

Source	COCO (train)	CC (all)	SBU (all)	Flicker30k (train)	VQA (train)	GQA (bal-train)	VG-QA (train)	Total
Image/Text	113k/609k	3.0M/3.0M	980k/980k	29k/29k	83k/444k	82k/946k	48k/484k	4.1M/6.5M