CPL: Counterfactual Prompt Learning for Vision and Language Models

1 Visualization of Sampled Images

We visualize the sampled image pairs via random sampling and BERTScore sampling for image classification as shown in Figure 1, image-text retrieval as shown in Figure 2, and visual question answering as shown in Figure 3.

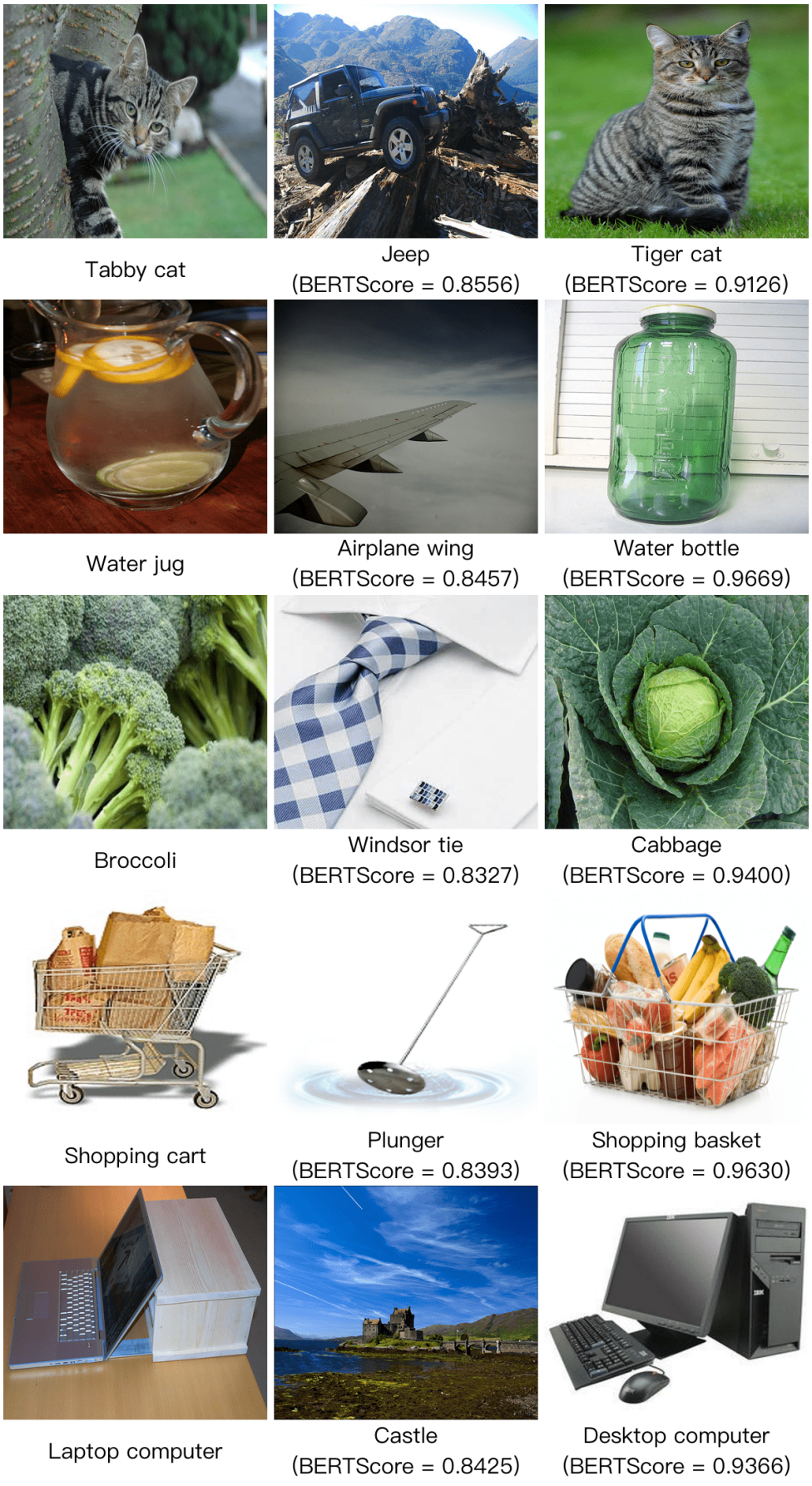

Figure 1: Comparison of sampled images from the ImageNet dataset via random sampling and BERTScore sampling. The first column is original positive examples. The second column is randomly sampled images. The third column is BERTScore sampled images.

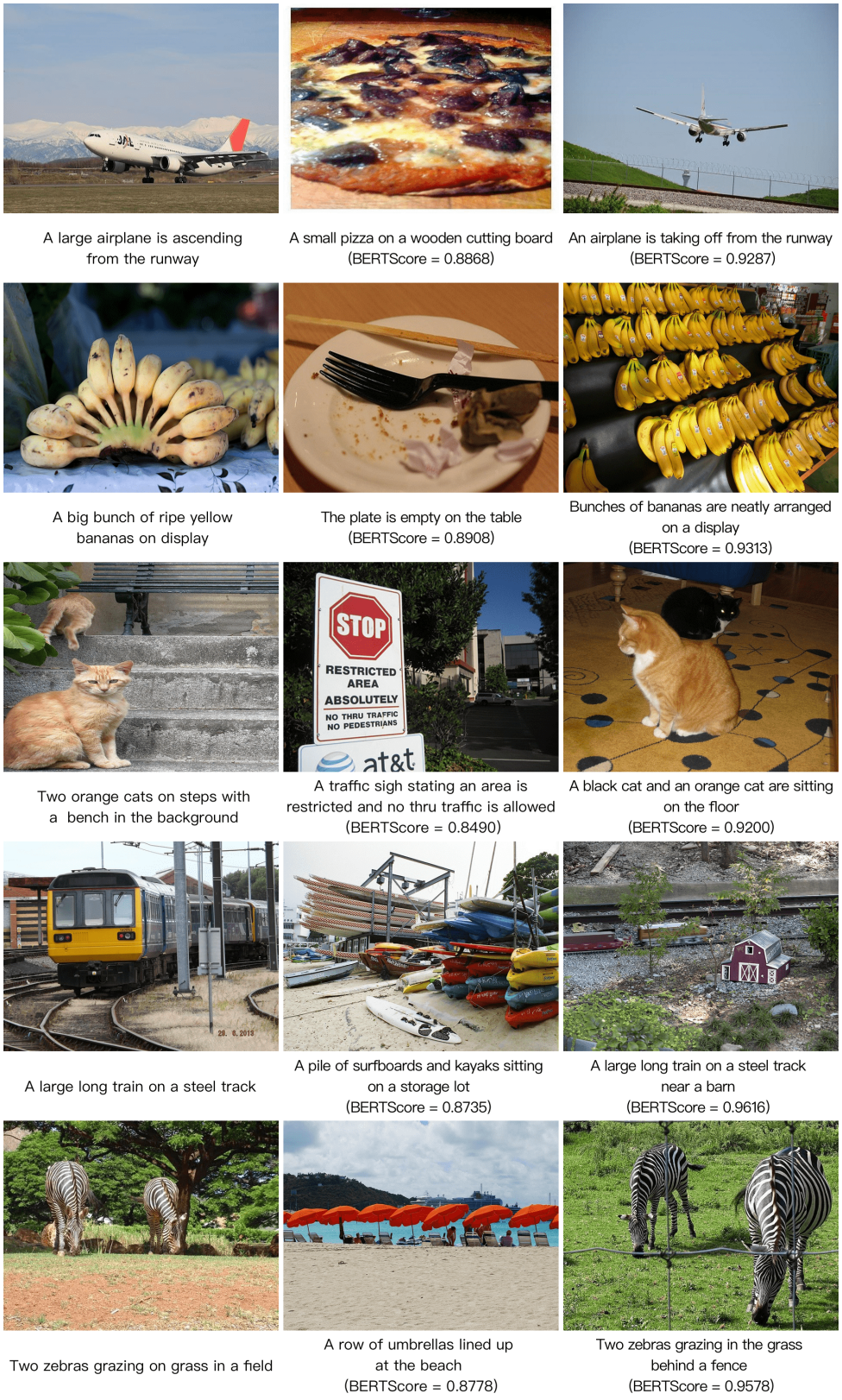

Figure 2: Comparison of sampled images from the COCO dataset via random sampling and BERTScore sampling. The first column is original positive examples. The second column is randomly sampled images. The third column is BERTScore sampled images.

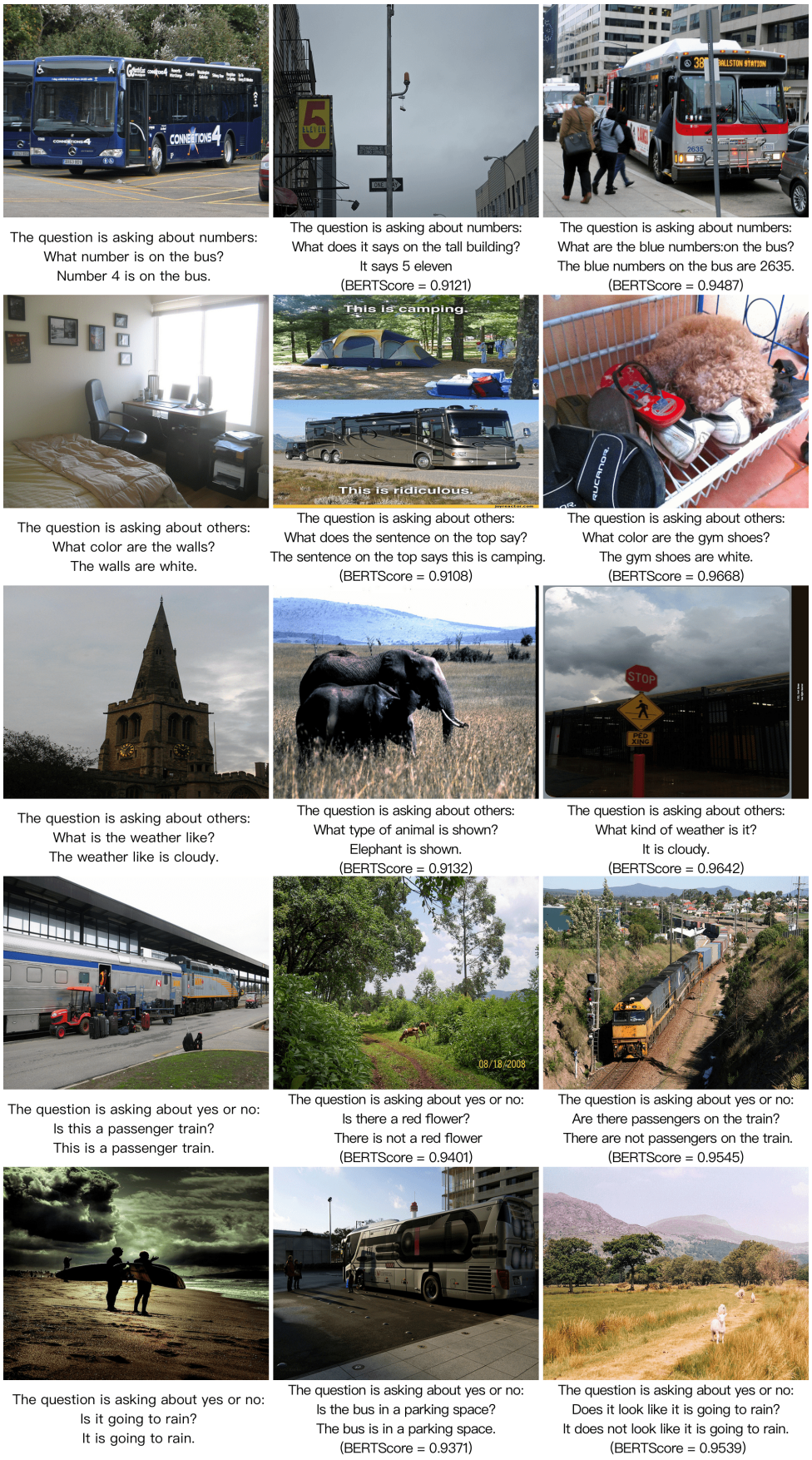

Figure 3: Comparison of sampled images from the VQAv2 dataset via random sampling and BERTScore sampling. The first column is original positive examples. The second column is randomly sampled images. The third column is BERTScore sampled images.