EcoTrack: Few-Shot Species Re-Identification for Automatic Wildlife Census in Vietnam

COMP SCI 3315 – Computer Vision

Assignment 3 Report

Manh Ha Nguyen - a1840406

The University of Adelaide

1. Problem definition and application scenarios

The chosen computer-vision problem is the combined task of animal-species detection and individual re-identification in camera-trap imagery. Vietnam's protected forests still harbour highly threatened mammals such as the Saola, Indochinese tiger and Asian elephant, but rugged terrain and limited conservation budgets make collaring or manual surveys impractical. Passive infrared camera traps can capture rich behavioural data, yet parks often accumulate millions of unlabelled photographs such as Idaho Department of Fish and Game database (Beery et al., 2019).

Solving the coupled detection + re-ID problem unlocks several practical scenarios. First, conservationists can compute reliable population baselines from periodic camera deployments instead of speculative guesswork. Second, migration-corridor analysis becomes feasible: if the same tiger is detected by cameras placed kilometres apart, the corridor is confirmed and can be prioritised for protection. Third, crop-raiding alerts are possible: a camera near farmland can warn rangers when a known elephant approaches. Finally, the same technology can be repurposed for Southeast Asian neighbours that face similar biodiversity threats, making the solution regionally scalable.

2. Proposed solution

The proposed system, EcoTrack, follows a two-stage pipeline grounded in recent advances in vision transformers and few-shot metric learning:

1. Species detector. A lightweight transformer backbone built on the MetaFormer design converts each frame into fixed-size visual tokens. A DETR-style prediction head outputs bounding boxes and coarse species labels. Transformers are selected because their global receptive field helps suppress dense foliage and faint night-vision artefacts (Yu et al., 2022).
2. Few-shot re-identification head. Each detected crop is passed to an embedding network trained episodically with prototypical loss. During training, the network learns to cluster images of the same individual together while pushing different individuals apart (Snell et al., 2017). At inference time, EcoTrack maintains a small gallery of prototype vectors. A new detection is assigned the identity of its closest gallery vector unless the cosine distance exceeds a threshold, in which case a new identity is created. This design lets the system "grow" as unfamiliar animals appear, an ability that is critical in open-set wildlife monitoring.

Prior to fine-tuning, both modules are self-supervised on unlabelled camera-trap footage, leveraging modern contrastive objectives to learn habitat-specific cues without manual annotation. Only a modest set of expert-labelled crops is then required to adapt the model to Vietnamese fauna. Similar transfer strategies have already demonstrated large accuracy gains on African and South American trap datasets (Chen et al., 2023).

3. Advantages and Limitations

3.1. Advantages

EcoTrack is non-invasive: no collars or tranquilisation are needed, minimising stress to animals and cost to field teams. The few-shot framework means conservationists can start with only a handful of photographs per known individual and still obtain useful matches. The transformer backbone, once quantised, can also run on inexpensive edge devices, allowing preliminary filtering in remote ranger stations before data are synchronised to the cloud.

3.2. Limitations

Infrared motion blur and partial occlusion reduce detector recall, especially for small nocturnal species. Look-alike ungulates (such as sambar and muntjac) risk identity switches if the similarity threshold is set too loosely. Additionally, visual appearance changes with season, mud or injuries; regular gallery maintenance and optional human verification remain necessary. Finally, the model may underperform when transferred to a biome with radically different vegetation colours or lighting conditions; a small round of local fine-tuning is then required (Schneider et al., 2019).

4. Performance metric and trade-off discussion

A two-level evaluation metric is proposed:

- Detection stage: Performance is assessed with mean average precision (mAP), which counts a prediction as correct only when its bounding box overlaps at least 50 percent of the true animal area and the species label is accurate.

- Re-ID stage: rank-1 accuracy—the proportion of queries whose top match is the correct individual—captures practical tracking success.

In deployment, these metrics interact through the similarity-threshold trade-off. A strict threshold lowers the chance of merging two distinct tigers into one ID but risks splitting the same tiger into multiple identities, inflating population counts. A lower threshold has the inverse risk. Therefore, the optimal point must be chosen according to conservation priorities: under-counting critically endangered species could mask population collapse, whereas slight over-counting is often acceptable when the goal is broad trend monitoring.

5. Conclusion

EcoTrack illustrates how modern computer vision can transform Southeast Asian wildlife conservation. By detecting animals and recognising individuals from a handful of examples, rangers gain continuous, non-intrusive insight into “who is where and how often.” While motion blur, look-alikes and domain shifts remain challenges, they are manageable through data

augmentation, careful threshold tuning and targeted human oversight. With minimal hardware and annotation effort, the proposed pipeline can thus provide the evidence base required to protect Vietnam's unique biodiversity.

Total Word Count (including title to conclusion): 800

References

Beery, S., Morris, D., & Yang, S. (2019). Efficient Pipeline for Camera Trap Image Review. ArXiv:1907.06772 [Cs]. https://arxiv.org/abs/1907.06772

Chen, H., Lindshield, S., Ndiaye, P. I., Ndiaye, Y. H., Pruetz, J. D., & Reibman, A. R. (2023). Applying Few-Shot Learning for In-the-Wild Camera-Trap Species Classification. AI, 4(3), 574–597. https://doi.org/10.3390/ai4030031

Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods in Ecology and Evolution, 10(4), 461–470. https://doi.org/10.1111/2041-210x.13133

Snell, J., Swersky, K., & Zemel, R. S. (2017). Prototypical Networks for Few-shot Learning. ArXiv.org. https://arxiv.org/abs/1703.05175

Yu, W., Luo, M., Zhou, P., Si, C., Zhou, Y., Wang, X., Feng, J., & Yan, S. (2022, July 4). MetaFormer Is Actually What You Need for Vision. ArXiv.org. https://doi.org/10.48550/arXiv.2111.11418