

Models	#Trainable Params	NoCaps Zero-shot (validation set)								COCO Fine-tuned Karpathy test	
		in-domain		near-domain		out-domain		overall		B@4	C
		C	S	C	S	C	S	C	S		
OSCAR (Li et al., 2020)	345M	-	-	-	-	-	-	80.9	11.3	37.4	127.8
VinVL (Zhang et al., 2021)	345M	103.1	14.2	96.1	13.8	88.3	12.1	95.5	13.5	38.2	129.3
BLIP (Li et al., 2022)	446M	114.9	15.2	112.1	14.9	115.3	14.4	113.2	14.8	40.4	136.7
OFA (Wang et al., 2022a)	930M	-	-	-	-	-	-	-	-	<b>43.9</b>	<u>145.3</u>
Flamingo (Alayrac et al., 2022)	10.6B	-	-	-	-	-	-	-	-	-	138.1
SimVLM (Wang et al., 2021b)	~1.4B	113.7	-	110.9	-	115.2	-	112.2	-	40.6	143.3
BLIP-2 ViT-g OPT <sub>2.7B</sub>	1.1B	<u>123.0</u>	<u>15.8</u>	117.8	<u>15.4</u>	123.4	<b>15.1</b>	119.7	<u>15.4</u>	<u>43.7</u>	<b>145.8</b>
BLIP-2 ViT-g OPT <sub>6.7B</sub>	1.1B	<b>123.7</b>	<u>15.8</u>	<u>119.2</u>	15.3	<u>124.4</u>	14.8	<u>121.0</u>	15.3	43.5	145.2
BLIP-2 ViT-g FlanT5 <sub>XL</sub>	1.1B	<b>123.7</b>	<b>16.3</b>	<b>120.2</b>	<b>15.9</b>	<b>124.8</b>	<b>15.1</b>	<b>121.6</b>	<b>15.8</b>	42.4	144.5