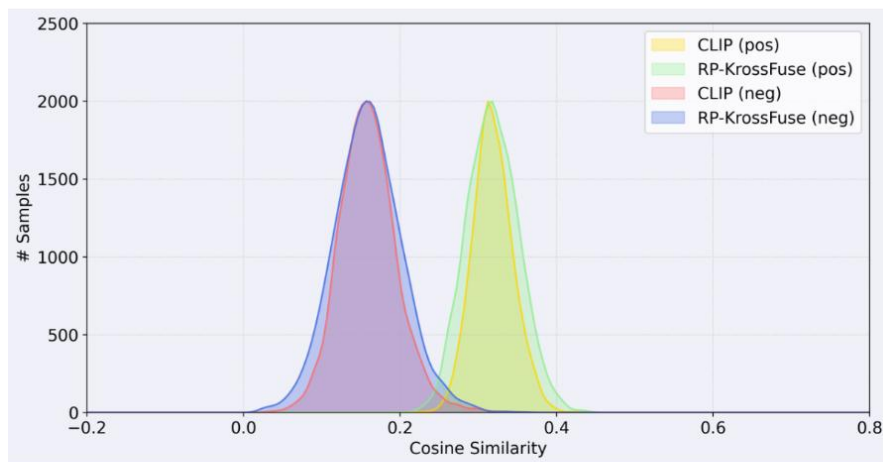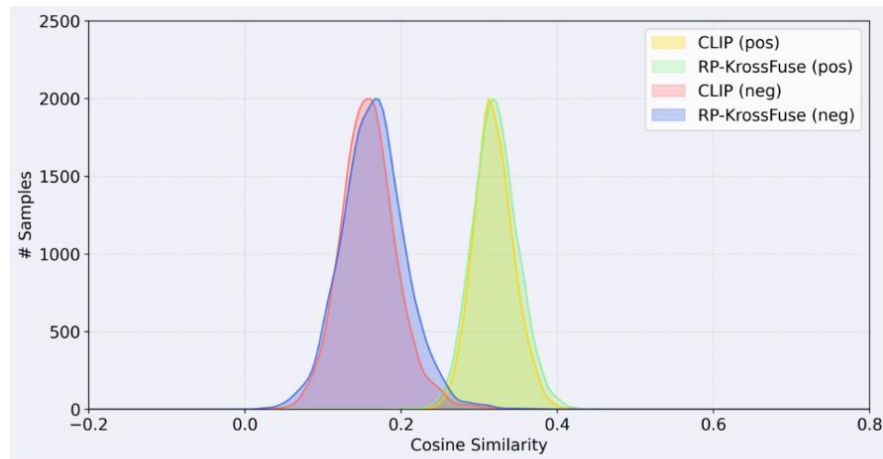C=0.1

C=0.5

C=1.0
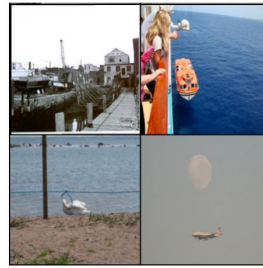
Fig1. Pairwise Cosine Similarities of Positive/Negative Samples in COCO with Different Values of Hyperparameter C. (Response to Review 29Mf, Item 7.)

Joint Cluster 1 (mismatched image and text)

01. There is a phone on top of a calculator
02. A infant holding a baby toothbrush in his hand looking at it.
03. A little boy is playing with a suitcase.
04. A young boy reaching up to grab a red apple.

Joint Cluster 2 (image only, missing text)

Joint Cluster 3 (text only, missing image)

01. A herd of sheep grazing on a lush green field.
02. A dog driving an SUV in an open grass covered field.
03. A man that is standing in the grass near animals.
04. A cow grazing on the grass on the side of a side walk.
05. A lush green forest filled with lots of trees.
06. A horse standing in a lush green grass field.
07. A mother and it's calf standing in a field.
08. a number of horses in a field near a building
09. Two horses grazing on grass in the sun

Joint Cluster 4 (text only, missing image)

01. there is a cat that is drinking out of the toilet
02. A cat sitting in a bathroom sink under a faucet.
03. A small cat sits perched on the back of a toilet.
04. A black cat is sitting in a bathroom sink.
05. A black cat sits in a bathroom sink.
06. a cat that is standing on a toilet and next to a sink
07. A calico cat drinking from a sink faucet.
08. a cat sitting in a bathtub not looking amused
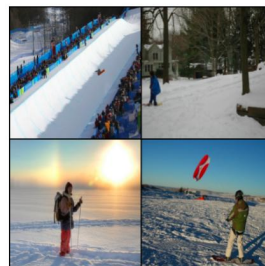09. An orange and white cat laying on top of a sink.

Fig2. The Failure Case of Naive Concatenation in Joint Clustering Results on the COCO Dataset. (Response to Review M1uF, Item 1.)

Joint Cluster 1

01. A man in yellow shirt and black shorts playing a game of tennis.
02. A woman standing on a tennis court holding a racquet.
03. A woman wearing a red tennis outfit holding onto a racquet.
04. A tennis ball and a woman with a racket on a tennis court.

Joint Cluster 2

01. A cross country skier on the slopes on a sunny day
02. Two skiers flipping in the air with people in the background.
03. A group of people standing on ski's in the snow.
04. A man that is standing on ski's in the snow.

Joint Cluster 3

01. A female horse rider mounted near a body of water and mountains.
02. Two jockeys ride their horses in a race at the track.
03. a group of people are with a horse in a stable
04. A horse drawn carriage parked on top of a field.

Joint Cluster 4

01. A group of people riding waves in the ocean.
02. a person parachute surfing on a large body of water
03. A blue, yellow and green surfboard sticking out of a sandy beach.
04. A person on a surf board riding a wave.

Fig3. RP-KrossFuse Demonstrates Full Modality Alignment in Joint Clustering on the COCO Dataset (Response to Review M1uF, Item 1.)