

Input features	Attentional fusion	BLEU1	BLEU2	BLEU3	BLEU4	METEOR	ROUGE_L	CIDEr
QA	-	0.236	0.142	0.094	0.065	0.101	0.257	0.595
QA + Captions	-	0.245	0.152	0.103	0.073	0.109	0.271	0.705
QA + VGG16	-	0.231	0.141	0.095	0.067	0.102	0.259	0.618
QA + I3D	no	0.246	0.153	0.104	0.073	0.109	0.269	0.680
QA + I3D	yes	0.250	0.157	0.108	0.077	0.110	0.274	0.724
QA + I3D + VGGish	no	0.249	0.155	0.106	0.075	0.110	0.275	0.701
QA + I3D + VGGish	yes	0.256	0.161	0.109	0.078	0.113	0.277	0.727