Model	MSCOCO (5K test set)						Flickr30K (1K test set)					
	Image → Text			Text → Image			Image → Text			Text → Image		
	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
FLAVA	42.74	76.76	-	38.38	67.47	-	67.7	94.0	-	65.22	89.38	-
CLIP	58.4	81.5	88.1	37.8	62.4	72.2	88.0	98.7	99.4	68.7	90.6	95.2
BEiT-3	**84.8**	**96.5**	**98.3**	**67.2**	**87.7**	**92.8**	**98.0**	**100.0**	**100.0**	**90.3**	**98.7**	**99.5**
S-SMKE	53.54	81.1	89.52	35.65	66.0	77.77	70.9	92.1	96.0	52.72	80.2	87.46
S-SMKE finetuned	56.2	83.3	91.1	39.8	69.2	79.8	82.0	95.4	98.0	64.6	87.5	93.1

Table 1: