Shreyus Puthiyapurail
014577264

Short Story Proposal:

Deep learning in video multi object tracking – a survey

In this article we describe how deep learning models are used to solve the task of Multiple object tracking on single camera videos. Four main steps involved in MOT algorithm are identified and reviewed how learning was employed in each of those steps. Similarities among the top performing methods are presented with some future research directions.

Multiple object tracking is a computer vision task which aims to analyze the videos in order to identify and track different objects without prior knowledge and those may belong to different categories such as cars, pedestrians or animals and the output is a target id associated rectangular bounding boxes identified by their coordinates, height and width. Many of the problems such as autonomous cars, recognition of crowd behavior analysis can be benefitted from this high quality tracking algorithm.

This article gives a brief introduction of MOT algorithm and steps followed. Also talks about various metrics used test and compare the algorithms and provides the description of some of the most important MOT challenge benchmark datasets which have been used since past few years.

Further the article describes how various deep learning methods are used in four MOT stages. Various DL steps in detection stage such as faster R-CNN, SSD, and other detectors are explained in brief.
DL steps involved in feature extraction and motion prediction such as autoencoders, CNN's as visual feature extractors and motion predictors, Siamese networks and some more complex approaches for visual feature extractions.
DL steps in affinity computation which uses Recurrent Neural networks ,LSTMs, Siamese LSTMs, CNNs, Siamese CNNs and other recurrent networks.
DL in association/tracking steps which describes the use of Recurrent Neural Networks, deep multi-layer perceptrons and deep reinforcement learning agents.
Also describes some other uses of DL in MOT.

The above mentioned DL techniques are analyzed and compared as how the setup is done and the results are discussed briefly to give an idea of comparative performance.
The article also describes about the best approaches and finally concluded along with future directions.