Technical Report 150

Deep Learning Methods to Leverage Traffic Monitoring Cameras for Pedestrian Data Applications

Research Supervisor: Natalia Ruiz Juri
Center for Transportation Research

Project Title: Transit Policy in the Context of New Transportation Paradigms

September 2019

Data-Supported Transportation Operations & Planning Center (D-STOP)

A Tier 1 USDOT University Transportation Center at The University of Texas at Austin

D-STOP is a collaborative initiative by researchers at the Center for Transportation Research and the Wireless Networking and Communications Group at The University of Texas at Austin.

Technical Report Documentation Page

1. Report No. D-STOP/2019/150	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Deep Learning Methods to Leverage Traffic Monitoring Cameras for Pedestrian Data Applications		5. Report Date May 2019 6. Performing Organization Code
7. Author(s) Weijia Xu, Natalia Ruiz-Juri, Ruizhu Huang (The University of Texas at Austin); Jennifer Duthie, Joel Meyer, John Clary (City of Austin Transportation Department)		8. Performing Organization Report No. Report 150
9. Performing Organization Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, Texas 78701		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRT13-G-UTC58
12. Sponsoring Agency Name and Address United States Department of Transportation University Transportation Centers 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes
Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.
Project Title: Transit Policy in the Context of New Transportation Paradigms

16. Abstract
Transportation agencies often own extensive networks of monocular traffic cameras, which are typically used for traffic monitoring by officials and experts. While the information captured by these cameras can also be of great value in transportation planning and operations, such applications are less common due to the lack of scalable methods and tools for data processing and analysis. This paper exemplifies how the value of existing traffic camera networks can be augmented using the latest computing techniques. We use traffic cameras owned by the City of Austin to study pedestrian road use and identify potential safety concerns. Our approach automatically analyzes the content of video data from existing traffic cameras using a semi-automated processing pipeline powered by the state-of-art computing hardware and algorithms. The method also extracts a background image at analyzed locations, which is used to visualize locations where pedestrians are present, and display their trajectories. We also propose quantitative metrics of pedestrian activity which may be used to prioritize the deployment of pedestrian safety solutions, or evaluate their performance.

17. Key Words Pedestrian safety, road usage, video recognition and analysis	18. Distribution Statement No restrictions. This document is available to the public through NTIS (http://www.ntis.gov): National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161		
19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages 14	22. Price

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgements

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers.

This work is based on data provided by the City of Austin, which also provided partial support for this research. The authors are grateful for this support. We would like to thank Kenneth Perrine and Chris Jordan for their help in setting up video recording environment. We would also like to thank City of Austin staff Joel Meyer and John Clary for their feedback during the project. The computation of all experiments was supported by the National Science Foundation, through Stampede2 (OAC-1540931), and XSEDE (ACI-1953575) awards.

Paper ID # AM-TP2323

Deep learning methods to leverage traffic monitoring cameras for pedestrian data applications

Weijia Xu [1], Natalia Ruiz-Juri [2*], Ruizhu Huang[1], Jennifer Duthie[3], Joel Meyer[3], John Clary[3]

1. Texas Advanced Computing Center, The University of Texas at Austin, USA

2. Center of Transportation Research, The University of Texas at Austin, USA

3. Austin Transportation Department, City of Austin, USA

Abstract

Transportation agencies often own extensive networks of monocular traffic cameras, which are typically used for traffic monitoring by officials and experts. While the information captured by these cameras can also be of great value in transportation planning and operations, such applications are less common due to the lack of scalable methods and tools for data processing and analysis. This paper exemplifies how the value of existing traffic camera networks can be augmented using the latest computing techniques. We use traffic cameras owned by the City of Austin to study pedestrian road use and identify potential safety concerns. Our approach automatically analyzes the content of video data from existing traffic cameras using a semi-automated processing pipeline powered by the state-of-art computing hardware and algorithms. The method also extracts a background image at analyzed locations, which is used to visualize locations where pedestrians are present, and display their trajectories. We also propose quantitative metrics of pedestrian activity which may be used to prioritize the deployment of pedestrian safety solutions, or evaluate their performance.

Keywords:

Pedestrian safety, road usage, video recognition and analysis

Introduction

Incorporating Internet of Things (IoT) and smart devices within an intelligent transportation system (ITS) usually comes with substantial up-front costs for installation and deployment. At the same time, advances in algorithm development and software design bring new opportunities to increase utilization of existing transportation infrastructure. In this paper, we present an approach that utilizes existing traffic monitoring cameras within an intelligent transportation system to understand pedestrian movement patterns and safety.

Due to their low maintenance and operational cost, video sensors, such as pan-tilt-zoom (PTZ) cameras, are commonly installed along freeways and arterial streets [1]. However, the use of video data from these cameras for system performance/safety assessment or strategic planning is not widespread. Transportation Management Centers (TMCs) primarily use traffic video data from roadside cameras to identify incidents, prepare the response for emergency situations, manage traffic in special events, and

dispatch technicians for maintenance [2]. The video data is also used to manually conduct traffic studies, including collecting traffic counts by mode, turning movement counts for traffic signal timing applications, and conducting safety analysis by observing the behavior of traffic in weaving zones [3]. Such applications are usually labor intensive, and impractical for large-scale implementation.

While traffic video data analysis software tools exist, they are mostly used to support real-time traffic operations, commonly focusing on one type of analysis, and often deployed in dedicated, specialized hardware. Examples of video data use include safety analysis for intersections and corridors [4–6], identification of unusual events on corridors, such as wrong-way driving and stalled vehicles [7], generation of traffic statistics including counts and queue lengths, and for vehicular emission analysis by estimating traffic speeds [8].

While possible, the analysis of historical video camera data is not common in practice due to the significant storage and computing resources required to support it. Traffic monitoring data is often discarded after pre-specified time periods ranging from one day to one year, depending on the recording purpose[2].

In this paper we propose a flexible framework for collecting and analyzing videos from existing traffic monitoring cameras. We present a prototype pipeline for traffic camera video content recognition and analysis, and explore its use to support pedestrian safety analysis. The proposed framework is more general than traditional traffic video analysis tools, typically designed to accomplish a single type of analysis. Further, our approach separates the expensive computational steps of object recognition from the subsequent data intensive analysis, allowing using different hardware and software resources at each stage for maximum efficiency.

The proposed use case is selected because of the significant challenges in systematically studying and evaluating pedestrian safety and activity patterns in the transportation system; the latter is critical for transportation planners and policy makers. As an example, transportation agencies often make substantive changes to a wide range of built environment features seeking to foster physical activity. Walking is one of the most sustainable modes of transportation, and promoting walking can contribute to the development of healthy and livable communities. However, pedestrians are the most vulnerable group among all non-motorized modes, and endure the highest share of fatal road collisions.

Pedestrian safety analysis involves identifying factors leading to unsafe conditions at a particular location, and has traditionally been conducted based on the judgment and experience of traffic safety professionals. The collection and analysis of video data at critical locations provides an opportunity to capture and analyze traffic conflicts based on a permanent, verifiable account of road user behavior. The former reduces the need to rely on ad hoc decision making [9]. However, if analyses are conducted by human observers, there is a limitation in the number of locations and analysis periods that may be

considered. Automated approaches to effectively recognize, analyze and store pedestrian activities over time are needed. The technical challenges associated with pedestrian activities analysis using traffic monitoring video data are different from those faced when conducting traffic flow analyses. Regular roadside cameras are installed to have wide and deep fields of view. Pedestrian activities only occupy a small portion of the view, and at many locations are only present sporadically. Further, pedestrians are smaller than cars, and are more frequently subject to obstruction from other objects within the scene.

This paper describes an effort to extend a framework already tested for traffic analysis [10] to the study of pedestrian travel. The prototype application analyzes video recordings over time and generates two types of visual summaries of pedestrian activities: a visualization of locations where pedestrians are present, and a display of their trajectories. The software tool capabilities and potential applications are exemplified using camera data gathered from actual locations in the City of Austin.

Methodological Approach and Implementation

We have proposed a framework that separates the video analysis process into two distinct parts: object recognition and analysis of recognized objects [10]. The approach uses convolutional neural networks to detect and track the motion of objects from each frame in the video stream, and then store and process information using Spark programming framework for scalability [10]. By combining the best practice of object recognition through deep learning and big data processing through Spark the framework can efficiently process large volume traffic video data and meet evolving analytic needs over time.

Video collection and processing pipeline

To implement the framework, we have set up a multi-systems cross-domain video aggregation and analysis pipeline (Figure-1). Raw videos are originated from IP cameras in the City of Austin (CoA) private network, which has limited accessibility. To overcome this, the CoA set up a proxy server to forward selected video feeds from the IP cameras to a storage cluster hosted at the Texas Advanced Computing Center (TACC). The recorded video can be then be processed by another high-performance computing cluster at TACC. Processed data is saved in the storage server, which is accessed by our project server for results dissemination purposes. The project server also hosts tools and scripts to schedule video recording and processing tasks.

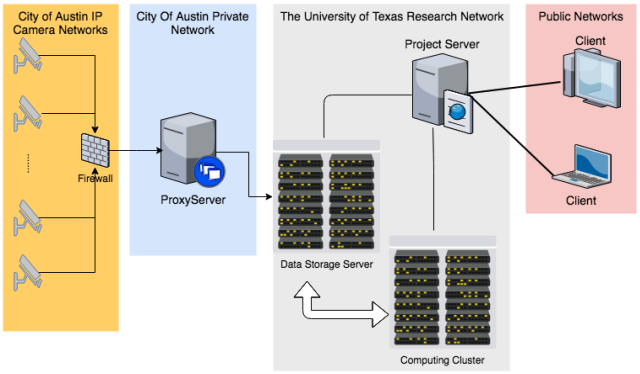

Figure-1 Camera access and processing pipeline overview

Pedestrian recognition and activity detection

The proposed processing approach consists of two main steps: the *video content recognition* step identifies and labels all physical objects from original input video files using a deep-learning based algorithm; the second step is *object tracking*, which "follows" each recognized object across all frames in the input video.

Our video content recognition process is based on Darknet, an open source library of image recognition [10, 11]. The core algorithm utilizes a convolution-neural-network-based object detection system, YOLOv2, to analyze each frame of an input video [11]. For each frame, the algorithm outputs a list of objects including their location in the frame, class label, and confidence of recognition. We have limited recognition to seven class labels that are most relevant, including person, car, bus, truck, bicycle, motorcycle, and traffic light. To improve algorithmic performance and maximize utilization of multi-node computing clusters, we have also adapted the YOLO implementation for parallel execution [10]. Our implementation enables parallel object recognition on multiple frames using pthread within individual compute nodes, and using MPI for inter-node communication. Specifically, one thread is used to pre-fetch *n* frames, while *n* extra worker threads are assigned to labeling. Since each worker thread is independent, near-ideal linear scaling can be achieved for longer videos [10]. For video recordings from different times/locations, multiple video files can be processed independently across multiple nodes concurrently. A non-maximum suppression (NMS) algorithm with the locally maximal confidence measure is used to remove unnecessary/duplicated objects. In addition to content recognition, the framework outputs a background image (i.e. non-moving features) from each video recording. For more details about the original YOLO algorithm and our implementation, please refer to [11] and [10], respectively.

To track object across video frames, we compare recently recognized objects with previously recognized objects. Particularly, we use background subtraction techniques to differentiate moving objects from still ones or background. Redundant objects are also filtered out within this component.

Pseudo code for Pedestrian Tracking
Input: N = {n_{ij} | i: frame index, j: object index} as the set of recognized objects found in each frame
Output: T= { t_{ij} | i: trajectory index, j: object index within this trajectory} as the set of objects stored by a list of trajectories
1: Initialize *T* with each object found in the starting frame
2: for each n_{ij} in *N*
3: for each t_k in *T*
4: dists <= distance(n_{ij}.location, pred(t_k, *i*))
5: if min(dists) < *threshold*
6: add n_{ij} to $t_{argMin(min\_dists)}$
7: else add n_{ij} as a new trajectory

Figure-2 Pseudo code for tracking pedestrians.

To track pedestrians in particular, we propose an approach based on predicted positions of objects from previous frames (Figure-2). The algorithm is initialized with the set of recognized "person" objects in each frame. For each recognized object in the first frame, we initialize a trajectory for that object. Recognized objects in the subsequent frame are associated to the closest objects from the previous frame. Once a trajectory has more than two distinct positions, direction and velocity of the trajectory can be estimated. In subsequent frames, we compute the distance between all identified objects and the predicted positions of existing trajectories at that frame. If the distance between an object and a trajectory is larger than a pre-defined threshold, the algorithm will generate a new trajectory. Otherwise, the object position is added to the trajectory whose predicted position is the closest.

A complete list of all tracked objects with corresponding detailed information is stored in a structured data file for further study. The result files are subject to additional analysis and visualization. The pedestrian safety case study presented below was conducted using a Spark program to read and process results files from multiple video recordings. The detection and tracking of pedestrians can also be exported as a delimited file for further analysis. Figure-3 shows an example of pedestrian-crossing- road events. Each row represents a pedestrian track detected through the algorithm, and includes information on file names, size of the tracks, start and end frames, and start and end locations on the video. For the crossing event detection, there is an additional column appended at the end to indicate if the track is considered as a pedestrian crossing event or not.

file	size	end_frame	end_xmin	end_ymin	end_xmax	endy_max	from_frame	from_xmin	from_ymin	from_xmax	from_ymax	isCrosing
/work/03076/rhuang/maverick/CTR/A	222	26732	962	386	977	431	26196	946	370	960	419	FALSE
/work/03076/rhuang/maverick/CTR/A	74	25128	315	391	331	437	24536	158	472	179	526	FALSE
/work/03076/rhuang/maverick/CTR/A	52	24553	168	689	222	719	24305	30	542	55	608	FALSE
/work/03076/rhuang/maverick/CTR/A	209	23056	0	549	28	645	22789	20	524	51	606	FALSE
/work/03076/rhuang/maverick/CTR/A	292	22832	26	521	62	606	21967	190	458	209	523	FALSE
/work/03076/rhuang/maverick/CTR/A	443	11802	995	386	1014	463	10250	60	468	84	517	TRUE
/work/03076/rhuang/maverick/CTR/A	691	5594	948	376	964	429	3830	980	456	1007	499	FALSE
/work/03076/rhuang/maverick/CTR/A	339	3808	980	458	1005	501	2935	1011	482	1035	538	FALSE
/work/03076/rhuang/maverick/CTR/A	116	2198	1028	519	1055	558	1769	1040	543	1073	617	FALSE
/work/03076/rhuang/maverick/CTR/A	70	1670	1047	540	1082	639	1500	1052	553	1085	631	FALSE
/work/03076/rhuang/maverick/CTR/A	109	1760	1038	544	1071	619	1491	1051	556	1080	608	FALSE
/work/03076/rhuang/maverick/CTR/A	139	1346	1059	574	1099	667	1125	1058	592	1102	673	FALSE
/work/03076/rhuang/maverick/CTR/A	694	1618	1052	546	1083	645	628	1114	620	1154	688	FALSE
/work/03076/rhuang/maverick/CTR/A	315	818	1099	610	1139	690	153	993	405	1013	456	FALSE

Figure-3 tracking result examples

Case study: understanding the location of frequent pedestrian street-crossing.

The use case analyzed for this application consists of identifying locations where pedestrians frequently

cross a street. This is an important step towards understanding the impact of measures designed to promote street crossing on designated safe areas, such as crosswalks. We selected several camera locations in Austin, Texas, and used the video aggregation pipeline to record video segments throughout the daylight time. These locations include sites where a pedestrian hybrid beacon (PHB) will be or has been recently installed, and where data collection can support impact assessment.

Based on our processing method results, we propose a quantitative metric and qualitative visual data representations to support a better understanding of time-dependent pedestrian activity patterns.

Quantifying pedestrian activity: The Activity Index

Equation 1 defines the *activity index*, which is computed based on the detection results with the goal of summarizing pedestrian presence at the analyzed location over a pre-defined time period.

$$Activity\ Index\ (AI) = \frac{Number\ of\ Person\ Idenfication}{Number\ of\ Frames\ of\ videos} \quad [1]$$

The activity index can be used as a singular numerical indicator of pedestrian presence. A higher value indicates more pedestrian activity in a video recording per unit time. The measure can be used for comparison purposes across different locations and times of day. Figure 4 shows the evolution of the activity index between 5 a.m. and 9 p.m. on Lamar and 24th Street.

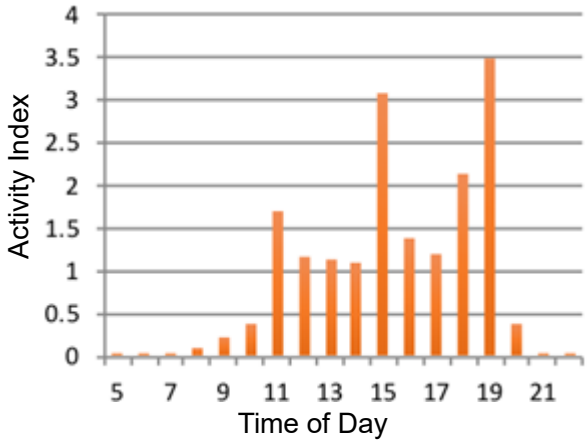

Figure-4 Activity Index by time of day for Lamar & 24th Street on April 28, 2018.

Visual comparison of pedestrian activity patterns over time.

To help users intuitively understand pedestrian road use over time, we propose the visual representation presented in Figure-5. The analyze data was collected on Oct. 1, 2018 at the intersection of Ashdale Drive and Burnet Road, using a camera located at Burnet Road and Anderson Lane.

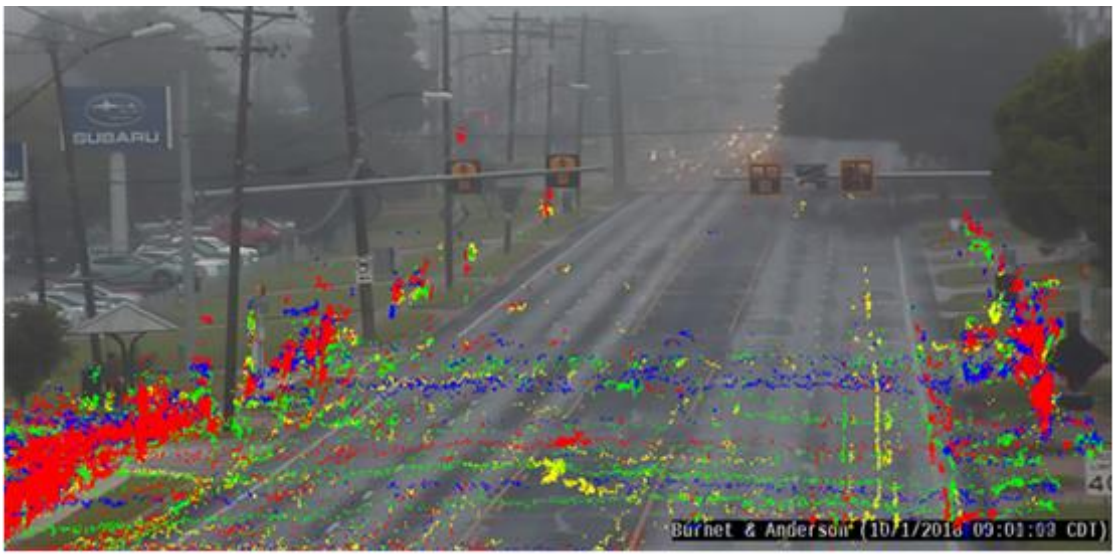

Figure-5 visual summary of pedestrian activity patterns by time of day on Oct 1. 2018 at Burnet and Ashdale.

Figure-5 presents the location of detected pedestrians for four time periods, represented using different colors. Colors yellow and red correspond to the a.m. and p.m. peak periods, respectively. In the legend we also present the actual number of crossing events identified during each time period.

Visual representations of pedestrian activity level.

Figure-5 suggests that there is more pedestrian activity on sidewalks than crossing events. The former is partly explained by the presence of two bus stops on each side of the road at the selected location. To understand pedestrian activity level by location, we propose the heatmap presented in (Figure-6).

Figure-6 heatmap view to indicate where pedestrians have been detected the most over Oct.1 and Oct. 2 2018

In Figure-6, the frequency of pedestrian appearances (pedestrian activity level, defined as the fraction of frames where a pedestrian object is detected on any given pixel) is indicated through a color map. The results shown are consistent with expectations of where people spend time at bus stops. Figure-6 also shows a small region in the middle of the road with high level of pedestrian activity, which is consistent with pedestrians stopping in the middle of the road during road-crossing events. The latter is a potential safety concern that requires further monitoring.

Discussion

Artificial intelligence technologies can greatly reduce the effort involved in analyzing video data, and frameworks such as the one presented here can facilitate research traditionally based on manual video data analysis, and promote further work on video data applications and integration. A unique advantage of our framework is to convert video recordings into queryable information, which can accommodate multiple subsequent use cases without re-processing [12]. While the framework and specific applications are still under development, we have exemplified their potential to support useful analyses with minimal effort compared to manual processing.

The approach provides a space-saving alternative for raw video data storage, as the output of recognized objects can be much smaller than the raw video files. The storage requirement is significantly reduced when the raw video is no longer needed, and the data becomes anonymized, since identifiable information is not stored with recognized objects. Our method facilitates the preservation of useful key traffic information for large regions in the long term.

An additional benefits of the proposed approach is that processed data can be combined with other

datasets to conduct more complex analyses. For example, video data may be combined with loop detector data and signal timing data to understand pedestrian compliance with traffic signals. Traffic data from Bluetooth or Wavetronix sensors may support a more comprehensive assessment of pedestrian behavior by providing contextual information including prevalent vehicle speeds and traffic volumes.

The use cases presented in this work illustrate the benefits and limitations of the proposed methodology. Our video aggregation pipeline has the potential to support long-term road usage monitoring. The flexibility of the data selection and filtering capabilities is expected to enables further applications. In addition to the visual summaries described in this study, quantitative outputs can be generated to facilitate the comparison of conditions across different locations or time ranges, and to evaluate the impact of infrastructure changes and construction scenarios, among others.

References

[1] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," Image and Vision Computing, 2003.

[2] S. Kuciemba and K. Swindler, "Transportation Management Center Video Recording and Archiving Best General Practices," 2016.

[3] S. Zangenehpour, L. F. Miranda-Moreno, and N. Saunier, "Automated classification based on video data at intersections with heavy pedestrian and bicycle traffic: Methodology and application," Transportation Research Part C: Emerging Technologies, 2015.

[4] W. Hu, X. Xiao, D. Xie, T. Tan, and S. Maybank, "Traffic accident prediction using 3-D model-based vehicle tracking," IEEE Transactions on Vehicular Technology, 2004.

[5] P. St-Aubin, L. Miranda-Moreno, and N. Saunier, "An automated surrogate safety analysis at protected highway ramps using cross-sectional and before-after video data," Transportation Research Part C: Emerging Technologies, 2013.

[6] P. St-Aubin, N. Saunier, and L. Miranda-Moreno, "Large-scale automated proactive road safety analysis using video data," Transportation Research Part C: Emerging Technologies, 2015.

[7] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," IEEE Transactions on Circuits and Systems for Video Technology. 2008.

[8] B. T. Morris, C. Tran, G. Scora, M. M. Trivedi, and M. J. Barth, "Real-time video-based traffic measurement and visualization system for energy/emissions," IEEE Transactions on Intelligent Transportation Systems, 2012.

[9] T. Sayed, M. H. Zaki, and J. Autey, "Automated safety diagnosis of vehicle–bicycle interactions using computer vision analysis," Safety science, vol. 59, pp. 163–172, 2013.

[10] L. Huang, W. Xu, S. Liu, V. Pandey, and N. R. Juri, "Enabling versatile analysis of large scale traffic video data with deep learning and HiveQL" *in proceedings of Big Data (Big Data), 2017 IEEE International Conference on*, 2017, pp. 1153–1162.

[11] D. Impiombato et al., "You Only Look Once: Unified, Real-Time Object Detection," in IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[12] W. Xu, N. R. Juri, R. Huang, J. Duthie, and J. Clary. " Automated pedestrian safety analysis using data from traffic monitoring cameras" In *proceedings of 1st ACM/EIGSCC Symposium On Smart Cities and Communities (SCC '18)*, June 20--22, 2018, Portland, OR, USA, ACM, New York