What Do You See in Common?

Learning Hierarchical Prototypes over Tree-of-Life to Discover Evolutionary Traits

arXiv

Harish Babu Manogaran[1], M.Maruf[1], Arka Daw[6], Kazi Sajeed Mehrab[1], Caleb Charpentier[1], Josef Uyeda[1], Wasila M Dahdul[2], Matthew J Thompson[5], Elizabeth G Campolongo[5], Kaiya L Provost[5], Paula Mabee[3], Hilmar Lapp[4], Anuj Karpatne[1]

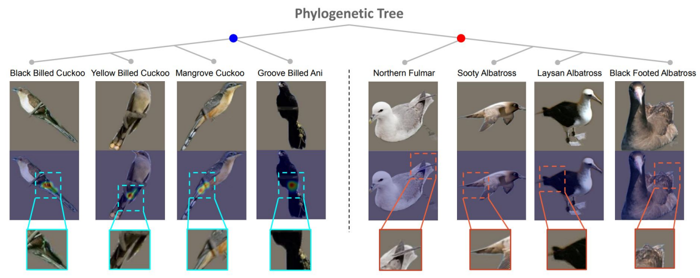
[1]Virginia Tech, [2]UC Irvine, [3]Battele, [4]Duke University, [5]Ohio State University, [6]Oak Ridge National Laboratory

Abstract

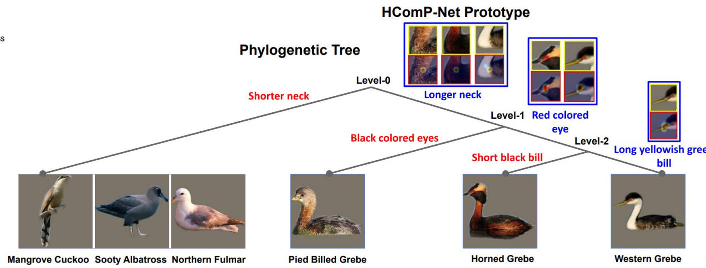
We introduce the framework of **H**ierarchy aligned **Com**monality through **P**rototypical **Net**works (**HComP-Net**)

HComP-Net learns hierarchical prototypes with a known hierarchy such as phylogeny

The prototypes represent traits shared by all the leaf descendants with common ancestor that are not present in other species



What do you see in common between descendant species derived from same ancestor node, that is not present in other species that diverged and evolved differently? (Hypothesized trait: long tail)



Problem Statement

Existing Hierarchical Prototypical network (HPnet) [2] based on ProtoPNet [3] is prone to **over-specificity** and **lack of semantic correspondence**

At internal nodes, HPnet can learn prototypes that represent traits of one or few species rather than what is common to all leading to **over-specificity**

A learned prototype can represent different visual concept in different images leading to **lack of semantic correspondence** making interpretations unreliable.

Proposed Solution

HComP-Net learns common traits shared by all descendant species of an internal node and avoids the learning of over-specific prototypes using a novel **over-specificity loss**

We further introduce a **masking module** to identify and ignore over-specific prototypes

We also adopt contrastive learning approach from PIPNet [1] to improve semantic correspondence and make interpretations reliable

Methodology

Contrastive learning is achieved through alignment and tanh losses as introduced in PIPNet

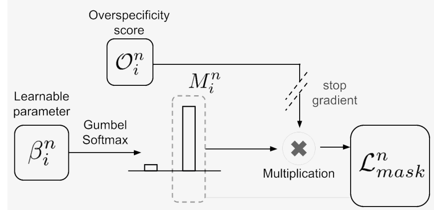
$$\mathcal{L}_A^n = -\frac{1}{HW}\sum_{(h,w)\in H\times W}\log(\hat{\mathbf{z}}'^n_{h,w}\cdot\hat{\mathbf{z}}''^n_{h,w})$$

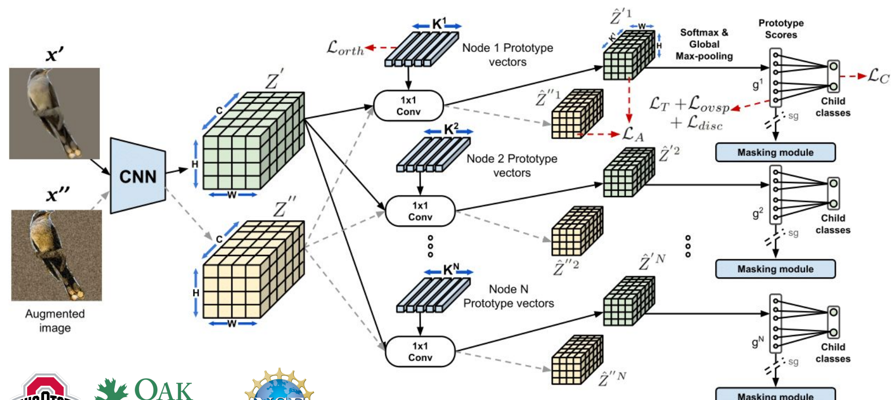
$$\mathcal{L}_T^n = -\frac{1}{K^n}\sum_{i=1}^{K^n}\log(\tanh(\sum_{b=1}^{B} g_{b,i}^n))$$

Over-specificity loss is formulated as follows

$$\mathcal{L}_{ovsp}^n = -\frac{1}{K^n}\sum_{i=1}^{K^n}\sum_{d=1}^{D_i}\log(\tanh(\sum_{b\in B_d} g_{b,i}^n))$$

Masking module is used to identify the prototypes that do not represent a shared trait. Masking module is attached to every prototype. The learned mask indicates whether the prototype is over-specific or not





Results

Classification performance

Model	Hierarchy	% Fine-grained accuracy: Bird	Fish
ResNet-50	No	74.18	86.63
INTR		69.22	86.73
HPnet	Yes	36.18	77.51
HComP-Net		70.01	90.80

Part purity: Prototypes that represent the same part (eye, beak, tail, etc.) in Top-10 closest images have part purity close to 1

Model	Part purity
HPnet	0.14 ± 0.09
HComP-Net w/o $\mathcal{L}_{ovsp}$	0.68 ± 0.22
HComP-Net w/o $\mathcal{L}_{ovsp}$ with mask applied	0.75 ± 0.17
HComP-Net with $\mathcal{L}_{ovsp}$	0.72 ± 0.19
HComP-Net with $\mathcal{L}_{ovsp}$ with mask applied	**0.77** ± **0.16**

References

[1] Nauta, M., et al.: Neural prototype trees for interpretable fine-grained image recognition. In: CVPR. pp. 14933–14943 (2021)
[2] Peter Hase, Chaofan Chen, Oscar Li, and Cynthia Rudin. Interpretable image recognition with hierarchical prototypes.In AAAI, pages 32–40, 2019.
[3] Chen, Chaofan, et al. "This looks like that: deep learning for interpretable image recognition." Advances in neural information processing systems 32 (2019).

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. 2118240.