Self-Supervised Contrastive Learning with NNCLR

Abstract:

Self-supervised learning algorithms based on instance discrimination train encoders to be invariant to pre-defined transformations of the same instance. While most methods treat different views of the same image as positives for a contrastive loss, we are interested in using positives from other instances in the dataset.

Our method, Nearest Neighbour Contrastive Learning of visual Representations (NNCLR), samples the nearest neighbours from the dataset in the latent space, and treats them as positives. This provides more semantic variations than pre-defined transformations. We find that using the nearest-neighbour as positive in contrastive losses improves performance significantly on ImageNet classification using ResNet-50 under the linear evaluation protocol, from 71.7% to 75.6%, outperforming previous state-of-the-art methods.

On semi-supervised learning benchmarks, we improve performance significantly when only 1% ImageNet labels are available, from 53.8% to 56.5%. On transfer learning benchmarks our method outperforms state-of-the-art methods (including supervised learning with ImageNet) on 8 out of 12 downstream datasets. Furthermore, we demonstrate empirically that our method is less reliant on complex data augmentations. We see a relative reduction of only 2.1% ImageNet Top-1 accuracy when we train using only random crops.

Existing System:

Clustering is a class of unsupervised learning methods that has been extensively applied and studied in computer vision. Little work has been done to adapt it to the end-to-end training of visual features on large scale datasets. In this work, they present Deep Cluster, a clustering method that jointly learns the parameters of a neural network and the cluster assignments of the resulting features. Deep Cluster iteratively groups the features with a standard clustering algorithm, k-means, and uses the subsequent assignments as supervision to update the weights of the network. They apply Deep Cluster to the unsupervised training of convolutional neural networks on large datasets like ImageNet and YFCC100M. The resulting model outperforms the current state of the art by a significant margin on all the standard benchmarks.

Proposed System:

Even without being told explicitly what a dodo is, we will likely form associations between the dodo and other similar semantic classes; for instance, a dodo is more similar to a chicken or a duck than an elephant or a tiger. This act of contrasting and comparing new sensory inputs with what one has already experienced happens subconsciously and encoder image nearest neighbour in support set.

We propose a simple self-supervised learning method that uses similar examples from a support set as positives in a contrastive loss. might play a key role in how humans are able to acquire concepts quickly. In this work, we show how an ability to find similarities across items within previously seen examples improves the performance of self-supervised representation learning. A particular kind of self-supervised training – known as instance discrimination has become popular recently. Models are encouraged to be invariant to multiple transformations of a single sample.

This approach has been impressively successful at bridging the performance gap between self-supervised and supervised models. In the instance discrimination setup, when a model is shown a picture of a dodo, it learns representations by being trained to differentiate between what makes that specific dodo image different from everything else in the training set.

Software Tools:

1. TensorFlow
2. Keras
3. Matplotlib
4. VS Code
5. Jupyter Notebook
6. Python3

Hardware Tools:

1. Laptop
2. Operating System: Windows 11
3. RAM: 16GB
4. ROM: 8GB
5. Fast Internet Connectivity

Applications:

1. More similar images can be easily distinguished by using this architecture.
2. Robotics, Deep AI can be benefited by using architecture.