FID3018 - Motivation

Large-scale labelled data are generally required to train deep neural networks to obtain better performance in visual feature learning from images or videos for computer vision applications. However, collection and annotation of large-scale datasets are time-consuming and expensive. To avoid extensive cost of collecting and annotating large-scale datasets, self-supervised learning methods are proposed to learn general image and video features from large-scale unlabelled data without using any human-annotated labels. I will focus on three papers published in very reputable conferences ICCV, ICLR and CVPR: the first one is the first attempt towards self-supervision in computer vision. The second focuses on a very simple yet very successful approach for self-supervision. Lastly an empirical evaluation of different approaches showing different behavior and patterns.

1. C. Doersch, A. Gupta and A. A. Efros, "Unsupervised Visual Representation Learning by Context Prediction," *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1422-1430, doi: 10.1109/ICCV.2015.167.

2. Nikos Komodakis, Spyros Gidaris. "Unsupervised representation learning by predicting image rotations." *International Conference on Learning Representations (ICLR)*, Apr 2018, Vancouver, Canada.

3.

4. Alexander Kolesnikov, Xiaohua Zhai, Lucas Beyer; "Revisiting Self-Supervised Visual Representation Learning" *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern* Recognition (CVPR), 2019, pp. 1920-1929