

CLIP ViT-B/32 similarity score mean: 0.254, median: 0.247, min: -0.055, max: 0.520

CLIP ViT-L/14 similarity score mean: 0.208, median: 0.203, min: -0.114, max: 0.524