Modeling Semantic Correlation and Hierarchy for Real-world Wildlife Recognition

Dong-Jin Kim[1], Zhongqi Miao[2], Yunhui Guo[3], Stella X. Yu[2,4], Kyle Landolt[5], Mark Koneff[6], Travis Harrison[5]

[1]Hanyang University [2]UC Berkeley / ICSI [3]University of Texas at Dallas [4]University of Michigan [5]United States Geological Survey [6]US Fish and Wildlife Service

Abstract

We explore the challenges of human-in-the-loop frameworks to label wildlife recognition datasets with a neural network. In wildlife imagery, the main challenges for a model to assist human annotation are two-fold: (1) the training dataset is usually imbalanced, which makes the model's suggestion biased, and (2) there are complex taxonomies in the classes. We establish a simple and efficient baseline, including the debiasing loss function and the hyperbolic network architecture, to address these issues. Moreover, we propose leveraging the semantic correlation to train the model more effectively by adding a co-occurrence layer to our model during training. We demonstrate the efficacy of our method in both our real-world wildlife areal survey recognition dataset and the public image classification dataset, CIFAR100-LT and CIFAR10-LT.

	Test accuracy (%)					
Species	**Vanilla**	**+LDAM**	**+LA**	**+LA+H**	**+LA+Corr**	**+LA+H+Corr (Ours)**
Common Eider	**99.1**	93.8	92.0	86.8	91.5	91.8
Black Scoter	95.4	92.6	97.2	87.0	97.2	**98.0**
White-winged Scoter	38.1	71.4	85.7	61.9	85.7	**85.9**
Long-tailed Duck	0.0	100.0	**100.0**	**100.0**	**100.0**	**100.0**
Non-target Species	39.5	68.4	78.9	**84.1**	81.6	81.9
Average accuracy (%)	54.1	85.3	90.8	82.9	91.2	**91.5**

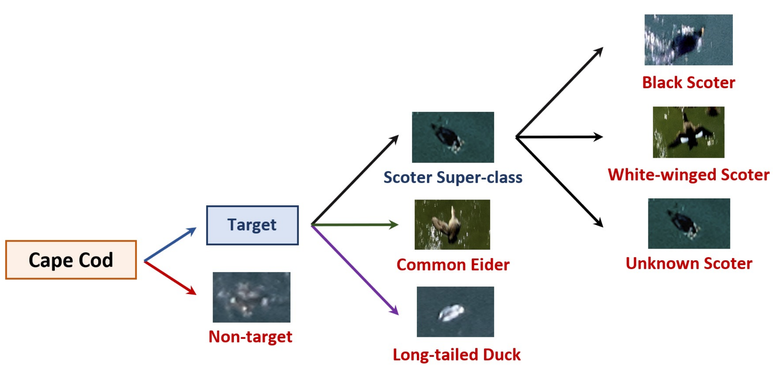
Our model with logit adjustment loss (+LA), hyperbolic module (+H), and our semantic co-occurrence method (+Corr) shows the best performance.

Introduction

- Processing **areal remote sensing** datasets is expensive.
- **Human-in-the-loop** frameworks to label **wildlife** datasets.

Main challenges:

1. Class **imbalance**.→ Debiasing Loss Function
2. Complex **taxonomies**.→ Hyperbolic Module & Co-occurrence



The 6 classes in our dataset have a **hierarchical** relationship [1].

Modeling Semantic Correlation and Hierarchy

Our model consists of 3 components.

1. Debiasing Loss Function: Logit Adjustment [1]

$$\mathcal{L}_{LA} = -\log \frac{\exp(f_y(x) + \tau \log \pi_y)}{\sum_{j\in[L]} \exp(f_j(x) + \tau \log \pi_j)}$$

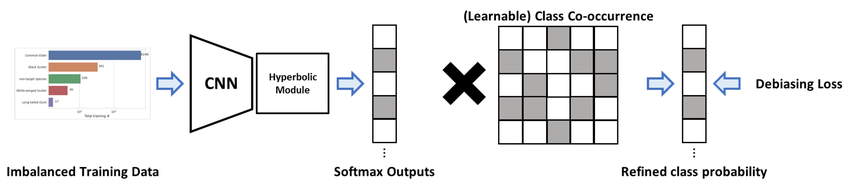
2. Hyperbolic Network Architecture [3]

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

3. New Method: Semantic Correlations [4,5]

- A **co-occurrence matrix** to refine the class probability:

$$P(Y = i|X) = \sum_{j\in\mathcal{Y}} p(Y = i|Y' = j) * p(Y' = j|X).$$



Our framework including a **hyperbolic module**, debiasing **loss function**, and our **co-occurrence layer**.

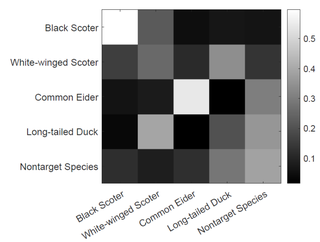
Experiments

Species	**Train #**	**Test #**
Common Eider	6,246	3,172
Unknown Scoter	466	114
Black Scoter	341	108
White-winged Scoter	45	21
Long-tailed Duck	17	5
Non-target Species	108	38

The number of samples in our dataset.

Methods	CIFAR100-LT	CIFAR10-LT
Vanilla	38.3	70.4
LDAM [1]	42.0	77.0
LA [16]	43.9	77.7
Ours	**45.3**	**79.0**

Accuracy on CIFAR10-LT and CIFAR100-LT dataset.



The visualization of the co-occurrence matrix.

Acknowledgement

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

References

[1] Zhongqi Miao et al., Challenges and solutions for automated avian recognition in aerial imagery. Remote Sensing in Ecology and Conservation 2022.

[2] Aditya Krishna Menon et al., Long-tail learning via logit adjustment. ICLR 2021.

[3] Yunhui Guo et al., Clipped hyperbolic classifiers are super-hyperbolic classifiers. CVPR 2022.

[4] Dong-Jin Kim et al., Detecting human-object interactions with action co-occurrence priors. ECCV 2022.

[5] Dong-Jin Kim et al., Acp++: Action co-occurrence priors for human-object interaction detection. IEEE TIP 2021