Category	Detail	Accuracy
Attention direction	I ← Q	60.95
	I → Q	62.63
	I ↔ Q*	62.94
Memory size (K)	1	62.53
	3*	62.94
	5	62.83
Number (h) of parallel attention maps	2	62.82
	4*	62.94
	8	62.81
Number (L) of stacked layers	1	62.43
	2	62.82
	3*	62.94
	4	62.67
Attention in answer prediction layer	Attention used*	62.94
	Avg of features	61.63
Attention in image extraction layer	Attention used*	62.94
	Only last conv layer	62.39