

Table 1: Comparison of c-SKIT and PCBM on AwA2.

Model	Importance	Accuracy	f_1 score
CLIP:RN50	PCBM	94.52%	0.45
	c-SKIT	98.55%	0.60
CLIP:ViT-B/32	PCBM	95.48%	0.48
	c-SKIT	99.43%	0.63
CLIP:ViT-L/14	PCBM	94.79%	0.51
	c-SKIT	99.80%	0.64
OpenClip:ViT-B-32	PCBM	93.84%	0.53
	c-SKIT	99.53%	0.54
OpenClip:ViT-L-14	PCBM	95.84%	0.49
	c-SKIT	100.0%	0.54
FLAVA	PCBM	96.15%	0.48
	c-SKIT	99.37%	0.52
ALIGN	PCBM	96.45%	0.42
	c-SKIT	99.76%	0.51
BLIP	PCBM	93.76%	0.54
	c-SKIT	99.53%	0.55
average	PCBM	95.10%	0.48
	c-SKIT	99.50%	0.55

Figure 1: SKIT results on Imagenette.

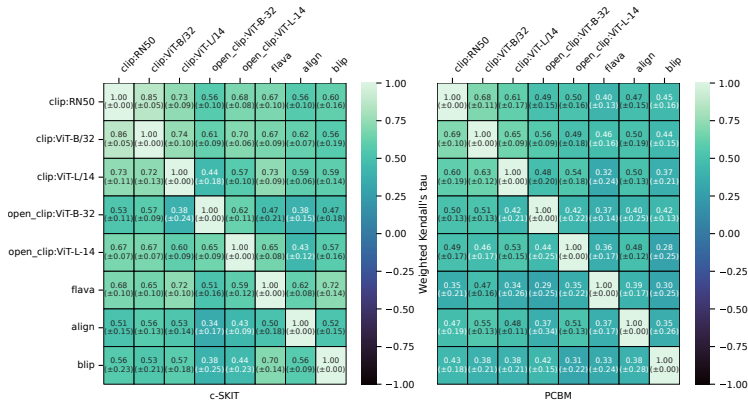


Figure 2: Comparison of c-SKIT and PCBM on Imagenette.

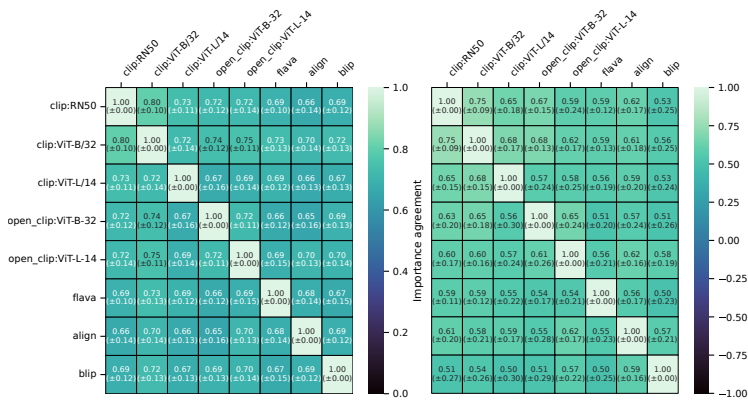


Figure 3: x-SKIT results on Imagenette.

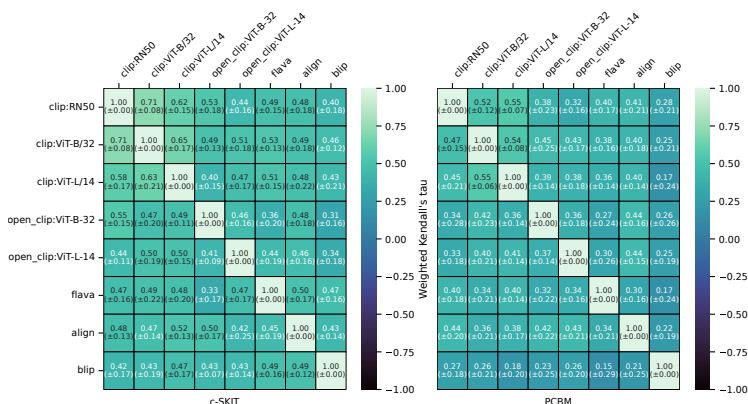


Figure 4: Comparison of c-SKIT and PCBM on AwA2.

Figure 5: x-SKIT results on CUB-200-2011.

Table 2: x-SKIT results on CUB-200-2011.

Model	Accuracy	f_1 score		
		$s = 1$	$s = 2$	$s = 4$
CLIP:RN50	81.52%	0.82 ± 0.25	0.83 ± 0.26	0.84 ± 0.22
CLIP:ViT-B/32	89.49%	0.85 ± 0.22	0.86 ± 0.20	0.86 ± 0.19
CLIP:ViT-L/14	97.30%	0.84 ± 0.22	0.87 ± 0.23	0.83 ± 0.23
OpenClip:ViT-B-32	95.17%	0.84 ± 0.20	0.86 ± 0.23	0.82 ± 0.23
OpenClip:ViT-L-14	97.67%	0.86 ± 0.20	0.89 ± 0.17	0.87 ± 0.15
FLAVA	93.17%	0.84 ± 0.26	0.87 ± 0.24	0.84 ± 0.22
ALIGN	86.23%	0.85 ± 0.24	0.88 ± 0.18	0.82 ± 0.22
BLIP	75.61%	0.85 ± 0.25	0.85 ± 0.24	0.80 ± 0.30
average	/	0.84	0.86	0.83

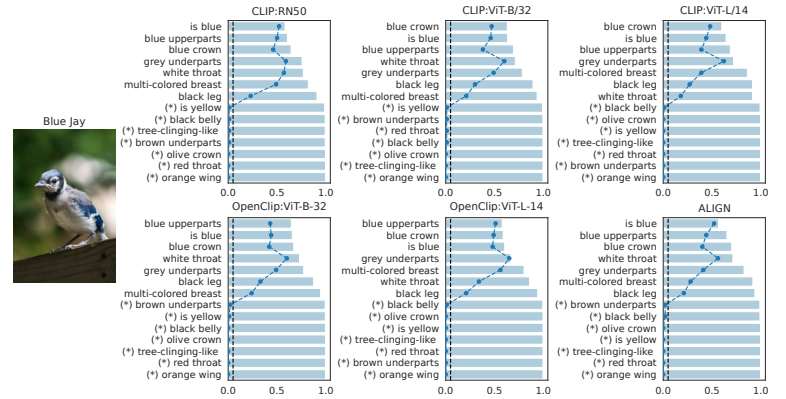


Figure 6: Example x-SKIT results on CUB-200-2011. Absent concepts are annotated with an asterisk (*) symbol.