Learning Structured Representations with Hyperbolic Embeddings

Aditya Sinha[1*], Siqi Zeng[1*], Makoto Yamada[2], Han Zhao[1]

University of Illinois, Urbana-Champaign[1] Okinawa Institute of Science and Technology[2]

OIST

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

NEURAL INFORMATION PROCESSING SYSTEMS

/UIUCTML/HypStructure

Goal: Accurately and explicitly embed the label hierarchy into the representation space

Motivation

- **Hierarchical label structures** → widely exist in real world datasets - CIFAR100, Imagenet-1k, ... 🌲
- **Most representation learning methods** → ignore hierarchical semantic relationships between classes in the feature space ✘
- **Structured Representation Learning** → Hierarchy informed representations incorporating semantic context ✓ [Zeng et. al, 2023], but cannot embed some trees in the Euclidean (ℓ_2) space exactly ✘

ℓ_2- Cophenetic Correlation Coefficient (CPCC)

- [Zeng et. al, 2023] → ℓ_2-CPCC for structural regularization based on label hierarchy
- $$\mathrm{CPCC}(d_{\mathcal{T}}, \rho) := \frac{\sum_{i<j}(d_{\mathcal{T}}(v_i,v_j) - \overline{d_{\mathcal{T}}})(\rho(v_i,v_j) - \overline{\rho})}{\sqrt{\sum_{i<j}(d_{\mathcal{T}}(v_i,v_j) - \overline{d_{\mathcal{T}}})^2}\sqrt{\sum_{i<j}(\rho(v_i,v_j) - \overline{\rho})^2}}$$
- $\rho(v_i, v_j) :=$ Euclidean dist. b/w two **class centroids**
- $d_{\mathcal{T}}(v_i, v_j) :=$ **Shortest tree distance** in the hierarchy
- Loss: $\mathcal{L}(\mathcal{D}) = \sum_{(\boldsymbol{x},y)\in\mathcal{D}} \ell_{\mathrm{Flat}}(\boldsymbol{x}, y, \theta, w) - \alpha \cdot \mathrm{CPCC}(d_{\mathcal{T}}, \rho)$

Our Contribution: HypStructure

- label-hierarchy → structured learning in the hyperbolic space →interpretable *tree-like* features
- Combine with any loss, beneficial across tasks
- Formal analysis of hierarchical representations

KEY IDEAS

- **`HypStructure`**: Combination of two losses (1) <u>Hyp</u>erbolic <u>Cop</u>henetic <u>C</u>orrelation <u>C</u>oefficient Loss (**`HypCPCC`**) and (2) <u>Hyp</u>erbolic <u>Cen</u>tering Loss (**`HypCenter`**)
- **`HypCPCC`**: extend ℓ_2-CPCC [Zeng et. al, 2023] to the hyperbolic space
  1. map Euclidean vectors to Poincáre disk
  2. compute class prototypes
  3. use Poincáre distance for CPCC computation
- **`HypCenter`**: Inspired from Sarkar's construction
  - place root node at the origin [2012]
  - ℓ_{center} loss → minimize the norm of the hyperbolic representations of the root
- Learn hierarchy-informed representations by minimizing:

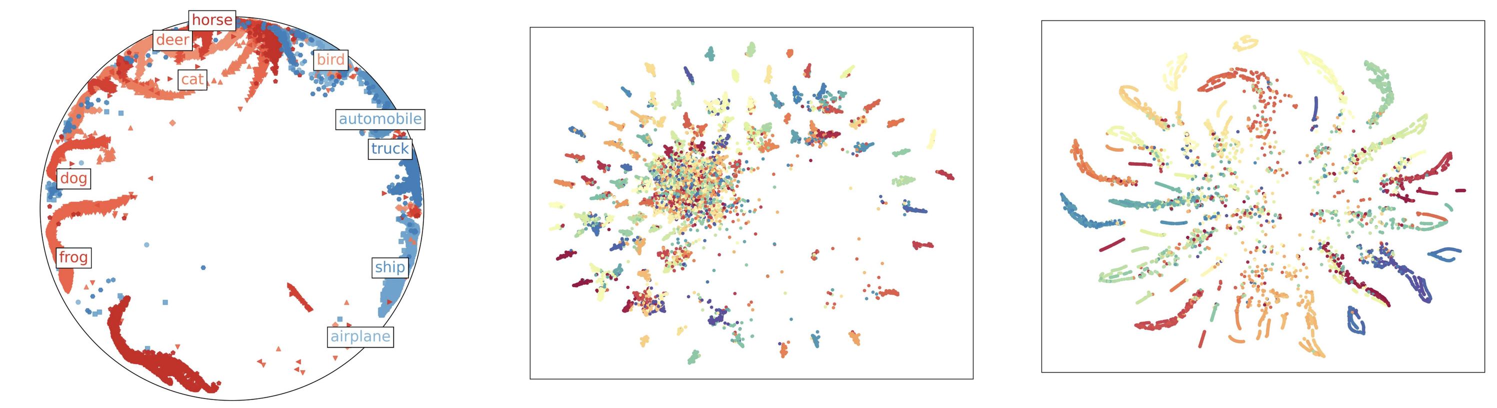
$$\mathcal{L}(\mathcal{D}) = \sum_{(\boldsymbol{x},y)\in\mathcal{D}} \ell_{\mathrm{Flat}}(\boldsymbol{x}, y, \theta) - \alpha \cdot \mathrm{HypCPCC}(d_{\mathcal{T}}, d_{\mathbb{B}_c}) + \beta \cdot \ell_{\mathrm{center}}(\boldsymbol{x}, \theta)$$

Results: Classification and Embedding Hierarchy

- Experiments: CIFAR10, CIFAR100, ImageNet100
- Compared to Flat and ℓ_2-CPCC
  - Reduced distortion in embedding the hierarchy
  - Improved Coarse and Fine Classification Accuracies

Dataset (Backbone)	Method	Distortion of Hierarchy		Classification Accuracy	
		δ_{rel} (↓)	CPCC (↑)	Fine (↑)	Coarse (↑)
CIFAR10 (ResNet-18)	Flat	0.232 (0.001)	0.573 (0.002)	94.64 (0.12)	99.16 (0.04)
	ℓ_2-CPCC	0.174 (0.002)	0.966 (0.001)	94.47 (0.13)	98.91 (0.02)
	HypStructure	**0.094 (0.003)**	**0.992 (0.001)**	**94.79 (0.14)**	**99.18 (0.04)**
CIFAR100 (ResNet-34)	Flat	0.209 (0.002)	0.534 (0.119)	74.96 (0.14)	84.15 (0.19)
	ℓ_2-CPCC	0.213 (0.006)	**0.779 (0.002)**	76.07 (0.19)	85.28 (0.32)
	HypStructure	**0.127 (0.016)**	0.766 (0.007)	**76.68 (0.22)**	**86.01 (0.13)**
ImageNet100 (ResNet-34)	Flat	0.168 (0.003)	0.429 (0.002)	90.01 (0.07)	90.77 (0.11)
	ℓ_2-CPCC	0.213 (0.009)	0.834 (0.002)	89.57 (0.38)	90.34 (0.28)
	HypStructure	**0.134 (0.001)**	**0.841 (0.001)**	**90.12 (0.01)**	**90.84 (0.02)**

Visualization: Learnt Representations

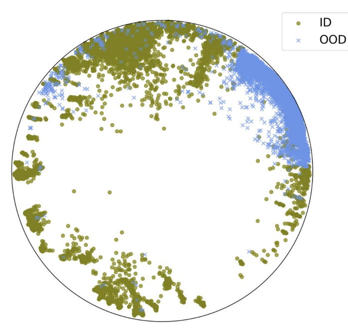


- **`HypStructure`**: Sharper and discriminative feats.
- Fine classes of the same coarse parent → closer

Results: OOD Detection

- Experiments: 3 ID datasets on 9 OOD datasets
- **`HypStructure`**:
  - Improvement in OOD detection AUROC
  - Improved ID vs OOD separation in Poincare disk

Method	AUROC	Method	AUROC	Method	AUROC
CIFAR10		CIFAR100		ImageNet100	
SSD+	97.38	SSD+	85.90	SSD+	92.46
KNN+	97.22	KNN+	86.14	KNN+	92.74
ℓ_2-CPCC	76.67	ℓ_2-CPCC	85.26	ℓ_2-CPCC	91.33
HypStructure	**97.75**	HypStructure	**88.21**	HypStructure	**93.83**



Theoretical Analysis

- **`HypStructure`** with Mahalanobis OOD Score

$$s(\boldsymbol{x}) = (f(\boldsymbol{x}) - \mu)^\top \Sigma^{-1}(f(\boldsymbol{x}) - \mu)$$

- **Theorem 5.1**: Existence of eigenvalue gaps between each level of hierarchy with CPCC. Can be generalized to arbitrary label tree.
- Kernel Matrix K = ZZ^T, eigenspectrum of K