

Momentum Contrast for Unsupervised Visual Representation Learning

Unsupervised learning has seen a tremendous growth in the development of visual tasks. Various unsupervised methods in deep learning consist of self-supervision, or contrastive learning, wherein the loss function is defined on a pretext task consisting of true labels as intrinsically generated entities. This aids in efficient recognition of images from large datasets and successfully transfer the learned features to downstream tasks such as object detection. The work presents a contrastive learning algorithm namely Momentum Contrast (MoCo). MoCo models contrastive learning as a dictionary-lookup task with the query being the input to the model and the keys being augmentations of the input image. The encoded query is contrasted against the momentum-encoded keys which aid in learning rich representations which are transferable to downstream tasks. MoCo is shown to outperform various supervised learning algorithms on the ImageNet dataset and 7 detection tasks from other large-scale benchmarks.