Review-2

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 from a queue 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 bechmarks.

The MoCo algorithm comprises of a query encoder and a key encoder. The query encoder encodes the visual query and compares it against the encoded keys using the similarity dot product. Keys are generated from augmentations of the input image and encoded using a momentum encoder which is a slow moving average of the query encoder. Utilizing momentum of the query encoder leads to consistency in keys of the dictionary which has proven to be a hindrance in the memor-bank approach and previous works on contrastive learning. Moreover, MoCo makes use of a queue for comparing dynamically-generated keys with the query. This leads to the construction of a larger dictionary in comparison to the simpler end-to-end method which can only accomodate smaller batch-sizes due to memory constraints. Additionally, MoCo makes use of shuffling batch-normalization which shuffles the sample order of the mini-batch before distributional execution. Shuffling batch-normalization prevents the leakage of information in comparison to its standard counterpart in the ResNet architecture.