Datasets	Methods	Image to Sentence				Sentence to Image			
		R@1	R@5	R@10	Med r	R@1	R@5	R@10	Med r
Flickr 8K	DeViSE	4.8	16.5	27.3	28	5.9	20.1	29.6	29
	SDT-RNN	4.5	18.0	28.6	32	6.1	18.5	29.0	29
	DeFrag^{+O}	12.6	32.9	44.0	14	9.7	29.6	42.5	15
	Kiros et al. ^{A}	13.5	36.2	45.7	13	10.4	31.0	43.7	14
	Kiros et al. ^{V}	18	40.9	55	8	12.5	37	51.5	10
	m-RNN^{A}	14.5	37.2	48.5	11	11.5	31.0	42.4	15
	Mind's Eye^{V}	17.3	42.5	57.4	7	15.4	40.6	50.1	8
	DeepVS$^{+O,V}$	16.5	40.6	54.2	7.6	11.8	32.1	44.7	12.4
	NIC^{G}	20	-	60	6	19	-	**64**	5
	Bi-LSTM^{A}	21.3	44.7	56.5	6.5	15.1	37.8	50.1	9
	Bi-S-LSTM^{A}	19.6	43.7	55.7	7	14.5	36.4	48.3	10.5
	Bi-F-LSTM^{A}	19.9	44.0	56.0	7	14.9	37.4	49.8	10
	Bi-LSTM^{V}	**29.3**	**58.2**	**69.6**	**3**	**19.7**	**47.0**	60.6	**5**
Flickr 30K	DeViSE	4.5	18.1	29.2	26	6.7	21.9	32.7	25
	SDT-RNN	9.6	29.8	41.1	16	8.9	29.8	41.1	16
	Kiros et al. ^{A}	14.8	39.2	50.9	10	11.8	34.0	46.3	13
	Kiros et al. ^{V}	23.0	50.7	62.9	5	16.8	42.0	56.5	8
	LRCN^{A}	14	34.9	47	11	-	-	-	-
	NIC^{G}	17	-	56	7	17	-	57	8
	m-RNN^{A}	18.4	40.2	50.9	10	12.6	31.2	41.5	16
	Mind's Eye^{V}	18.5	45.7	58.1	7	16.6	42.5	**58.9**	8
	DeFrag $^{+O}$	16.4	40.2	54.7	8	10.3	31.4	44.5	13
	DeepVS$^{+O,V}$	22.2	48.2	61.4	4.8	15.2	37.7	50.5	9.2
	Bi-LSTM^{A}	18.7	41.2	52.6	8	14.0	34.0	44.0	14
	Bi-S-LSTM^{A}	21	43.0	54.1	7	15.1	35.3	46.0	12
	Bi-F-LSTM^{A}	20	44.4	55.2	7	15.1	35.8	46.8	12
	Bi-LSTM^{V}	**28.1**	**53.1**	**64.2**	**4**	**19.6**	**43.8**	55.8	**7**
MSCOCO	DeepVS$^{+O,V}$	16.5	39.2	52.0	9	10.7	29.6	42.2	14.0
	Bi-LSTM^{A}	10.8	28.1	38.9	18	7.8	22.4	32.8	24
	Bi-S-LSTM^{A}	13.4	33.1	44.7	13	9.4	26.5	37.7	19
	Bi-F-LSTM^{A}	11.2	30	41.2	16	8.3	24.9	35.1	22
	Bi-LSTM^{V}	**16.6**	**39.4**	**52.4**	**9**	**11.6**	**30.9**	**43.4**	**13**