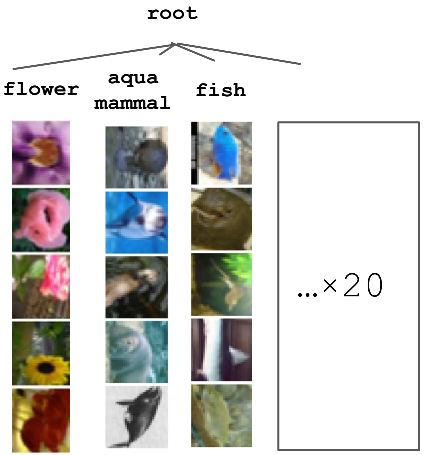
Learning Structured Representations by Embedding Class Hierarchy

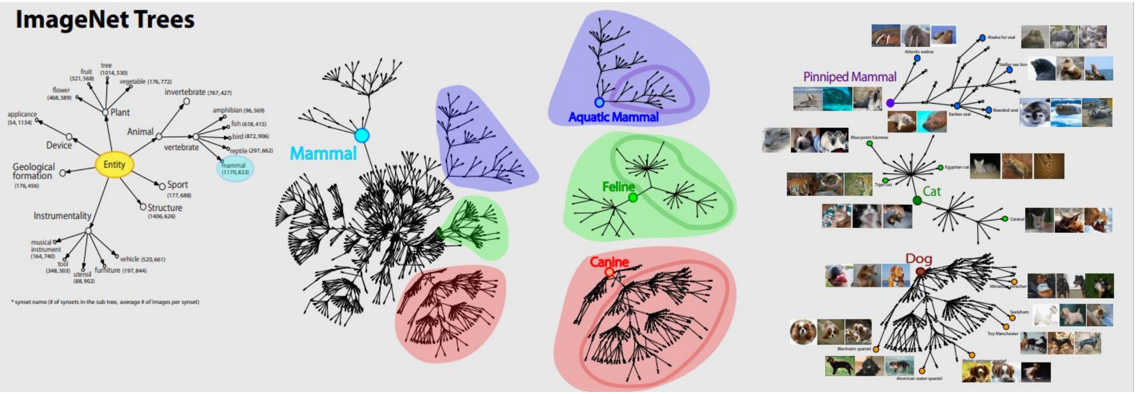
Siqi Zeng, Remi Tachet des Combes, Han Zhao

Hierarchical label structures widely exist in many real-world datasets…



CIFAR100 Tree

Isshiki 2020;
Krizhevsky, 2009

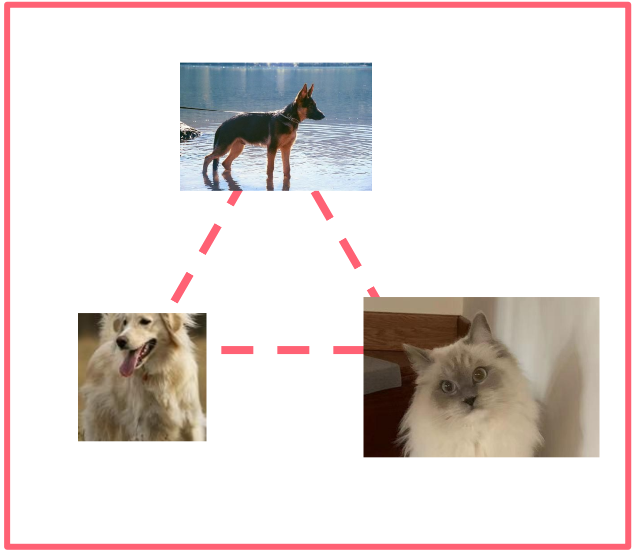


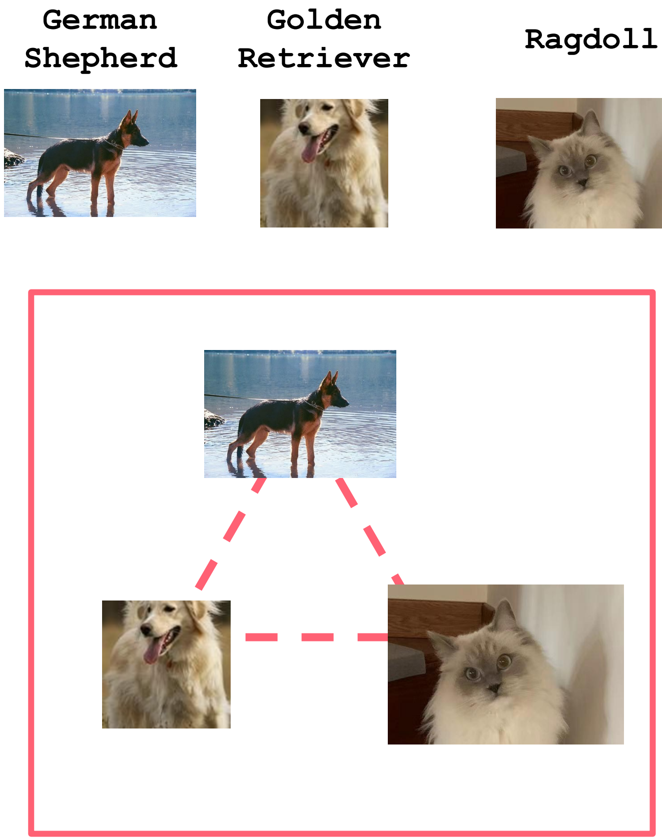
Tabin & Mohammad 2016; Deng et al. 2009

German Shepherd

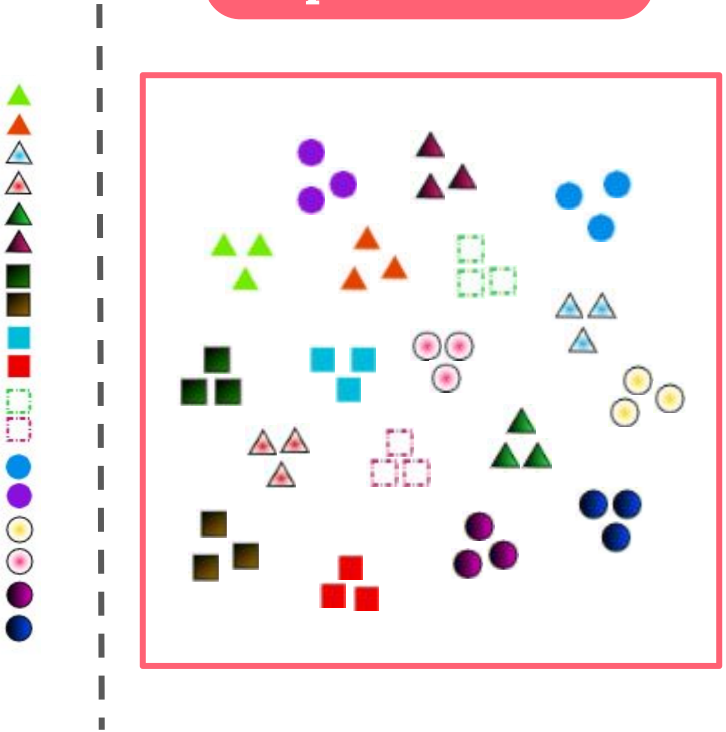
Golden Retriever

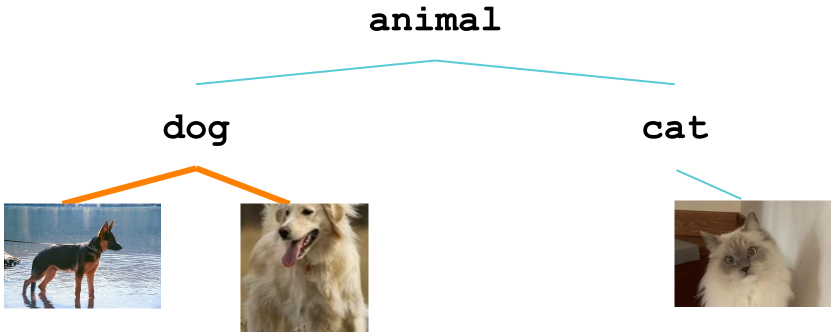
Ragdoll

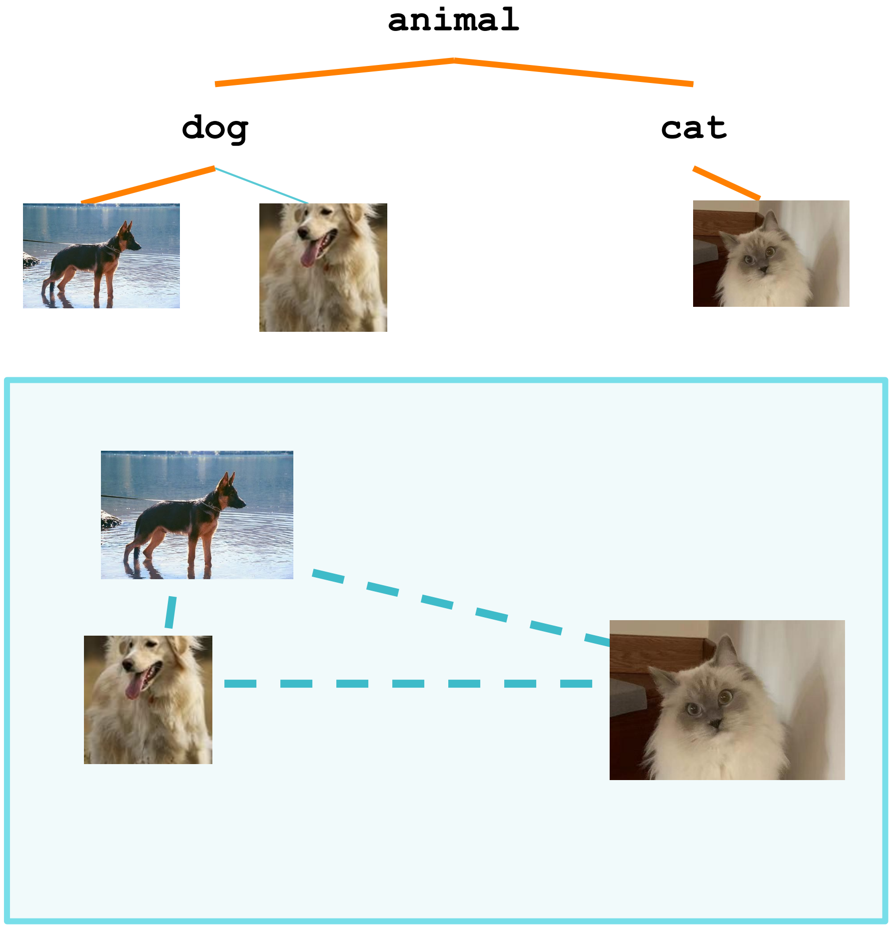


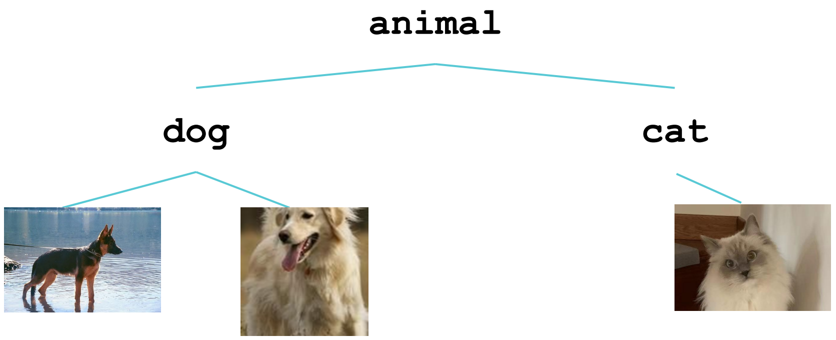


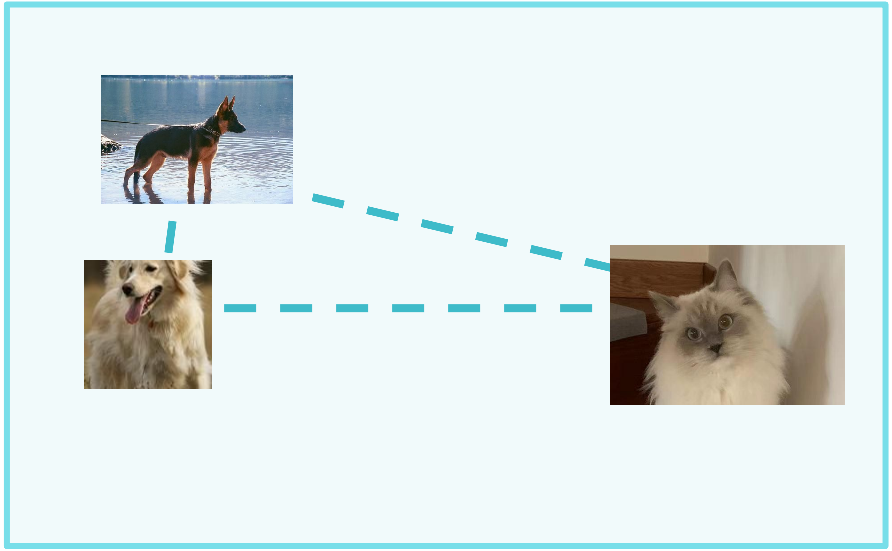
Permutation Invariant Representation

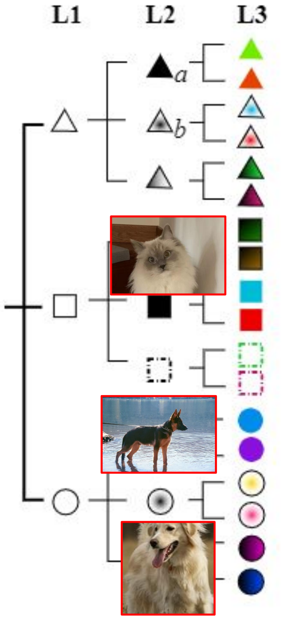



animal
dog
cat

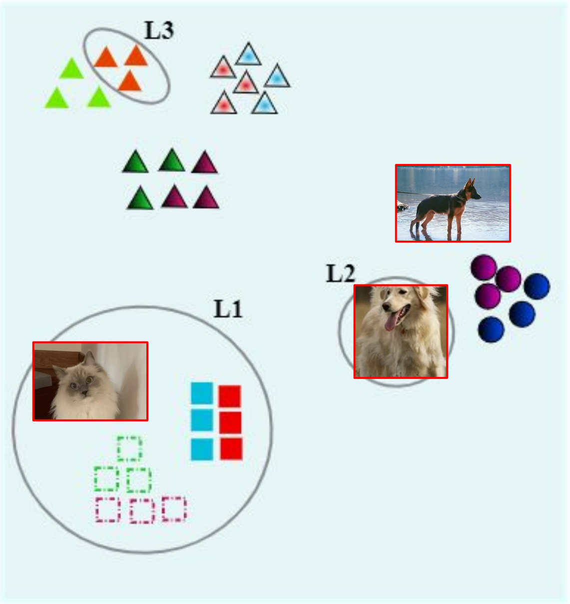

animal
dog
cat



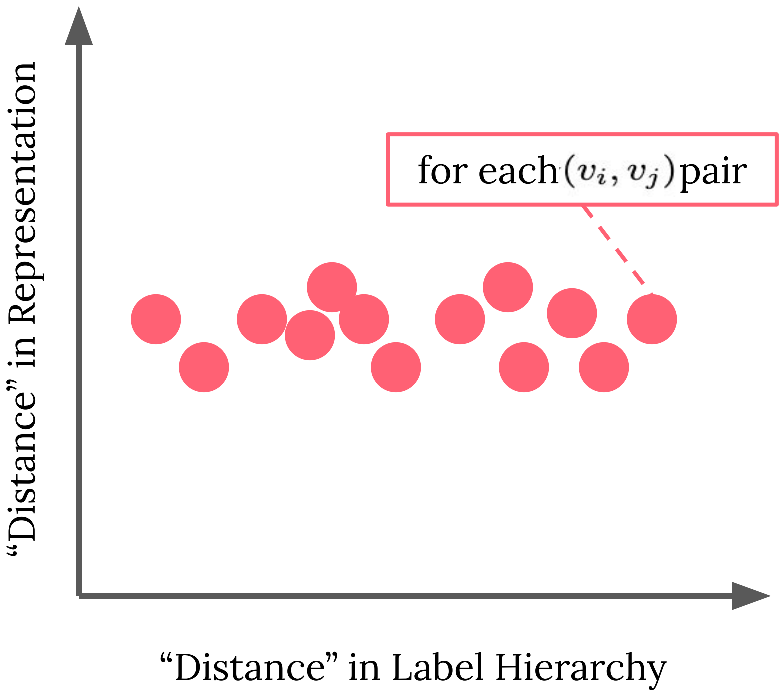




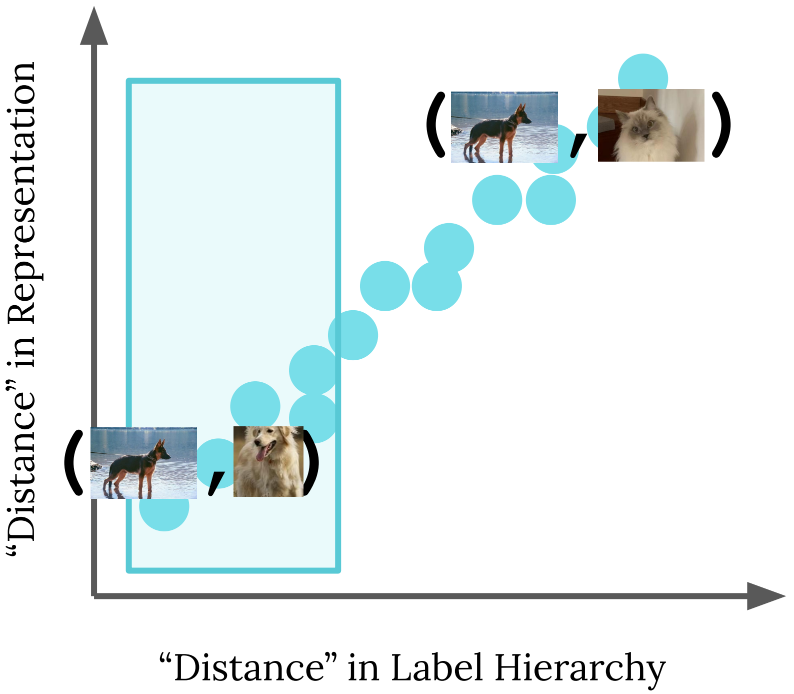
Structured Representation

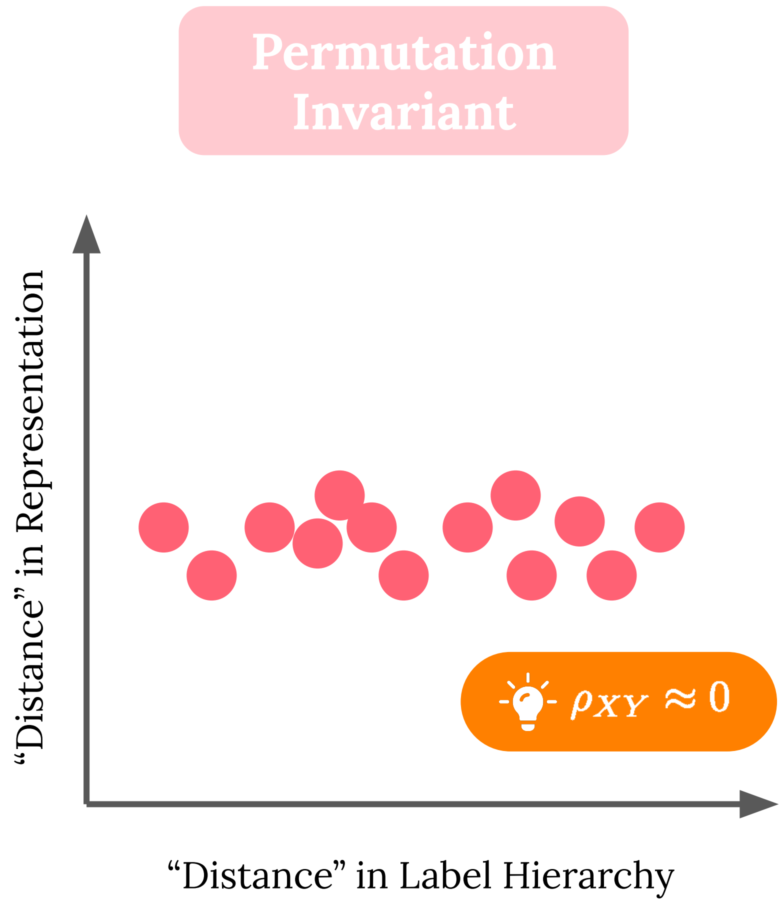


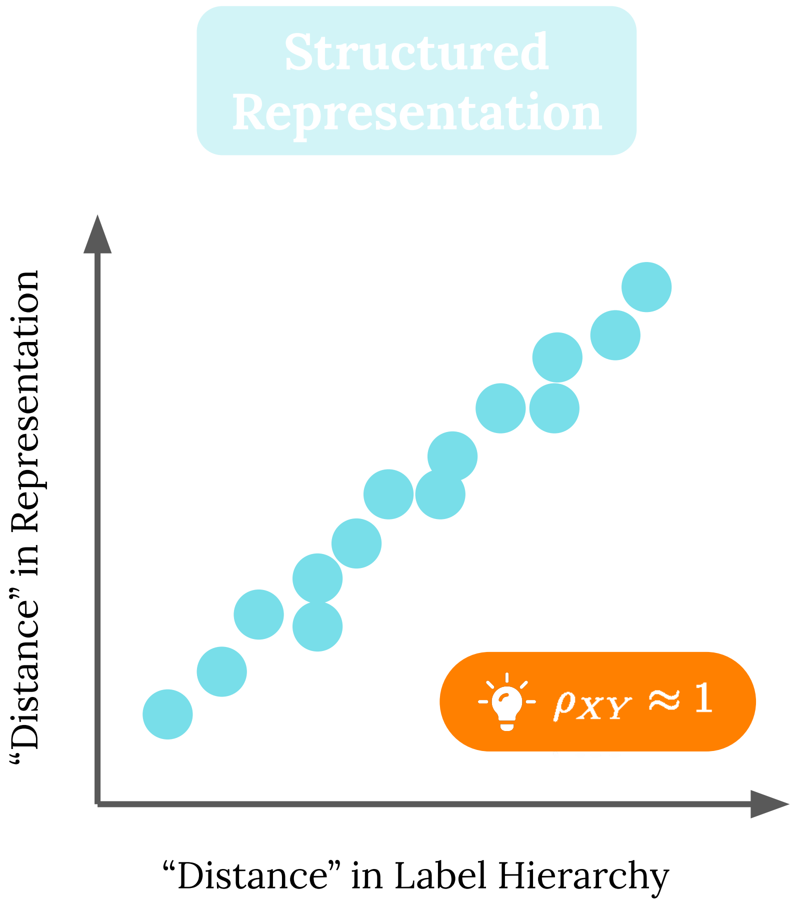
Permutation Invariant



Structured Representation




Permutation Invariant
“Distance” in Representation
$\rho_{XY} \approx 0$
“Distance” in Label Hierarchy


Structured Representation
“Distance” in Representation
$\rho_{XY} \approx 1$
“Distance” in Label Hierarchy

Cophenetic Correlation Coefficient (**CPCC**) [Sokal & Rohlf (1962)]

$$\mathrm{CPCC}(d_{\mathcal{T}}, \rho) := \frac{\sum_{i<j}(d_{\mathcal{T}}(v_i, v_j) - \overline{d_{\mathcal{T}}})(\rho(v_i, v_j) - \overline{\rho})}{\sqrt{\sum_{i<j}(d_{\mathcal{T}}(v_i, v_j) - \overline{d_{\mathcal{T}}})^2}\sqrt{\sum_{i<j}(\rho(v_i, v_j) - \overline{\rho})^2}}$$

$\rightarrow \rho(v_i, v_j) :=$ The Euclidean distance between **two class centroids**, where vi and vj are classes.

$\rightarrow d_{\mathcal{T}}(v_i, v_j) :=$ The **shortest distance** between two vertices.

With CPCC as a regularizer...

$$\mathcal{L}(\mathcal{D}) = \sum_{(x,y)\in\mathcal{D}} \ell(y, \hat{y}) - \lambda \cdot \mathrm{CPCC}(d_{\mathcal{T}}, \rho)$$

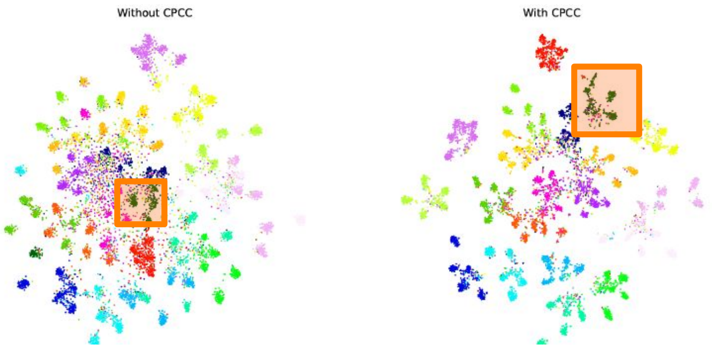
With CPCC as a regularizer...

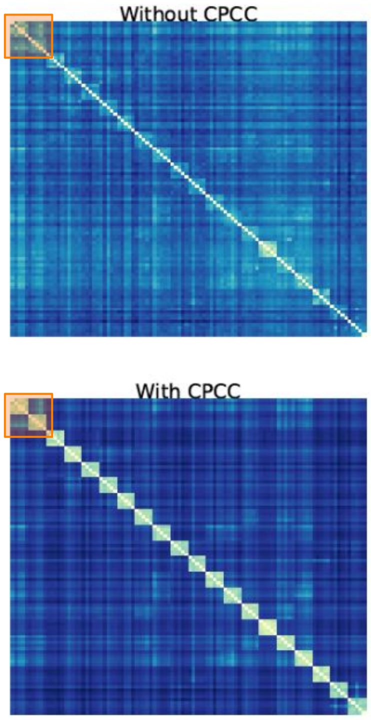
$$\mathcal{L}(\mathcal{D}) = \sum_{(x,y)\in\mathcal{D}} \ell(y, \hat{y}) - \lambda \cdot \mathrm{CPCC}(d_{\mathcal{T}}, \rho)$$

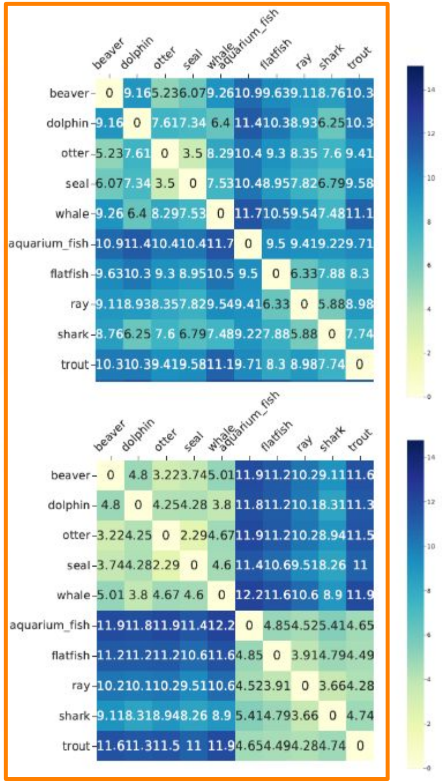
CPCC is flexible!

- ✓ Replace $\ell(y, \hat{y})$ with **any** flat/hierarchical loss functions
- ✓ Applicable to trees with **any** height
- ✓ Computationally Efficient

Structure of Learned Representations







Fine classes from the same coarse classes tend to be closer, and coarse classes tend to be further apart.

*Matrix values are Euclidean distance between two fine class centroids.

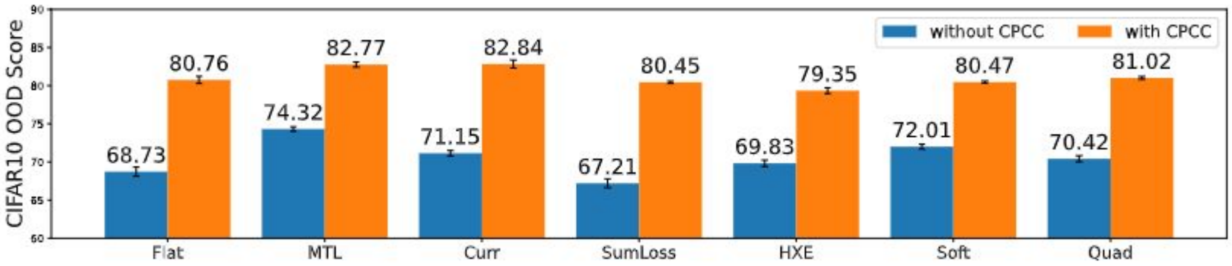
Why Structured Representation?

Better Generalization

… to **unseen classes** and **levels**

Dataset	Objective	CPCC	Silhouette	FineAcc	MidAcc	CoarseAcc	CoarserAcc
MNIST	Flat	10.80 (1.49)	13.97 (0.72)	99.05 (0.23)	**99.38 (0.04)**	99.49 (0.08)	N/A
	FlatCPCC	**99.96 (0.01)**	**61.33 (0.42)**	**99.28 (0.08)**	**99.38 (0.03)**	**99.61 (0.03)**	N/A
CIFAR100	Flat	24.38 (0.57)	5.59 (0.02)	76.82 (0.30)	80.27 (0.35)	85.59 (0.35)	86.85 (0.27)
	FlatCPCC	84.20 (0.39)	34.40 (0.11)	77.47 (0.27)	81.30 (0.14)	86.95 (0.17)	88.17 (0.17)
	MTL	39.75 (0.33)	8.09 (0.08)	76.56 (0.20)	80.17 (0.22)	85.79 (0.20)	87.11 (0.14)
	MTLCPCC	84.88 (0.58)	31.58 (0.23)	76.90 (0.32)	80.91 (0.29)	87.11 (0.19)	88.39 (0.19)
	Curr	23.81 (0.60)	5.25 (0.11)	76.84 (0.20)	80.40 (0.17)	85.72 (0.16)	87.02 (0.18)
	CurrCPCC	**85.32 (0.51)**	34.08 (0.23)	**77.48 (0.44)**	**81.42 (0.32)**	**87.15 (0.19)**	**88.44 (0.20)**
	SumLoss	29.85 (0.63)	4.93 (0.07)	76.78 (0.20)	80.47 (0.22)	85.88 (0.25)	87.11 (0.26)
	SumLossCPCC	84.78 (0.64)	31.16 (0.13)	77.26 (0.12)	81.17 (0.18)	86.99 (0.07)	88.26 (0.02)
	HXE	25.40 (0.68)	8.31 (0.05)	76.58 (0.27)	80.17 (0.24)	85.67 (0.15)	87.02 (0.16)
	HXECPCC	85.13 (0.22)	**35.84 (0.18)**	76.57 (0.33)	80.63 (0.24)	86.48 (0.20)	87.77 (0.20)
	Soft	55.95 (0.67)	14.48 (0.11)	76.82 (0.06)	80.41 (0.07)	85.84 (0.16)	87.16 (0.07)
	SoftCPCC	85.23 (0.24)	35.80 (0.16)	77.11 (0.16)	81.02 (0.13)	86.63 (0.17)	87.93 (0.14)
	Quad	25.08 (0.26)	6.75 (0.06)	76.40 (0.28)	80.05 (0.27)	85.30 (0.11)	86.67 (0.14)
	QuadCPCC	84.65 (0.32)	34.79 (0.23)	77.10 (0.16)	80.92 (0.12)	86.78 (0.09)	88.04 (0.09)

OOD Detection



Takeaway

✓ CPCC successfully creates a structured representation
✓ CPCC is flexible and lightweight
✓ CPCC leads to better generalization in some scenarios, and can be applied to even more settings (subpopulation shift, OOD detection, ...)

Paper: https://openreview.net/forum?id=7J-30ilaUZM

Code: https://github.com/hanzhaoml/HierarchyCPCC

Contact: siqiz@andrew.cmu.edu

Thanks for listening!