

Reviving the Context: Camera Trap Species Classification as Link Prediction on Multimodal Knowledge Graphs

Vardaan Pahuja¹, Weidi Luo¹, Yu Gu¹, Cheng-Hao Tu¹, Hong-You Chen¹,
Tanya Berger-Wolf¹, Charles Stewart², Song Gao³, Wei-Lun Chao¹, Yu Su¹

¹The Ohio State University, ²Rensselaer Polytechnic Institute,

³University of Wisconsin-Madison

Camera Trap Wildlife Recognition Challenges



Illumination: Animals are not always salient

Camera Trap Wildlife Recognition Challenges



Motion blur: common with poor illumination at night

Camera Trap Wildlife Recognition Challenges



Size of the region of interest (ROI): Animals can be small or far from the camera

Camera Trap Wildlife Recognition Challenges



Occlusion: e.g. by bushes or rocks

Camera Trap Wildlife Recognition Challenges



Camouflage: decreases saliency
in animals' natural habitat

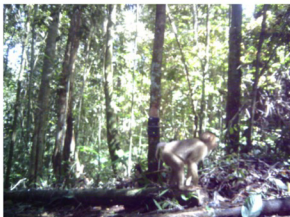
Camera Trap Wildlife Recognition Challenges



Perspective: Animals can be close to the camera, resulting in partial views of the body.

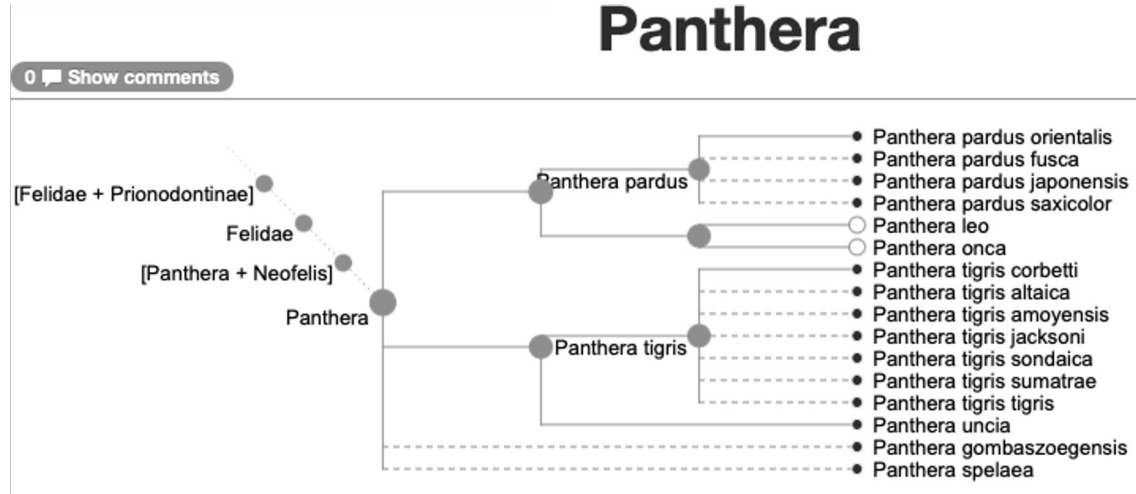


Camera Trap Wildlife Recognition Challenges

Train			Test (OOD)
$d = \text{Location 1}$	$d = \text{Location 2}$	$d = \text{Location 245}$	$d = \text{Location 246}$
			
Vulturine Guineafowl	African Bush Elephant	...	Wild Horse
			
Cow	Cow	Southern Pig-Tailed Macaque	Great Curassow

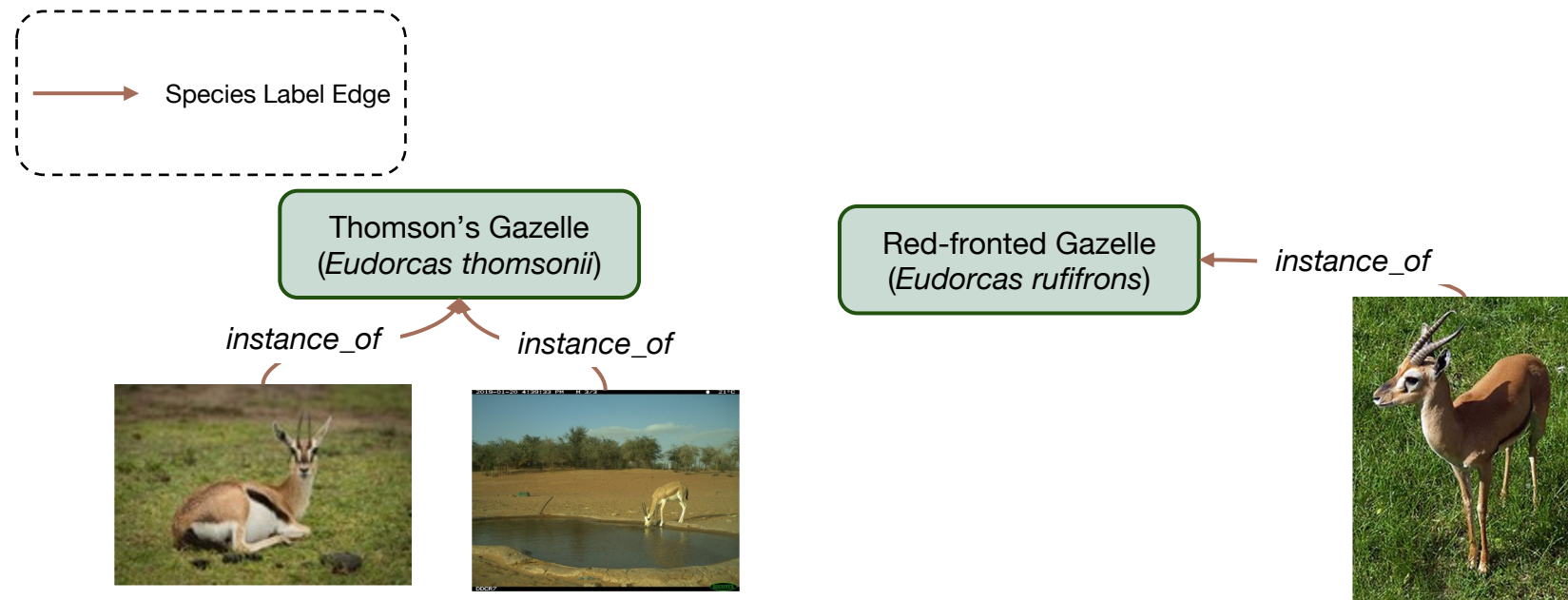
Wildlife Taxonomy

- Phylogenetic taxonomy from **Open Tree of Life (OTT)**



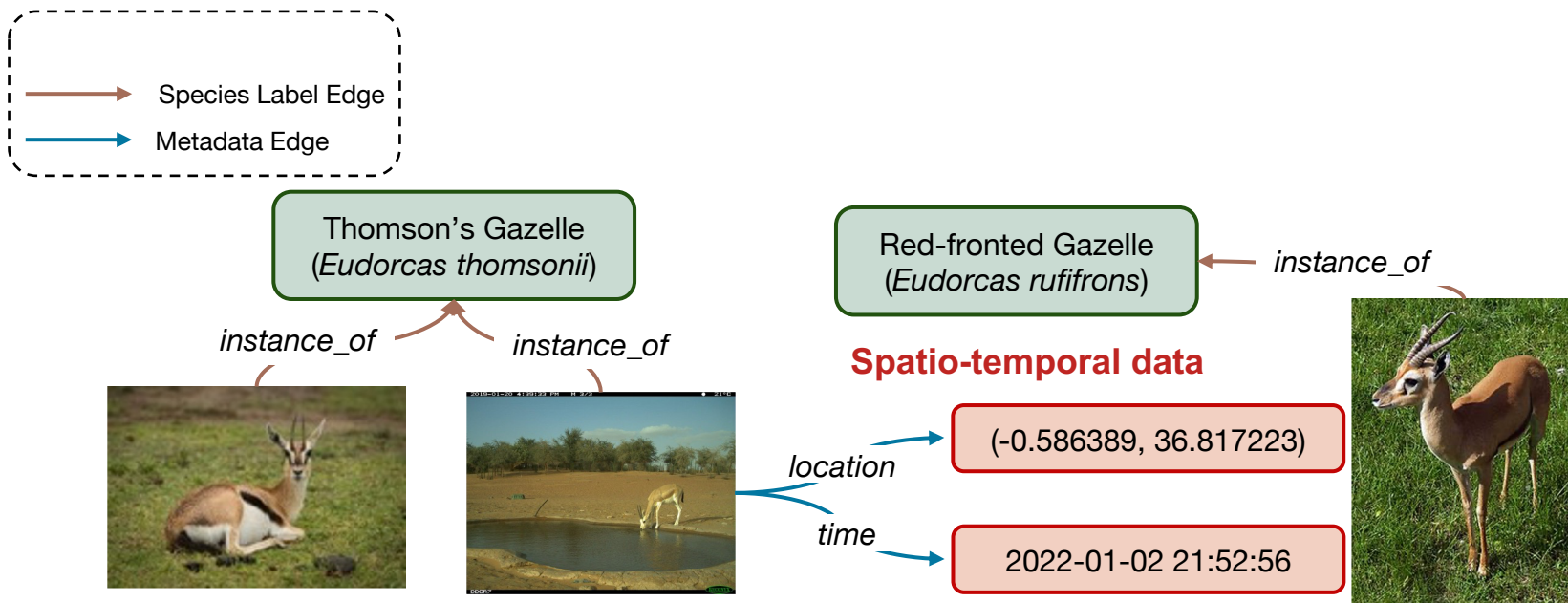
Multimodal KG

- Structured representation for semantic relationships b/w wildlife species, images, attributes and taxonomy.



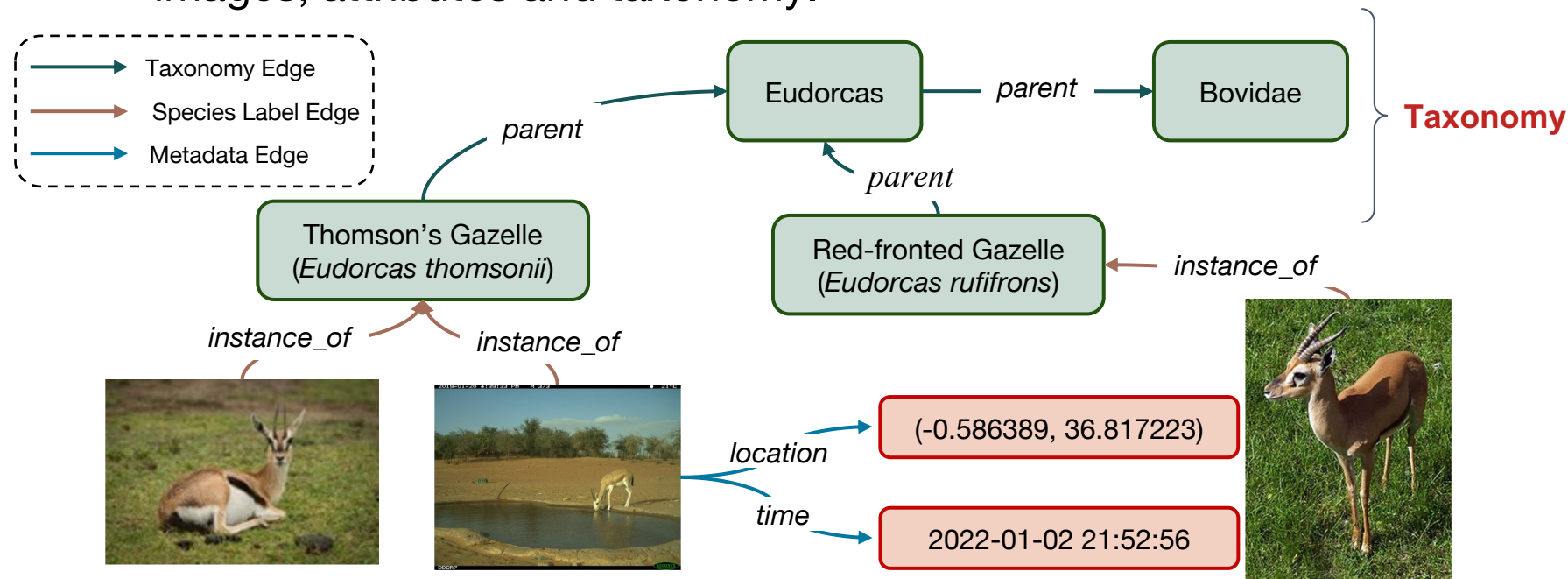
Multimodal KG

- Structured representation for semantic relationships b/w wildlife species, images, attributes and taxonomy.



Multimodal KG

- Structured representation for semantic relationships b/w wildlife species, images, attributes and taxonomy.



Multimodal KG

- Entity types:
 - **Categorical** (e.g. Species labels, taxons)
 - Encoder: Entity lookup table

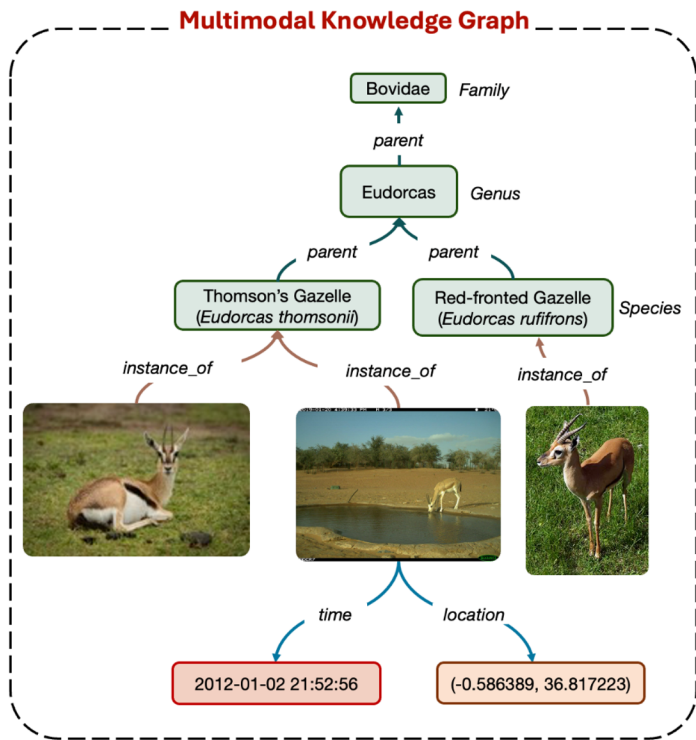
Multimodal KG

- Entity types:
 - Categorical (e.g. Species labels, taxons)
 - Encoder: Entity lookup table
 - Visual
 - Encoder: CNN/Vision Transformer

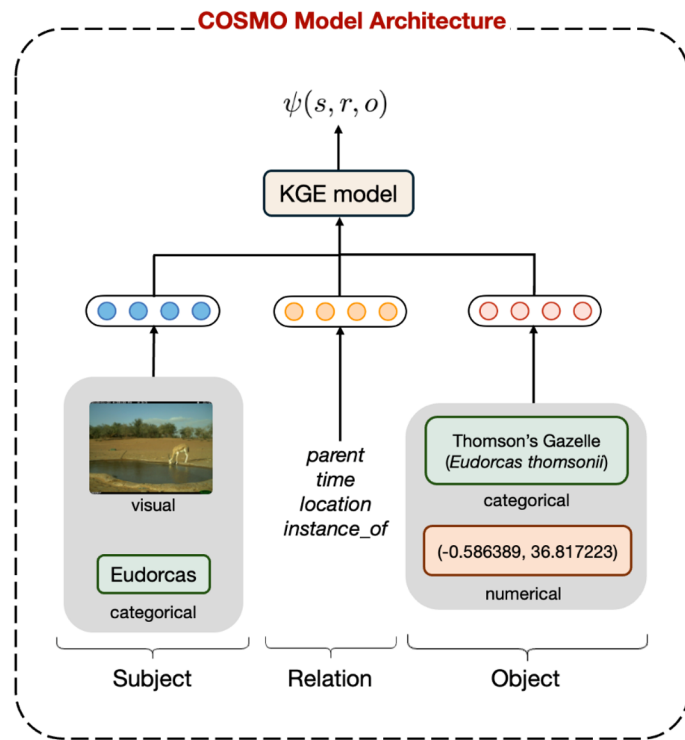
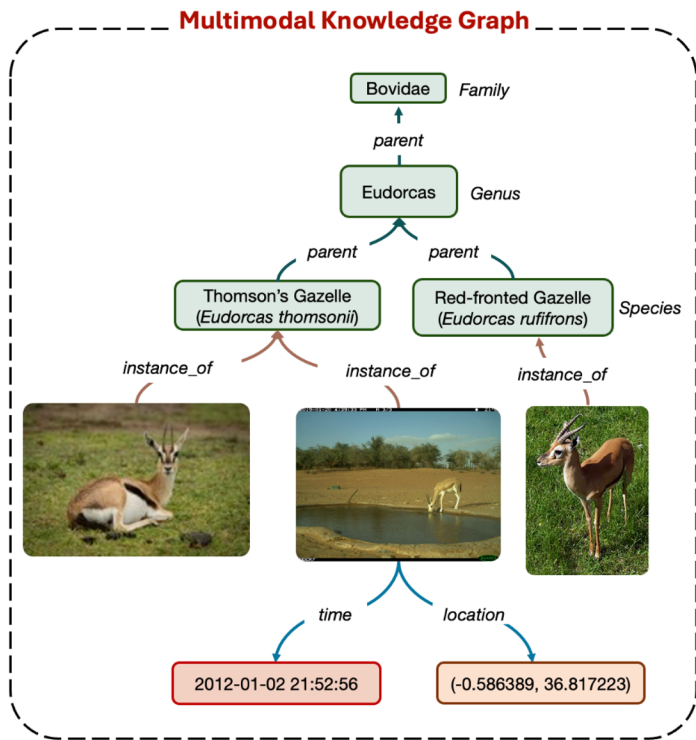
Multimodal KG

- Entity types:
 - Categorical (e.g. Species labels, taxons)
 - Encoder: Entity lookup table
 - Visual
 - Encoder: CNN/Vision Transformer
 - Numerical (e.g. location, time)
 - Encoder: MLP

Multimodal KG Embedding Model



Multimodal KG Embedding Model

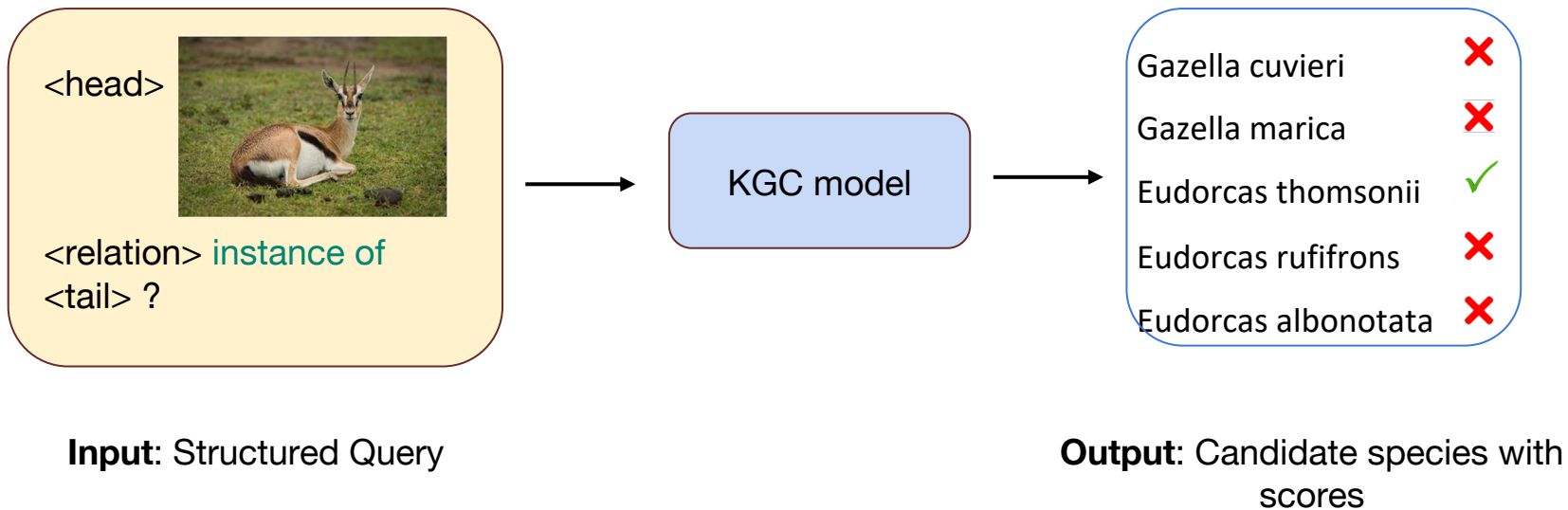


Task: Species Classification

- Task: Given a camera trap image, predict the species label.

Task: Species Classification

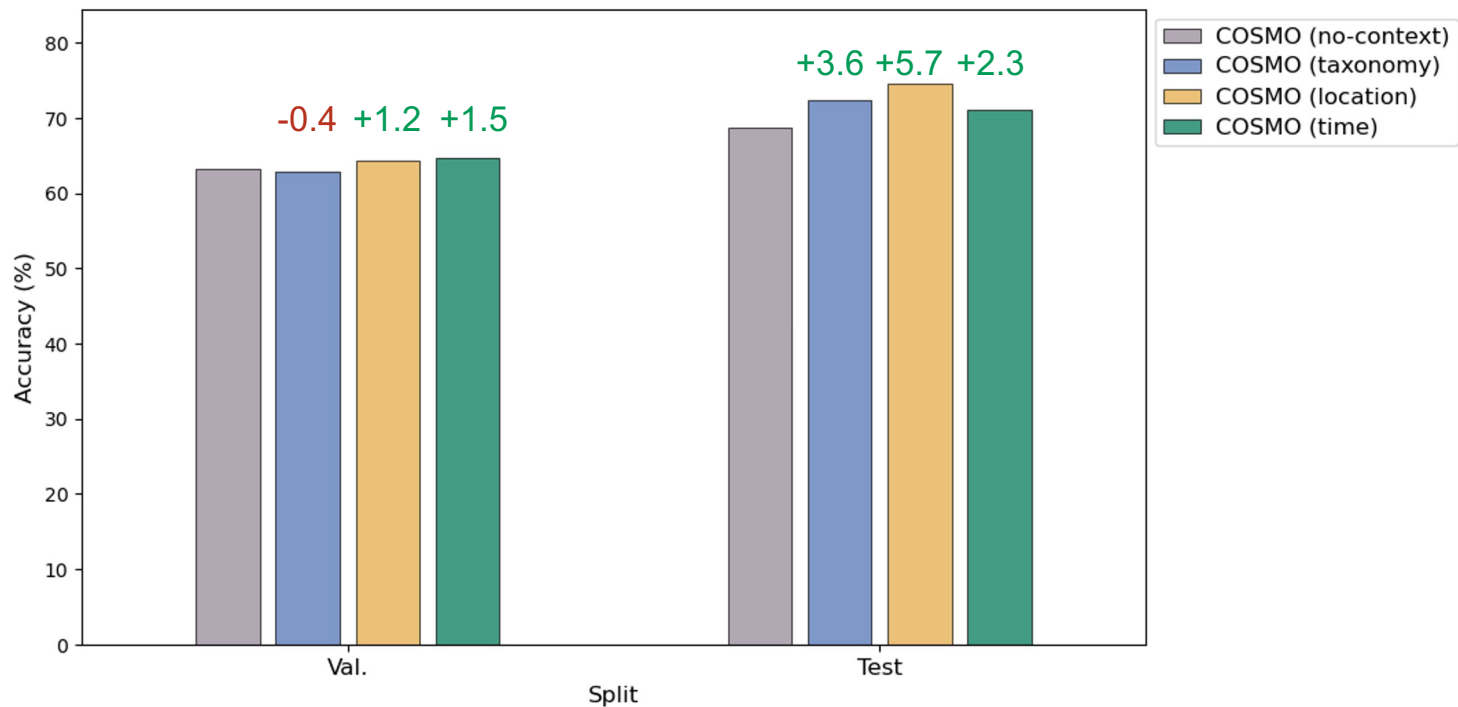
- Task: Given a **camera trap image**, predict the **species label**.



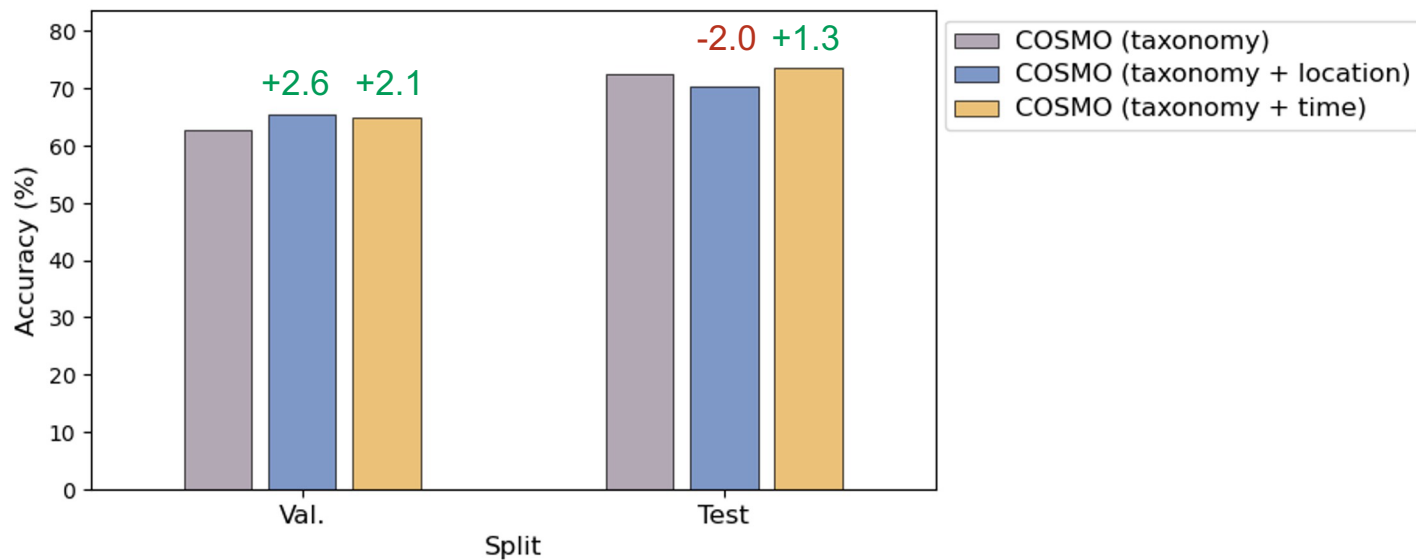
Training

- **Categorical attributes:** Multi-class classification problem
- **Numerical attributes:** Multi-class multi-label classification problem.
- Sequentially train on each set of triples types.

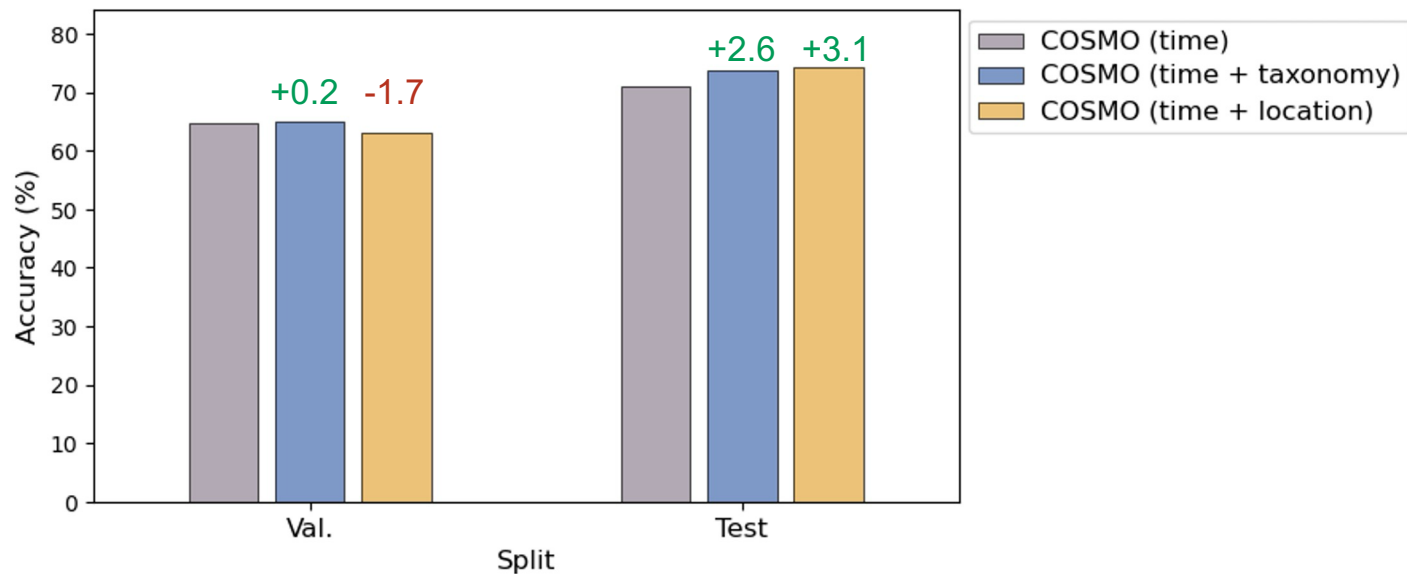
Results (iWildCam2020-WILDS)



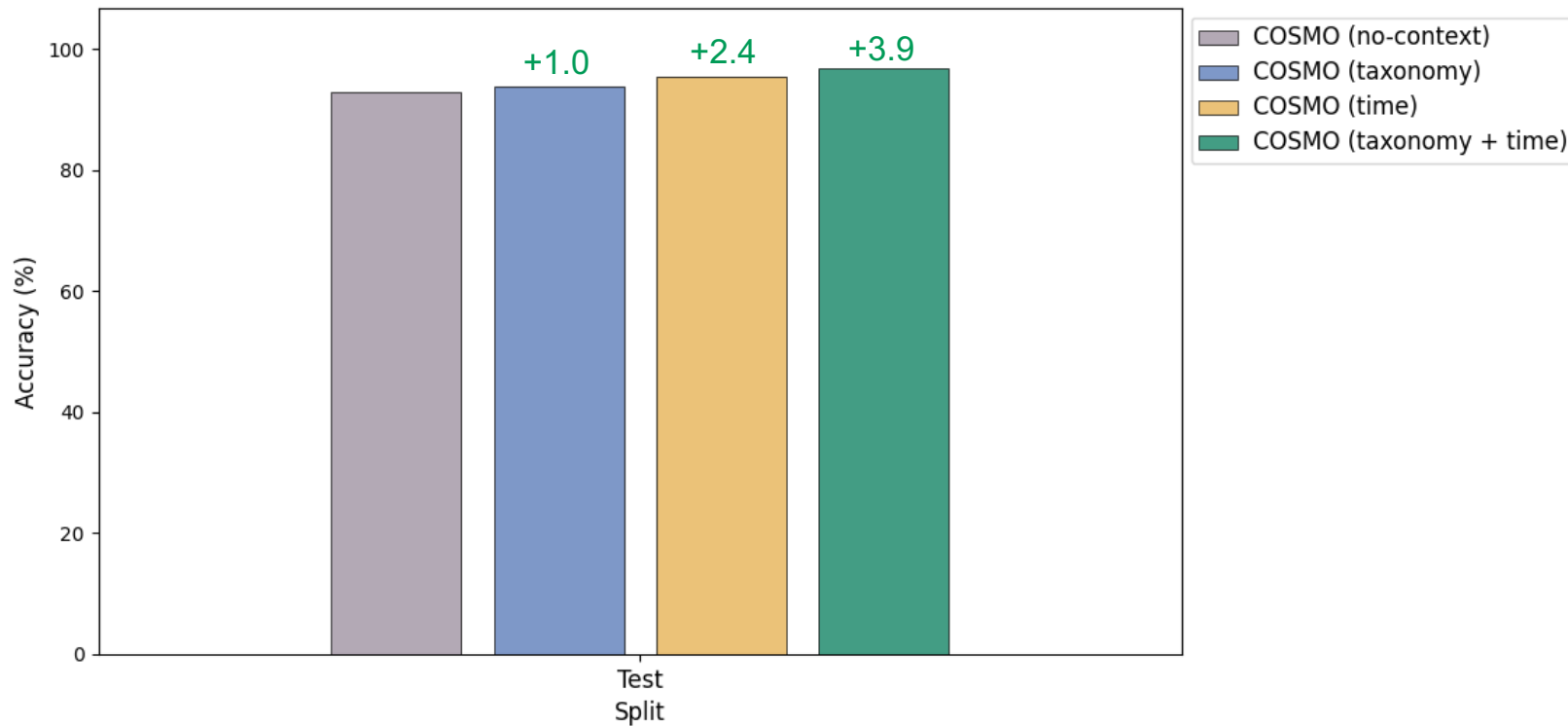
Results (Multiple contexts)



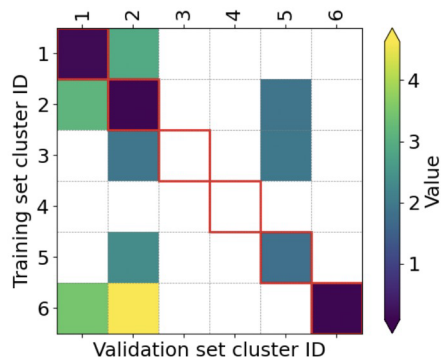
Results (Multiple contexts)



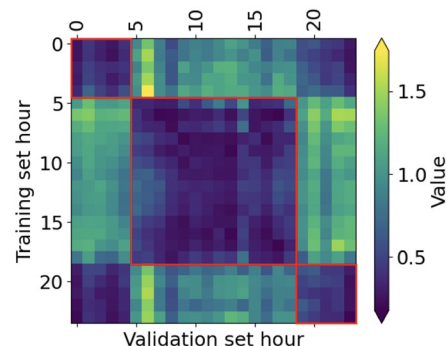
Results (Snapshot Mountain Zebra)



Metadata Analysis



(a) 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 (highlighted in red)³.



(b) 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 (highlighted in red).

- Similarity peaks along the diagonal for the location attribute, as well as for the day/night categorization of the time attribute.
- These metadata give a prior for species class distribution

Key Takeaways

- Unified way to leverage various forms of multimodal context for species classification in camera traps.
- Our framework achieves superior out-of-distribution generalization and competitive performance with state-of-the-art.

Reviving the Context: Camera Trap Species Classification as Link Prediction on Multimodal Knowledge Graphs

Contact: pahuja.9@osu.edu

Code: <https://github.com/OSU-NLP-Group/COSMO>



Thank You!