ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

Multimodal Contrastive Learning

Representation learning has been greatly improved with the advance of contrastive learning methods with the performance being closer to their supervised learning counterparts. Those methods have greatly benefited from various data augmentations that are carefully designated to maintain their identities so that the images and text transformed from the same instance can still be retrieved.

A stream of "novel" self-supervised learning algorithms have set new state-of-the-art results in AI research. Although recent advances in Contrastive Learning have achieve high performance for text and vision modality, few studies have looked at the representation of the pair image-text in the context of vision and language tasks.

In this project, we aim to understand the latent relation of image-text pair by generate representations of pairs such that similar pairs are near each other and far from dissimilar ones. We will study the effect of different type of augmentation on the representation and study the effect on the performance of different tasks.

You will have access to powerful GPUs, and weekly discussions with two experienced PhD students in deep learning.

Requirements: Strong motivation, proficiency in Python, ability to read papers and work independently. Prior knowledge in deep learning is preferred.

Interested? Please contact us for more details!

Contact

- Zhao Meng: zhmeng@ethz.ch, ETHZ G61.3