

We study how to leverage off-the-shelf visual and linguistic data to cope with out-of-vocabulary answers in visual question answering task. Existing large-scale v