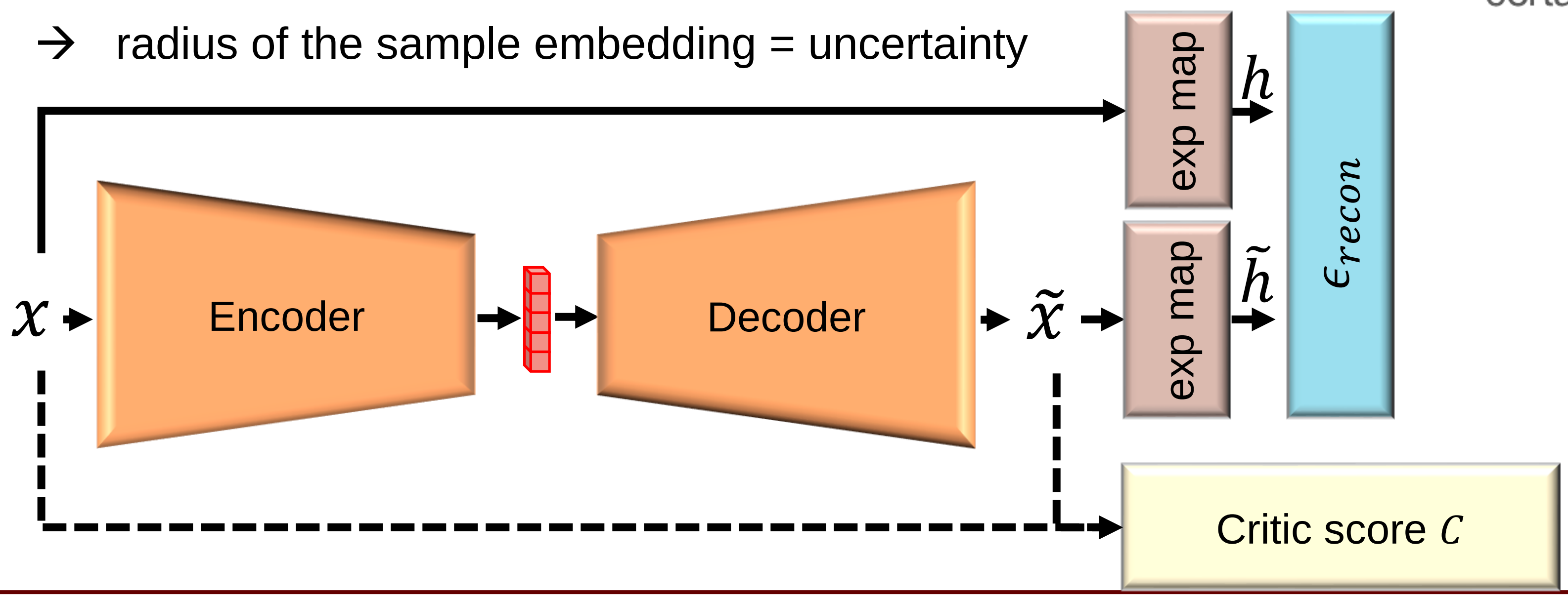
Proposed: HypAD
Hyperbolic Uncertainty for Anomaly Detection



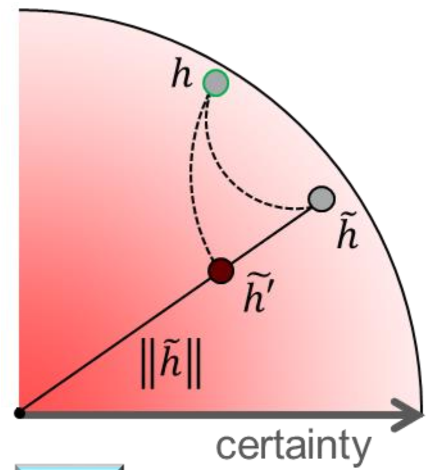
- Minimize reconstruction errors in Hyperbolic space
  - Map x and the reconstructed $\tilde{x}$ into the Poincaré ball via exp map [Ganea et al. NeurIPS'18]
  - Compare the embeddings via the Poincaré distance:

$$Z_{RE}(x) = \cosh^{-1}\left(1 + 2\frac{\|h - \widetilde{h}\|^2}{\left(1 - \|h\|^2\right)\left(1 - \|\widetilde{h}\|^2\right)}\right)$$

→ radius of the sample embedding = uncertainty



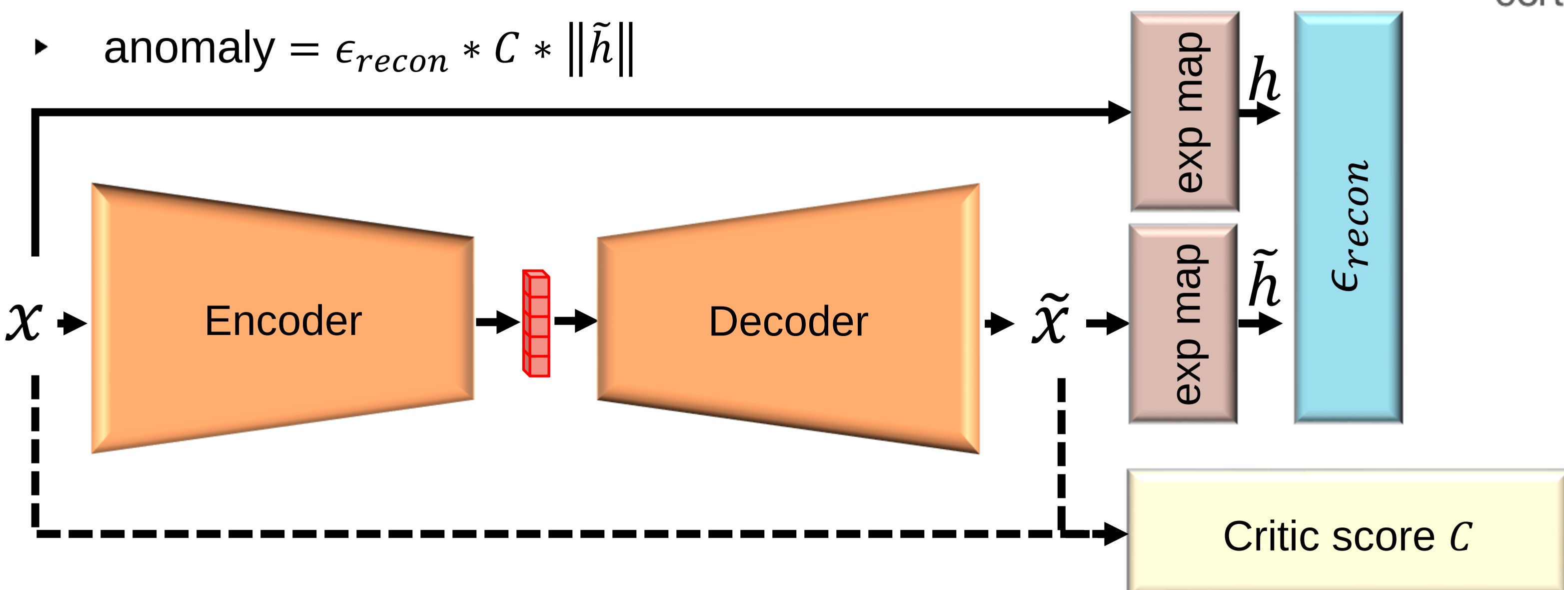
Proposed: HypAD
Hyperbolic Uncertainty for Anomaly Detection



- Minimize reconstruction errors in Hyperbolic space
  - Map x and the reconstructed $\tilde{x}$ into the Poincaré ball via exp map [Ganea et al. NeurIPS'18]
  - Compare the embeddings via the Poincaré distance:

$$Z_{RE}(x) = \cosh^{-1}\left(1 + 2\frac{\|h - \tilde{h}\|^2}{\left(1 - \|h\|^2\right)\left(1 - \|\tilde{h}\|^2\right)}\right)$$

  - anomaly = $\epsilon_{recon} * C * \|\tilde{h}\|$



PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

Evaluation of HypAD

- Surpass SoA on established anomaly detection benchmarks:

	NASA		YAHOO					NAB				
	MSL	SMAP	A1	A2	A3	A4	Art	AdEx	AWS	Traf	Tweets	F1 ($\mu \pm \sigma$)
TadGAN [Geiger et al.(2020)]	0.623	0.680	0.668	0.820	0.631	0.497	0.667	0.667	0.610	0.455	0.605	0.629 ± 0.123
AE	0.199	0.270	0.283	0.008	0.100	0.073	0.283	0.100	0.239	0.088	0.296	0.176 ± 0.099
LstmAE	0.317	0.318	0.310	0.023	0.097	0.089	0.261	0.130	0.223	0.136	0.299	0.200 ± 0.103
ConvAE	0.300	0.292	0.301	0.000	0.103	0.073	0.289	0.129	0.254	0.082	0.301	0.212 ± 0.096
TadGAN*	0.500	0.580	**0.620**	**0.865**	**0.750**	**0.576**	0.420	0.550	**0.670**	0.480	0.590	0.600 ± 0.115
HypAD (proposed)	**0.565**	**0.643**	0.610	0.670	0.670	0.470	**0.777**	**0.663**	0.630	**0.570**	**0.670**	$\mathbf{0.631 \pm 0.075}$

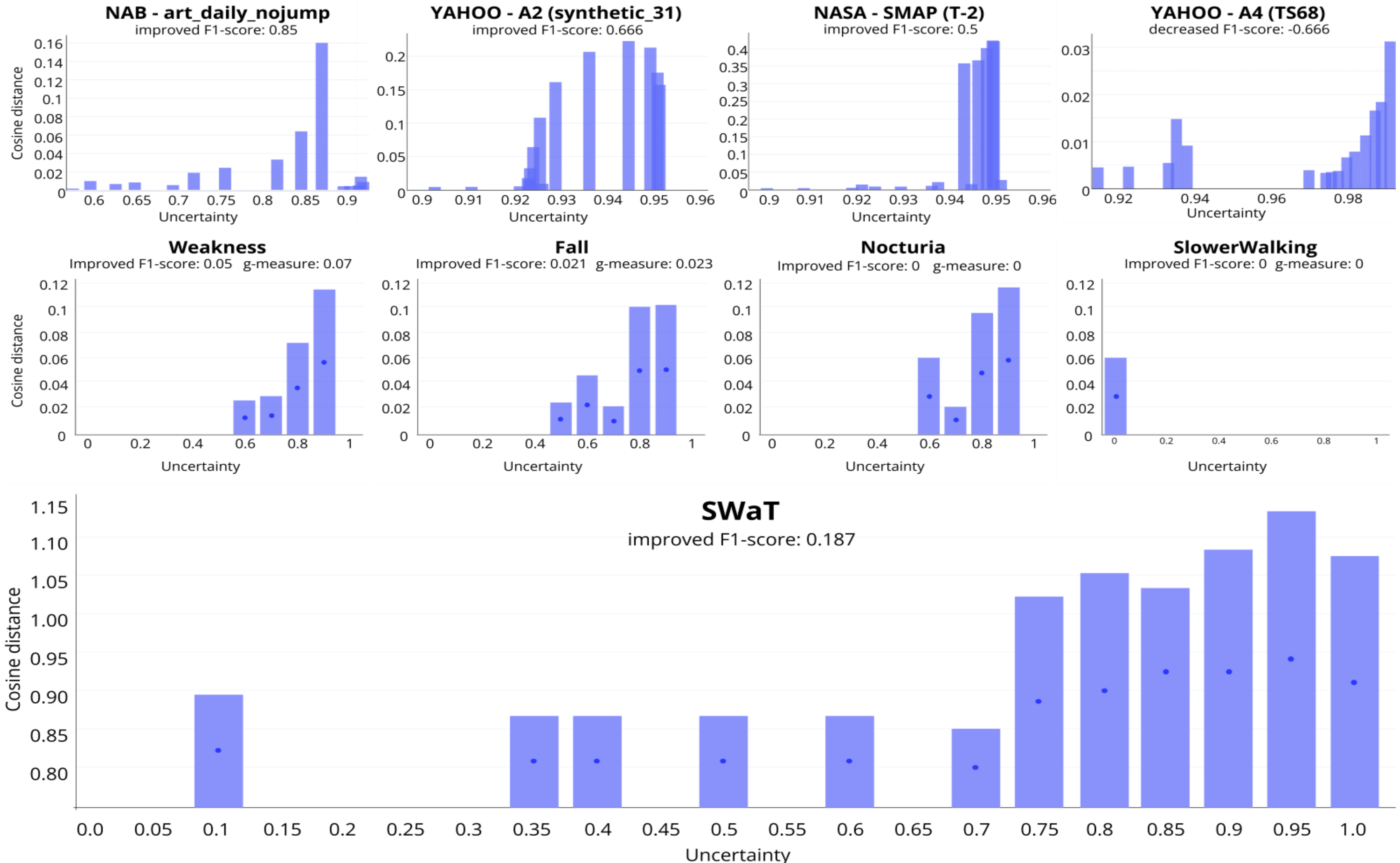
- and on medical datasets (anomaly in elderly daily activities)

	Fall		Weakness		Nocturia		SlowerWalking		MoreTimeInChair			
	g-measure	F1	g-measure	F1	g-measure	F1	g-measure	F1	g-measure	F1	g-measure ($\mu \pm \sigma$)	F1 ($\mu \pm \sigma$)
LstmAE	0.085	0.014	0.182	0.108	0.000	0.000	0.158	0.049	0.133	0.035	0.112 ± 0.064	0.041 ± 0.037
AE	0.139	0.127	0.033	0.027	0.116	0.103	0.000	0.000	0.158	0.049	0.089 ± 0.062	0.061 ± 0.047
ConvAE	0.086	0.014	0.284	0.150	0.251	0.119	0.158	0.048	0.134	0.035	0.183 ± 0.074	0.073 ± 0.052
TadGAN*	0.222	0.267	0.570	0.555	0.000	0.000	**0.630**	**0.570**	0.267	0.222	0.338 ± 0.233	0.323 ± 0.216
HypAD (proposed)	**0.447**	**0.333**	**0.660**	**0.610**	**0.447**	**0.333**	0.470	0.364	**0.577**	**0.5**	$\mathbf{0.520 \pm 0.095}$	$\mathbf{0.428 \pm 0.123}$

- Geiger et al. (2020). “TADGAN: Time series anomaly detection using generative adversarial networks”. In IEEE Int. Conf. on Big Data’20.
- Diane J Cook, Aaron S Crandall, Brian L Thomas, and Narayanan C Krishnan. Casas: A smart home in a box. Computer, 46(7):62–69, 2012.

More on HypAD

- Larger uncertainty mainly when the reconstruction is not correct



Overview

- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning

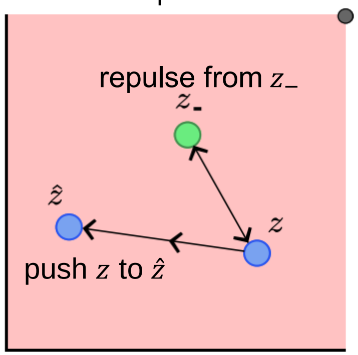
Franco, **Mandica**, Munjal, Galasso (2023). Hyperbolic Self-paced Learning for Self-supervised Skeleton-based Action Representations. In Proc. *ICLR*

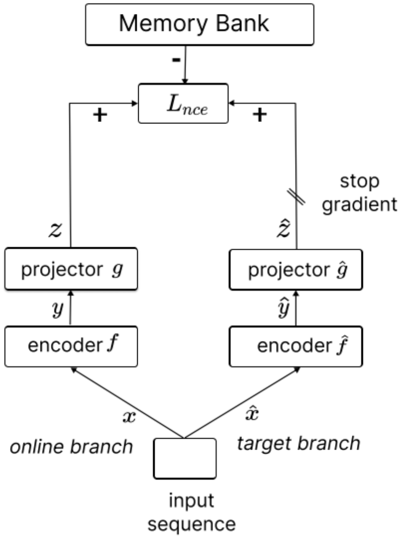
PINLAB
Perception and Intelligence

SAPIENZA
UNIVERSITÀ DI ROMA

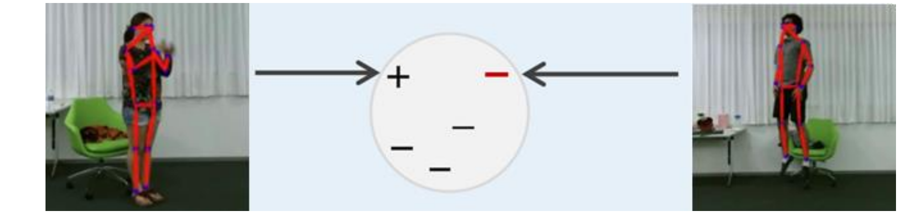
Skeleton-based SSL for action representations

- SoA builds on SkeletonCLR [Li et al. CVPR'21]



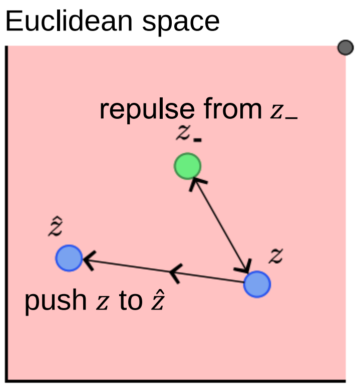


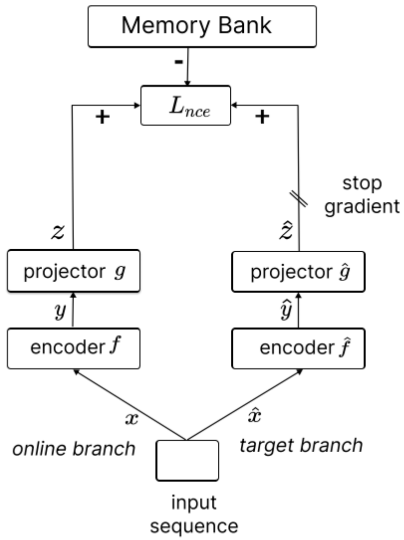
$(a) SkeletonCLR$



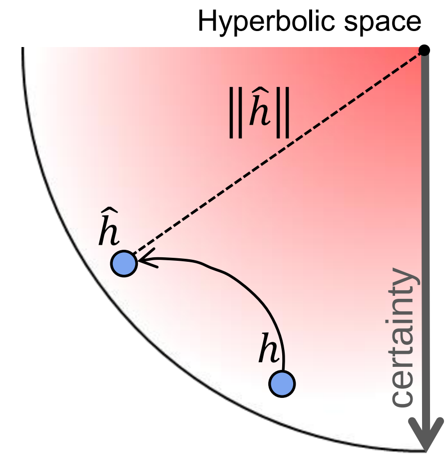
Hyperbolic Self-paced SSL (HYSP) ICLR’23

- SoA builds on SkeletonCLR [Li et al. CVPR’21]





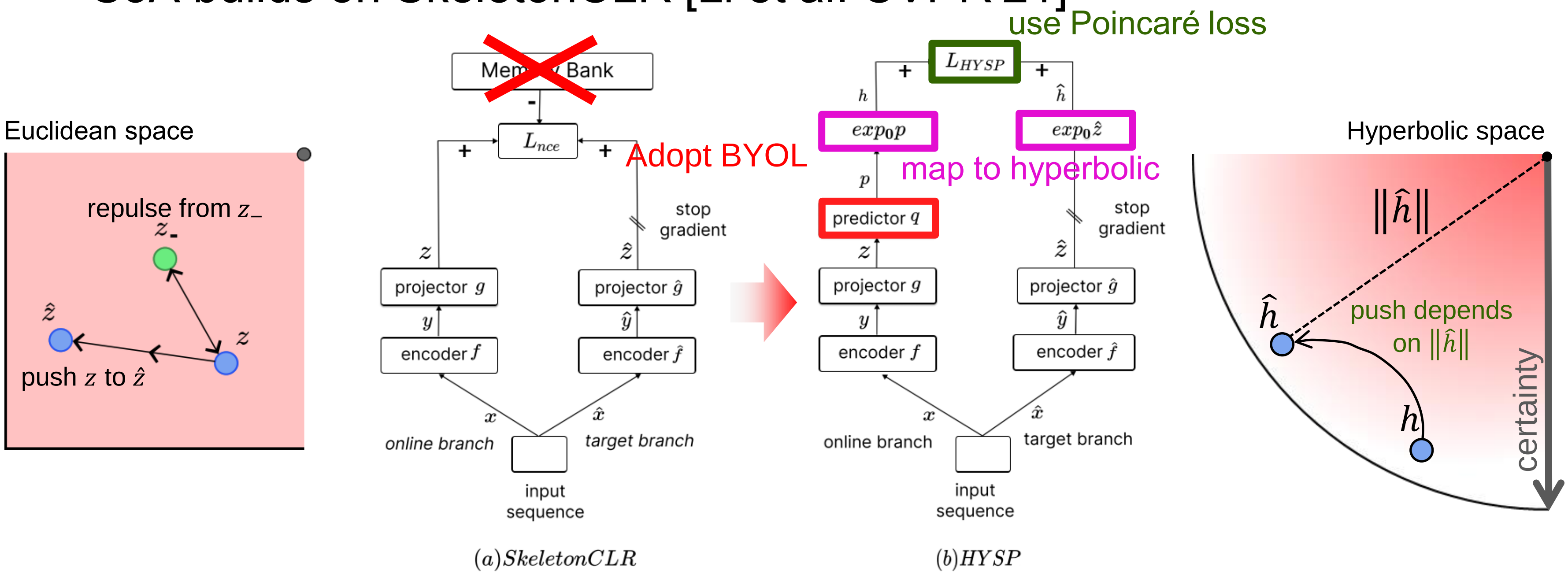
$(a) SkeletonCLR$



- Proposition: use hyperbolic uncertainty to self-pace SSL (*HYSP*)
  - More certain samples should drive learning more predominantly

Hyperbolic Self-paced SSL (HYSP) ICLR'23

- SoA builds on SkeletonCLR [Li et al. CVPR'21]



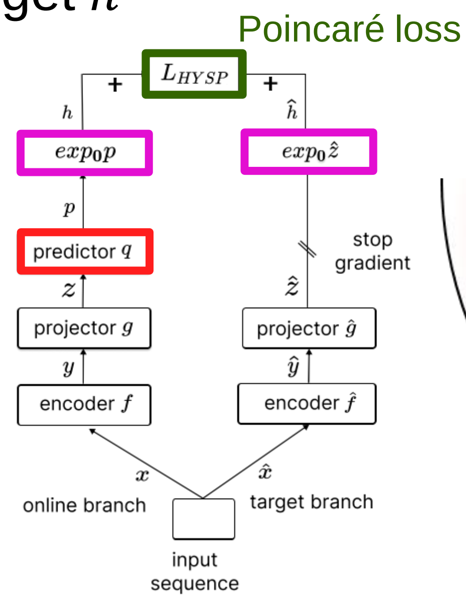
$(a) SkeletonCLR$ $(b) HYSP$

- Proposition: use hyperbolic uncertainty to self-pace SSL
  - More certain samples should drive learning more predominantly
- Hyperbolic Self-paced Self-Supervised Learning (HYSP)

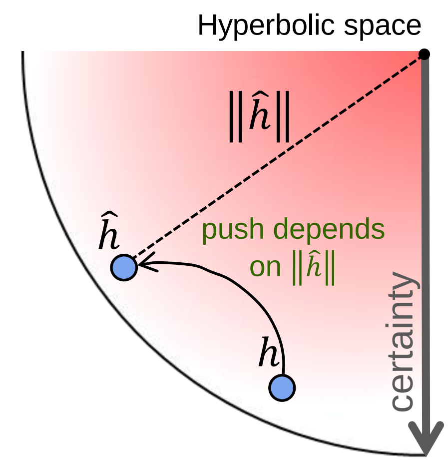
PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

Hyperbolic Self-paced SSL (HYSP) ICLR'23

- Learning: online h to match the target $\hat{h}$
- HYSP learning is self-paced
  - Poincaré loss gradient changes according to the certainty of the target $\hat{h}$
  - The larger the radius of $\hat{h}$, the more certain $\hat{h}$, the stronger are the gradients
  - More certain $\hat{h}$ drive learning more



$(b) HYSP$



$$L_{poin}(h, \hat{h}) = \cosh^{-1}\left(1 + 2\frac{\|h - \hat{h}\|^2}{\left(1 - \|h\|^2\right)\left(1 - \|\hat{h}\|^2\right)}\right)$$

$$\nabla L_{poin}(h,\hat{h}) = \frac{(1-\|h\|^2)^2}{2\sqrt{(1-\|h\|^2)(1-\|\hat{h}\|^2)+\|h-\hat{h}\|^2}}\left(\frac{h-\hat{h}}{\|h-\hat{h}\|} + \frac{h\|h-\hat{h}\|}{1-\|h\|^2}\right)$$

HYSP
Comparison with the state-of-the-art (SoA)

- Best on almost all SSL protocols on three established benchmarks

Datasets:

NTU60, NTU120, PKUMMD

Evaluation (NTU60):

- **Linear Protocol** **(+1.4%)**
- **Semi-supervised** **(+3.8%)**
- **Fine-Tuning** **(+2.4%)**

Results of linear, semi-supervised and finetuning protocols on NTU-60 and NTU-120

	Linear eval.				Semi-sup. (10%)		Finetune				Additional Techniques				
	NTU-60		NTU-120		NTU-60		NTU-60		NTU-120				Extra	Extra	
Method	*xsub*	*xview*	*xsub*	*xset*	*xsub*	*xview*	*xsub*	*xview*	*xsub*	*xset*	*3s*	*Neg.*	*Aug.*	*Pos.*	*ME*
P&C *Su et al. (2020)*	50.7	76.3	42.7	41.7	-	-	-	-	-	-					
MS^{2}L *Lin et al. (2020b)*	52.6	-	-	-	65.2	-	78.6	-	-	-		✓			
AS-CAL *Rao et al. (2021)*	58.5	64.8	48.6	49.2	-	-	-	-	-	-		✓			
SkeletonCLR *Li et al. (2021)*	68.3	76.4	56.8	55.9	66.9	67.6	80.5	90.3	75.4	75.9		✓			
MCC[§] *Su et al. (2021)*	-	-	-	-	55.6	59.9	83.0	89.7	79.4	80.8		✓			
AimCLR *Guo et al. (2022a)*	74.3	79.7	63.4	63.4	-	-	-	-	-	-		✓	✓	✓	
ISC *Thoker et al. (2021)*	76.3	**85.2**	**67.9**	**67.1**	65.9	72.5	-	-	-	-		✓	✓		✓
HYSP *(ours)*	**78.2**	82.6	61.8	64.6	**76.2**	**80.4**	**86.5**	**93.5**	**81.4**	**82.0**			✓		
3s-ST-GCN	-	-	-	-	-	-	85.2	91.4	77.2	77.1	✓				
3s-SkeletonCLR *Li et al. (2021)*	75.0	79.8	-	-	-	-	-	-	-	-	✓	✓			
3s-Colorization *Yang et al. (2021)*	75.2	83.1	-	-	71.7	78.9	88.0	94.9	-	-	✓				
3s-CrossSCLR *Li et al. (2021)*	77.8	83.4	67.9	66.7	74.4	77.8	86.2	92.5	80.5	80.4	✓	✓		✓	
3s-AimCLR *Guo et al. (2022a)*	78.9	83.8	**68.2**	**68.8**	78.2	81.6	86.9	92.8	80.1	80.9	✓	✓	✓	✓	
3s-HYSP *(ours)*	**79.1**	**85.2**	64.5	67.3	**80.5**	**85.4**	**89.1**	**95.2**	**84.5**	**86.3**	✓		✓		

Results of linear, semi-supervised and finetuning protocols on PKU-MMD I

	Linear eval.				Semi-sup. (10%)				Finetune			
Method	*Joint*	*Bone*	*Motion*	*3s*	*Joint*	*Bone*	*Motion*	*3s*	*Joint*	*Bone*	*Motion*	*3s*
MS^{2}L *Lin et al. (2020b)*	64.9	-	-	-	70.3	-	-	-	85.2	-	-	-
SkeletonCLR *Li et al. (2021)*	80.9	72.6	63.4	85.3	-	-	-	-	-	-	-	-
ISC *Thoker et al. (2021)*	80.9	-	-	-	72.1	-	-	-	-	-	-	-
AimCLR *Guo et al. (2022a)*	83.4	82.0	**72.0**	87.8	-	-	-	86.1	-	-	-	-
HYSP *(ours)*	**83.8**	**87.2**	70.5	**88.8**	**85.0**	**87.0**	**77.8**	**88.7**	**94.0**	**94.9**	**91.2**	**96.2**

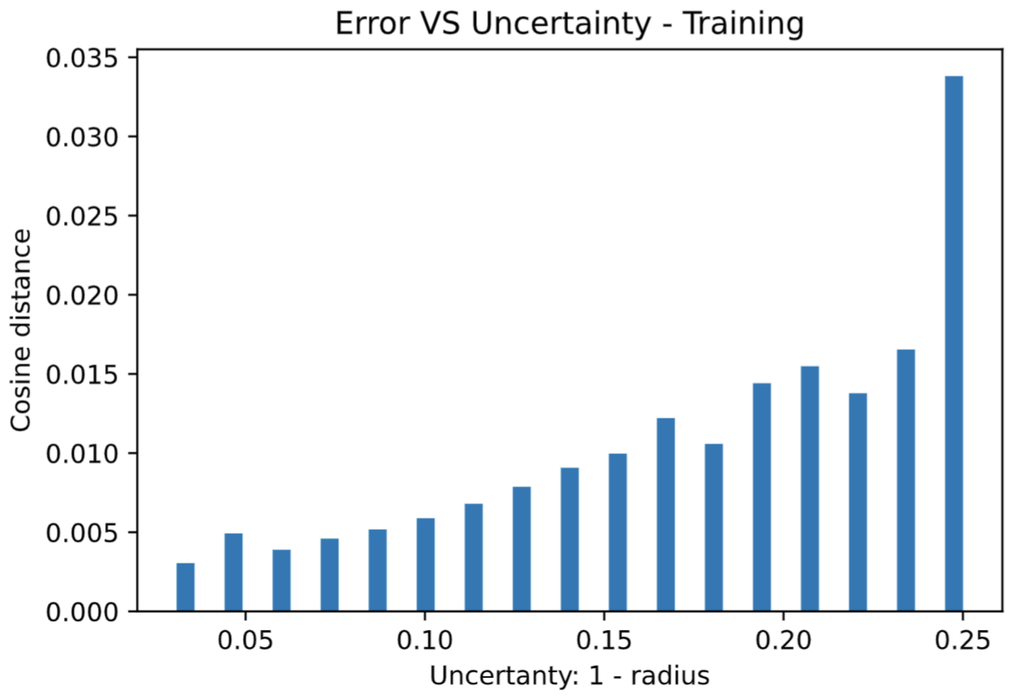
Results of transfer learning on PKU-MMD II

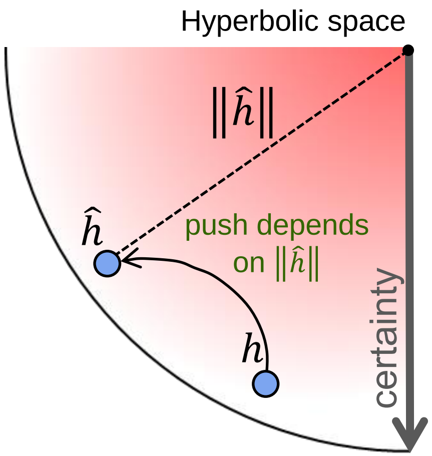
	Transfer Learning (PKU-MMD II)	
Method	*PKU-MMD I*	*NTU-60*
S+P *Zheng et al. (2018)*	43.6	44.8
MS^{2}L *Lin et al. (2020b)*	44.1	45.8
ISC *Thoker et al. (2021)*	45.1	45.9
HYSP *(ours)*	**50.7**	**46.3**

PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

HYSP after training

- End-to-end trained uncertainty matches the intuition
  - Large sample uncertainty corresponds to larger prediction errors (larger cosine distance)
  - Learn larger uncertainty for more ambiguous actions

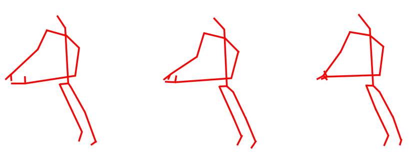




HYSP - Qualitative results

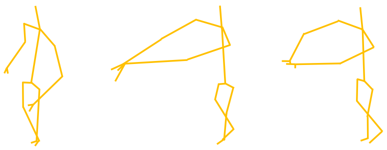
Inter-Class Variability:

small movements, **ambiguous movements,** **larger unambiguous movements**



Playing with phone (Radius 0.7679)

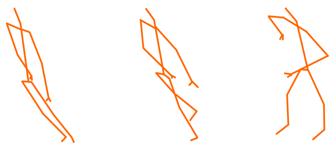
Take off hat (Radius 0.8143)



Pushing (Radius 0.9697)

Intra-Class Variability: larger unambiguous motions get larger radii

Staggering (Radius 0.7725)



Staggering (Radius 0.873)

Staggering (Radius 0.9754)

Overview

- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

Open Research Perspectives on the Hyperbolic Radius

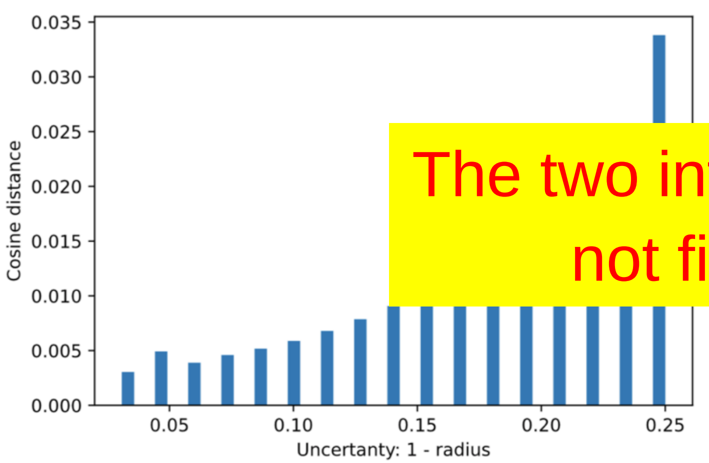
Franco, Mandica, Kallidromitis, Guillory, Li, Galasso (2023). Hyperbolic Active Learning for Semantic Segmentation under Domain Shift. ArXiv:2306.11180 pre-print

PINLAB Perception and Intelligence

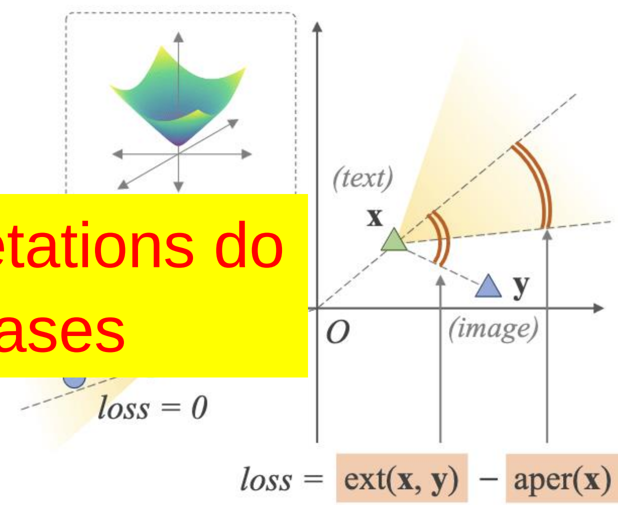
SAPIENZA UNIVERSITÀ DI ROMA

Recall the Leading Interpretations of the Hyperbolic radius

- Radius → certainty
  - Weight samples in SSL via the hyperbolic radius
    - **Franco et al. ICLR'23**
  - Detect anomalies and abstain if uncertain
    - **Flaborea et al. CVPR'23 Wks**
  - Larger error when more uncertain



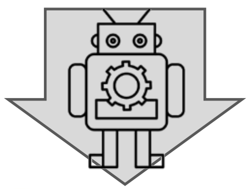
- Radius → hierarchy (parent-to-child)
  - In hierarchical classification, parent classes have lower radii
    - Ghadimi Atigh et al. CVPR'22
  - Enforce image-text hierarchies by an entailment loss
    - Desai et al. PMLR'23



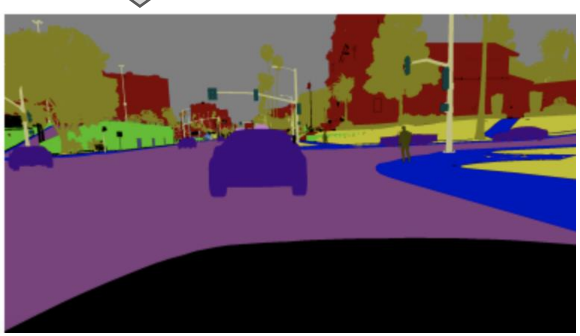
The two interpretations do not fit all cases

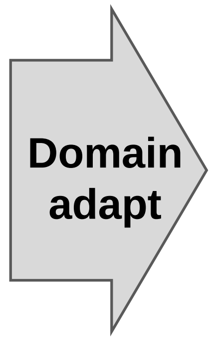
Active Domain Adaptation for Semantic Segmentation

GTA V (synthetic source dataset)

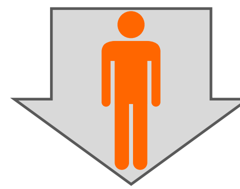


Automatically
annotate
all pixels





Cityscapes (real target dataset)

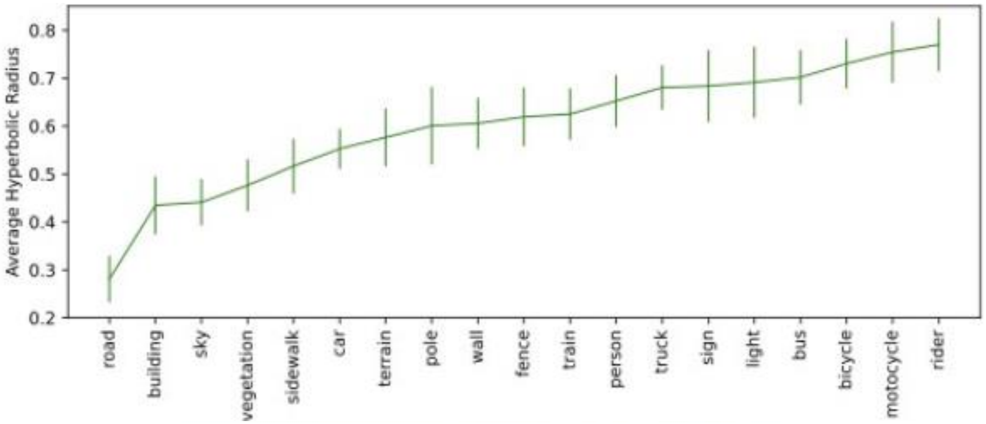


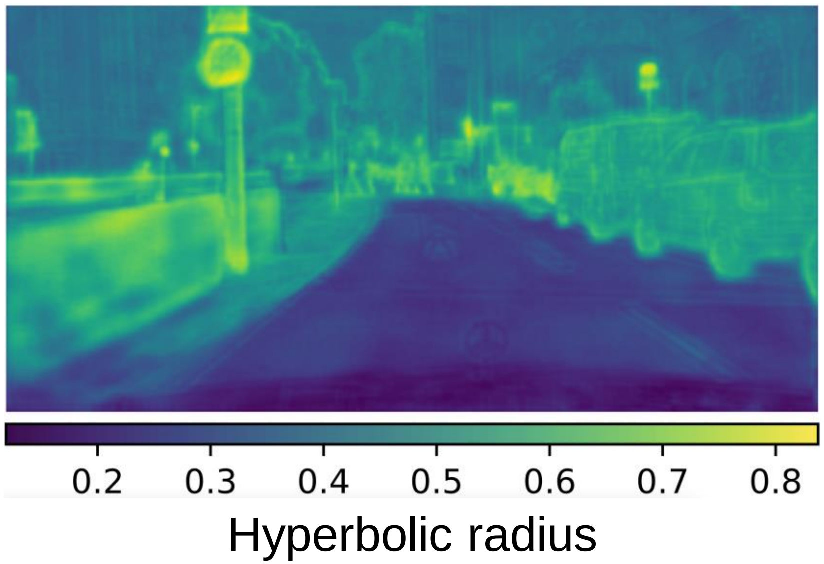
AL→ Manually label
a few pixels in
labelling rounds

PINLAB Perception and Intelligence
SAPIENZA UNIVERSITÀ DI ROMA

Surprising hyperbolic radius

- Learn SoA a semantic segmenter w/o enforcing hierarchies
  - Learned pixel embeddings via hyperbolic multinomial logistic regression

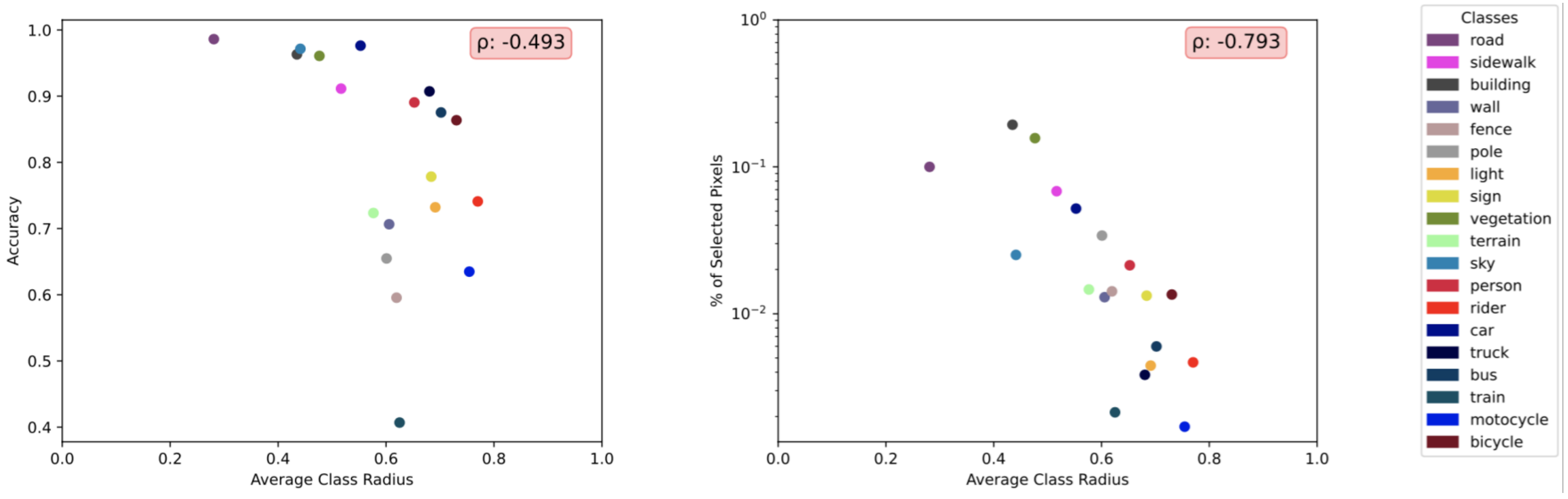




1. An object hierarchy do not emerge
2. The radius does not correspond to the pixel/class uncertainty
3. The model associates characteristic radii to each class
   → **classes are compact**

Bottom-up from data statistics and a novel interpretation of hyperbolic radius

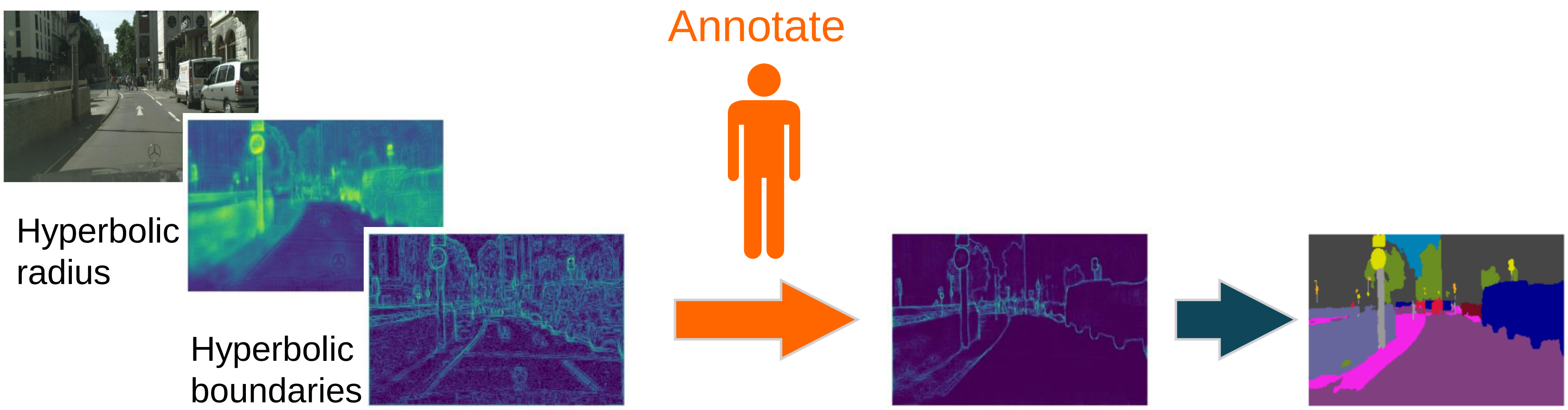
- The average class radius negatively correlates with accuracy
  - Larger hyperbolic radius → larger error
- The average class radius negatively correlates with the number of selected pixel
  - More labelled pixels → lower hyperbolic radius



- Radius → class complexity (intrinsic complexity + data scarcity)

Hyperbolic Active Learning Optimization (HALO) for semantic segmentation under domain shift

- Radius variation → class boundaries → label for active learning



- Active domain adaptation for semantic segmentation

GTA V → Cityscape

Method	road	sidew.	build.	wall	fence	pole	light	sign	veg.	terr.	sky	pers.	rider	car	truck	bus	train	motor	bike	mIoU
RIPU (2.2%) [57]	96.5	74.1	89.7	53.1	51.0	43.8	53.4	62.2	90.0	57.6	92.6	73.0	53.0	**92.8**	**73.8**	**78.5**	62.0	55.6	**70.0**	69.6
Ours (2.2%)	**97.4**	**79.8**	**90.5**	**53.6**	**53.4**	**49.9**	**57.7**	**67.6**	**90.5**	**59.7**	**93.0**	**74.4**	**54.6**	92.3	61.9	76.2	**62.9**	**56.4**	69.5	**70.6**
RIPU (5%)$^\sharp$ [57]	97.0	77.3	90.4	**54.6**	53.2	47.7	55.9	64.1	90.2	59.2	93.2	75.0	54.8	92.7	73.0	79.7	68.9	55.5	70.3	71.2
Ours (5%)$^\sharp$	**97.7**	**81.5**	**91.3**	53.5	53.6	**56.4**	**62.7**	**71.9**	**91.3**	**59.4**	**94.3**	**77.9**	**58.0**	**93.9**	**77.9**	**83.9**	**70.6**	58.9	**72.5**	**74.1**
Eucl. Fully Supervised	96.8	77.5	90.0	53.5	51.5	47.6	55.6	62.9	90.2	58.2	92.3	73.7	52.3	92.4	74.3	77.1	64.5	52.4	70.1	70.2
Hyper. Fully Supervised	97.3	79.0	89.8	50.3	51.8	43.9	52.0	61.8	89.8	58.0	92.6	71.3	50.5	91.8	65.6	78.3	64.9	52.4	67.7	68.8
Eucl. Fully Supervised$^\sharp$	97.4	77.9	91.1	54.9	53.7	51.9	57.9	64.7	91.1	57.8	93.2	74.7	54.8	93.6	76.4	79.3	67.8	55.6	71.3	71.9
Hyper. Fully Supervised$^\sharp$	97.6	81.2	90.7	49.9	53.2	53.5	58.0	67.2	91.0	59.1	93.9	74.2	52.6	93.1	76.4	81.0	67.0	55.0	70.8	71.9

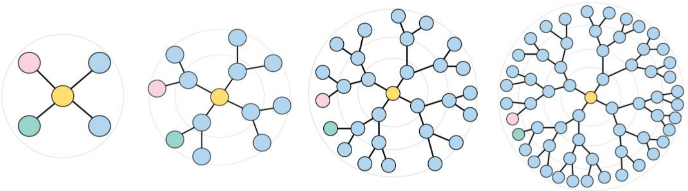
Overview

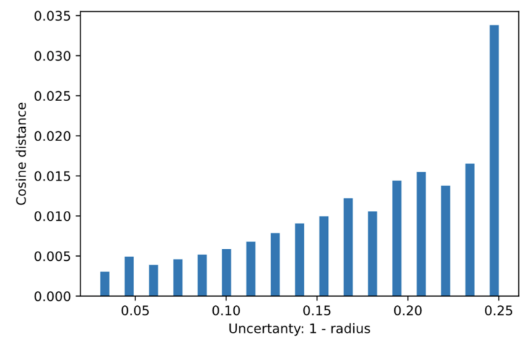
- What is hyperbolic geometry?
- Overview of the field
- From Euclid to Hyperbolic Deep Learning
- Leading Interpretations of the Hyperbolic Radius
- Hyperbolic Uncertainty for Anomaly Detection
- Hyperbolic Uncertainty for (Self-Paced) Self-Supervised Learning
- Open Research Perspectives on the Hyperbolic Radius
- Closing remarks

Closing remarks

Hyperbolic learning: a new field in computer vision

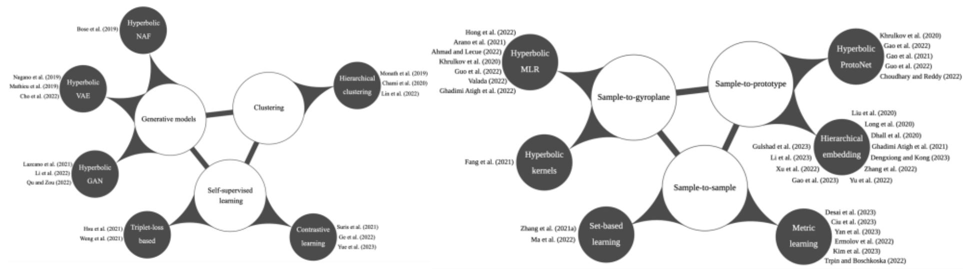
- Visual data and labels are commonly hierarchical
  - Euclidean space distorts hierarchies
  - Hyperbolic space is compact and a natural hierarchical geometry





- Uncertainty estimation is a long-standing goal of ML
  - Hyperbolic geometry allows end-to-end estimation of uncertainty

- Exponential paper growth at top-tier CV and ML conferences and several open theoretic and experimental challenges



Thank you

Our open-source code for HypAD and HYSP available at
https://github.com/paolomandica/hysp
https://github.com/aleflabo/HypAD

Our code for HALO (https://arxiv.org/abs/2306.11180)
will be available soon on https://www.pinlab.org/

Slide credits to Pascal Mettes, Max van Spengler, Yunhui Guo, and Stella Yu and to their tutorial on Hyperbolic Deep Learning for Computer Vision at CVPR'23 (https://sites.google.com/view/hdlcv-cvpr23tutorial)

Recommended further reading material:

- Mettes et al. (2023). Hyperbolic Deep Learning in Computer Vision: A Survey. Pre-print arXiv:2305.06611.
- Ganea Becigneul Hofmann NeurIPS 2018 Hyperbolic neural networks
- Shimizu Mukuta Harada ICLR 2021 Hyperbolic Neural Networks++
- Blog post by Keng: https://bjlkeng.io/posts/hyperbolic-geometry-and-poincare-embeddings/

PINLAB Perception and Intelligence

SAPIENZA UNIVERSITÀ DI ROMA