Data efficient image recognition with contrastive predictive coding

Summary

1. Good representations should make the spatio-temporal variability in natural signals more predictable.

2. CPC-trained representation can require 2-5x labels to achieve the same performance as randomly initialized rep.

3. Linear classification on top of the learned representation is not a good metric to measure the quality of the learned rep.

from CPC v1 to v2

1. Increasing model capacity

2. Layer normalization :
   - Using batch normalization allows the model to find trivial solutions to the prediction task of CPC by relying on the batch statistics.
   - Layer normalization allows for training large architecture without this problem.

3. Prediction lengths and directions : predict a patch from four directions (above, below , left, right)

4. Patch-based augmentation :
   - low-level patterns can allow the network to solve the CPC tasks without learning semantically meaningful content.
   - Augmenting the low-level variability can remove such cues.