

Table 1: Ablations on effects of different LLMs used to generate hard negative captions on performance.

Models	SugarCrepe								Retrieval				ImageNet 1k	
	Replace			Swap		Add		Overall	Text Retrieval (R@5)		Image Retrieval (R@5)		top-1	top-5
	Object	Attribute	Relation	Object	Attribute	Object	Attribute		MSCOCO	Flickr30k	MSCOCO	Flickr30k		
Gemma	63.68	58.25	51.85	51.43	50.45	56.35	59.97	56.00	5.70	12.30	6.89	12.90	4.44	12.09
Phi	66.56	69.92	65.15	52.25	56.33	57.76	60.55	61.22	6.20	13.50	6.45	12.54	3.93	10.94
Mistral	64.77	66.12	65.93	55.51	55.41	59.65	64.45	61.69	5.80	11.20	6.19	10.24	3.84	10.52

Table 2: Performance comparison of CLIP and CyCLIP with and without TripletLoss.

Models	SugarCrepe								Retrieval				ImageNet 1k	
	Replace			Swap		Add		Overall	Text Retrieval (R@5)		Image Retrieval (R@5)		top-1	top-5
	Object	Attribute	Relation	Object	Attribute	Object	Attribute		MSCOCO	Flickr30k	MSCOCO	Flickr30k		
CLIP	63.68	54.70	50.65	53.06	49.40	55.48	55.82	55.11	6.4	12.70	6.73	12.90	4.63	12.58
TripletCLIP	74.33	71.95	71.12	58.78	59.16	65.66	65.75	66.68	11.50	24.50	12.58	24.74	7.72	19.95
CyCLIP	55.68	54.31	50.99	53.88	53.45	55.72	55.35	54.62	5.30	10.90	6.53	12.64	4.63	13.01
CyCLIP + TripletLoss	66.34	65.74	62.09	44.49	52.70	60.33	58.82	58.64	9.90	20.20	10.79	20.38	7.27	19.05

Table 3: Finetuning base models performance comparison on various benchmarks.

Models	SugarCrepe								Retrieval				ImageNet 1k	
	Replace			Swap		Add		Overall	Text Retrieval (R@5)		Image Retrieval (R@5)		top-1	top-5
	Object	Attribute	Relation	Object	Attribute	Object	Attribute		MSCOCO	Flickr30k	MSCOCO	Flickr30k		
CLIP (org)	90.92	80.08	69.13	61.22	64.26	77.16	68.64	73.06	74.9	94.60	55.92	83.38	63.31	88.22
CLIP (finetuned)	90.92	79.69	64.01	60.82	64.26	84.67	78.76	74.73	68.9	88.40	53.50	81.10	49.95	79.16
NegCLIP	91.53	83.25	73.97	72.24	67.72	86.95	88.44	80.59	66.00	88.60	53.41	81.12	48.85	78.34
Baseline [3]	93.22	84.39	67.35	62.04	70.12	88.31	79.48	77.84	81.4	96.0	67.49	89.84	61.40	88.10
CoN-CLIP [5]	93.58	80.96	63.3	87.29	79.62	59.18	65.16	75.58	-	-	-	-	-	-
TSVLC (RB) [7]	91.34	81.34	64.15	68.16	69.07	79.49	91.33	77.84	71.70	93.00	62.01	87.12	58.81	85.97
TSVLC (LLM+RB) [7]	88.13	76.78	62.73	64.08	66.67	75.80	81.07	73.61	71.82	92.50	62.24	87.46	59.77	87.02
DAC [8]	94.43	89.48	84.35	75.10	74.17	89.67	97.69	86.41	54.5	79.60	63.51	87.84	51.02	81.22
TripletCLIP	94.43	85.53	80.94	69.80	69.82	90.40	86.27	82.46	55.6	82.60	53.32	80.88	45.92	75.54