CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

Clipped Hyperbolic Classifiers Are Super-Hyperbolic Classifiers

Yunhui Guo

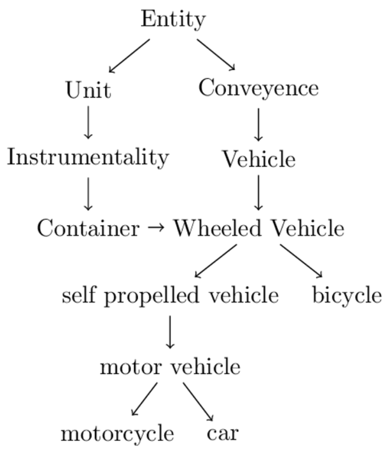
Xudong Wang

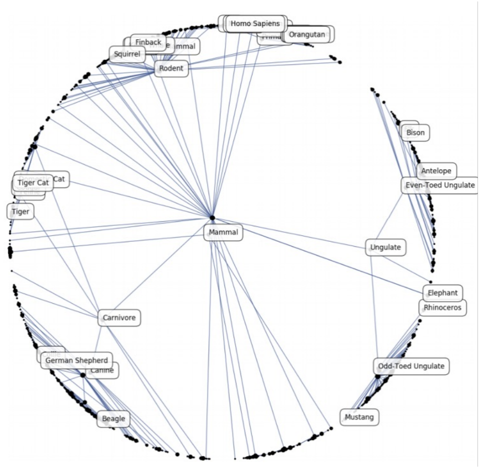
Yubei Chen

Stella X. Yu

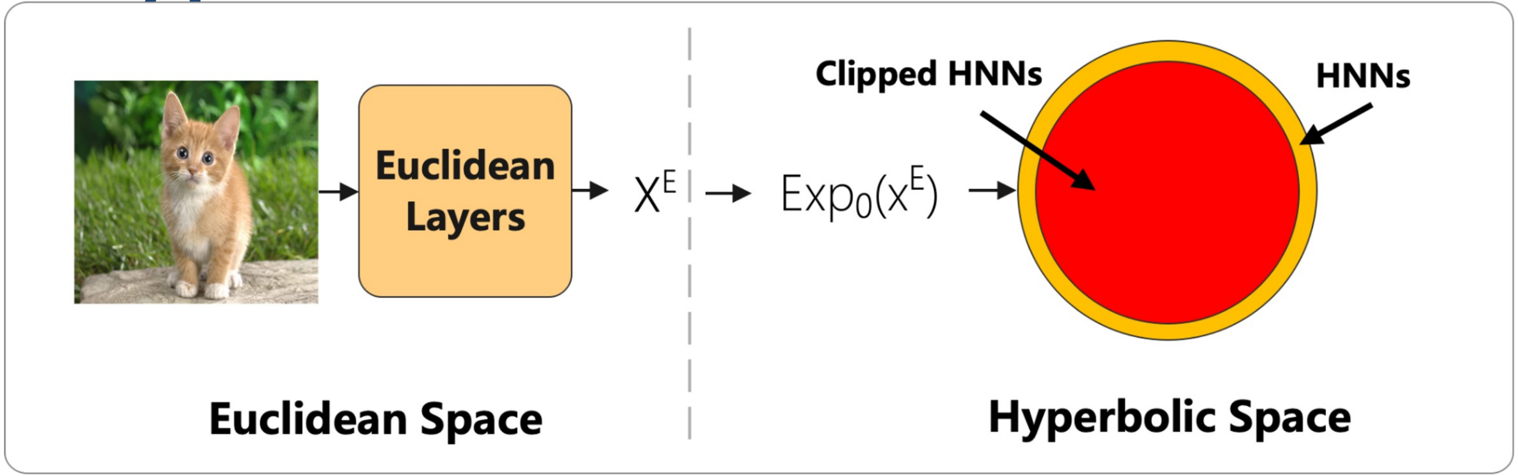
Why Hyperbolic (Non-Euclidean) Space?

Hyperbolic space can embed tree metric with low distortion, unlike Euclidean space.





Hyperbolic Neural Network (HNN)



They are made of Euclidean feature extractor + hyperbolic classifier

Exponential Map

Softmax

$$\mathrm{Exp}_{\mathbf{0}}^c(\mathbf{v}) = \tanh(\sqrt{c}\|\mathbf{v}\|)\frac{\mathbf{v}}{\sqrt{c}\|\mathbf{v}\|}$$

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

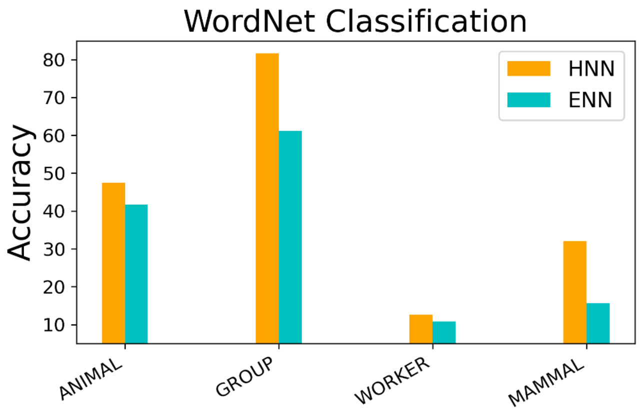
Better on hierarchical classification and few-shot learning

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

HNN Better than ENN on Hierarchical Tasks Only

Known to outperform standard Euclidean neural net (ENN) **only** for hierarchical categorization tasks, greatly limiting its applicability.

HNN > ENN

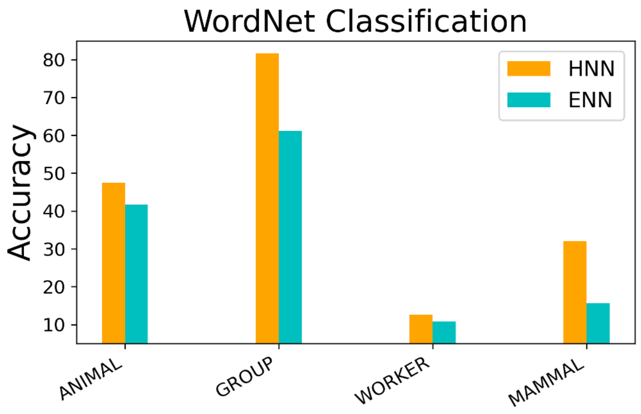


Ganea, Octavian, et al. "Hyperbolic neural networks." NeurIPS 2018.

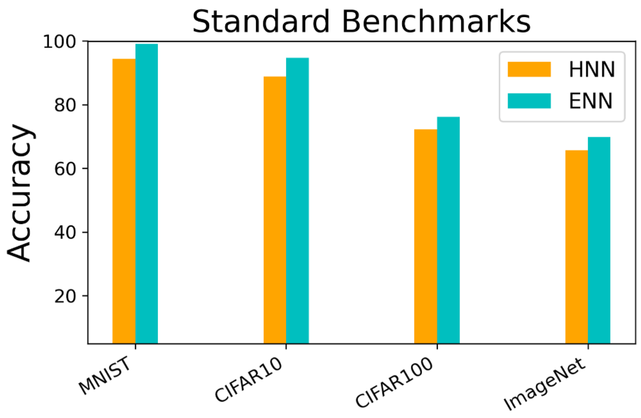
HNN Better than ENN on Hierarchical Tasks Only

Known to outperform standard Euclidean neural net (ENN) **only** for hierarchical categorization tasks, greatly limiting its applicability.

HNN > ENN



HNN < ENN



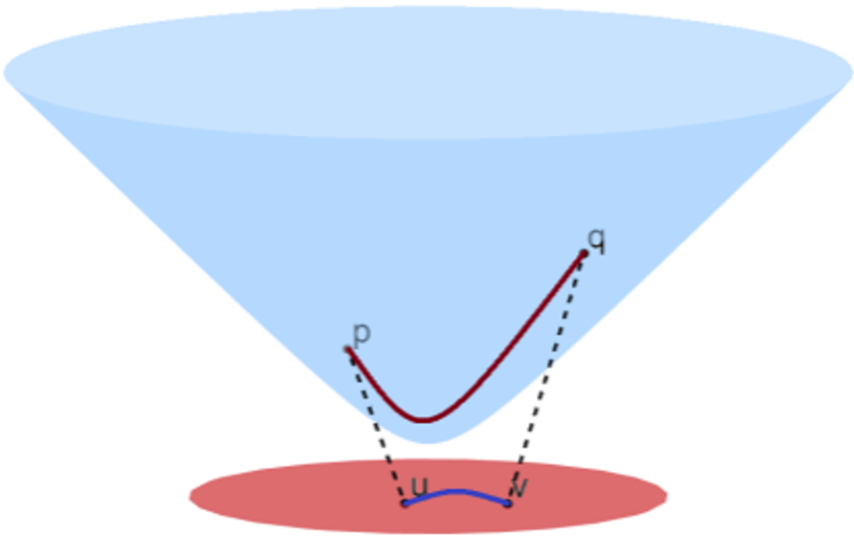
Ganea, Octavian, et al. "Hyperbolic neural networks." NeurIPS 2018.

Hyperbolic Distance in Poincaré Ball Increases Exponentially

Stereographic Projection

Hyperboloid Model

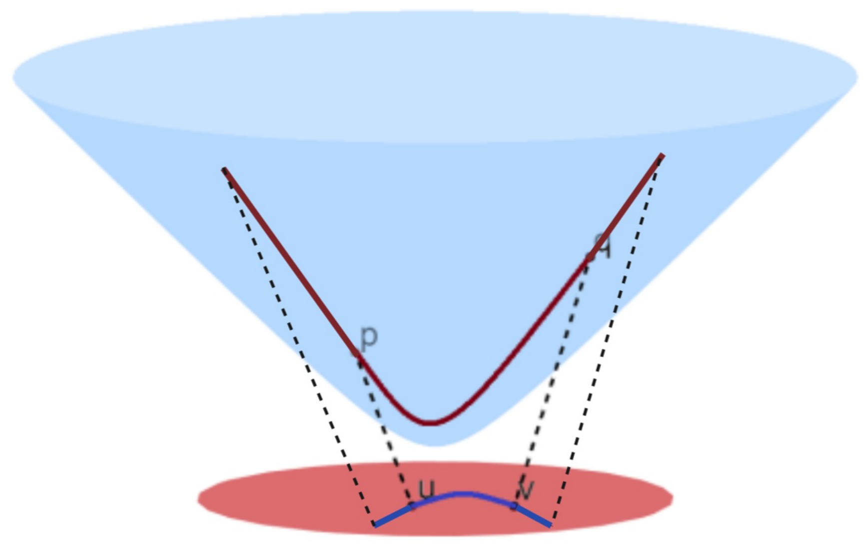
Poincaré ball Model



Hyperbolic Distance in Poincaré Ball Increases Exponentially

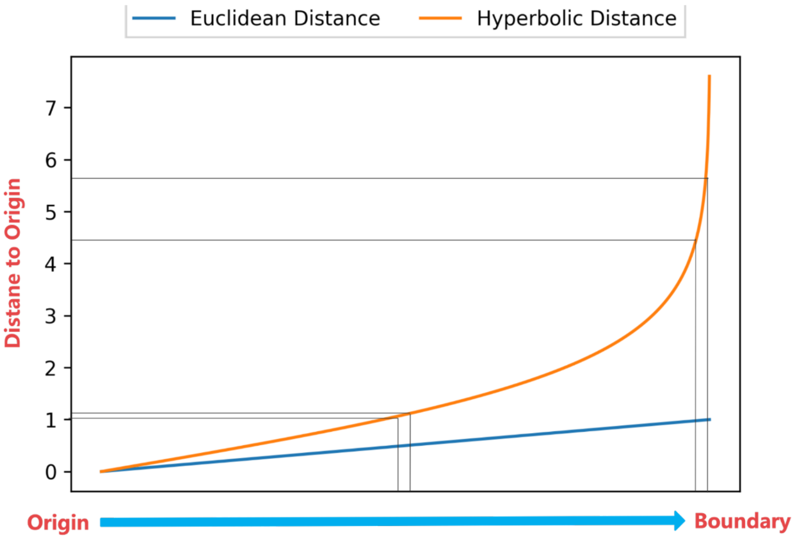
Stereographic Projection

Hyperboloid Model

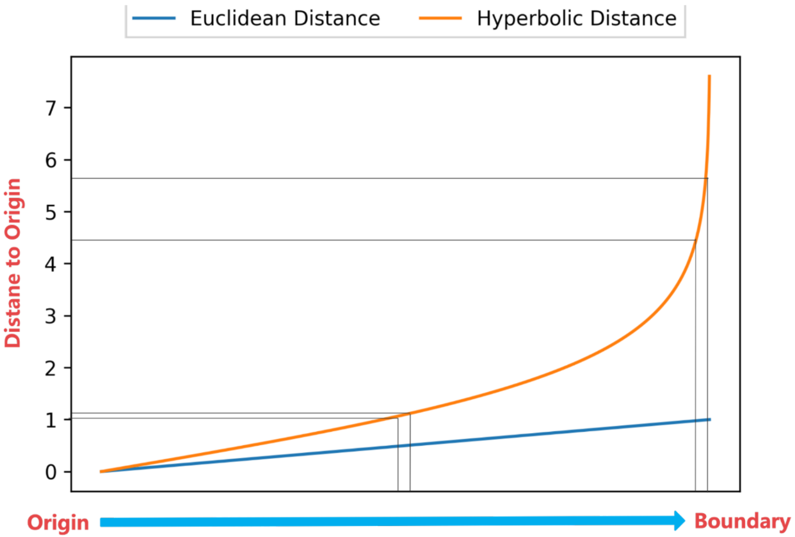


Poincaré ball Model

Non-trivial Optimization Challenges of HNNs



Non-trivial Optimization Challenges of HNNs

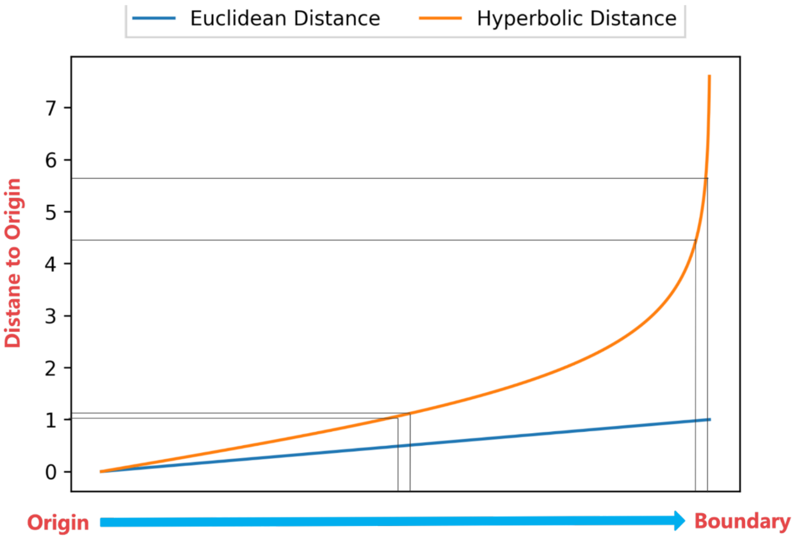


Optimizing HNNs end-to-end with backpropagation

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

Gradients w.r.t hyperbolic embedding

Non-trivial Optimization Challenges of HNNs



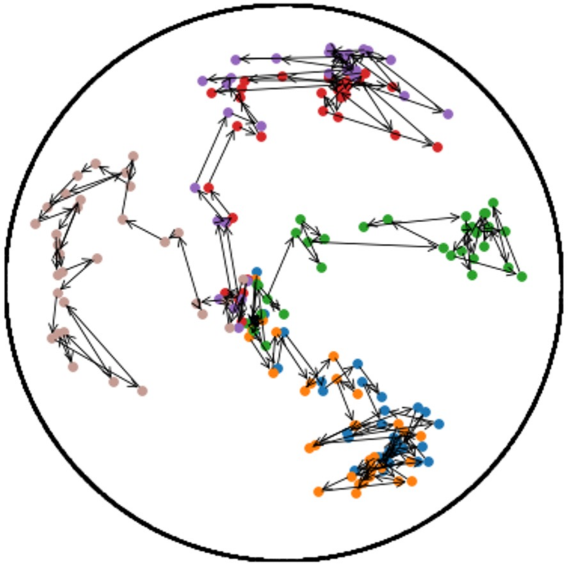
Optimizing HNNs end-to-end with backpropagation

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

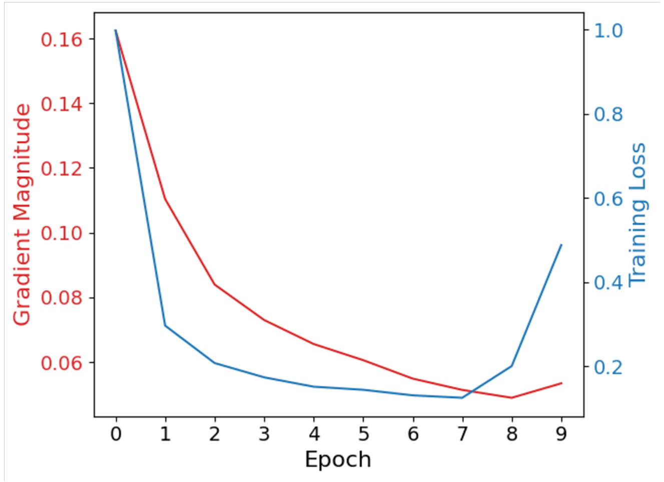
Gradients w.r.t hyperbolic embedding

Hyperbolic embedding close to the boundary -> Gradients vanish

Consequence: Gradients Vanish for Larger Features



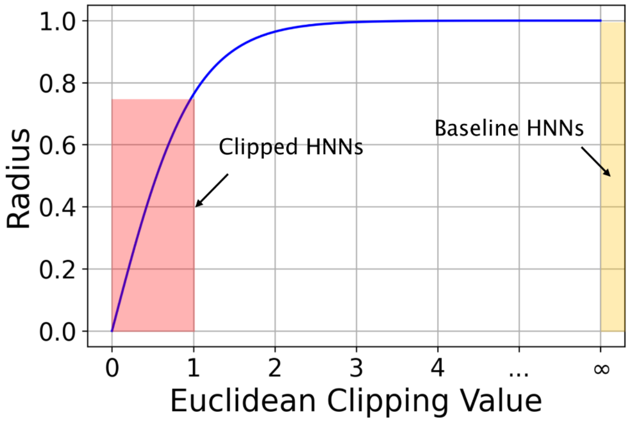
The trajectories of six sampled points during training



Training loss increases at the end of training

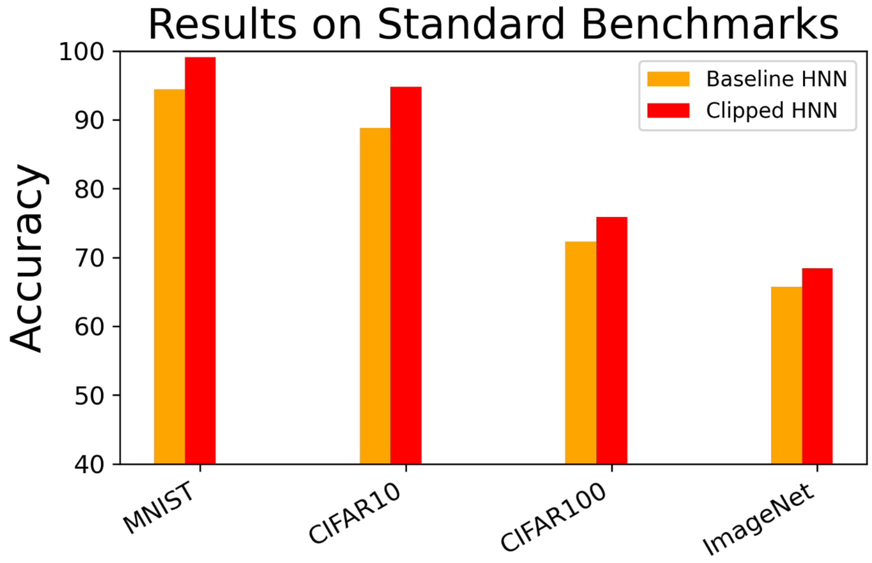
Our Solution: Feature Clipping

Avoid the ill-conditioned region while well preserving the hyperbolic property

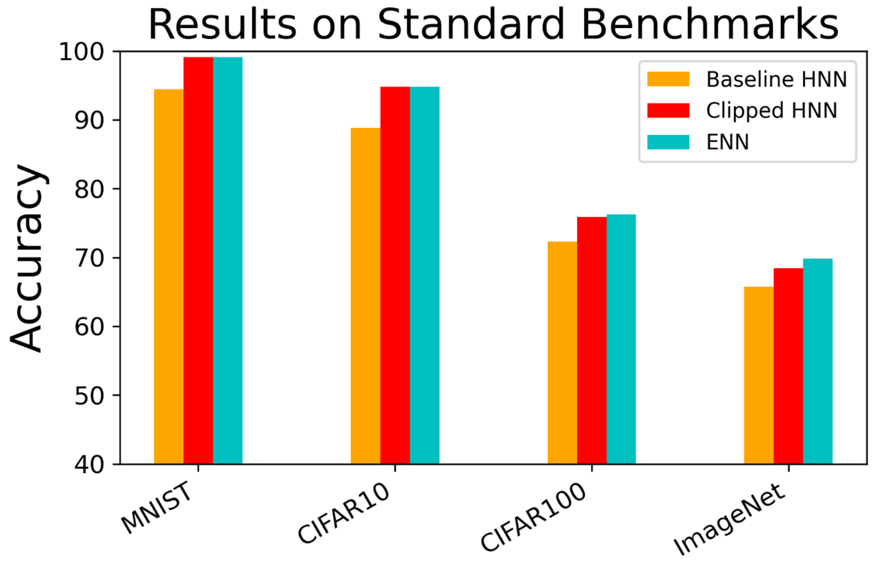


$$\mathrm{Clip}(\mathbf{x}^E; \mathrm{Clip\ Value}) = \begin{cases} \mathbf{x}^E & \text{if } \|\mathbf{x}^E\| \leq \mathrm{Clip\ Value} \\ \frac{\mathrm{Clip\ Value}}{\|\mathbf{x}^E\|} \cdot \mathbf{x}^E & \text{if } \|\mathbf{x}^E\| > \mathrm{Clip\ Value} \end{cases}$$

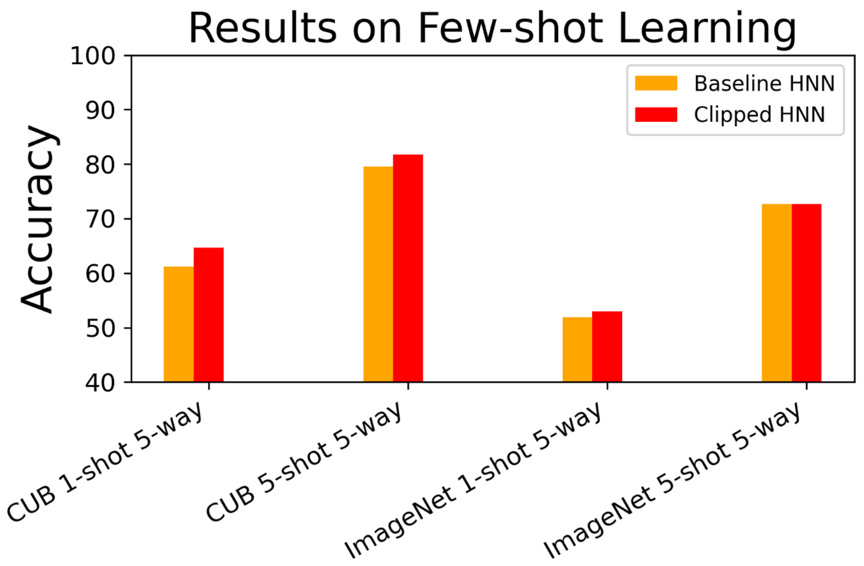
Clipped HNNs Are Super HNNs!



Clipped HNNs $\longrightarrow$ ENN on Flat Categorization

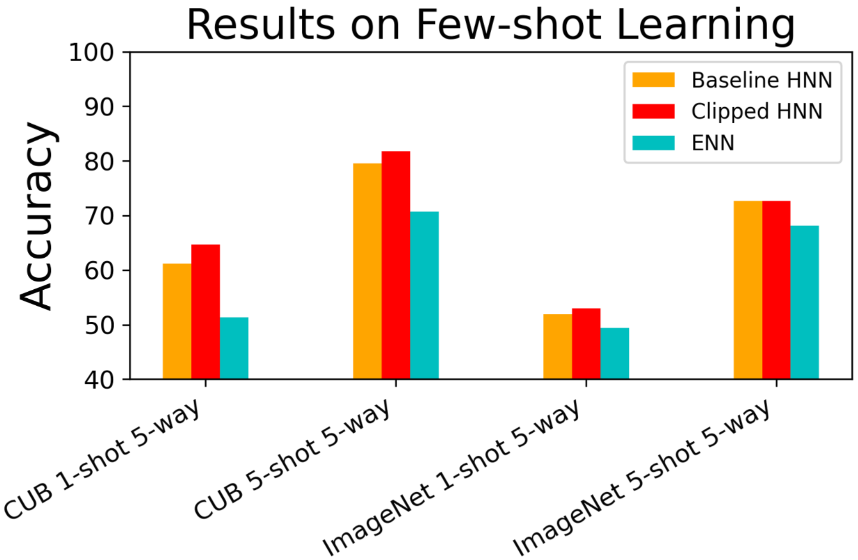


Super HNNs on Few-shot Learning



Khrulkov, Valentin, et al. "Hyperbolic image embeddings." CVPR 2020.

Super HNNs on Few-shot Learning

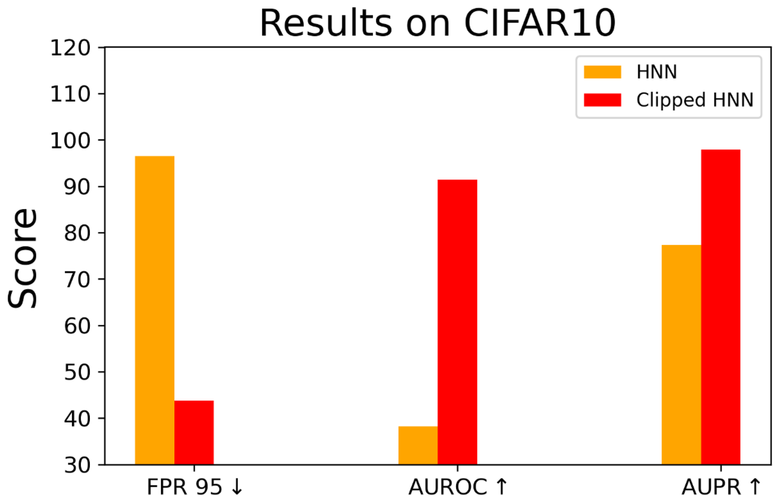


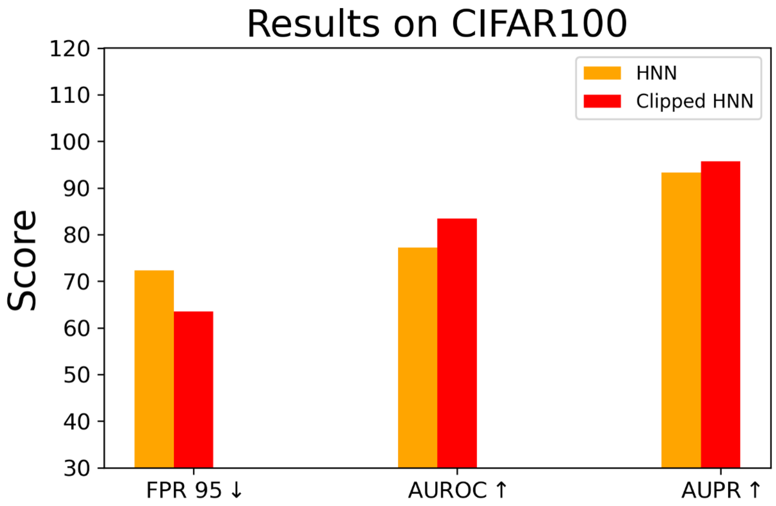
Khrulkov, Valentin, et al. "Hyperbolic image embeddings." CVPR 2020.

Generalize to Out-of-Distribution Detection

In-distribution data: CIFAR10 and CIFAR100

Out-of-distribution data: ISUN, Place365, Texture, SVHN, LSUN-Crop and LSUN-Resize

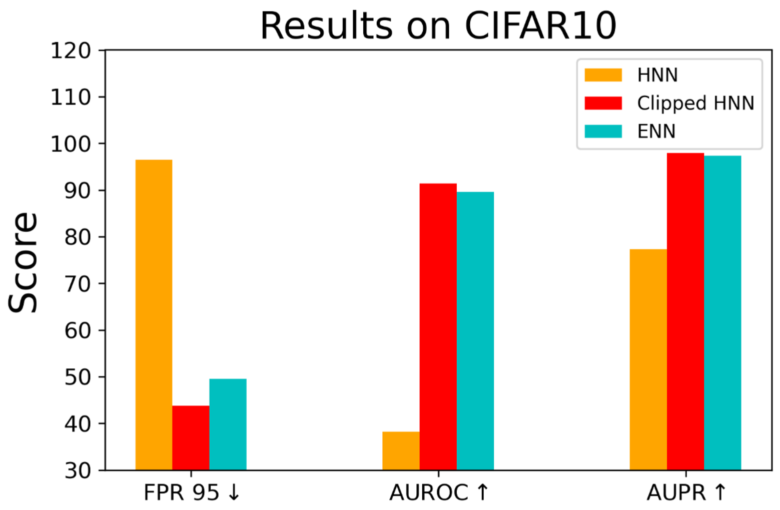


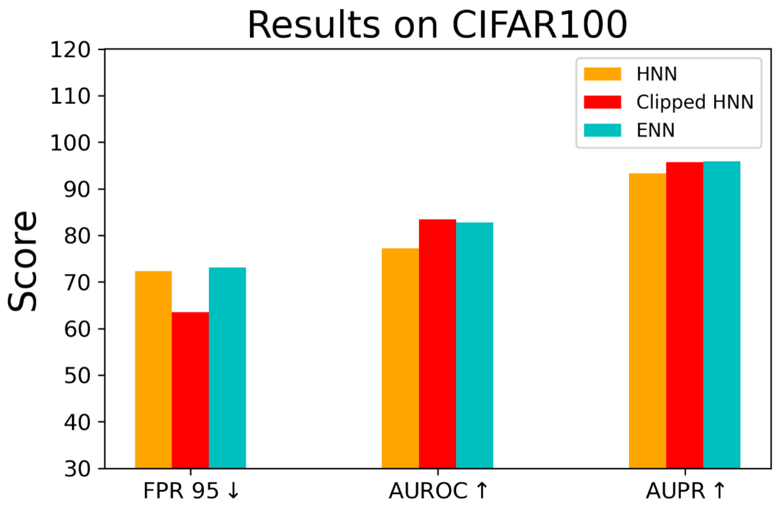


Generalize to Out-of-Distribution Detection

In-distribution data: CIFAR10 and CIFAR100

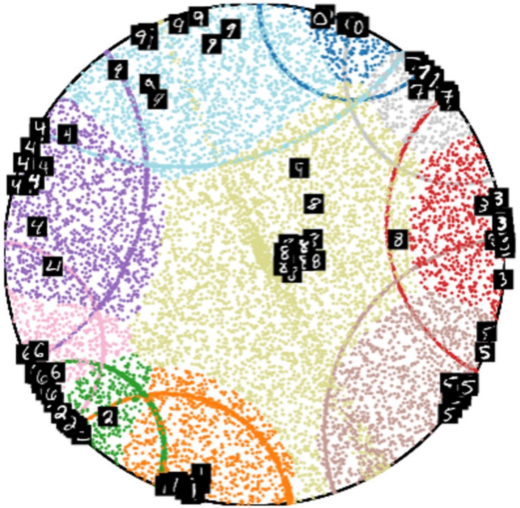
Out-of-distribution data: ISUN, Place365, Texture, SVHN, LSUN-Crop and LSUN-Resize

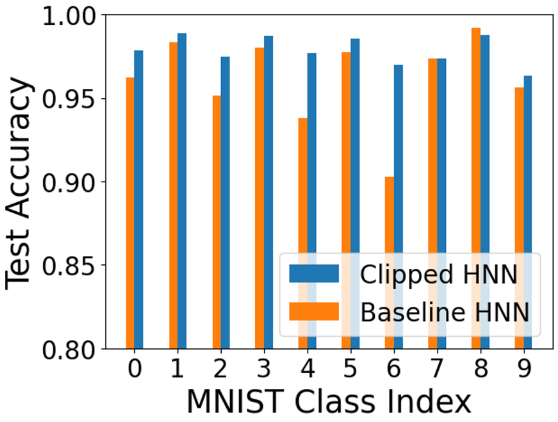




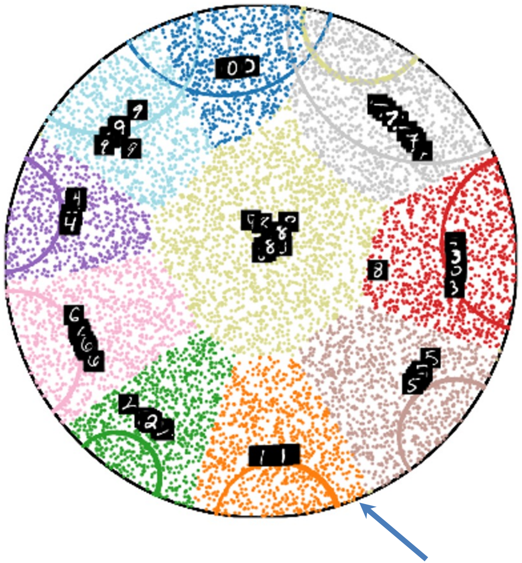
Impact on the Learned Hyperbolic Feature

HNNs



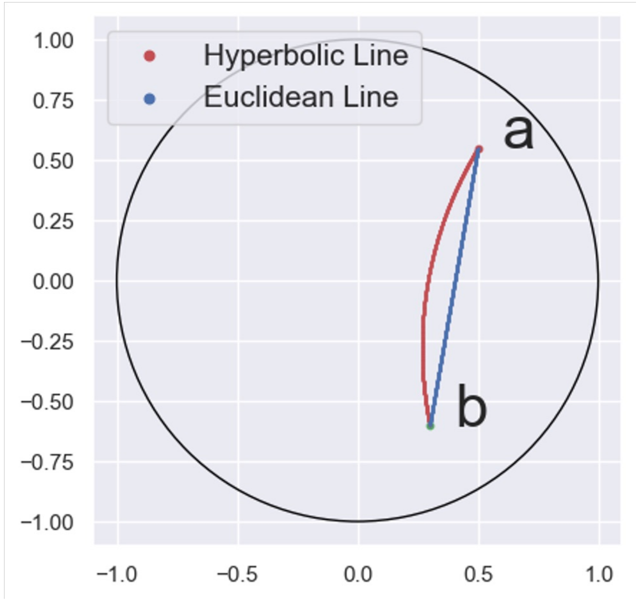


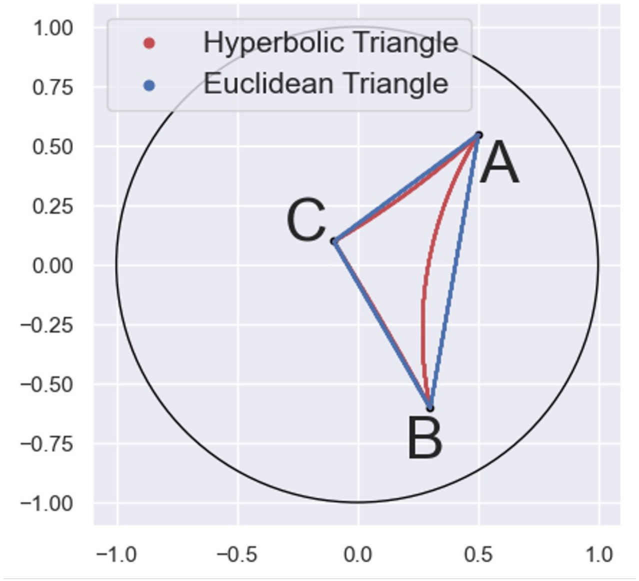
Clipped HNNs



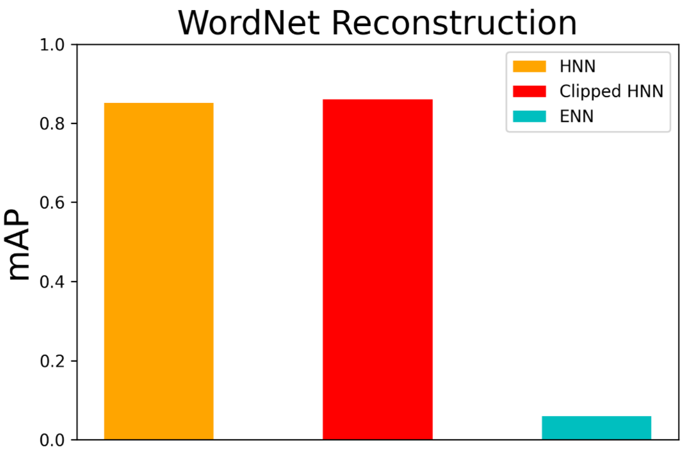
Poincaré
Hyperplane

A Clipped Hyperbolic Space Is Still Hyperbolic

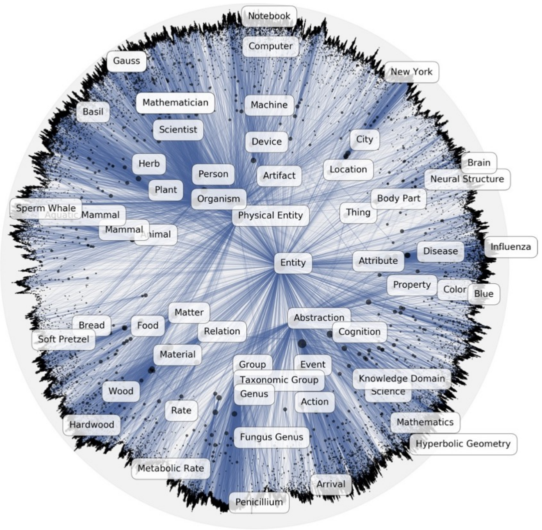




Learning Hyperbolic Word Embeddings



Hyperbolic embeddings



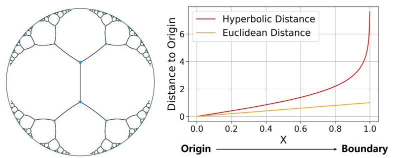
$$\mathcal{L}(\Theta) = \sum_{(u,v)\in\mathcal{D}} \log \frac{e^{-d(\boldsymbol{u},\boldsymbol{v})}}{\sum_{\boldsymbol{v}'\in\mathcal{N}(u)} e^{-d(\boldsymbol{u},\boldsymbol{v}')}}$$

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

Clipped Hyperbolic Classifiers Are Super-Hyperbolic Classifiers

Yunhui Guo Xudong Wang Yubei Chen Stella X. Yu

Hyperbolic Space



Hyperbolic Distance

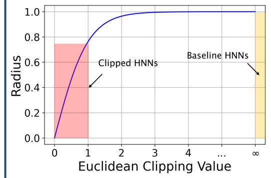
$$d_{\mathbb{B}^n}(\boldsymbol{u},\boldsymbol{v}) = \operatorname{arcosh}\left(1+2\frac{\|\boldsymbol{u}-\boldsymbol{v}\|^2}{(1-\|\boldsymbol{u}\|^2)(1-\|\boldsymbol{v}\|^2)}\right)$$

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

Contributions

Feature Clipping: Clip the Euclidean embedding before the exponential map

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

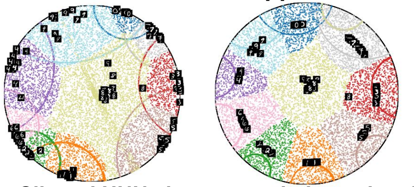


- Escape the ill-conditioned region

- Preserving the hyperbolic property

Hyperbolic Feature Space

HNNs **Clipped HNNs**



Clipped HNNs learn more balanced and discriminative feature space

OOD Detection

CIFAR10

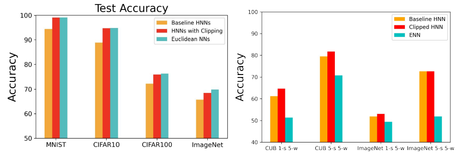
OOD Dataset	FPR95 ↓	AUROC ↑	AUPR ↑
ISUN	46.30±0.78	91.50±0.16	98.16±0.05
	45.28±0.65	91.61±0.21	98.09±0.06
Place365	51.09±0.92	87.56±0.37	96.76±0.15
	54.77±0.76	86.82±0.41	96.17±0.20
Texture	65.04±0.91	82.80±0.35	94.59±0.20
	47.12±0.62	89.91±0.20	97.39±0.09
SVHN	71.66±0.84	86.58±0.21	97.06±0.06
	49.89±1.03	91.34±0.22	98.13± 0.06
LSUN-Crop	22.22±0.78	96.05±0.10	99.16±0.03
	23.87±0.73	95.65±0.22	98.98±0.07
LSUN-Resize	41.06±1.07	92.67±0.16	98.42±0.04
	41.49±1.24	92.97±0.24	98.46 ±0.07
Mean	49.56	89.53	97.36
	43.74	**91.38**	**97.87**

CIFAR100

OOD Dataset	FPR95 ↓	AUROC ↑	AUPR ↑
ISUN	74.07±0.87	82.51±0.39	95.83±0.11
	68.37±0.90	81.31±0.43	94.96±0.20
Place365	81.01±1.07	76.90±0.45	94.02±0.15
	79.66±0.69	76.94±0.28	93.91±0.18
Texture	83.67±0.68	77.52±0.32	94.47±0.10
	64.91±0.80	83.26±0.25	95.77±0.08
SVHN	84.56±0.78	84.32±0.22	96.69±0.07
	53.11±1.04	89.53±0.26	97.71±0.07
LSUN-Crop	43.46±0.79	93.09±0.23	98.58±0.05
	51.08±1.17	87.21±0.39	96.83±0.13
LSUN-Resize	71.50±0.73	82.12±0.40	95.69±0.13
	63.86±1.10	82.36±0.42	95.16±0.13
Mean	73.05	82.74	**95.88**
	63.50	**83.43**	95.72

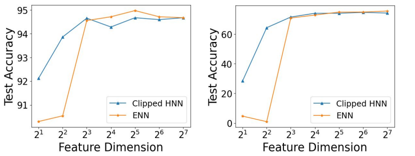
HNNs show stronger OOD detection ability than ENNs

Standard Benchmarks and Few-shot Learning



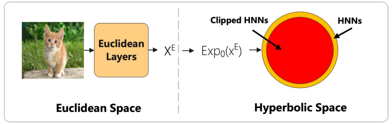
Clipped HNNs show better results compared with baseline HNNs

Impact of Dimensionality

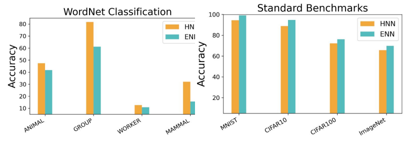


HNNs outperforms ENNs when the feature dimensionality is low

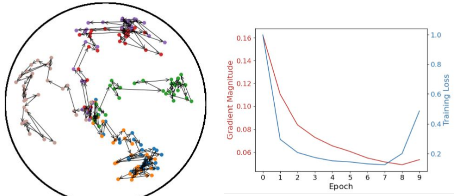
HNNs Underperform ENNs on Standard Benchmarks



HNNs consist of ENN feature extractors and hyperbolic classifiers



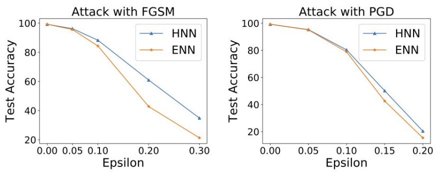
Training Dynamics of HNNs



$$\mathbf{x}^H_{t+1} = \mathbf{x}^H_t + C(E(\mathbf{w}^E_t)^T)\frac{(1-\|\mathbf{x}^H_t\|^2)^2}{4}\frac{\partial \ell}{\partial \mathbf{w}^E}$$

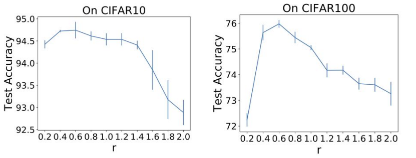
Gradients vanish when the embedding is close to the boundary

Adversarial Robustness



HNNs show stronger adversarial robustness

Impact of Clip Value



The clipping value should not be too large (causing vanishing gradient problem) or too small (no enough capacity).