

Models	#Trainable Params	Open- sourced?	Visual Question Answering	Image Captioning		Image-Text Retrieval	
			VQAv2 (test-dev)	NoCaps (val)		Flickr (test)	
			VQA acc.	CIDEr	SPICE	TR@1	IR@1
BLIP ( <a href="#">Li et al., 2022</a> )	583M	✓	-	113.2	14.8	96.7	86.7
SimVLM ( <a href="#">Wang et al., 2021b</a> )	1.4B	✗	-	112.2	-	-	-
BEIT-3 ( <a href="#">Wang et al., 2022b</a> )	1.9B	✗	-	-	-	94.9	81.5
Flamingo ( <a href="#">Alayrac et al., 2022</a> )	10.2B	✗	56.3	-	-	-	-
BLIP-2	188M	✓	<b>65.0</b>	<b>121.6</b>	<b>15.8</b>	<b>97.6</b>	<b>89.7</b>