Open Internship position: Benchmarking self-supervised video representation

Description

Self-supervised learning is a form of unsupervised learning where the data provide the supervision, typically by withholding some information about the data and the network's task is to predict it. Over the past few years, self-supervised learning from visual data has shown impressive progress both on images [1,2,3] and on videos [4,5,6,7,8]. Several self-supervised objectives have been proposed for videos, ranging from contrastive learning on video clips [4,5], to colorization [6], clip order ranking [7] or future frame prediction [8]. The obtained models are then evaluated on only a couple of specific downstream tasks (i.e. usually action recognition in short video clips and retrieval). However, it is not clear how these models perform on other video tasks and on which aspects of the video content (scene, humans, object, motion, texture, shape, etc.) they tend to focus. The primary goal of this internship is to address this issue. Specifically, the project includes benchmarking different methods on various video tasks that require different understanding of the video content (eg. understanding motion, context, etc.) and studying whether their properties come from the self-supervised objective, the training data or the data augmentation. High-end downstream tasks include scene or object recognition, temporal action detection, etc. The secondary goal of this internship is to exploit the fundings of the benchmark and develop a novel method for self-supervised video representation that leverages the best properties of existing approaches.

Advisors

The internship will be co-supervised by Vicky Kalogeiton, Assistant Professor in the GeoViC team, at the Computer Science Laboratory of Ecole Polytechnique, where the internship will take place, and Philippe Weinzaepfel, Research Scientist in the Computer Vision team at Naver Labs Europe.

Starting date

March-May 2021

Duration

5-6 months

Required skills

– In pursuit of Masters or Doctoral degree in a relevant field
– Proven experience in python
– High level of innovation and motivation
– Hands-on experience with deep learning frameworks
– Communication skills in English

Application

To apply, please contact Vicky Kalogeiton (vicky.kalogeiton@polytechnique.edu) and Philippe Weinzaepfel (philippe.weinzaepfel@naverlabs.com) and include (1) your CV, (2) a short motivation letter, and if applicable (3) your graduate/undergrad transcripts, (4) we may also request the email of two referees. We particularly encourage applications from women, and from underrepresented groups in academia.

References

[1] Momentum contrast for unsupervised visual representation learning. He et al. CVPR'20.
[2] Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. Caron et al. NeurIPS'20.
[3] Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. Noroozi & Favaro. ECCV'16.
[4] TCLR: Temporal Contrastive Learning for Video Representation. Dave et al. arXiv'21.
[5] Can Temporal Information Help with Contrastive Self-Supervised Learning? Bai et al. arXiv 2021.
[6] Tracking Emerges by Colorizing Videos. Vondrick et al. ECCV'18.
[7] Self-supervised Spatiotemporal Learning via Video Clip Order Prediction. Xu et al. CVPR'19.
[8] Unsupervised Video Representation Learning by Bidirectional Feature Prediction. arXiv'20.