



(a) Input Image

Q-I: What is written on this blue **sign**?



Q-2: How many **buses** are there in the image?



Q-3: What color is this **roof**?



(b) Ours. Text-guided Visual Sparsification



(c) Existing Methods. Text-agnostic Visual Sparsification