

dataset	modality	# pairs	avg. text len.	# classes
Wikipedia	img./text	2,866	640.5 words	10
Flickr30k	img./sentence	31,783	5 sentences	n.a
IAPR TC-12	img./text	20,000	23.1 words	n.a
Pascal VOC'12	img./tags	11,540	1.4 tags	20
NUS-WIDE	img./tags	269,648	2.4 tags	81