

Method	Visual7W	VQA
Our Baseline	65.6	
+POS tag guided attention (POS-Att)	66.3	
+Convolutional N-Gram (Conv N-Gram)	66.2	
+POS-Att +Conv N-Gram	66.6	
+POS-Att +Conv N-Gram +Triplet attention-Q	66.8	
+POS-Att +Conv N-Gram +Triplet attention-A	67.0	
+POS-Att +Conv N-Gram +Triplet attention-Q+A	67.3	
+POS-Att +Conv N-Gram +Triplet attention-Q+A + Structured Learning Triplet	67.5	