Supplementary Material

In the supplementary material, we provide: (i) additional dataset statistics in Sec. 1; (ii) a description of how synthetic background videos are created in Sec. 2; (iii) foreground-background video pair visualisations (see Fig 6) and; (iv) a visualisation showcasing a small fraction ($\sim 0.05\%$) of the total available frames (see Fig. 7)

1. Dataset Statistics

We present additional dataset statistics for the country, research site, camera locations and behaviours. Specifically, (i) the distribution of countries and the corresponding research sites are displayed in Fig. 1; (ii) a comparison of the behaviour distribution for the overlapping and disjoint datasets is displayed in Fig. 2; (iii) time interval standardisation statistics are shown in Fig 3 and; (iv) the accumulative proportion of videos contributed by each camera is shown in Fig. 4.



Figure 1. **Proportion of videos from each country and research site**. The inner ring displays the proportion of videos extracted from each country, while the outer ring represents individual research sites. Each research site segment is a unique shade derived from its corresponding country's colour. All proportions are shown in brackets. Note that research site names are replaced with letters to protect the location of the chimps.

2. Synthetic Background Generation

We generated synthetic background videos using SAM2 [1] and mean pixel value filling. Specifically, we prompted the SAM2.1-Large model using a single spatial coordinate indicating the location of the chimpanzee to produce an initial segmentation mask. We then leveraged the automatic mask propagation functionality of SAM2 to create spatiotemporal masklets for the full video. Note that spatial coordinates were produced manually. Then, we filled the area indicated by the segmentation mask with the mean pixel value for the frame (see Fig. 5 for examples).



Figure 2. Comparison of the proportion of videos containing each behaviour between overlapping and disjoint datasets. Behaviours are ordered from highest to lowest proportion, with exact values displayed above each bar for easy comparison.



Figure 3. **Time interval standardisation between foreground and background video pairs**. The majority of foreground background video pairs are sampled within one hour of each other.



Figure 4. Accumulative proportion of videos contributed by each camera. The y-axis represents the accumulative proportion of videos, with individual cameras arranged by their contribution on the x-axis. The red dashed line divides the cameras into two groups: those contributing the first 50% of the data (left) and the remaining 50% (right).



Figure 5. Synthetic Background Video Examples. Three example video clips with the original segmentation mask generated by one-shot prompting of SAM2 overlaid and the corresponding mean pixel value filled frame.



Figure 6. Foreground-Background Video Pair Examples. Shown are 18 pairs of still frames (two pairs per row) extracted from foreground-background video pairs.

		La la		Ver		E.S.	KA_4		300	R		ANT	and the second	
KAN		1	14		n.		V. Alexandre	i de la			1 al	ant a second		-24A
					-		1.19		PLLI	1.16	17/14		NO.S	
								MASS OF	300		7100			KIN 41
16	Size of	-						//V						
		SEA.					14 cm	1 Martin				J.L. K.		
				1						A THE				
			No. Com		(ALL)		In Index			9			A Bill	
Contract of the second second			5 A		A.U.				- R - P				153 W. Linker	
	AMART			23 - WILLIA										
				ALC: NO										
		2 / p 4								- 2		Towned M		
2														
				CONTRACTOR		21.								
							er an			2	A AN AN AN			
	ESTATE STORE						and a	E.	ALC: NO STATE	ar na still ann				
A AN			8											
				R				T A					2410	1-14
				and a la		E PART MAN	AL SOM						20	-
Carl .	A					\$- 1987	- 8						A Company	COAST.
			113		L.	and the second	See 18	And and a second	30 M				at here	
1 par					E.			6	e.	A PAR			1 Jane	
	A series													
					S The						Page 1			
Pere		an the second				1		ti and						1/2
				Nu2	A		1 Y	MAN .			-			1 40
	SANG.							- NOA	MAR N					
			C. W.		. 19 -			A WIT						
			a t i j	**3				- 13/1						Res William
CHU /											alan Arili			
N		ALCO-	1.0 1				AIZ		T	NE YA				
												1/ And Alle		
				Nº 72			A ST	CONTRACTOR			No MICE			
								ARAN						
									a sait h		ins e			
			Y IN								de la com			
Para la							The second	THE R			- M/ - 2	TIP		
							Alle a					-N		
	States and		A AND A			No.	and the set	1.500		14 1 2 2 2	North Carlins	C. C. AND	Case on an and the A	

Figure 7. **Dataset Overview**. A small fraction ($\sim 0.05\%$) of the 1.8 million frames in the dataset are shown, highlighting its diversity with respect to exhibited behaviours, habitat, weather conditions, time of day, and more.

References

[1] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. 1