Bringing Back the Context: Camera Trap Species Identification as Link Prediction on Multimodal Knowledge Graphs

THE OHIO STATE UNIVERSITY

Vardaan Pahuja, Weidi Luo, Yu Gu, Cheng-Hao Tu, Hong-You Chen,
Tanya Berger-Wolf, Charles Stewart, Song Gao, Wei-Lun Chao, Yu Su

ICICLE DEMOCRATIZING AI

Problem Setup

- Species recognition for camera trap images amidst distribution shifts.
- The training and test sets comprise images obtained from disjoint camera traps, enabling the evaluation of out-of-domain generalization.
- During training, we use the multimodal KG to train our model, while we use just the image to make predictions for inference.

Multimodal KG

- The base KG consists of camera trap images linked with their species labels from the training set.
- The **taxonomy** edges connect distinct species to higher-order taxa.
- The **location** edges connect species images to the GPS coordinates of their source cameras.
- The **temporal** edges connect species images to the timestamp of image capture.

Training

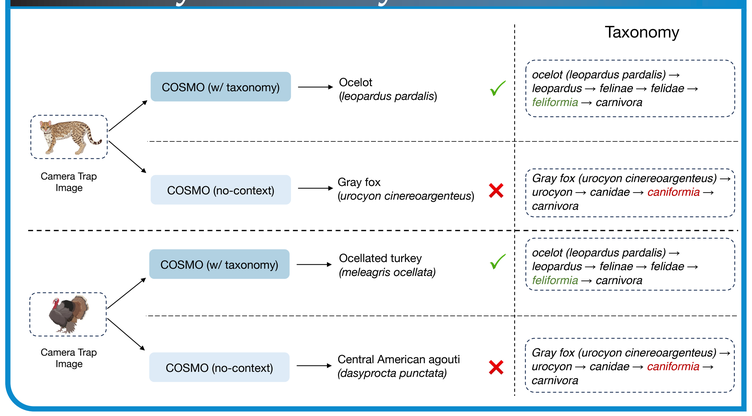
- **Categorical attributes** such as species ID and taxon ID:

$$\mathcal{L}(\mathcal{I}, \texttt{instance of}, s) = -\log \frac{\exp(\psi(\mathcal{I}, \texttt{instance of}, s))}{\sum_{s' \in S} \exp(\psi(\mathcal{I}, \texttt{instance of}, s'))}$$

- **Numerical attributes** such as location and time:

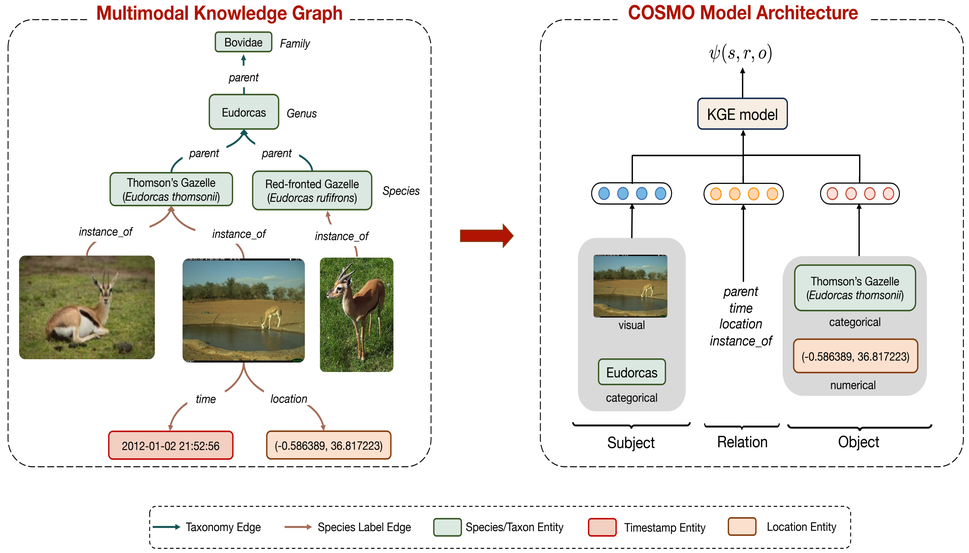
$$\mathcal{L}(\mathcal{I}, \texttt{time}, t) = -\sum_{t'} l_{t'}^{\mathcal{I},time} \cdot \log(\sigma(\psi(\mathcal{I}, \texttt{time}, t'))) + (1 - l_{t'}^{\mathcal{I},time}) \cdot (1 - \log(\sigma(\psi(\mathcal{I}, \texttt{time}, t'))))$$

Taxonomy Case Study



Model Architecture

- **DistMult** as backbone KG embedding model.
- **ResNet-50** pre-trained on ImageNet as image encoder.
- **MLP** for numerical metadata such as location and time.



Experiments

Model	Multi-modality: Taxonomy	Multi-modality: Time	Test Acc.
ERM	–		96.2 (0.6)
CORAL			96.6 (1.2)
Group DRO			93.4 (2.1)
ABSGD			93.4 (2.0)
COSMO (no-context)	–		92.9 (2.5)
COSMO	✓		93.9 (2.8) (+1.0)
		✓	95.3 (3.1) (+2.4)
	✓	✓	**96.8 (0.4)** (+3.9)

Table 1: Species Classification results on Snapshot Mountain Zebra dataset.

	Model	Multi-modality: Taxonomy	Multi-modality: Location	Multi-modality: Time	Val. Acc.	Test Acc.
	Empirical Risk Minimization (ERM)				62.7 (2.4)	71.6 (2.5)
	CORAL				60.3 (2.8)	73.3 (4.3)
	Group DRO		–		60.0 (0.7)	72.7 (2.0)
	Fish				58.0 (0.2)	63.2 (0.7)
	ABSGD				–	72.7 (1.8)
	COSMO (no-context)		–		63.2 (0.4)	68.8 (2.1)
Single context	COSMO	✓			62.8 (2.2) (-0.4)	72.4 (2.5) (+3.6)
			✓		64.4 (1.0) (+1.2)	**74.5 (3.6)** (+5.7)
				✓	64.7 (0.4) (+1.5)	71.1 (3.1) (+2.3)
Multiple contexts	COSMO	✓	✓		**65.4 (0.4)** (+2.2)	70.4 (2.1) (+1.6)
		✓		✓	64.9 (1.6) (+1.7)	73.7 (3.8) (+4.9)
			✓	✓	63.0 (2.1) (-0.2)	74.2 (2.2) (+5.4)
		✓	✓	✓	65.0 (1.6) (+1.8)	71.5 (2.8) (+2.7)

Table 2: Species Classification results on iWildCam2020-WILDS (OOD) dataset.

Spatiotemporal Correlation Analysis

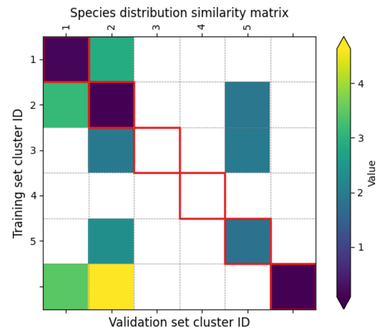


Figure 1: Each color square shows the distance between the corresponding validation cluster centroid on x-axis and the training cluster centroid on y-axis. The correlation peaks along the diagonal.

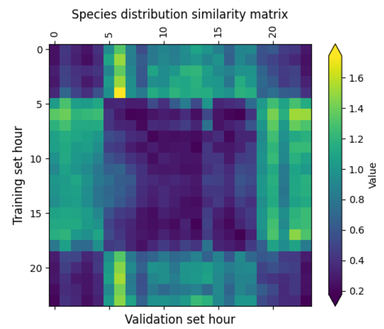


Figure 2: Each color square shows the distance between the corresponding training hour slot on x-axis and validation hour slot on y-axis. The correlation peaks for day-day and night-night hour slots.

Case Study on Under-represented Species

Model	Accuracy
ERM (ResNet-50)	16.3
COSMO	**19.0** (+2.7)

Table 3: Performance comparison on under-represented species classification (OOD Test set). The best performing COSMO model improves over the ERM baseline by a significant margin.

Contact Information

Paper

Code

Author Contact: pahuja.9@osu.edu