



Can Visual Language Models benefit Wildlife Re-Identification?

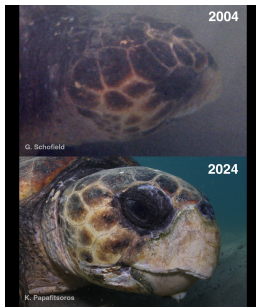
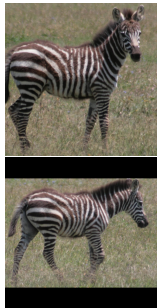
AI Collaborative Workshop, 4 November 2025

Kostas Papafitsoros

Queen Mary University of London

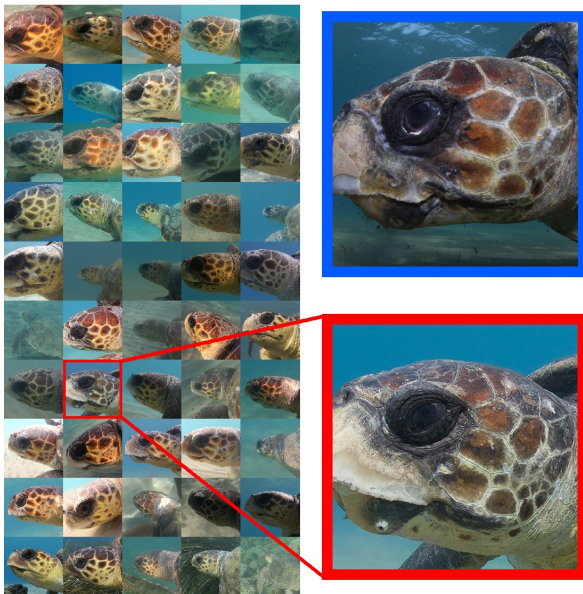
Wildlife Re-Identification

The process of identifying individual animals from images based on their unique, external morphological patterns (stripes, spots, scales and others)



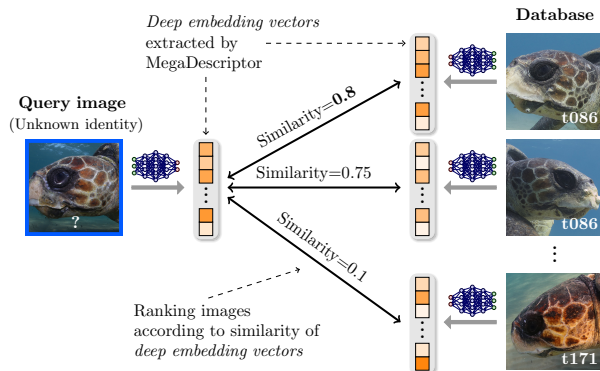
- Survival rates
- Abundance
- Behaviour (social, reproductive,...)
- Antropogenic pressure
- Movements
- Public Engagement

Image retrieval



Wildlife Re-Identification and deep metric learning

A deep neural network $\mathcal{N}$ is trained to map animal images to L -dimensional vectors in a feature space. The training is performed such that given two images u_i, u_j showing the same (different) individual animal, their feature vectors $v_i := \mathcal{N}(u_i), v_j := \mathcal{N}(u_j)$ have small (large) distance $d(v_i, v_j)$ under some suitable distance function d , or in other words high (low) similarity.



Can large Visual Language Models perform wildlife Re-Identification?



Can large Visual Language Models perform wildlife Re-Identification?

ChatGPT ▾

I will give you two photos of sea turtles and I would like you to tell me whether they show the same individual

Sure! Please upload both photos of the sea turtles, and I'll carefully compare them to see whether they appear to be the same individual.

🖼️ 👍 🗨️ ↻ 🔍 ...



Thanks for uploading both photos!

Conclusion (final judgement)

Putting the above together, the **head scute arrangement (the single most diagnostic feature)** and the matching carapace patching point to these two photos showing the **same individual**.

Confidence: **moderate-high (about 80–90%)**, limited mostly by difference in angle and lighting which could obscure small differences.

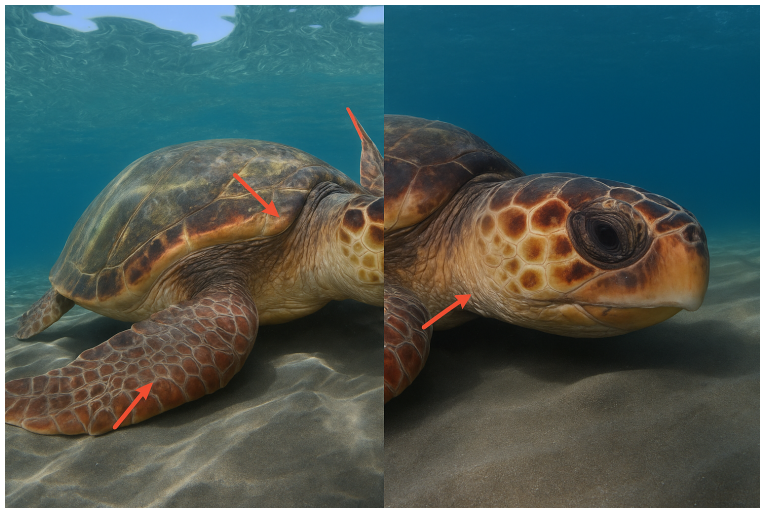
If you'd like, I can:

- Annotate both images with the matching points (head scutes, shell patch, scar) to make the match visual, or
- Produce a side-by-side overlay showing those landmarks.

Which would you prefer?

🖼️ 👍 🗨️ ↻ 🔍 ...

Can large Visual Language Models perform wildlife Re-Identification?



More systematic experiments with Gemini 2.5 pro / Outlook

Preliminary work jointly with Daniel Kühlwein:

- Used both naive and expert prompts
- Pairs of both easy and difficult, negative and positive matches
- **Gemini cannot really do wildlife re-identification!**

More systematic experiments with Gemini 2.5 pro / Outlook

Preliminary work jointly with Daniel Kühlwein:

- Used both naive and expert prompts
- Pairs of both easy and difficult, negative and positive matches
- **Gemini cannot really do wildlife re-identification!**

Outlook:

- Anyone who is interested/expert in VLMs has any ideas how to make them work?
- We have plenty of imaging data for this (but not texts!). Need for text-image datasets?
- It would be very useful to have a model that mimic the human decision making ("Yes" or "No")
- Useful for interpretability purposes