A. Text Retrieval Case Study

We also visualize the text retrieval results using our proposed GraDual Figure 1. We demonstrate the same text retrieval samples as in SCAN [1]. Same to SCAN, our GraDual retrieves texts (a)1-(a)5 and (b)1-(b)4 all correctly for Figure 1(a) and (b) but with slightly different order due to the variance of representation and alignment. For Figure 1(c), SCAN fails to retrieve the (c)5 correctly whereas our GraDual does it correctly after integrating the cross-modal contextual information into our initial modality representation via GraDual.

References