GorillaWatch
An Automated System for In-the-Wild Gorilla Re-Identification and Population Monitoring

Maximilian Schall[1], Felix Leonard Knöfel[1], Noah Elias König[1], Jan Jonas Kubeler[1], Maximilian von Klinski[1], Joan Wilhelm Linnemann[1], Xiaoshi Liu[1], Iven Jelle Schlegelmilch[1], Ole Woyciniuk[1], Alexandra Schild[1], Dante Wasmuht[2], Magdalena Bermejo Espinet[3], German Illera Basas[3], Gerard de Melo[1]
[1]Hasso Plattner Institute [2]Conservation X Labs [3]Sabine Plattner African Charities

WACV 2026 TUCSON, AZ 3/6 - 3/10

Because every gorilla deserves to be re-KONG-nized.

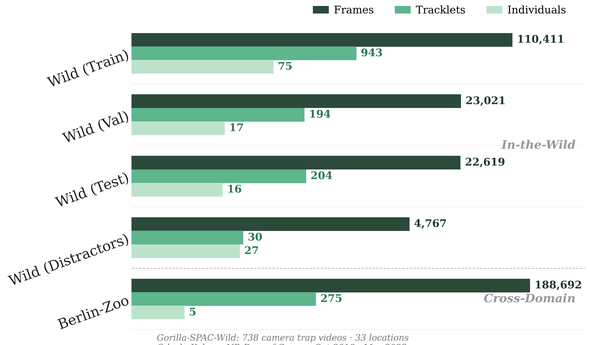
Abstract

Monitoring critically endangered western lowland gorillas **requires immense manual effort** to analyze vast camera trap archives. We present GorillaWatch, **an end-to-end pipeline** for **automated gorilla detection, tracking, and re-identification**, alongside three novel benchmark datasets. Using multi-frame self-supervised pretraining and spatiotemporal constrained clustering, our system enables scalable, non-invasive population monitoring without manual labels.

Contributions

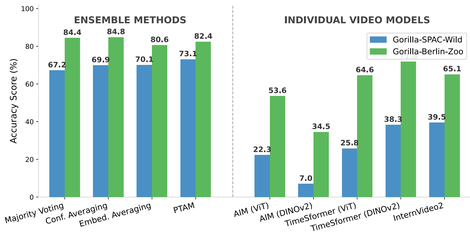
1. We introduce **three** novel datasets for open-set primate re-identification and tracking.
2. We propose a multi-frame self-supervised pretraining strategy which leverages temporal consistency in tracklets without manual labels.
3. We demonstrate that aggregating features from large-scale image backbones outperforms specialized video architectures.
4. We address unsupervised population counting via spatiotemporal constrained clustering, reducing over-segmentation.

Dataset



Gorilla-SPAC-Wild: 738 camera trap videos · 33 locations
Odzala-Kokoua NP, Rep. of Congo · Oct 2018 - Mar 2023
Gorilla-Berlin-Zoo: 153 videos · 3 cameras
Berlin Zoo, Germany · 3 months

Ensemble vs. Video



Ensemble methods outperform video models

AttnLRP - Adaption

Original Fine-Tuned

Constrained Clustering

Method	ARI ↑	AMI ↑	# Clusters
Unconstrained			
HAC	0.606	0.728	465
DBSCAN	0.379	0.625	341
HDBSCAN	0.114	0.510	113
Constrained			
HAC	**0.837**	**0.891**	**17 ✓**
DBSCAN	0.586	0.873	13
HDBSCAN	0.184	0.677	204

Further in the paper

1. **Tracker Comparison:** Detailed comparison of trackers on our Gorilla-SPAC-Multi-Object-Tracking dataset
2. **Backbone Zero-Shot Benchmark:** Systematic comparison of pre-trained embedding models .
3. **DINOv2 Scaling Analysis:** Accuracy vs. model size (Small → Giant) after fine-tuning
4. **Video Architecture Deep-Dive:** Full results, including the effect of replacing standard ViT with DINOv2.

Pipeline