**Self-supervised Method for Multi Object Tracking**

Dr. Leulseged Tesfaye and Dr. Mubarak Shah

**Introduction**

Multiple object tracking (MOT) is the task of determining the trajectories of all object instances in a video. Due to different factors such as occlusions, fast-moving objects, or moving camera platforms, MOT is a challenging problem. Several methods formulate tracking problem as a tracking-by-detection paradigm, that fundamentally comprises two separate modules such as object detection and object association. Recently, unlike the predominant tracking approaches, different methods tackle tracking problem with a unified deep model that solve both the detection and association tasks. However, these methods are fully supervised and heavily rely on manually annotated training data. Underlying the nature of the MOT problem and the recent advance in self-supervised methods [1], in this project, we will investigate the tracking problem in a self-supervised fashion.

**Task**

In this project, you will first learn how the state-of-the-art MOT method [2] tackles the tracking problem in Waymo Dataset [3]. Towards this end, you first understand Waymo tracking challenges [3] and the dataset [3]. You then integrate a recently proposed contrastive based representation learning SimCLR [1], self-supervised method, in [2]. Here is the main motivation; in a difficult scenario, the objects in the upcoming video frames are either partially occluded, distorted, or changed in illumination. Thus, if we apply distinct augmentations on objects in time t - n, there will be a high chance that the new instances will resemble the upcoming object in frame t. As can be observed in Figure 2, you will employ the clustering method to determine the right association among objects in a pair of frames. To do that, you first generate a new set of instances for each object by applying different augmentations as in SimCLR [1]. Then, you will apply clustering on the bipartite graph that is built by using the newly generated instance of objects in the given pair of frames. The ultimate goal of this project is to avoid the use of manually labeled data and endowing a competitive result comparing to the counter supervised MOT models.





Figure 2. Illustration of the proposed work

**Project Steps**

- You first understand SimCLR [1] and [2], including their implementations.

- Then, you will understand the Waymo tracking dataset and challenge.

- Finally, you will adopt the concept of SimCLR [1] in [2] as illustrated in Figure 2.

**Implementation**

-The project will be implemented in Pytorch, deep learning library.

**References**

1. Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey E. Hinton. **A Simple Framework for Contrastive Learning of Visual Representations,** *arXiv preprint*[arXiv:2002.05709](https://arxiv.org/abs/2002.05709)**, 2020**
2. Xingyi Zhou, Vladlen Koltun, Philipp Kr. **Tracking Objects as Points**, *arXiv preprint* [arXiv:2004.01177](https://arxiv.org/abs/2004.01177), 2020
3. Waymo tracking Challenge: https://waymo.com/open/challenges/3d-tracking/