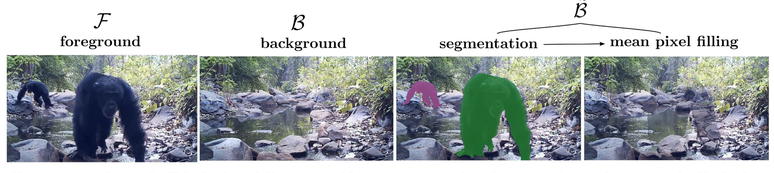
PanAf-FGBG: Understanding the Impact of Backgrounds in Wildlife Behaviour Recognition

CVPR 2025 Best Paper Award Candidate

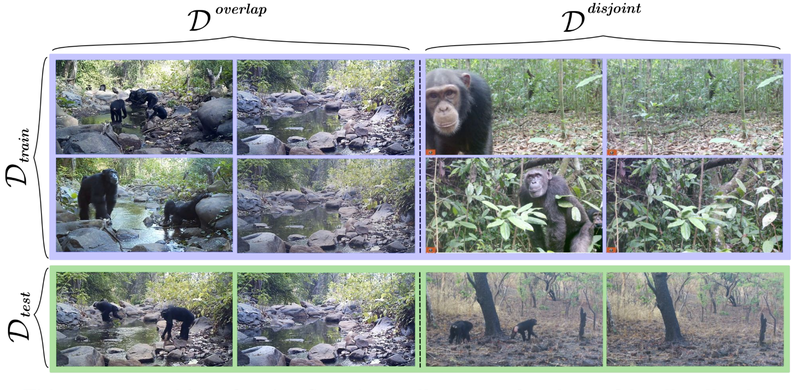
Otto Brookes[1,2†*], Maksim Kukushkin[3,4*], Majid Mirmehdi[1], Colleen Stephens[5], Paula Dieguez[5], Thurston Hicks[6], Sorrel Jones[5], Kevin Lee[5], Maureen McCarthy[5], Amelia Meier[5], Emma Normand[2], Erin Wessling[7], Roman Wittig[5,8], Kevin Langergraber[9], Klaus Zuberbühler[10], Lukas Boesch[2], Thomas Schmid[3,11], Mimi Arandjelovic[5], Hjalmar Kühl[5,12], Tilo Burghardt[1]

[1]University of Bristol, [2]Wild Chimpanzee Foundation, [3]Martin Luther University Halle-Wittenberg, [4]Leipzig University, [5]Max Planck Institute for Evolutionary Anthropology, [6]University of Warsaw, [7]Harvard University, [8]University of Lyon, [9]Arizona State University, [10]University of St Andrews, [11]Lancaster University in Leipzig, [12]Senckenberg Museum of Natural History

Overview: We present the PanAf-FGBG dataset, featuring **40 hours of paired video** comprising chimpanzees in the wild. Each sample consists of a foreground, background, and synthetically generated background video.

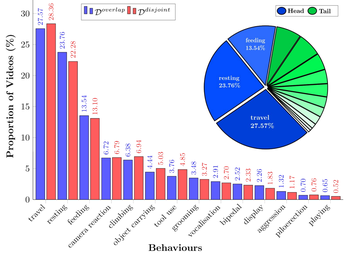


Overlapping & Disjoint Views: We present both **overlapping and disjoint views** of dataset.

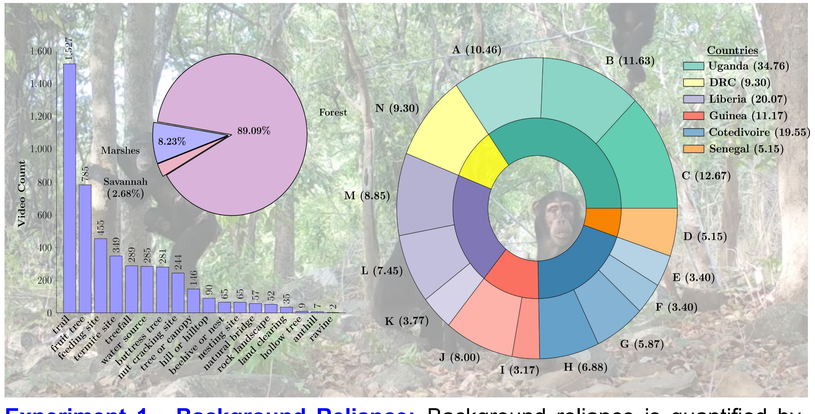


This setup enables, for the first time, **direct evaluation of in-distribution and out-of- distribution conditions**, and for the **impact of backgrounds on behaviour recognition models** to be quantified.

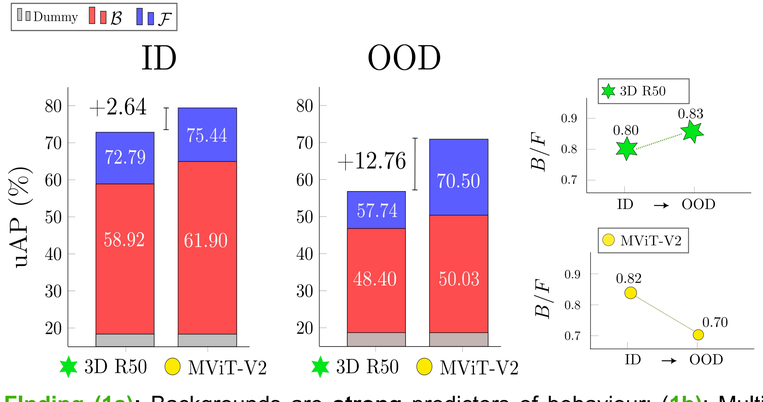
Behaviour Distribution: Videos are annotated with **14 co-occurring behaviours**, exhibiting a long-tailed distribution. The class distribution across configurations remains approximately consistent facilitating full cross-view comparison.



Habitat & Metalocation: Footage is collected from **389 individual camera locations across 14 national parks in 6 African countries**.

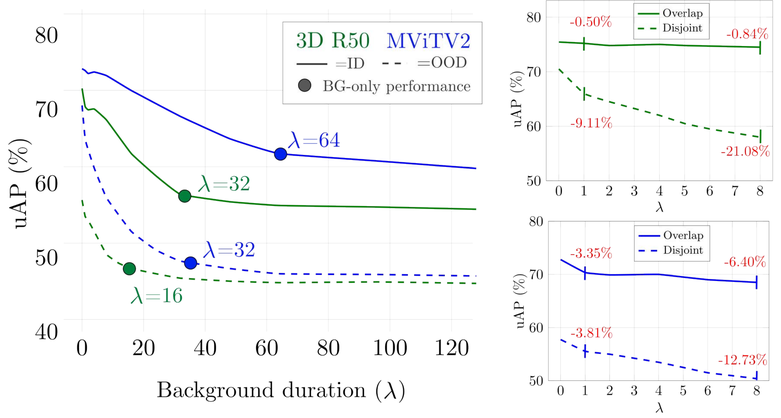


Experiment 1 - Background Reliance: Background reliance is quantified by training model architectures on the foreground, background, and synthetic background videos.



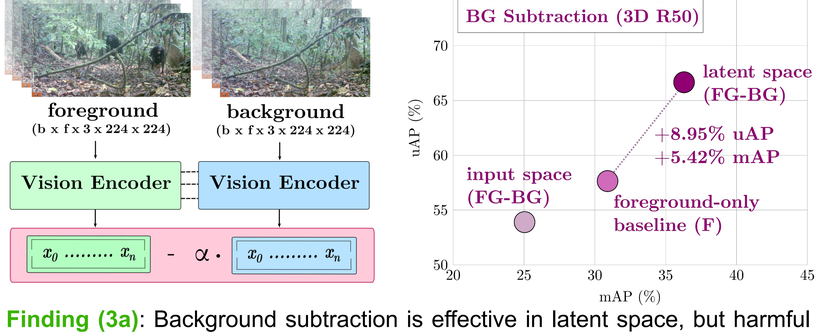
FInding (1a): Backgrounds are **strong** predictors of behaviour; **(1b)**: Multi-scale Vision Transformers **rely less** on backgrounds; **(1c)**: Background performance is **not** a subset of foreground performance; **(1d)**: Synthetic backgrounds are a good proxy for real-world backgrounds but a residual activity signal still remains.

Experiment 2 - Background Duration: We simulate **increasing background duration** by appending background frames to foreground video to quantify its effect on model performance.



Finding (2a): 3D-R50 is more sensitive to increases in background durations than MViT-V2 when evaluated in-distribution; **(2b):** MViT-V2 is *more* sensitive to increases in background durations when evaluated OOD.

Experiment 3 - Background Subtraction: Analysing the impact of background subtraction operations in both input and latent spaces.



Finding (3a): Background subtraction is effective in latent space, but harmful in input space.