

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.



Tabby cat



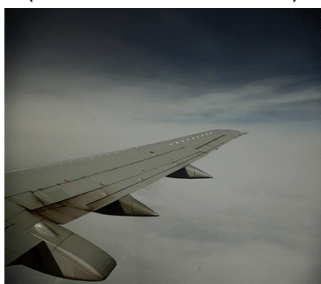
Jeep
(BERTScore = 0.8556)



Tiger cat
(BERTScore = 0.9126)



Water jug



Airplane wing
(BERTScore = 0.8457)



Water bottle
(BERTScore = 0.9669)



Broccoli



Windsor tie
(BERTScore = 0.8327)



Cabbage
(BERTScore = 0.9400)



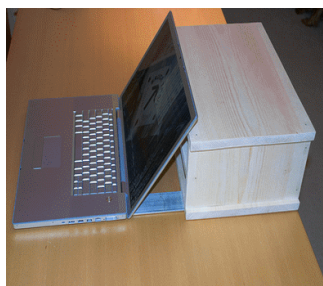
Shopping cart



Plunger
(BERTScore = 0.8393)



Shopping basket
(BERTScore = 0.9630)



Laptop computer



Castle
(BERTScore = 0.8425)



Desktop computer
(BERTScore = 0.9366)

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.



A large airplane is ascending from the runway



A small pizza on a wooden cutting board (BERTScore = 0.8868)



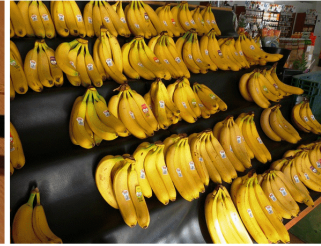
An airplane is taking off from the runway (BERTScore = 0.9287)



A big bunch of ripe yellow bananas on display



The plate is empty on the table (BERTScore = 0.8908)



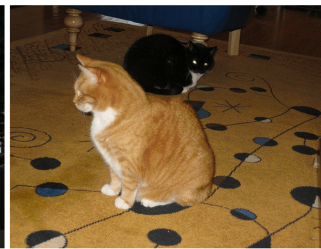
Bunches of bananas are neatly arranged on a display (BERTScore = 0.9313)



Two orange cats on steps with a bench in the background



A traffic sign stating an area is restricted and no thru traffic is allowed (BERTScore = 0.8490)



A black cat and an orange cat are sitting on the floor (BERTScore = 0.9200)



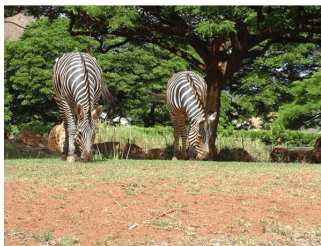
A large long train on a steel track



A pile of surfboards and kayaks sitting on a storage lot (BERTScore = 0.8735)



A large long train on a steel track near a barn (BERTScore = 0.9616)



Two zebras grazing on grass in a field



A row of umbrellas lined up at the beach (BERTScore = 0.8778)



Two zebras grazing in the grass behind a fence (BERTScore = 0.9578)

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.



The question is asking about numbers:
What number is on the bus?
Number 4 is on the bus.



The question is asking about numbers:
What does it say on the tall building?
It says 5 eleven
(BERTScore = 0.9121)



The question is asking about numbers:
What are the blue numbers on the bus?
The blue numbers on the bus are 2635.
(BERTScore = 0.9487)



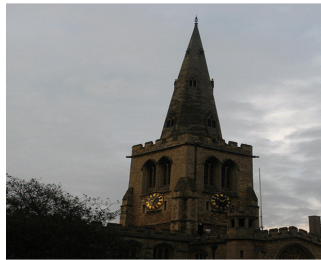
The question is asking about others:
What color are the walls?
The walls are white.



The question is asking about others:
What does the sentence on the top say?
The sentence on the top says this is camping.
(BERTScore = 0.9108)



The question is asking about others:
What color are the gym shoes?
The gym shoes are white.
(BERTScore = 0.9668)



The question is asking about others:
What is the weather like?
The weather like is cloudy.



The question is asking about others:
What type of animal is shown?
Elephant is shown.
(BERTScore = 0.9132)



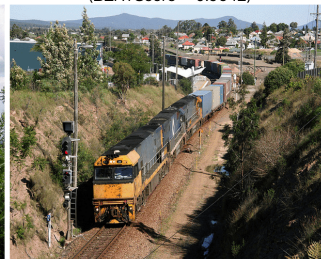
The question is asking about others:
What kind of weather is it?
It is cloudy.
(BERTScore = 0.9642)



The question is asking about yes or no:
Is this a passenger train?
This is a passenger train.



The question is asking about yes or no:
Is there a red flower?
There is not a red flower
(BERTScore = 0.9401)



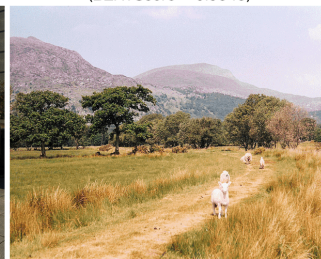
The question is asking about yes or no:
Are there passengers on the train?
There are not passengers on the train.
(BERTScore = 0.9545)



The question is asking about yes or no:
Is it going to rain?
It is going to rain.



The question is asking about yes or no:
Is the bus in a parking space?
The bus is in a parking space.
(BERTScore = 0.9371)



The question is asking about yes or no:
Does it look like it is going to rain?
It does not look like it is going to rain.
(BERTScore = 0.9539)

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.