

Towards Real-Time Systems for Vehicle Re-Identification, Multi-Camera Tracking and Anomaly Detection

Neehar Peri¹, Pirazh Khorramshahi¹, Sai Saketh Rambhatla¹, Vineet Shenoy¹, Saumya Rawat¹,
Jun-Cheng Chen² and Rama Chellappa¹

¹Center for Automation Research, UMIACS, University of Maryland, College Park

²Research Center for Information Technology Innovation, Academia Sinica

{peri, pirazhkh, rssaketh, vshenoy, saumya, rama}@umi.acs.umd.edu,

pul1pul1@ci.ti.sinica.edu.tw

Abstract

Vehicle re-identification, multi-camera vehicle tracking, and anomaly detection are essential for city-scale intelligent transportation systems. Both vehicle re-id and multi-camera tracking are challenging due to variations in aspect-ratio, occlusion, and orientation. Robust re-id and tracking systems must consider small scale variations in a vehicle's appearance to accurately distinguish among vehicles of the same make, model, and color. Scalability is critical for multi-camera systems, as the number of objects in a scene is not known a-priori. Anomaly detection presents a unique challenge due to a dearth of annotations and varied video quality. In this paper, we address the task of vehicle re-id by introducing an unsupervised excitation layer to enhance representation learning. We propose a multi-camera tracking pipeline leveraging this re-id feature extractor to compute a distance matrix and perform clustering to obtain multi-camera vehicle trajectories. Lastly, we leverage background modeling techniques to localize anomalies such as stalled vehicles and collisions. We show the effectiveness of our proposed method on the NVIDIA AI City Challenge, where we obtain 7th place out of 41 teams for the task of vehicle re-id, with an mAP score of 66.68% and achieve state-of-the-art results on the Vehicle-ID dataset. We also obtain an IDF1 score of 12.45% on multi-camera vehicle tracking, and an S4 score of 29.52% for task of anomaly detection, ranking in the top 5 for both tracks.

1. Introduction

In recent years, there has been great demand to develop automated and intelligent transportation systems for smart cities that can facilitate dynamic traffic routing, traffic plan-

ning, gathering vehicle-specific analytics like speed [17], and traffic anomaly detection. Moreover, the development of Deep Convolutional Neural Networks (DCNNs) has enabled the development of effective solutions to these challenges. For the past three years, NVIDIA AI City Challenge has pushed the boundaries of intelligent transportation systems. In this paper, we present a deep learning-based algorithm for the task of vehicle re-identification (re-id), and end-to-end pipelines for Multi-Camera Tracking (MTC) and anomaly detection.

Vehicle re-id refers to the task of identifying all true matches of a given vehicle identity in a large gallery set composed of images of different vehicles that are captured under diverse conditions, e.g., different image quality, orientation, weather condition and lightening. Therefore, learning robust representations able to handle the aforementioned conditions is of great importance. At the same time, a representation learning algorithm should be both real-time and scalable to adapt to a large number of vehicles and traffic cameras in the wild. To this end, we propose the fast and accurate Excited Vehicle Re-identification (EVER) model to meet these challenges. Recent work has shown the importance of attending to local regions, vehicle key-points, [13, 37] and part bounding boxes [9] to create robust deep features. However, generating key-point annotations and part bounding boxes is costly and will not scale across different domains. [14] has proposed a novel self-supervised model to generate residual maps that act as pseudo-attention maps. In this work, we take advantage of the residuals generated from [14] to excite intermediate feature maps during the course of training and encourage the feature extraction model to learn robust representations.

Multi-Camera Tracking aims to determine the position of objects under consideration, at all times from video streams taken by multiple cameras. The resulting multi-camera trajectories enable applications including visual an-

The first three authors equally contributed to this work.

