

# **Computer Vision and Multi-object Tracking for Traffic Measurement from Campus Monitoring Cameras**

**Yalong Pi,<sup>1</sup> Nick Duffield,<sup>2</sup> Amir H. Behzadan,<sup>3</sup> and Tim Lomax<sup>4</sup>**

<sup>1</sup>Operation Data Scientist, Institute of Data Science, Division of Research, Texas A&M University, College Station, TX 77843; e-mail: piyalong@tamu.edu

<sup>2</sup>Professor, Department of Electrical and Computer Engineering & Institute of Data Science, Texas A&M University, College Station, TX 77843; e-mail: duffieldng@tamu.edu

<sup>3</sup>Associate Professor, Department of Construction Science, Texas A&M University, College Station, TX 77843; e-mail: abehzadan@tamu.edu

<sup>4</sup>Research Fellow, Texas A&M Transportation Institute, Texas A&M University System, College Station, TX 77843-3135; e-mail: t-lomax@tti.tamu.edu

## **ABSTRACT**

Conventional methods of traffic monitoring primarily rely on sensors, on-board GPS, and human observation, which are generally resource-intensive, slow, and subject to implementation limits. Computer vision (particularly convolutional neural networks) is a more flexible, convenient, and cost-effective alternative to extract traffic information from traffic videos. This paper presents a methodology to extract traffic volume counts from cameras on Texas A&M University (TAMU) campus. Particularly, an object detector model, namely YOLOv5 (You Only Look Once), is integrated with an object tracking algorithm (Deep-SORT) to track vehicles in camera views. These tracks are projected using homography transformation from the local coordinate system of the camera onto an orthogonal map (world coordinates) along with information such as unique vehicle ID, type, position, and time stamp. This data is used to count vehicles that cross the pre-defined study zones (lanes) on different directions. A total of 60 samples (each 1 minute long) recorded during a 24-hour period on TAMU campus are manually annotated and used to benchmark model predictions, resulting in a traffic volume count accuracy ranging from 68% to 93%. Moreover, the best input pixel resolution is determined to be 1280\*1280 by measuring the average precision for each class of interest.