Segmenting and Tracking Every Pixel

Andreas Geiger

University of Tübingen

The task of assigning semantic classes and track identities to every pixel in a video is called video panoptic segmentation. In this talk, I will present recent work that targets this task in a real-world setting requiring dense interpretation in both spatial and temporal domains. As the ground-truth for this task is difficult and expensive to obtain, existing datasets are either constructed synthetically or only sparsely annotated within short video clips. To overcome this, we introduce a new benchmark encompassing two datasets, KITTI-STEP, and MOTChallenge-STEP. The datasets contain long video sequences, providing challenging examples and a test-bed for studying long-term pixel-precise segmentation and tracking under real-world conditions. We further propose a novel evaluation metric Segmentation and Tracking Quality (STQ) that fairly balances semantic and tracking aspects of this task and is more appropriate for evaluating sequences of arbitrary length. Finally, we provide several baselines to evaluate the status of existing methods on this new challenging dataset.