2023

Development,Validation, and Integration of AI-Driven Computer Vision System and Digital-twin System for Traffic Safety Dignostics

Ou Zheng
University of Central Florida

DEVELOPMENT, VALIDATION, AND INTEGRATION OF AI-DRIVEN COMPUTER VISION AND DIGITAL-TWIN SYSTEMS FOR TRAFFIC SAFETY DIAGNOSTICS

By

OU ZHENG

B.S., Stetson University,2017

M.S., University of Central Florida, 2019

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering

in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term, 2023

Major Professor

Mohamed Abdel-Aty

ABSTRACT

The use of data and deep learning algorithms in transportation research have become increasingly popular in recent years. Many studies rely on real-world data. Collecting accurate traffic data is crucial for analyzing traffic safety. Still, traditional traffic data collection methods that rely on loop detectors and radar sensors are limited to collect macro-level data, and it may fail to monitor complex driver behaviors like lane changing and interactions between road users. With the development of new technologies like in-vehicle cameras, Unmanned Aerial Vehicle (UAV), and surveillance cameras, vehicle trajectory data can be collected from the recorded videos for more comprehensive and microscopic traffic safety analysis. This research presents the development, validation, and integration of three AI-driven computer vision systems for vehicle trajectory extraction and traffic safety research: 1) A.R.C.I.S, an automated framework for safety diagnosis utilizing multi-object detection and tracking algorithm for UAV videos. 2)N.M.E.D.S., A new framework with the ability to detect and predict the key points of vehicles and provide more precise vehicle occupying locations for traffic safety analysis. 3)D.V.E.D.S applied deep learning models to extract information related to drivers' visual environment from the Google Street View (GSV) images. Based on the drone video collected and processed by A.R.C.I.S at various locations, CitySim: a new drone recorded vehicle trajectory dataset that aim to facilitate safety research was introduced. CitySim has vehicle interaction trajectories extracted from 1140- minutes of video recordings, which provide a large-scale naturalistic vehicle trajectory that covers a variety of locations, including basic freeway segments, freeway weaving segments, expressway segments, signalized intersections, stop-controlled intersections, and unique intersections without sign/signal control. The advantage of CitySim over other datasets is that it contains more critical safety events in quantity and severity and provides supporting

scenarios for safety-oriented research. In addition, CitySim provides digital twin features, including the 3D base maps and signal timings, which enables a more comprehensive testing environment for safety research, such as autonomous vehicle safety. Based on these digital twin features provided by CitySim, we proposed a Digital Twin framework for CV and pedestrian in-the-loop simulation, which is based on Carla-Sumo Co-simulation and Cave automatic virtual environment (CAVE). The proposed framework is expected to guide the future Digital Twin research, and the architecture we build can serve as the testbed for further research and development.

ACKNOWLEEGMENT

First of all, I would like to express my deepest gratitude to my PhD advisor, Dr. Mohamed Abdel-Aty, for their unwavering support and guidance throughout my academic journey. His expertise, encouragement, and encouragement have been invaluable in shaping my research and guiding me to success. And I am forever grateful for their wisdom, dedication, and generosity, and I am proud to have had the privilege of working with them. Thank you for everything. The second thing I need to appreciate Dr. Aty is his support for my exploring with unfunded idea. Whenever I have new idea out of scope of my exiting research and project, he is always there supporting me.

Secondly, I would like to appreciate my committee members: Dr. Carolina Cruz-Neira, Dr. Lishengsha Yue, Dr. Natalia Barbour and Dr. Mohamed H. Zaki, for their encouragement and suggestions.

Additionally, I want to acknowledge Dr.Yina Wu, Dr. Qing Cai and Zijin Wang, who has provided unlimited ideas and valuable help for my research.

I would also like to acknowledge my colleagues. Thanks to my kind colleagues, Siyuan Tang, Jiahao Zhu, Shengxuan Ding , Dr. Jinghui Yuan, Dr. Nada Mahmoud, Dr. Amr Abdelraouf, Dr. Zubayer Islam, Jorge Ugan, Dr. Patricia Tice, Md. Rakibul Islam

Then, I want to thank my family, my mom and my father. It is your support that encourages me to go further. You always give me one hundred percent of support, to let me chase my dream without worrying about other issues.

Last but not least, I would like to thank my wife Jielan Zheng and daughter Hathaway Zheng for being my best colleague and my best friend. I would also like to thank my lovely cats Baozi and Huajuan for staying with me day and night.

TABLE OF CONTENTS

LIST OF FIGUTES

LIST OF TABLES

CHAPTER 1 : INTRODUCTION

1.1 Overview

Data-driven intelligent transportation has developed as a popular research issue[1-3], due to the rapid development of deep learning algorithms[4-6]. It is vitally important to collect traffic data before doing any traffic safety analysis. When analyzing the traffic conditions for specific areas or road segments, thousands of detectors such as loop detectors or radar sensors located at fixed locations help the traffic data collections. In 2018, researcher analyzed rear-end crash risk for individual vehicles through a radar sensor on a freeway location [7]. However, such analysis is still limited to certain locations that have installed detectors; the detectors could not monitor many detailed driver behaviors, such as lane changing, merging, interaction between road users, etc. Due to the development of various new technologies such as in-vehicle GPS, surveillance cameras, and Unmanned Aerial Vehicle (UAV) trajectory data could be collected and utilized in traffic safety analysis in recent years [8-10].To determine the relationship between crash risk and driver behavior, some studies utilize in-vehicle devices to collect data and some studies extracted road users' trajectories by surveillance cameras. By reviewing the data from surveillance cameras, the safety conditions for the selected areas can be calculated through surrogate safety measures, such as time-to-collision (TTC), post-encroachment time (PET)) [11-13].

Another necessary part of conflict analytics is scenario-based traffic safety evaluation. And this method is also used when testing the effectiveness of automated driving systems. However, this approach heavily relies on real-world data to calibrate the model. Driving tests and naturalistic driving studies (NDS) are this study's most common data sources. Most of this data was from a different floating vehicle equipped with a senor or infrastructure sensor installed with a roadside unit. Due to the physical limitations of the sensor visibility, those two approaches cannot give

accurate naturalistic behavior invariably. Meanwhile, UAVs equipped with 4K cameras have the advantage of being able to capture traffic from a top-down perspective with great longitudinal and lateral accuracy. From the drone video, as shown in FIGURE 1 SNAPSHOOT OF , all the vehicle GPS positions, velocity, acceleration, deceleration, and the gap between vehicles can be directly obtained and plotted on a 2D map based on detection and tracking results without considering the impact of depth of field like in-vehicle or conventional surveillance cameras.

FIGURE 1 SNAPSHOOT OF UAV VIDEO

However, collecting and processing such a large amount of drone data will be expensive and time-consuming. In addition, most researchers may not have the resources to do it themselves. Therefore, the need for Open datasets is significant in traffic safety research because they allow for greater collaboration and transparency among researchers and increase the data access of many researchers. As a result, open datasets can lead to more efficient and effective research and a greater understanding of traffic safety issues. Additionally, open datasets can help to ensure

that research findings are reproducible, which is essential for building trust in the research community and advancing the field.

Another emerging technology that has attracted increasing attention in academia and industry is Digital Twin based study, which use various techniques to replicate real-world entities in a virtual space for different types of study. Researchers are investigating the potential of Digital Twin technology in the transportation industry to aid in developing Intelligent Transportation Systems (ITS). Vehicles and pedestrians are among the major traffic participants in ITS. Although the Digital Twin concept has been applied to vehicles, drivers, and other traffic participants, many simulation or field test efforts have been made for it. Still, there is a lack of systematic framework and research tool that considers vehicle and pedestrian in-the-loop for a Digital Twin environment.

1.2 Objectives

1.2.1 Development Of Ai-Driven Computer Vision Systems

1.2.1.1 Automated Roadway Conflicts Identification System (A.R.C.I.S)

UAVs with 4K camera to monitor traffic was already investigated [14, 15]. However, the majority of the research was to extract macroscopic statistics such as traffic density, flow, and speed[16-19]. The resultant trajectories are not acceptable for the safety certification of conflict research because the positions of road users were not recovered with decimeter-accuracy.

In this study, an automated framework for safety diagnosis to extract accurate trajectory data and vehicle information from UAV videos based on the pixel-to-pixel manner predicted masks will be proposed. And safety diagnostics based on PET values should be conducted for each pixel in the UAV images.

1.2.1.2 Near Miss Identification and Analysis System

The study of near-miss identification and analysis has been prevalent in recent years. Video data that could provide high-resolution trajectories is an essential source for identifying near misses that happen on roadways. Closed-Circuit Television (CCTV) cameras, which has been widely installed in the current roadway system, should be a cost-effective sensor to monitor traffic safety in the within-intersection area. Thus, it would be essential to investigate how to conduct traffic analysis using CCTV cameras, especially road safety analysis. Unlike the UAVs' video images that could observe the vehicles' occupying areas (vehicle sizes) based on bird's-eye views, CCTV cameras at intersections are generally at low positions. The methods used for detecting vehicles in UAV videos could not describe the actual locations and sizes in the 3D scene. Key points detection of vehicles such as headlights and taillights would be necessary for vehicle localization if no in-depth information could be obtained from other sensors (e.g., Lidar, radar).There is needed to fill the gap by proposing a framework named "Near Miss Event Detection System (NMEDS)" for CCTV cameras. To be specific, the proposed system contributes to the literature for safety analysis in three parts:

(1) In adaptation to regular detection and tracking with a bounding box, A new method with the ability to detect and predict the key points of vehicles, including right-front headlight, left-front headlight, right-back taillight, and left-back taillight. The key points could provide more precise vehicle occupying locations than the bounding boxes generated based on regular deep learning detection.

(2) proposing a method to modify the occluded points in the real-world coordinate system considering the relation of identified key points;

(3) introducing a grid-based spatial autocorrelation analysis to identify significant hotspots in the within-intersection area.

A case study is presented at a typical intersection. The results should indicate the proposed framework's performance for vehicle localization. It is expected that the proposed methods could help diagnose road safety problems using roadside video cameras.

1.2.1.3 Driving Visual Environment Detection System (D.V.E.D.S)

Speeding is one of the major factors impacting traffic safety. According to the National Highway Traffic Safety Administration (NHSTA), nearly a third of fatal crashes in the United States have been designated as "speeding-related" in the last decade[20]. On urban arterials, the speed limit violation could significantly increase the severity levels of pedestrian and bicycle crashes [21]. A lot of studies have been conducted to examine the contributing factors for the crash occurrence and speeding behavior. The factors include traffic volume, roadway geometric design, land use, socio-demographic characteristics, weather, etc. For example,[22] developed grouped random parameter models to examine the crash occurrence on segments and intersections considering the roadway attributes such as speed limit and the zonal level effects.[23] categorized the speeding behavior into three levels by proportions based on the speed camera data. It was found that high speed limits are highly associated with moderate speed limit violations, compared to minor or major speed limit violations. Besides, the study also revealed that a divided median and higher functional class could lead to more major speed limit violations. [24] explored factors contributing to operating speeds on arterial roads and revealed the significant effects of inside shoulder width, speed limit, and number of signalized intersections per mile on operating speeds.[25] firstly investigated the effects on monthly weather variations on crash occurrence and revealed that the driving environment such as weather could have significant effects on traffic safety. When driving on roads, the visual environment could be also a major contributor to drivers' speeding and traffic safety[26-28] .

However, the existing data to reflect drivers' visual environment is limited or less detailed. Thus, it is important to explore a new method to collect extensive data to represent drivers' visual environment on roads and explore its effects on speeding crashes. This study will proposing a novel method to obtain drivers' visual environment from GSV images and explore the effects of the visual environment on speeding crashes. To this end, deep learning models were applied to obtain the cluster and depth information from GSV images. The coordinate transformation was conducted to quantify drivers' visual environment in the real world. Then, the important variables reflecting drivers' visual environment were extracted and their effects on speeding crashes were explored by developing explainable machine learning models.

1.2.2 A Drone-Based Vehicle Trajectory Dataset for Safety Oriented Research and Digital Twins

The traffic safety research domain has rapidly progressed in the past few years. Example applications include autonomous vehicle safety, proactive traffic safety, and connected traffic safety applications based on Vehicle-to-Everything (V2X) and Infrastructure-to-Everything (I2X) communication paradigms. Increasing interest in safety research stimulates a rising demand for vehicle interaction datasets, in particular, birds-eye-view video-based trajectory datasets Currently, a significant portion of related studies were formulated based on several video-based trajectory datasets such as NGSIM,[29] highD[30] and InD[31] etc. These studies include driver behavior analysis[32-34], autonomous vehicle virtual testing [35-37], crash mitigation and avoidance system design [26, 38, 39], advanced autonomous control algorithm development ([40-43]), surrogate safety measures[44-47], and many other safety applications. NGSIM was proposed in 2002. Due to the technological limitations in early years, some dataset contains many trajectory errors and most existing dataset has a low density of aggressive driving and very few near-collision events [48-50].Consequently, safety research that depends on critical driving trajectories might not

be able to collect sufficient samples from those dataset. Another major issue with those datasets is that the rotated bounding box information of the vehicles and their trajectories are not provided. As further demonstrated in later sections of this study, accurate vehicle geometric representation is essential for robust safety-related measurements and applications.

In addition, the most recent autonomous vehicle safety research is looking at the digital twin concept that leverages the advancement of virtual simulation. By providing a virtual environment that is an exact copy of the real world, the autonomous vehicle can be tested more thoroughly with virtual on-board sensors. While, neither NGSIM, highD nor InD has features to support the digital twin concept.

To address this issue, the following feature of a new open dataset were identified:

1) Accurate vehicle trajectories. Detailed introduction of the algorithms and data processing methods are presented to give user a clear mind about the strict quality control of the dataset.
2) Full observation of trajectory. Drone have a flight height limit of 400 feet, this dataset need to covers a much larger area compared to other datasets which only used one drone. A stitching method is required used two or more drones hovering over target areas (freeways and intersections) and provides comprehensive results,
3) More critical safety events. This dataset should selected locations that contain more intensive vehicle interactions which has more critical safety events compared with other datasets, in terms of both event severity level and sample size
4) More accurate critical safety events. This dataset need provides highly accurate bounding box information for each detected vehicle, which enables a more accurate estimation of the safety level when compared with only using the vehicle central point information.

5) Digital twin features: To facilitate for researchers to verity their vehicle safety products in a more high-fidelity virtual environment. This dataset need provides high resolution of 3D maps and physical models for each of the collected locations, and it provides the signal timing information as well.

1.2.3 Digital Twin Based Simulation Framework

Digital Twin is an emerging technology that replicates real-world entities into a digital space. It has attracted increasing attention in the transportation field and many researchers are exploring its future applications in the development of Intelligent Transportation System (ITS) technologies. Connected vehicles (CVs) and pedestrians are among the major traffic participants in ITS. Although the Digital Twin concept has been applied to vehicles, drivers and other traffic participants, and many simulation or field test efforts have been made for it, to the best of our knowledge, there is a lack of systematic framework and research tool that considers both CV and pedestrian in-the-loop for a Digital Twin environment. As they are among the two most important traffic participants on the road, it is necessary to build a Digital Twin architecture to facilitate the research and development of the next generation of ITS technology.

In this study, a Digital Twin framework for CV and pedestrian in-the-loop simulation is proposed. The proposed framework consists of the physical world, the digital world, and data transmission in between. The features for the entities (CV and pedestrian) that need digital twining are divided into external state and internal state, and the attributes in each state are described. We also demonstrate a sample architecture under the proposed Digital Twin framework, which is based on Carla-Sumo Co-simulation and Cave automatic virtual environment (CAVE). A case study that investigates Vehicle Pedestrian (V2P) warning system is conducted to validate the effectiveness of the presented architecture. The proposed framework is

expected to provide guidance to the future Digital Twin research, and the architecture we build can serve as the testbed for further research and development of ITS applications on CV and pedestrians.

1.3 Source of Data

- **Crash data**: The crash data is obtained from the Florida Crash Analysis Reporting System (CARS) database. The CARS database provides the information about the time of a crash, crash location, crash type, crash severity level, driver age, etc.
- **Drone Data:** the drone video is obtained from UCF SST team and private contractor. From the drone video, the vehicle GPS position, velocity, acceleration, deceleration, gap and conflict can be extracted.
- **CCTV Data:** the drone video is obtained from FDOT Center to Center (c2c) connection. From those CCTV video, the vehicle GPS position, velocity, acceleration, deceleration, gap and conflict can be extracted.
- **Google Street view Data:** The data used in this study were collected from urban arterials in Central Florida. The urban arterials of nearly 75 miles were included, and around 15,000 GSV images were downloaded and processed to get the indexes about drivers' visual environment

1.4 Dissertation Organization

The rest of the dissertation is organized as follows: Chapter 2 covers the literature review for each research objective. Chapter 3, discusses the system architecture of UCF SST Automated Roadway Conflicts Identification System (A.R.C.I.S)[51], Near Miss Event Detection System

(N.M.E.D.S)[52] and Driving Visual Environment Detection System (D.V.E.D.S)[53]. Chapter 4 describes the development of CitySim: A Drone-Based Vehicle Trajectory Dataset for Safety Oriented Research and Digital Twins[54]. In Chapter 5 explores framework of Digital Twin based Co-Simulation Pedestrian and Connected Vehicle In-the-loop simulation[55]. Chapter 6 summarizes the dissertation's key findings and contributions.

CHAPTER 2: LITERATURE REVIEW

2.1 Computer Vision In Traffic Safety

Various methods could be utilized to conduct road safety analysis. The traditional crash count analysis uses highly aggregated data to evaluate road safety for certain situations [56]. However, one of the limitations of using highly aggregated data is that it could not consider the heterogeneity traffic or the surrounding environment, which could have an impact on the estimation accuracy; In the recent decades, the wide employment of traffic infrastructure-based sensors could generate big data in real-time that enables researchers to conduct crash likelihood analysis by aggregating it into shorter time interval. By aggregating the data from nearby traffic detectors (e.g., MVDS, Loop Detectors) and the other infrastructures (e.g., weather stations, signals) at certain time intervals, precursors of crash occurrence could be identified, which could be used to investigate road safety situations and identify corresponding strategies to prevent crashes[57-60].

Although the deployment of traffic infrastructure brings big data and enables the analysis of real-time crash risk, the rarity of crashes leads to some limitations to safety research and practice. The use of conflicts could overcome the limitation of crash-based safety analysis and provide compelling explanation to crash causations while considering the interactions between road users and other behavioral factors by obtaining individual road users' data. First, only after crashes have been observed can researchers and practitioners evaluate the safety performance of the studied locations and situations. The use of conflicts provides an alternative method to investigate road safety conditions that overcomes the limitations of crash analysis [61]. Second, although risky situations are present, crashes may still be prevented through evasive actions,

such as emergency brake, lane changing, etc. However, the traditional infrastructure-based traffic data could barely capture information of individual road users' behavior, and the crash prediction model could be prone to providing false alarms [62]. Moreover, as human factors are solely or in conjunction with other factors contribute to around 95% of crash occurrence, traffic data from detectors may not be sufficient to identify crash precursors for many situations [63]. The conflict events could be observed by calculating the surrogate safety measures between road users. Time-to-collision (TTC) and post-encroachment time (PET) are two of the most prevalent surrogate safety measures that use road users' trajectory data to investigate safety conditions. TTC is widely employed to estimate rear-end conflicts. Meanwhile, PET describes the time separation between two road users with consideration of the effect of road users' evasive conflict avoidance behaviors and is more commonly utilized when using large datasets or evaluating safety conditions when turning movements are prevalent. Researcher utilized PET to investigate vehicle- bicycle conflicts at signalized intersections[64]. In 2020, researcher proposed a framework to conduct road safety diagnostics by automated extracting vehicles' trajectory using Mask-RCNN detection and Channel and Spatial Reliability Tracking (CSRT). The trajectory data are utilized to calculate PET values to identify conflict situations based on Unmanned Aerial Vehicle (UAV) video images [51].

2.1.1 Objection Detection, Tracking And Classification

Detection and tracking of road users are one of the central applications in computer vision studies[65, 66]. Thus, in recent years, it has attracted more attention to solve video-based detection and tracking issues for both automated driving applications and road monitoring applications. Some car manufactures integrate cameras with Lidars in automated vehicles to obtain depth information and determine the boundaries of surrounding vehicles using sensor fusion

technology. The improvement of vehicle boundary detection would be beneficial for real-time safety estimation including calculating TTC between the automated vehicles and their surrounding road users to avoid potential collisions. As for vision-based detection, the conventional vehicle detection methods (e.g. background subtraction, optical flow) tend to only work under simple traffic scenes such as uninterrupted traffic flow and have limitations to detect the precise locations of vehicles from UAV images. Meanwhile, the results could be sensitive to the environment like vehicle color, vehicle orientation, shadows, background motion, intricate ground conditions. Moreover, these methods may have difficulties to detect and track slow-moving or stopped vehicles. Thus, these methods could not be employed to analyze congested area or intersections. In recent years, multiple Regional-CNN (R-CNN) detection methods have been proposed based on deep learning approaches, which include R-CNN, fast R-CNN, and faster R-CNN. These methods have been applied for vehicle detection and have shown better performance than conventional approaches [67-69]

2.1.2 UAV Video

Several studies have been conducted UAV video, to obtain traffic parameters through detecting and tracking road users' positions and movements from UAV videos. Most of the previous studies focused on obtaining traffic flow data. In details, the data includes speed, volume, and density from UAV videos. For example, researcher proposed a framework for traffic flow parameters estimation based on UAV videos through ensemble classifier (Haar cascade + convolutional neural network) and optical flow, and UAV videos from a freeway segment was utilized to test the system performance[70].Zhao et al. collected speed, density, and volume data for uninterrupted flow corridors based on an aerial camera array mounted on an airplane [71].Yamazki extracted vehicle speeds based on detecting vehicles from two consecutive digital aerial images *[65]*. Meanwhile,

some studies utilized UAV videos to investigated traffic safety related issues, such as incident detection. In 2014, researcherused UAV to detect traffic incidents for low-volume roads[72].In 2015, Lee proved the applicability of using UAV to conduct real-time incident monitoring through pilot tests [73]. Even though some studies have been conducted to evaluate traffic safety situations using UAV[8, 74], to the authors' best knowledge, few studies have been conducted to propose methods to specifically diagnose traffic safety conditions based on UAV videos other than incident detection. Although UAV videos with high resolution and frame frequency could capture adequate ground details and vehicles' movements, most of the current studies focus on calculating the traffic flow characteristic and aggregating the data into certain time intervals.

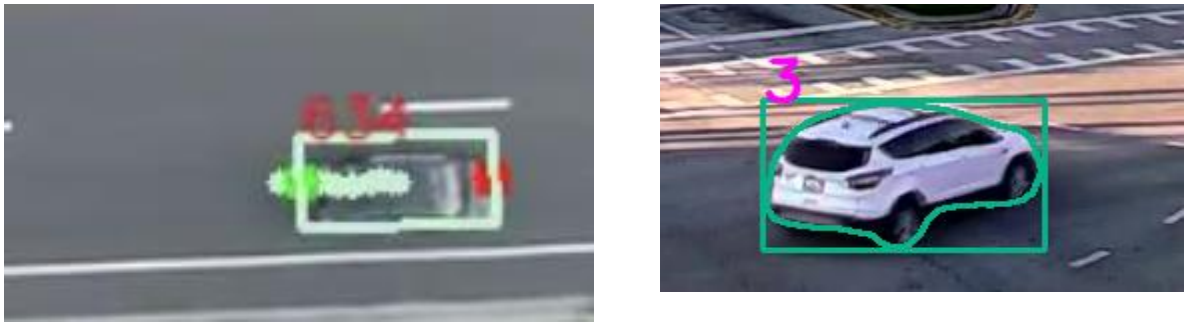
2.1.3 Roadside Video

Computer-vision technologies play a vital role in video-based traffic analysis, which is employed for object classification, localization, and trajectory extraction. Various techniques have been utilized to obtain data from video images including feature-based detection & tracking, background subtraction, and optical flow [12, 70]. In recent years, some studies employed deep learning approaches to extract traffic parameters from videos that could be utilized to overcome the limitations of the traditional approaches under uninterrupted traffic flow conditions or the changes in environments (e.g., shadows, intricate ground conditions)[75]. Besides the trajectory extraction, vehicle localization for the occupying areas also has a significant influence on road safety diagnostics when calculating surrogate safety measures including TTC and PET. The video data used for traffic analysis could be collected from multiple types of video cameras that include UAVs and video cameras at the roadside. For example, the UAVs could be employed for data collection and provide high flexibility of locations, angle of view, and time selection[9, 41, 76] . Moreover, the bird's-eye view video images collected from UAVs could reduce the efforts needed

for data processes such as perspective transformation and vehicle localization. However, the issues of short battery life, matching images from multiple times of flies, and the concerns of safety are nonnegligible. In order to overcome the limitations, many studies utilized their own cameras to collect video data at the roadside. [77]developed conflict-based real-time safety models using multiple safety indicators that were calculated using surveillance camera data. [78]used massive vehicle trajectory data that were collected from 70-hour video data at two signalized intersections for proactive safety analytics. These studies were conducted by using their own cameras at the roadside, which could provide high-resolution video data. It is worth noting that CCTV cameras could collect video data with much longer duration without additional hardware installation or labor work, which could also be integrated with other infrastructure data and conduct conflict-based road analysis. Besides, CCTV cameras have been widely installed in the existing roadway system. Hence, with the consideration of cost and coverage, CCTV camera videos are generally preferred when relatively larger datasets in a wider area are needed for analysis. Thus, it would be essential to investigate how to conduct traffic analysis, especially road safety analysis, using CCTV cameras. Moreover, with the development of Connected-Vehicle (CV) techniques, it is expected that CCTV cameras would play an important role in the infrastructure-to-Vehicle (I2V) technologies development and deployment.

Hence, it is very important to get accurately the vehicles' locations and shapes from videos when calculating surrogate safety measures. Besides, the camera calibration is required considering the relationship between the video image coordinates to physical coordinates on a world map. In this way, the trajectory information in real-world coordinates instead of pixel-based coordinates is used for the surrogate safety measure calculation. Videos recorded by UAVs are in a bird's-eye view and vehicles can be treated as 2D objects directly (FIGURE 2a)). For the 2D viewpoint detection,

many methods using feature detection and Convolutional Neural Network (CNN) such as Mask R-CNN ([79]) have achieved a good detection performance of road user on roads. Also proposed a method to generate rotated bounding rectangles based on the pixel-to-pixel manner masks[80]. The method could improve Mask R-CNN detection at intersections where vehicles make turning movements.



(a) AN EXAMPLE FROM A UAV VIDEO

(b) AN EXAMPLE FROM A CCTV CAMERA

FIGURE 2 DETECTION AND TRACKING EXAMPLES IN UAV AND CCTV CAMERA VIDEOS

Different from the UAVs' video images that could observe the vehicles' occupying areas (vehicle sizes) based on bird's-eye views, CCTV cameras at intersections are generally at low positions. Objects such as vehicles are recorded by CCTV cameras through the extrinsic camera rotation to project 3D objects to the 2D videos (FIGURE 2(b)). Hence, the methods used for detecting vehicles in UAV videos could not describe the true locations and sizes in the 3D scene. Key points detection of vehicles such as headlights and taillights would be necessary for vehicle localization if no in-depth information could be obtained from other sensors (e.g., Lidar, radar) [81] The key points could be detected and tracked based on feature-based approaches [61].

However, in the within-intersection areas, vehicles in the video images from CCTV cameras usually have occluded areas due to the angle of video cameras or view obstruction by their nearby objects (e.g., vehicles, traffic signs). However, the occluded key points could not be identified from the feature-based approaches, and the occluded key points could be changed during the vehicle movement, which brings additional challenges for vehicle tracking and localization. Based on feature-based detection, the occluded points are difficult to be identified and used to estimate precise vehicle locations. Moreover, when turning movements are prevalent in the within-intersection area, the occluded points of vehicles could change during the turning movements, which brings more challenges for vehicle detection and tracking. For example, a vehicle moves from the left side of the video images to the right side of the video images, the occluded point of the vehicle may change from the rear-left point to the front-left point. If the front-left point is tracked by the feature-based tracking methods, the trajectory of the front-left point will be ended when it is occluded, which leads to the difficulties of vehicle localization and tracking. Reddy proposed a framework named "Occlusion-Net" for key points localization and vehicle key points estimation[82]. The points that are occluded due to the angle of view (self-occlusion) or the nearby objects (object-occlusion) could be predicted in the video images. The framework was proposed for the camera view from vehicles for the use of the driver-assist system, without the consideration of vehicle moving characteristics in the within-intersection area. Hence, to the authors' best knowledge, no study has been conducted to use the technology of key points location and vehicle key points estimation to improve the accuracy of vehicles' locations and occupying areas from CCTV cameras for safety diagnosis in the within-intersection area.

2.1.4 Google Street View (GSV) Images

Within the rapid development of deep learning and computer vision technology, detailed information including object clusters and depth could be obtained from images. In the era of transportation studies, computer vision has been applied to count traffic volume and detect traffic speed[83]. Besides, some studies applied detection and tracking algorithms to get vehicles' trajectory and calculate the surrogate safety measures [78, 80]. In these studies, researchers needed to use cameras to collect the video and image first. It might be time-consuming to collect data in a large study area. In the recent years, Google Street View (GSV) images have been used to analyze the relationship between the environment and traffic safety. For example, Mooney used GSV images to assess environmental contributions to the frequency of pedestrian crashes[84]. It was found that traffic islands, visual advertising, bus stops, and crosswalk infrastructures are significantly associated with the counts of pedestrian crashes. Kita and Kidziński manually labeled data about the conditions of the house and neighborhood from GSV images and developed a Generalized Linear Model (GLM) model to reveal the correlations between these factors and the risk of that house's residents getting involved in a car crash[84]. Recently, machine learning techniques were also used to explore traffic safety based on GSV images. Li developed deep learning algorithms to estimate and map the occurrence of sun glare for drivers using GSV images[85]. The study also estimated the time windows of sun glare by calculating the sun positions and the relative angles between drivers and the sun for different locations developed a distance-aware pixel accumulation to extract information about objects surrounding the spots on roads in the street view images[86]. The extracted characteristics were used to train fully connected neural networks to identify black spots on roads. While the machine learning methods could reach a high accuracy in traffic safety analysis, the relation between the

characteristics related to drivers' view extracted from GSV images and safety is unclear. Meanwhile, to the best of the authors' knowledge, several factors such as tree density and driving environment complexity which reflect drivers' visual environment have not been explored by using GSV images. Hence, the applications of using GSV images to enhance traffic safety might be limited.

Through a Google API, users could specify the location, heading, and vertical angle when downloading the image. Hence, it is possible to get a lot of images with drivers' views through the GSV images. Computer vision technology has been applied to process GSV images automatically instead of manually. Based on the computer vision technology, different information such as street-level morphology, urban feature composition, and urban greenery could be extracted from the GSV images [87-89]. [90]assessed the street-level urban greenery with the field of view (fov) as 60 degrees by using GSV images. The Red, Green, Blue (RGB) bands were detected from the images, and the difference among different bands was calculated to determine the area of green vegetation. Gong used the PSPNet model, which is a deep Convolutional Neural Network (CNN) model for the semantic segmentation of GSV images[87]. Several view factors, including sky, tree, and building view factors of street canyons, were quantified by using the photographic method. Researcher (derive street-level morphology and urban feature composition as experienced by a pedestrian from GSV images[88]. This study used the Caffe deep learning framework to segment GSV images into six classes: sky, trees, buildings, impervious surfaces, pervious surfaces, and non-permanent objects. While the previous studies could extract accurate information at each pixel from images by using computer vision methods, these studies focused on the feature segmentation in GSV images without considering depth information of features. In a GSV image, features should have different distances away from the

camera, reflecting objects at different locations drivers pass by along the road. Hence, depth information should be considered to reflect drivers' visual environment when driving along roads in the real world.

2.1.5 Surrogate Safety

Surrogate safety measures are widely employed for safety diagnostics using trajectory data to describe the safety situations for a certain time and spatial range. Among the previous studies, Time-to-collision (TTC) and post-encroachment time (PET) are two of the most prevalent surrogate safety measures that use road users' trajectory data to investigate safety conditions. TTC is widely employed to estimate rear-end conflicts. Meanwhile, PET describes the time separation between two road users considering the effect of road users' evasive conflict avoidance behavior and is more commonly utilized when using large datasets or evaluating safety conditions when turning movements are prevalent. Due to its ability to measure the proximity of conflicting road users to each other, PET has been recognized as an effective indicator of safety at intersections, where crossing events are frequently observed[11]. Researcher utilized PET to investigate vehicle-bicycle conflicts at signalized intersections[64]. Researcher proposed a framework to conduct road safety diagnostics by automated extraction of vehicles' trajectories using Mask-RCNN detection and Channel and Spatial Reliability Tracking (CSRT)[51]. The trajectory data are utilized to calculate PET values to identify conflict situations based on Unmanned Aerial Vehicle (UAV) videos. By using PET as an input to model crashes, Peesapati found that the PET could reflect the safety conditions considering the effects of characteristics such as sight distance, grade, and other parameters at intersections[91]. Other features of vehicles such as length and bumper location have also been considered to modify the

TTC and PET to better reflect the safety conditions[78, 80]. The thresholds to define risky conditions based on TTC and PET are normally between 1 and 3 seconds.

2.2 Drone-Based Vehicle Trajectory Dataset For Traffic Safety Research

Previous research efforts that utilized drone-video-based trajectories, particularly in the traffic safety domain, have outlined some limitations with the available datasets. Existing video-based trajectory datasets were not specifically designed for safety research. Therefore, they have not focused on capturing a significant number of safety-critical vehicle interactions. Currently, the most well-known and widely used video-based trajectory datasets are NGSIM (19) [29], highD[30], inD[31], and Interaction[92]. NGSIM was proposed in 2002. Due to the technological limitations in early years, the dataset contains many trajectory errors [48, 93, 94]. The dataset has a reported false negative issue which causes more than 10 percent of the vehicles' detection and tracking process to fail for several consecutive frames[49]. Consequently, previous work conducted filtering processes before using the dataset to remedy this issue[48].In addition, NGSIM has a low density of aggressive driving and very few near-collision events [92].

The highD, inD, and Interaction datasets were proposed much recently in 2018-2019, and they were able to use much more advanced computer vision technology to accurately extract vehicle trajectories. While, highD is mainly the videos from freeway segments under a free-flow condition, in which intensive vehicle interactions and critical safety events were inevitably limited[50]. Consequently, the vehicle safety research that depends on the critical safety scenarios might not be able to collect sufficient samples from highD. As for inD and Interaction, they are videos from intersections; the main issue is that the trajectories were measured only in the form of central point without the information about the true vehicle body profile, which

would cause a bias when calculating the safety metrics based on turning trajectories. Consequently, some critical safety events would be missed. The bias is further demonstrated in this dissertation.

2.3 Digital Twin based Simulation for Pedestrian safety and Connected Vehicle

Digital Twins have been discussed informally since the early 2000s [95]. Grieves introduced the first nomenclature in a 2003 presentation, which was later recorded in a white paper that laid the groundwork for the development of Digital Twins [95]. In 2012, the National Aeronautical Space Administration (NASA) published a document titled "The Digital Twin Paradigm for Future NASA and US Air Force Vehicles," which marked a turning point in the definition of Digital Twins. A digital twin is a computer representation of an existing or planned physical object. It could include but is not limited to, architectural plans, product designs, and development. The reference "Digital Twin" is created when data travels between an existing physical thing and a digital object, and they are fully integrated into both directions. When you change to a physical object in real-world, it automatically affects the digital object. Recently, Digital twins are becoming more popular in academics[96-101]. Due to rapid improvements in connectivity through IoT, the application and potential for Digital Twins to be highly successful inside a smart city are expanding year to year.

The more smart cities that are built, the more connected communities become, and the more Digital Twins are used. Not only that, but the more data we collect from IoT sensors implanted in our core city services will open the road for research aimed at developing sophisticated AI algorithms [102-104]The capacity of services and infrastructures in a smart city to be monitored using IoT devices and have sensors is extremely valuable for future-proofing. It can be utilized to

aid in the design and development of current smart cities and future smart city developments. Therefore, there are numerous advantages to preparing.

2.3.1 For Pedestrian Safety

Pedestrian safety is a serious concern. According to the National Highway Traffic Safety Administration (NHSTA), a pedestrian was killed every 85 minutes in traffic crashes in 2019 [105]. The pedestrian deaths accounted for 17% of total crash fatalities, 26% of the pedestrian fatalities happened at intersections [106]. Meanwhile, it was found that the obstruction of drivers' view is one of the most common crash causations for pedestrian crashes [107].

Many studies have been conducted that focused on pedestrian safety. In 2012, Zegeer & Bushell summarized the pedestrian crashes contributing factors into five categories: driver, vehicle, social-demographical or policy, pedestrian, and roadway factors [108]. Among those factors, vehicle speeds are found to have a significant impact on pedestrian safety for both crash occurrence and crash severity. Higher operating speed leads to a longer stopping distance. Thus, vehicles with higher speeds may not be able to stop completely and avoid crashes by emergency brakes. Previous research indicated that the odds of pedestrian fatality increase by 11% for a 1 km/h increment [109]. It was also found that if a pedestrian was hit by vehicles' bumpers, hoods, or the windshield area, the severity of the crash tends to be higher [106]. In 2014, Bertulis and Dulask investigated the relationship between vehicle speeds and yield rate [110]. The results indicated that vehicles with higher speeds are less likely to yield to pedestrians, which is consistent with other studies [111, 112]. Unexpected crossing behavior of pedestrians are also one of the common contributing factors for pedestrian safety critical situations [113], such as running and jaywalking. Meanwhile, it was found that some pedestrians have lower speeds and could not complete crossing before the onset of red signal [114], which could lead to potential conflicts with vehicles.

The advent of Autonomous Vehicle (AV) technologies provides tremendous opportunities to prevent crashes by mitigating human errors. One of these technologies is Automatic Emergency Braking (AEB), which is expected to benefit pedestrian safety by preventing pedestrian-related crashes. For pedestrian-related situations, if the vehicle sensors detect an imminent collision, the vehicle will begin to brake automatically. Some studied were conducted to estimate the safety benefits of AEB technologies. In 2013, Rosen investigated the impact of AEB systems on both pedestrian and cyclist crashes using real-world crash data. The results indicate that when the system is optimized, the effectiveness can be 52% and 31% for pedestrians and cyclists, respectively [115]. However, the effectiveness of AEB systems is highly influenced by other factors, such as sensors' Field of View (FoV) and the design of the systems [116]. According to previous research, the effectiveness of AEB decreases significantly with the decrease of FoV, as vehicles may not be able to detect dangerous situations in time and have a complete stop before a crash happens [117]. Yue proposed an augmentation function to estimate the crash risk given its time-space-distance relationship with a pedestrian [118]. The crash risk represents the probability of hitting the pedestrian given all the pedestrian's possible random trajectories in the near future. The study demonstrated that an FoV of 50o and a detection range of 40 m would be the minimum requirement to support the augmentation function. Similarly, Zhao reconstructed 40 crash cases based on the collected video data related to taxi-to-cyclists crashes [25]. The results illustrated that an increase of FoV from 50 ° to 90° could avoid 30% more cyclist-related crashes. Meanwhile, the effectiveness of AEB could be various based on when the vehicles start to decelerate automatically and what deceleration rate is employed. The start to brake decisions of the AEB systems usually depend on the Time-to-Collision (TTC) values. When the TTC is lower than the threshold, the

vehicle will start to decelerate to avoid the imminent collision. Thus, AEB systems with larger TTC thresholds are found to have higher safety benefits [119].

In recent years, many efforts have been conducted to evaluate the effects of various in-vehicle technologies. Some studies were conducted using naturalistic driving data that were collected in real-life driving conditions. This type of data could capture the drivers' behaviors and some crashes/near-crashes events [120, 121]. In 2020, Seacrist utilized SHRP 2 rear-end striking crashes data to evaluate the effectiveness of AEB for rear-end crashes. This study also found that the increase in vehicle operating speeds has a negative impact on the effectiveness of the AEB system [122]. One of the limitations of naturalistic driving-based studies is that the data collection method is found to be inefficient to obtain sufficient sample sizes for analysis, especially when interaction effects are explored. Hundreds of millions of miles driving may be necessary to assess AVs' safety performance [123]. Another type of method is the driving simulator experiment. Driving simulator experiments are widely utilized to investigate the effects of human factors. In 2017, a driving simulator experiment was conducted to investigate the differences in effects among various types of AEB systems under snow conditions and identified significant differences between males and females [124]. Although the abovementioned methods could be utilized to investigate AV's performance, they have limited capability to explore the relationship between the AV sensors' specifications and the driving performance. An open-source autonomous vehicle simulation platform named "CARLA" was developed in recent years and could be utilized to obtain simulated naturalistic driving data, which provides the flexibility of changing driving environments and sensors' specifications [125, 126]. Thus, it could be employed to identify the impact of AV technologies with different sensor-fusion techniques and AV control algorithms. In 2021, Feng utilized CARLA to build an environment and simulated life-like driving for AVs [123]. The

proposed method aims to accelerate the procedure of AV evaluation by reducing the required miles of naturalistic driving. Meanwhile, the CARLA platform can also be extended and employed to test Vehicle-to-Everything (V2X) or Cooperative Driving Automation (CDA) technologies [127, 128].

2.3.2 Connect Vehicle

There are numerous research focused on CV technology, including vehicle connectivity, cyber security, driving assistance, automotive control, and cooperative driving, etc. Most of the studies were conducted in a simulation environment as field test is expensive and time-consuming. Microscopic traffic simulation is widely used to study CV applications' benefits on safety[7, 129, 130], mobility [131-133], and energy consumption[134-136]. Virtual simulators, powered by game engine, have become increasingly popular for CV research, as they can simulate environment perception, vehicle dynamics, and control algorithms[125, 137-139]. Driving simulator is also an effective tool to test CV technology including safety warning system [140, 141], traveler information[142] , and cooperative driving[143, 144]. Recently, an increasing number of research adopted co-simulation technics for CV research. The co-simulation incorporates multiple simulation tools with different functionalities such as autonomous driving, traffic flow simulation, vehicular network and sensing technology [139, 145]

2.3.3 CAVE System

The research on applying intelligent transportation system applications to enhance pedestrian safety heavily relies on simulation. Various collision warning or avoidance systems have been proposed that powered by vehicle-to-pedestrian communication [146, 147], and tested in micro-simulation by modeling the vehicle and pedestrian behavior. The advent of virtual reality (VR) technology brings opportunities to create a high-fidelity environment to involve real pedestrians

for research and testing. VR headset is the most adopted equipment to display a virtual environment because of reasonable price and comprehensive technical support. However, the motion sickness issue is still a major problem that affects user experience[148] . Recently, some researchers used the Cave automatic virtual environment (CAVE) for pedestrian research [149-151] CAVE is a VR environment where projectors are directed to between three and six of the walls of a room-sized cube [152]. It has much less motion sickness compared to headsets and allows people to walk in the cube space freely.

2.3.4 Pedestrian Motion

Pedestrian motion, the way of simulating walking in the virtual environment, is another hard task for pedestrian-in-theloop research. The traditional method to simulate pedestrian crossing behavior is to perform "shout test": participants shout their crossing decision when they intend to cross the street. Walking simulators such as VR treadmills are also used in some research [153]. When conducting experiments, the participant is wearing VR goggles and walking in-place on a bowl-like treadmill, while tied with a band on his/her waist to prevent falling down. It could simulate unlimited walking distance, but the walking experience is inconsistent and still quite different from real-world walking. Recent studies have tried walking inhouse with a VR headset [154]. The issue with it is that the participant is easy to lose balance and prone to fall down. Locomotion, that pedestrian mark time, and use leg a band to detect pedestrian motion, is another method to simulate walking .[153]

2.3.5 Real-Digital World Hybrid Environment

The above-mentioned efforts are CV or pedestrian research based on simulation in designed experiments. The emerging Digital Twin technology allows researchers to develop new ITS applications in a real-virtual hybrid environment. There are already a few research studies that

explored the applications of Digital Twin on connected vehicles and drivers. A Mobility Digital Twin framework is proposed in [155], which consists of physical and digital space of three components: human, vehicle, and traffic. A concept of Driver Digital Twin is introduced with the aim of bridging the gap between existing automated driving system and driver digitization [156]. Also, deep learning for security in Digital Twin of cooperative ITS is investigated[157] . Although the Digital Twin concept has been applied to vehicles, drivers and other traffic participants, and many simulation or field test efforts have been made for it, to the best of our knowledge, there is a lack of systematic framework and research tool that considers both CV and pedestrian in-the-loop for a Digital Twin environment. As they are among the two most important traffic participants on the road, it is necessary to build a Digital Twin architecture to facilitate the research and development of the next generation of ITS technology.

CHAPTER 3: DEVELOPMENT OF AI-DRIVEN COMPUTER VISION SYSTEMS

3.1 Automated Roadway Conflicts Identification System (A.R.C.I.S)

3.1.1 Introduction

This study presents an automated traffic safety diagnostics solution using deep learning techniques to process traffic videos by Unmanned Aerial Vehicle (UAV). Mask R-CNN is employed to better detect vehicles in UAV videos after video stabilization. The vehicle trajectories are generated when tracking the detected vehicle by Channel and Spatial Reliability Tracking (CSRT) algorithm. During the detection process, missing vehicles could be tracked by the process of identifying stopped vehicles and comparing Intersect of Union (IOU) between the tracking results and the detection results. In addition, rotated bounding rectangles based on the pixel-to- pixel manner masks that are generated by Mask R-CNN detection, which are also introduced to obtain precise vehicle size and location data. Moreover, surrogate safety measures (i.e. post- encroachment time (PET)) are calculated for each conflict event at the pixel level. Therefore, conflicts could be identified through the process of comparing the PET values and the threshold. To be more specific, conflict types that include rear-end, head-on, sideswipe, and angle could be determined. A case study is presented at a typical signalized intersection, the results indicate that the proposed framework could notably improve the accuracy of the output data. Furthermore, by calculating the PET values for each conflict event, an automated traffic safety diagnostic for the studied intersection could be conducted. According to the research, rear-end conflicts are the most prevalent conflict type at the studied location, while one angle collision conflict is identified at the study duration. It is expected that the proposed method could help diagnose the safety problems efficiently with UAVs and appropriate countermeasures could be proposed after then.

3.1.2 Methodology

ARCIS has five steps (FIGURE 3). The five steps are video stabilization, object filtering, video stitching, enhanced error filtering and detection and tracking. These steps ensure the output (FIGURE 4) trajectory to be accurate as much as possible.

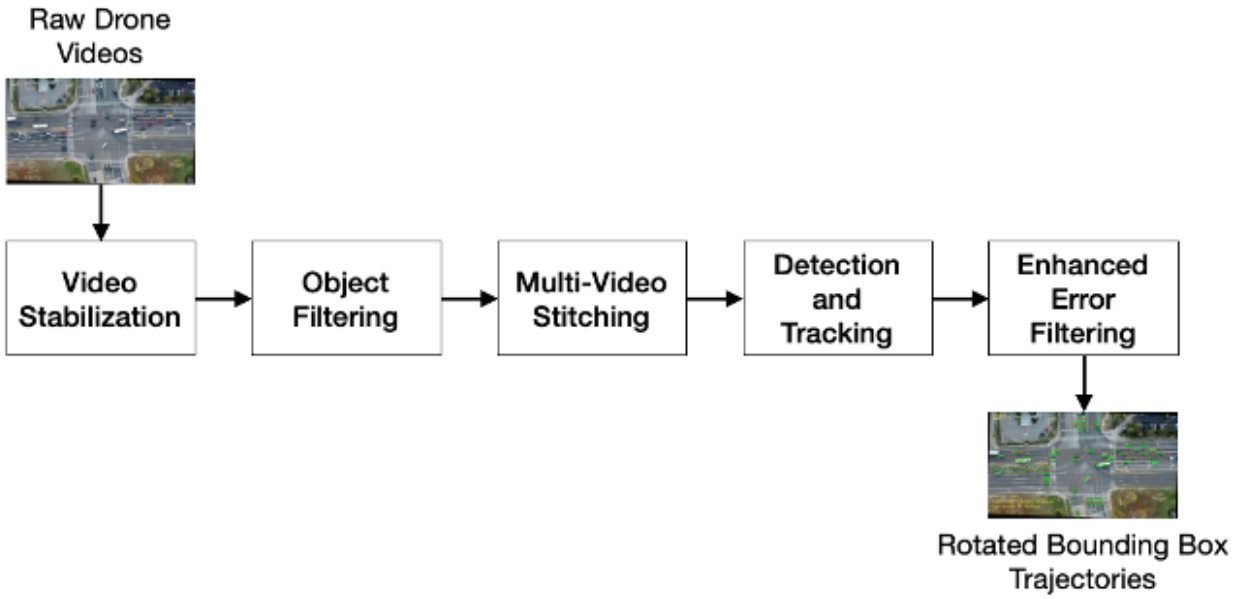
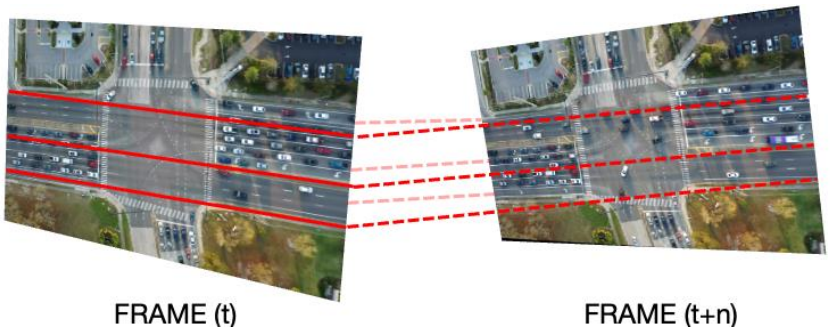


FIGURE 3 THE PRODUCTION PIPELINE OF ARCIS

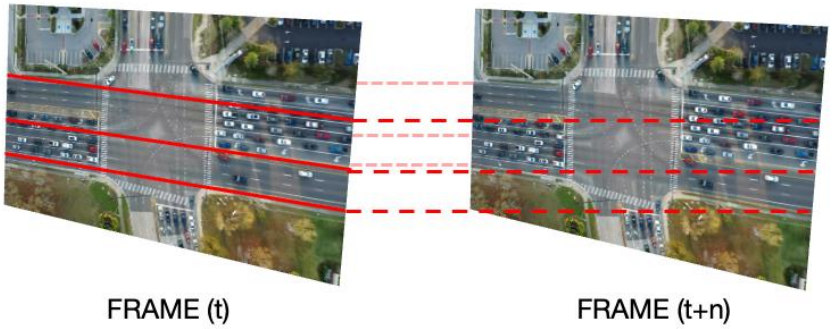
FIGURE 4 EXAMPLE OF ARCIS OUTPUT

3.1.2.1 Video Stabilization

Since the drone may be affected by unstable airflow and has vibration, it is necessary to stabilize the video. Firstly, the scale-invariant feature transform (SIFT) algorithm (*30*) was used to compare the first and last frames of the video, and a vehicle-free background feature map was generated. Further, the median pixel over 6000 random frames were calculated to build an accumulated weighted frame. Then, based on the weighted frame, each frame was mapped to the background feature map through the homography transformation. During the mapping, a CSRT tracker was used to resolve the failure mapping issue, and the position changes between frames were smoothed. Finally, the video was stabilized as shown in FIGURE 5



(A) BEFORE STABILIZATION



(B) AFTER STABILIZATION

FIGURE 5 (A) BEFORE AND (B) AFTER VIDEO STABILIZATION

3.1.2.2 Object Filtering

Since the video recordings may have some road markings or other objects that may cause a detection error; therefore, these objects have to be eliminated. Firstly, the vehicles were removed from the background by using a Gaussian-mixture-based algorithm (*31*); then the inpainting algorithm (*32*) was used to remove the unwanted objects in the background. FIGURE 6 shows the object filtering operation.

(A) BEFORE OBJECT FILTERING **(B) AFTER OBJECT FILTERING**

FIGURE 6 (A) BEFORE AND (B) AFTER OBJECT FILTERING FROM THE VIDEO RECORDING

3.1.2.3 Multi-Video Stitching

Since one drone can only cover a limited length of segment due to the flying height restriction, two or more drones were used to have a complete covering of the interested area. AS shown in FIGURE 7The multi-video recordings were stitched based on the histogram matching algorithm, SIFT algorithm and motion blur idea. These algorithms enable multi-videos to be of the same brightness level, color distribution and well-mapped feature points.

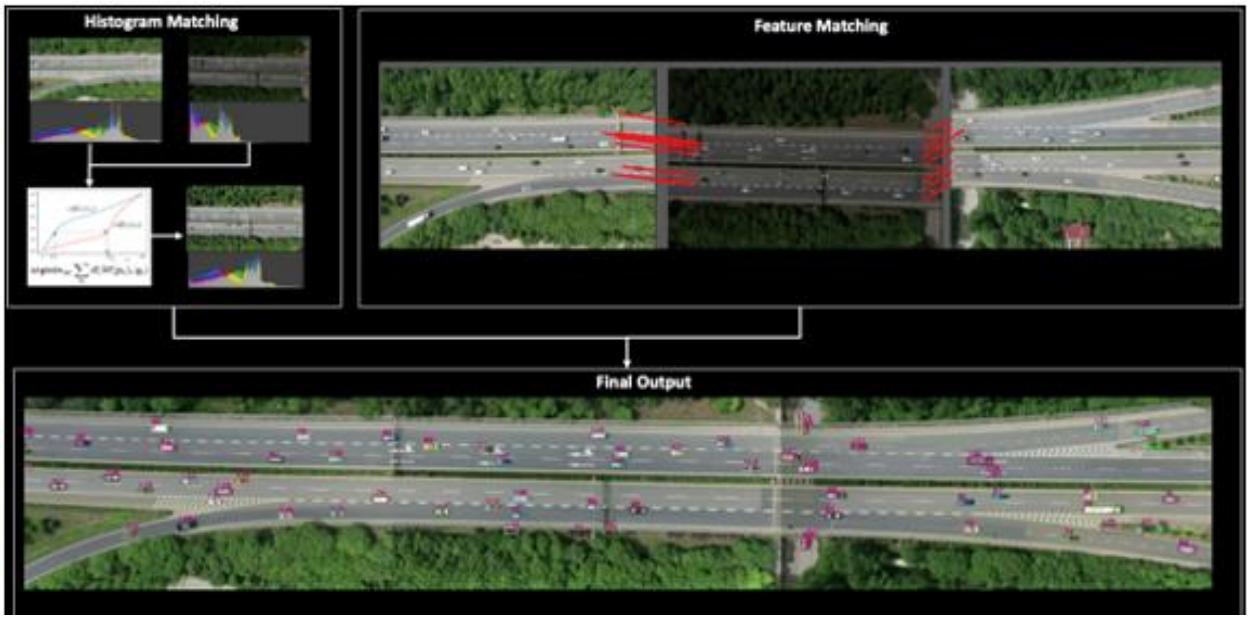


FIGURE 7 MULTI-VIDEO STITCHING FLOW CHART

3.1.2.4 Object Detection and Tracking

Mask R-CNN is used for vehicle detection in this research, which predicts segmentation mask in a pixel-to-pixel manner [5]. The pixel-to-pixel manner mask would be beneficial to obtain precise location of vehicles in UAV images. Mask R-CNN algorithm could provide both classification and masks as output. Meanwhile, since Mask R-CNN could generate precise masks for detected objects, rotated bounding rectangles can be obtained from the masks, which provide an alternative method to obtain vehicle sizes and more precise locations. The rotated bounding rectangles for the detected objects can be generated based on the mask as the smallest rectangle that could cover the predicted mask [158].Since the straight bounding rectangles of the vehicles from detection would not be rotated to align with the vehicles moving direction. Thus, the results of vehicle sizes tend to be larger when vehicles are on a curve or turn at an intersection. FIGURE 8illustrates an example of the differences between the detected object's mask, straight bounding rectangle, and rotated bounding rectangle. The area within the green line is the masked area of

the detected object. The red rectangle is the rotated bounding rectangle of the predicted mask, while the yellow area is the straight bounding rectangle. As shown in the figure below, the area covered by the straight bounding rectangle tends to be larger when the vehicle is conducting turning movement.

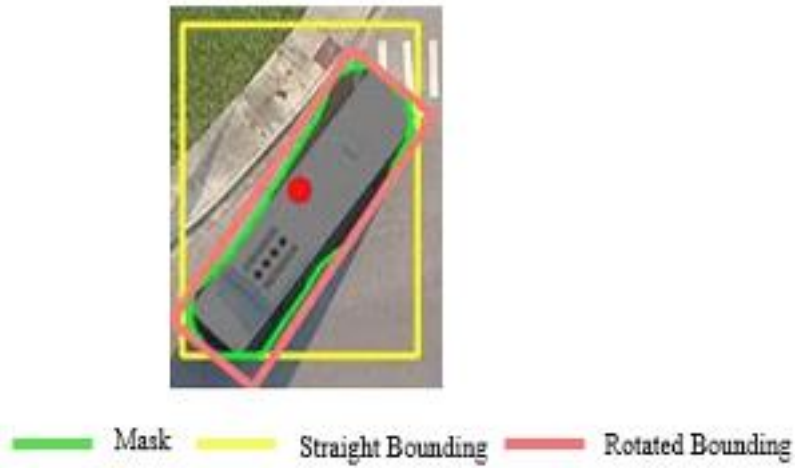


FIGURE 8 DIFFERENCES BETWEEN MASK, STRAIGHT BOUNDING BOX, AND ROTATED BOUNDING BOX

The first step of detection is to collect sample library for vehicle images from UAV videos. In this study, over 10,000 vehicle samples were manually collected from multiple UAV videos (1 image per 10 seconds) from different locations. Thus, each sample includes one vehicles without duplicated samples [159]. Three indicators are chosen to evaluate the detection accuracy, which include correctness, completeness, and quality.

$$correctness=TP/(TP+FP)$$

$$completeness=TP/(TP+FN)$$

$$quality=TP/(TP+FP+FN)$$

where, True Positive (TP) means the number of vehicles that detected correctly; False Positive (FP) means the number wrongly detected vehicles; False Negative (FN) means the number of missing vehicles. Over 2,000 samples with different vehicle images were utilized to test the performance of training results. The performance of the detection algorithm is as follows: The correctness of the sample is 98%; The completeness is 77%; The quality of the samples is 76%. It is worth noting that since detection will run multiple times during the detection & tracking process to find the untracked vehicles, the performance is expected to be better when processing continuous video images. Moreover, in order to illustrate the differences between rotated bounding rectangles and straight bounding rectangles. Intersection of Union (IOUgt) are calculated based on the following equation:

$$IOU_{gt} = \frac{DetectionResult \cap GroudTruth}{DetectionResult \cup GroudTruth}$$

For all the detected vehicles, the average IOU for rotated bounding rectangle area is 0.81, while the average IOU for straight bounding rectangle area is 0.62. Meanwhile, based on the results of paired t-test, there is a significant difference in IOU values between the rotated and straight bounding rectangle area (*p-value*<0.01). The results indicate that the rotated bounding rectangle could provide more precise data for the locations of the detected objects.

Chanel and Spatial Reliability Tracking (CSRT) is used for tracking the object that are detected based on Mask R-CNN algorithm [5], which provides high accuracy but lower speed based on two standard features (i.e. HoGs, Colornames) [158].One of the challenges for vehicle tracking is that vehicles may lose tracking due to the influence of shadows, light conditions, etc. Since the accuracy of the vehicle trajectories has a significant impact when calculating surrogate

safety measures and plays an importation role in safety analysis, this study proposes the following step to detect lost vehicles while tracking:

Step 0: Identify vehicle speed=0 at frame j;

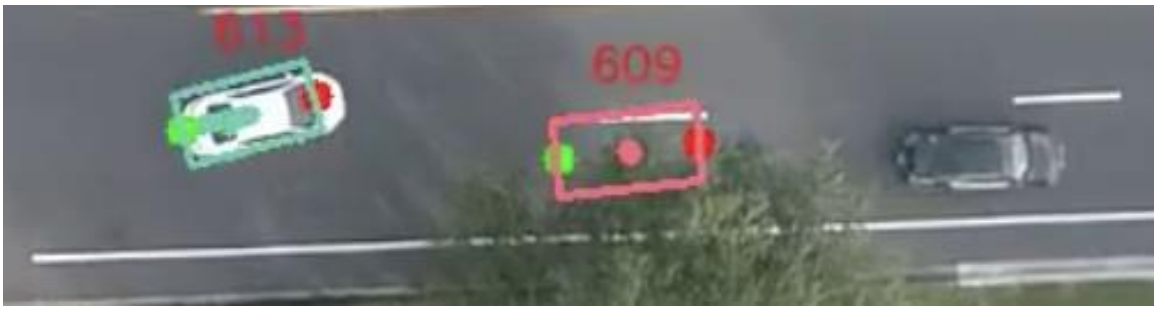
Step 1: Conduct detection for frame j;

Step 2: Calculate the intersection-over-union for tracking and detection areas based on the following equation.

$$IOU_{DT} = \frac{DetectionResult \cap TrackingResults}{DetectionResult \cup TrackingResults}$$

Step 3: if IOUDT<threshold, find a lost vehicle and start tracking for the lost vehicle. The threshold that we employed in this study is 0.5.

Meanwhile, vehicles' moving directions can be extracted based on the difference in locations from two consecutive frames. Thus, vehicles' bounding rectangles are rotated according to the moving direction to obtain more precise information of vehicles' occupied locations. **Error! Reference source not found.**(a) shows an example of a vehicle (#609) lost tracking when approaching a tree shadow area. The speed of the vehicle would be 0 after lost tracking. Thus, detection was conducted at that frame in order to find and continue tracking the lost vehicle. As it is shown FIGURE 9(b), the lost vehicle is detected and tracked again as vehicle #634. Then, by comparing the lost time and location of vehicle #609 and the detected time and location of vehicle #634, vehicle #609 and vehicle #634 could be identified as the same vehicle.



(a) LOST TRACKING

(B) TRACK THE LOST VEHICLE

FIGURE 9 AN EXAMPLE OF FIND LOST VEHICLE DURING TRACKING

3.1.2.5 Enhanced Error Filtering

The extracted trajectories may still have some bias due to factors such as the shade of trees or buildings, inevitable detection errors, and unobservable vehicle types in the training set. Therefore, we developed a data fix tool to further check potential errors. The data fix tool allows to manually delete wrong detection objects and adjust the size and heading of the bounding box. Then, the adjusted data was returned to the training dataset and participated in next rounds of training to increase the detection accuracy, using the active learning idea[160, 161]. FIGURE 10 shows an example of adjusting vehicle bounding box using the data fix tool.

(A) BEFORE ERROR FILTERING **(B) AFTER ERROR FILTERING**

FIGURE 10 USING DATA FIX TOOL TO ADJUST BOUNDING BOX ERROR (THE LIGHT BLUE BOUNDING BOX IS FIXED IN THE FIGURE).

3.1.2.6 Conflict Identification

FIGURE 11 shows the flowchart of conflict and conflict type identification. In order to calculate PET values, a vehicle occupancy table needs to be generated, which include the timestamps that each part of each vehicle (front, middle, rear) that arrive and leave each pixel that the vehicle has occupied in the video. Then, PET values could be calculated by comparing the timestamps for two consecutive vehicles at each pixel. The PET is calculated as time difference between the first vehicle leaving the pixel and the time that the second vehicle arrives at the pixel. Conflicts could be identified by filtering out the events that have PET less than the threshold. In this study, a sensitivity analysis was conducted for different thresholds. In order to find the precise locations and time of the identified conflicts, the earliest arrival time and pixel of the second vehicle are utilized as the corresponding conflict point and conflict time.

Then, moving direction of the vehicles, Intersect of Pixels (IOP), and vehicle occupancy table are utilized to identify conflict types (i.e., head on, angle, rear-end, sideswipe). IOP is defined as the percentage of the pixels that have been occupied by both vehicles before the conflict event happens over the pixels that have been occupied by either vehicle. The value is employed to identify if the

second vehicle is following the first vehicle, which could be used to determine if the conflict is rear-end collision conflict or not. If the angle between the two vehicles' moving direction is between 0 degree and 45 degree or between 315 degree and 360 degree, and the IOP is less than or equals to the threshold, the conflict is identified as a rear-end conflict; If the angle is between 0 degree and 45 degree or between 315 degree and 360 degree, and the IOP is greater than the threshold, the conflict would be sideswipe conflict; Meanwhile, if the angle is between 135 degree and 225 degree and the conflict parts are not the front parts of the vehicles, the conflict would also be identified as sideswipe conflict; If the angle is between 135 degree and 225 degree and the conflict parts are the front parts of the vehicles, the conflict would be a head-on conflict; Otherwise, the conflict would be identified as an angle conflict.

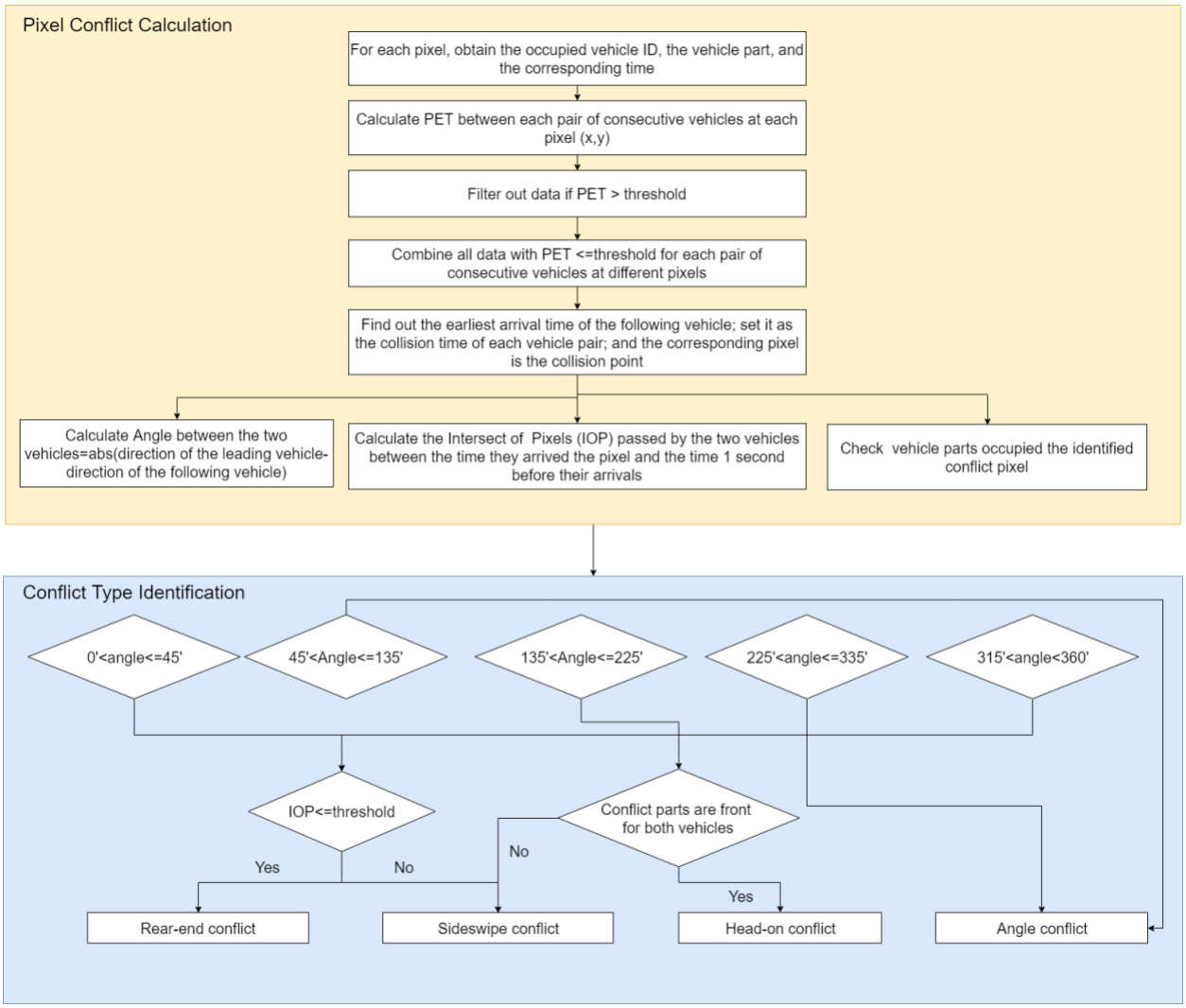


FIGURE 11 CONFLICT IDENTIFICATION

3.1.3 System Performance

To validate the proposed conflict diagnose framework, data collection was conducted on October 8th, 2018 from 8:30 AM to 8:50 AM at a typical 4-leg intersection at the University of Central Florida (UCF). A DJI Phantom 4 UAV was utilized to collect the data, and the video was captured by an optical camera with 1920 ×1080 resolution. Mask R-CNN detection was conducted at intervals of 0.5 seconds by Keras [162].Meanwhile, the position of the vehicles was tracked at intervals of every 115 second (15 frames per second mode) by OpenCV. All experiments are

conducted using Python implementation on a desktop computer with Intel i9-7980XE (18 cores and 36 threads) @ 4.2Hz, 64 GB DDr4 (3200MHz) memory and two Nvidia 2080ti GPUs.

3.1.3.1 Detection and Tracking Performance

In total, 1,588 vehicles were detected and tracked based on the UAV video. The accuracy of the proposed algorithm was evaluated based on IOU values. For each type of movement (i.e. left turn, right turn, straight), 20 vehicles were randomly selected from the UAV video. IOU values were calculated at the movement duration for each vehicle based on the outputs and the ground truths that were collected manually. Totally 3,541 video images were collected to calculate the IOU for the selected vehicles. Higher IOU values indicate higher accuracy of the detection and tracking results. Moreover, simple Mask R-CNN detection with CSRT tracker was conducted in order to compare with the performance of the proposed algorithm. As shown in

TABLE 1 COMPARISON OF AVERAGE IOUS, the proposed algorithm could significantly improve the performance for all type of movements, especially for turning movements (i.e. left turn, right turn). Also, the straight movements have the best performance for both methods.

TABLE 1 COMPARISON OF AVERAGE IOUS

	Simple Mask R-CNN detection and CSRT tracker	Proposed Algorithm
Left Turn	0.42	0.78
Right Turn	0.38	0.70
Straight	0.80	0.85
Average	0.53	0.76

3.1.3.2 PET Diagnostics

In order to calculate the PET values for each pixel data, vehicles' occupancy table for each pixel was generated based on the vehicles' trajectories. A table with 68,211,472 observations were generated, which includes arrival and leave timestamps for every pixel of each part of the vehicles, and PET values were calculated based on the table. The PET values were calculated between two consecutive vehicles that have occupied the same pixels in the video, and one value is returned for each conflict with the corresponding conflict location at the pixel level. The conflict events with the corresponding locations of the identified potential conflicts were obtained based on the PET values. Moreover, heatmap could be generated based on the locations of the conflict events to investigate the spatial distribution of the conflicts.

FIGURE 12 (a) shows the conflict heatmap that were identified based on vehicles' trajectories where PET values are less than 1.5 s. As it is shown in the figure, the right-turning lane of the northbound has the highest risk. The results are expected as the right-turning vehicles have more frequently stop-and-go behavior in order to yield to vehicles from the other directions and lead to higher rear-end collision risk. Since the data collection was conduct at a typical morning peak duration, many vehicles turn right at the northbound to enter the UCF campus area.

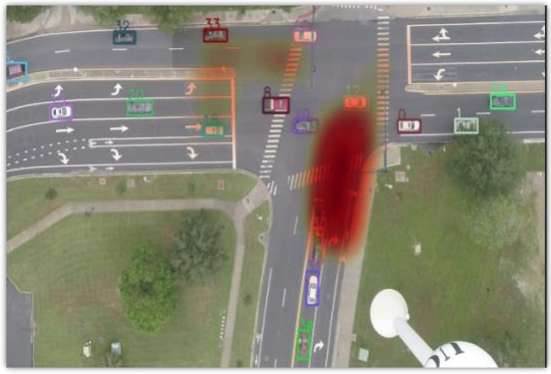
FIGURE 12(b) displays the number of conflicts based on different PET thresholds. Different PET values could be employed to determine conflict risk levels, while smaller PET values indicate higher risk (*29; 30*). Significant difference could be observed for the number of conflicts with the increase of PET thresholds. If the PET threshold is 1 s, only 2 conflicts were identified as high-risk conflicts. When the threshold increases to 1.5 s, the conflict count increased to 22. Moreover, the number of conflicts increases to 90 when the PET threshold increases to 2 s.

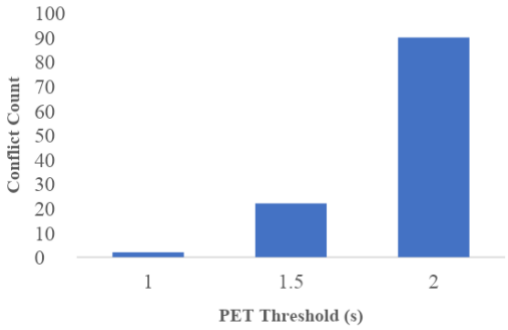
From the UAV video, most of the identified conflicts are rear-end collision conflicts, only 1 angle collision conflict was identified with PET value equal to 1.7 s.

FIGURE 12 (c) shows an example of angle collision conflict, where a left-turning vehicle (vehicle #948) has a potential conflict with the straight moving vehicle (vehicle #955).

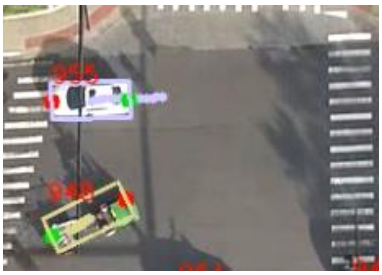
FIGURE 12(d) provides an example of rear-end collision conflict between vehicles #285 and #287 which is due to relatively small headway.

Based on the diagnosis results, countermeasures such as adding a dynamic message sign or beacon could be implemented at the upstream of the northbound approach to reduce the right-turn conflicts.





(A)CONFLICT HEATMAP **(B) CONFLICT COUNTS AT DIFFERENT PET THRESHOLDS**



(C) AN EXAMPLE OF ANGLE COLLISION CONFLICT. (D) AN EXAMPLE OF REAR-END COLLISION CONFLICT

FIGURE 12 SAFETY DIAGNOSTICS

3.1.4 Conclusions and Future Work

This research proposes an automated framework for safety diagnosis utilizing Mask R-CNN detection algorithm and Channel and Spatial Reliability Tracking (CSRT) multi-object tracking algorithm for UAV videos. A case study was conducted at a typical signalized intersection at UCF. The case study has validated the feasibility of investigating safety situation from UAV videos based on surrogate safety measures (i.e. PET). It also demonstrated that the proposed methods of using computer vision techniques to automatically extract vehicles' trajectories and

identify conflicts from UAV videos have better performance in terms of data accuracy, especially for turning vehicles (i.e., right turning, left turning). To the best of the authors' knowledge, this is the first study that proposes a framework to automatically identify conflict types from video-based trajectory data without using other types of data (e.g., road geometry, lane configuration). Sensitivity analysis for PET threshold was conduct in the study. The results of the identified conflicts indicate that rear-end conflicts is the most prevalent type of conflicts for the studied intersection, while only one angle collision conflict was identified between a left-turning vehicle and a through vehicle. Moreover, the right-turning lane at the northbound was found to have the highest risk where many vehicles turn right to enter the UCF campus and frequently stop-and-go behavior were present in order to yield to the vehicles from other directions.

In the future, different type of objects (e.g. pedestrian(FIGURE 13), cyclist) could be detected with enriched UAV datasets for detection. Also, other surrogate safety measures (e.g. TTC) could be utilized to further explore the safety conditions of the study area.

FIGURE 13 EXAMPLE OF PEDESTRIAN DECOCTION FROM DRONE VIDEO

3.2 Near Miss Event Detection System (N.M.E.D.S)

3.2.1 Introduction

Various methods could be utilized to conduct road safety analysis. The traditional crash count analysis uses highly aggregated data to evaluate road safety for certain situations[56]. However, one of the limitations of using highly aggregated data is that it could not consider the heterogeneity of traffic or the surrounding environment, which could have an impact on the estimation accuracy. In recent decades, the wide employment of traffic infrastructure-based sensors could generate big data in real-time that enables researchers to conduct crash likelihood analysis by aggregating it into shorter time intervals. By aggregating the data from nearby traffic detectors (e.g., MVDS, Loop Detectors) and the other infrastructure elements (e.g., weather stations, signals) at certain time intervals, precursors of crash occurrence could be identified, which could be used to investigate road safety situations and identify corresponding strategies to prevent crashes.

Although the deployment of traffic infrastructure sensors brings big data and enables the analysis of real-time crash risk, the rarity of crashes leads to some limitations to safety research and practice. The use of conflicts could overcome the limitation of crash-based safety analysis and provide a compelling explanation to crash causations while considering the interactions between road users and other behavioral factors by obtaining individual road users' data. First, only after crashes have been observed can researchers and practitioners evaluate the safety performance of the studied locations and situations. The use of conflicts provides an alternative method to investigate road safety conditions that overcome the limitations of crash analysis[61]. Second, although risky situations are present, crashes may still be prevented through evasive actions, such as emergency brake and lane changing. However, the traditional infrastructure-based traffic data could barely capture information of individual road users' behavior, and the crash prediction model could be prone to providing false alarms [62]. Moreover, as human factors are solely or in conjunction with other factors contribute to around 95% of crash occurrence, traffic data from detectors may not be sufficient to identify crash precursors for many situations[63] .

Many previous studies [22, 131, 163] have been conducted for the analysis of intersection safety by using total crashes at the intersection and intersection-related area (with a certain distance from the center of intersections). The effects of traffic, geometric design, and surrounding land use were identified for the whole intersection area. The intersection area could be divided into three types, which are intersection entrance area (where vehicles approach the stop line), within-intersection area (where vehicles make through, left-turn, right-turn, and U-turn movements), and intersection exit area (where vehicle pass the within-intersection area). Some recent studies attempted to explore traffic safety in the intersection entrance area by using traffic data of detectors at the entering approaches. For example, researcher used the Automated Traffic Signal Performance

Measures (ATSPM) data to predict crash risk at the entrance area in real-time. Researcher developed traffic conflict models to evaluate the safety of the entrance area at the cycle level based on vehicles' trajectory data from the camera. In the within-intersection area, a large number of conflict points are generated by turning vehicles of different movements and drivers are more prone to make mistakes, which could lead to severe crash outcomes. However, only limited studies have been conducted about the safety in the within-intersection area due to the lack of detectors. Some researcher developed models to predict crash risk in the within-intersection area by using the ATSPM data at the entering area, which may not capture the traffic interaction in the within-intersection area. This study attempts to analyze the safety in the within-intersection area by using Closed-Circuit Television (CCTV) Cameras. A new vehicle detection method is proposed to improve the trajectory accuracy for safety analysis. Also, a spatial analysis method is developed to identify safety hotspots.

3.2.2 Data Collection

The data collection was conducted on two different locations to validate the proposed framework (**FIGURE 14**):

1. October 30, 2019, from 8:30 am to 9:00 am at a Four-leg intersection in Lake Mary, Florida
2. May 24, 2021, from 9:30 am to 8:00 pm at a Four-leg intersection in, Casselberry, FL

The CCTV camera videos were collected for both locations. Besides, drone video was collected for Location 2, which will be used to further validate the proposed framework. The covered area by the camera is illustrated by a rectangle in Figure 9. Only vehicles' trajectories were collected to identify potential conflicts, and no confidential information was obtained. Mask R-CNN detection was conducted at intervals of 0.5 seconds to identify the bounding box, and the key points

detection was conducted at intervals of every 1/20 second (20 frames per second mode) using Occlusion-Net. Meanwhile, the position of the vehicles was tracked at intervals of every 1/20 second by OpenCV [164] The vehicle trajectories were transformed into the real-world coordinate system and modified to get the occluded points. All experiments are conducted using Python implementation on a desktop computer with Intel i9-7980XE (18 cores and 36 threads) @ 4.2Hz, 64 GB DDr4 (3200MHz) memory, and two Nvidia 2080ti GPUs.

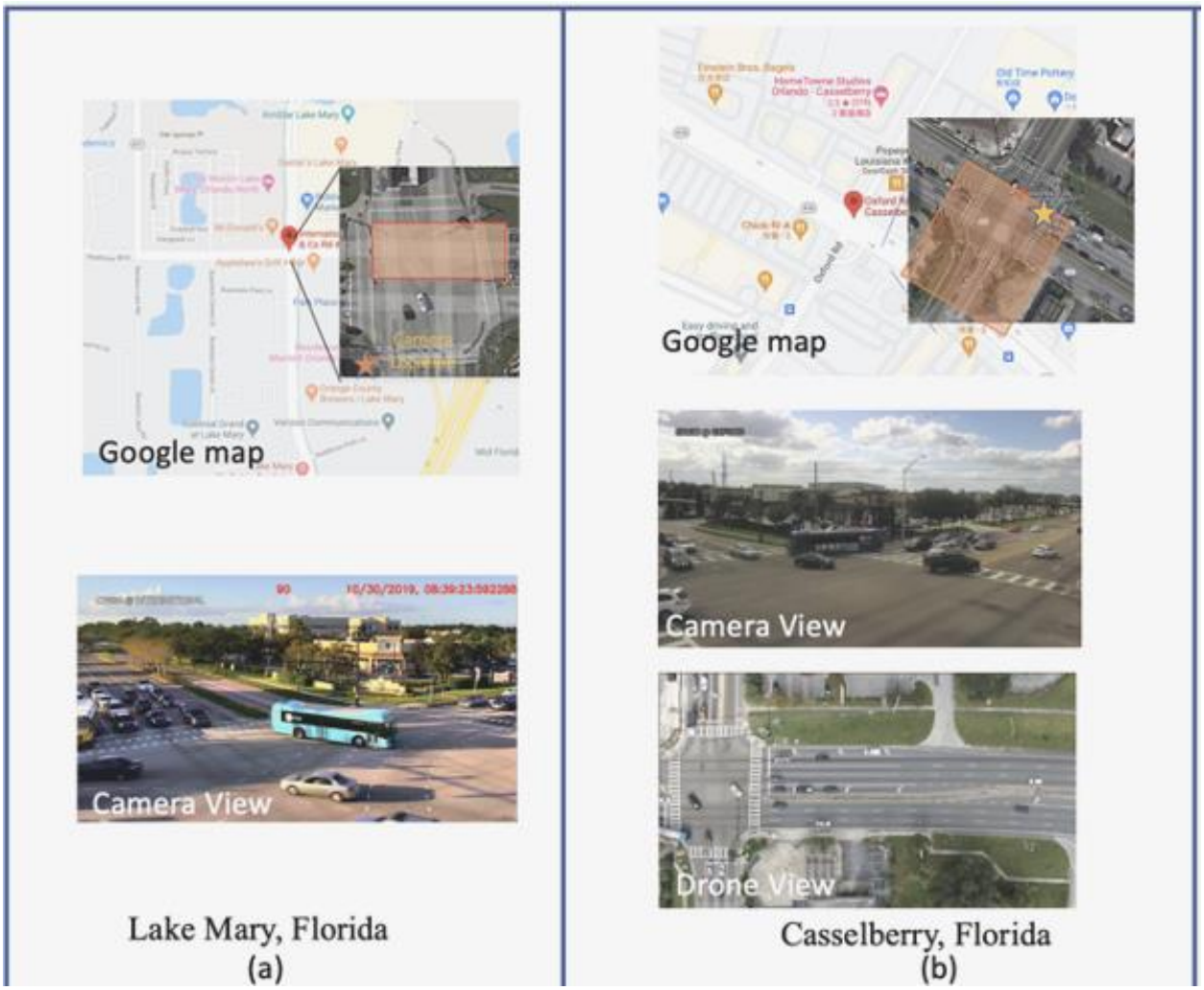


FIGURE 14 DATA COLLECTION LOCATION

3.2.3 Methodology

FIGURE 15 shows the framework of conducting safety diagnostics using CCTV cameras. The first step of the proposed framework is detecting vehicles from video images. Mask-RCNN is utilized as the detection algorithm in this study. The Mask-RCNN model could provide a classification of the objects in the video images and predict pixel-to-pixel manner segmentation masks. Meanwhile, a 2D bounding box is generated around each mask to indicate the locations of the detected objects. The four vertexes of the bounding boxes and the corresponding center points are obtained as the initial localization for each vehicle. The second step is vehicle key points detection and tracking. Researcher proposed a framework named "Occlusion-Net" to predict key points for objects including the occluded points that cannot be observed in the images due to the angle of view or the obstruction of other objects[82]. The key points could be utilized to provide more accurate localization. In this study, Occlusion-Net is employed to find the four key points of vehicles (i.e., right-front headlight, left-front headlight, right-back taillight, left-back taillight). Those key points are utilized to find more precise information about the vehicle occupying area, which would be important for calculating surrogate safety measures. Then, data filters will be applied to obtain the datasets that could be utilized to calculate surrogate safety measures (e.g., TTC, PET). The data filters include perspective transformation and data smoothing. The outputs from steps 1 & 2 are in a pixel format (x,y) based on the video images. Perspective transformation is employed to transfer the outputs to a real-world coordinate format, or a coordinate system based on the bird's-eye view of the study area[165]. Since the Occlusion-Net algorithm predicts the occluded key points based on the roadside view, occluded key points modification could be conducted using a 2D bird's-eye view. Finally, safety diagnostics could be conducted based on the outputs from step 3 by identifying conflicts through calculating surrogate safety measures.

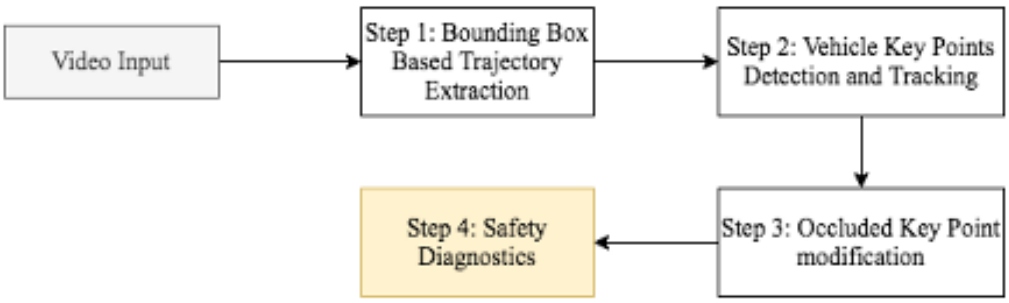


FIGURE 15 THE PROPOSED FRAMEWORK

Steps 1&2: Detection and Tracking

Bounding box vehicle detection and tracking

We proposed a framework named "Automated Roadway Conflict Identification System (ARCIS)" to detect and track vehicles from Unmanned Aerial Vehicle (UAV) images. Mask-RCNN was employed as the detection algorithm to find untracked vehicles from video images Meanwhile, Channel and Spatial Reliability Tracking (CSRT) algorithm was utilized to obtain the trajectories of the detected vehicles. Moreover, missing or lost vehicles could be found by comparing the Intersect of Union (IOU) between the detection results and the tracking results). In this study, the previous efforts are extended to videos from CCTV cameras. As a vast amount of data is needed to train a deep neural network object detection, a pre-trained weight using COCO 2017 dataset that includes 12,786 cars, 61,377 trucks, and 4,141 buses is utilized to find road users from video images. FIGURE 16 presents examples of the detection results based on Mask-RCNN. The rectangles are the bounding boxes of detected objects that were obtained based on the masks, and

the dots are the center points of the bounding box. It is worth noting that although most of the existing studies use one point of each object (e.g., center point) to represent vehicle locations and obtain vehicle trajectories, the length and width of vehicles should not be ignored, especially when conducting safety analysis. Meanwhile, some studies use bounding boxes to represent the occupying area of vehicles. However, the bounding box could not reflect the precise locations of vehicles, as the bounding boxes are obtained based on a 2D roadside view instead of 3D detection or 2D bird's-eye view.

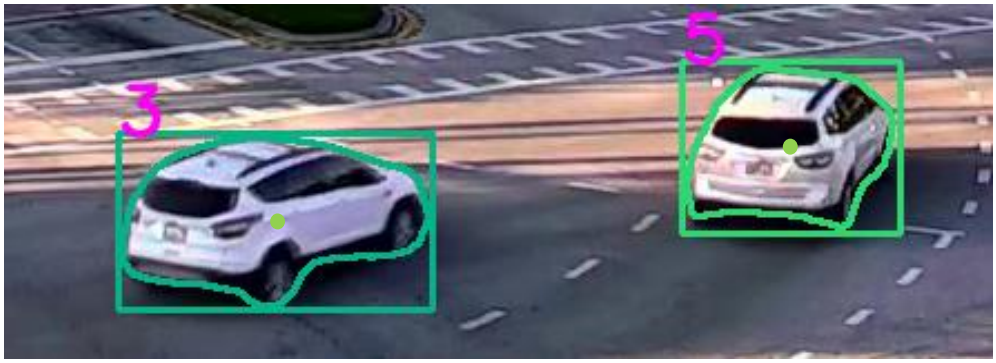


FIGURE 16 MASK-RCNN DETECTION

Vehicle key points detection

One of the methods to obtain more precise vehicle localization is using 3D reconstruction by identifying the key points on the objects. In this study, four key points' detection results of each vehicle are archived for each vehicle as outputs using the Occlusion-Net algorithms including right-front headlight, left-front headlight, right-back taillight, and left-back taillight at the detected frames. A pre-trained model developed by using open datasets is utilized in this study. While the keypoints detection could provide the 3D detection results, it could miss detections of some vehicles[82].Hence, this study combined the bounding box detection and key points detection to improve the accuracy, i.e., vehicle key points detection is conducted within the bounding box areas. Thus, key point detection results for the same vehicle at different frames could be identified by

matching the key point detection results with the bounding box detection and tracking results. The combined detection process could also improve the detection efficiency since the key points detection is only conducted in the bounding box area instead of the whole road. **FIGURE 17** illustrates an example of the detection and tracking results. FIGURE 17(a) shows the original image, and FIGURE 17(b) shows the generated vehicle bounding box for the two detected vehicles based on the Mask-RCNN algorithm (red rectangles) and the vehicle key points based on Occlusion-Net. It is worth noting that most of the vehicles from roadside views have occluded key points. The occlusion may be due to the angle of view (self-occlusion) or other nearby objects, which is very common in the within-intersection area. Thus, a modification method is introduced to get occluded key points in the following section.

(a) ORIGINAL IMAGE **(b) BOUNDING BOXES AND KEY POINTS**

FIGURE 17 VEHICLE DETECTION AND TRACKING

FIGURE 18 illustrates the vehicle detection and tracking procedures, which include bounding box vehicle detection & tracking and vehicle key points detection & tracking. The first four steps are used to ensure all vehicles could be detected and tracked. Then, the bounding box of each vehicle could be generated, where the key points detection is conducted. Two types of outputs could be

generated for all the vehicles in the video images at each timeframe: (1) bounding box data; (2) vehicle key points data.

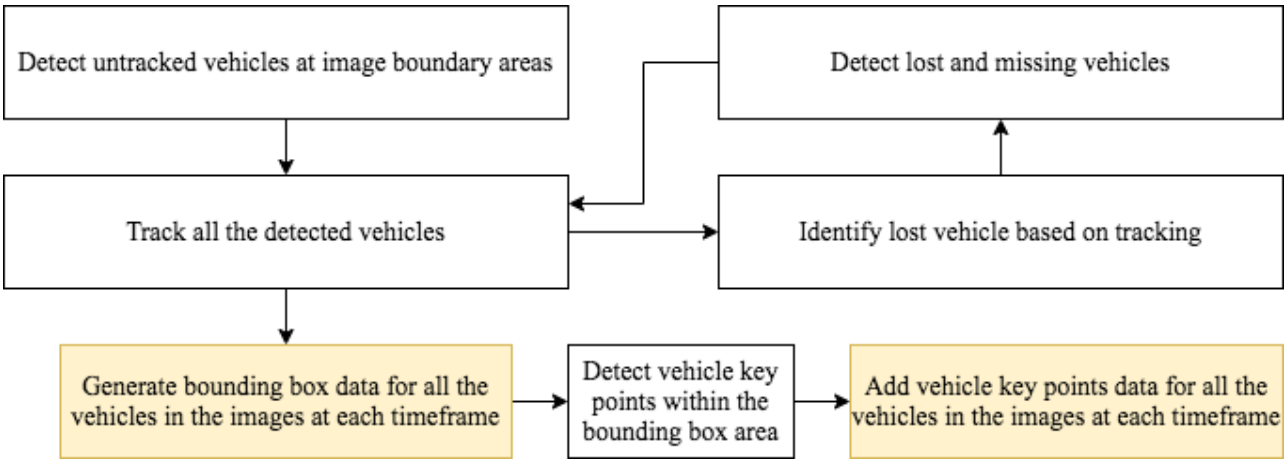


FIGURE 18 VEHICLE DETECTION AND TRACKING FLOW CHART

Step 3: Perspective Transformation and Occluded Key Points Modification

As illustrated in FIGURE 19, this step has two major parts: (1) key points transformation; (2) occluded key points modification. First, the output data collected from detection and tracking are transferred to coordinates format based on perspective transformation. The coordinates format data could be plotted on the Google satellite image at the selected location. Thus, the following data processing and analysis are based on the coordinate system of bird's-eye view (i.e., google satellite image) instead of the roadside view. After the perspective transformation, moving average is conducted for data smoothing and aggregated for a certain duration to make sure the trajectory of vehicles is smooth. In FIGURE 19, the red polygon is the bounding box after transforming to a 2D bird's-eye view, and the orange polygon is the key points detection results. The box should be outmost parts of the vehicle. It is shown that the bounding box areas tend to be much larger than the areas based on the key points detection. Thus, using the bounding box detection to conduct safety analysis could lead to significant errors since vehicle lengths and front/rear locations play important roles when calculating surrogate safety measures. Some studies even use center points

of the bounding boxes to estimate surrogate safety measures. However, after the perspective transformation, the center points are not the accurate center points of vehicles, and the vehicle's length estimation based on bounding boxes could tend to be longer than the true values. Moreover, it is known that each vehicle's shape should be close to a rectangle from the bird's-eye view. As noted above, most of the vehicles from roadside views have occluded key points. Occlusion-Net could predict occluded key points using the roadside view. After the perspective transformation, the occupying area of vehicles may not be a rectangle or a parallelogram due to the errors of occluded key points detection. Due to the detection errors, the detection for occluded points (the green point in FIGURE 19(d)) tend to be less accurate than the observed points and the shape of the vehicles may not be rectangular. Thus, the modification of occluded key points is needed in the 2D bird's-eye view. In this study, an occluded point modification method is proposed by using the locations of the observed points. A triangle is generated based on the observed key points, and a parallelogram could be generated based on the triangle. As shown in FIGURE 20, different parallelograms could be generated. The key point detection could provide the location information of three observed points corresponding to the vehicles. Hence, the correct parallelogram could be determined. Finally, the locations of the occluded points are modified based on the parallelogram (the red point in FIGURE 20(g)).

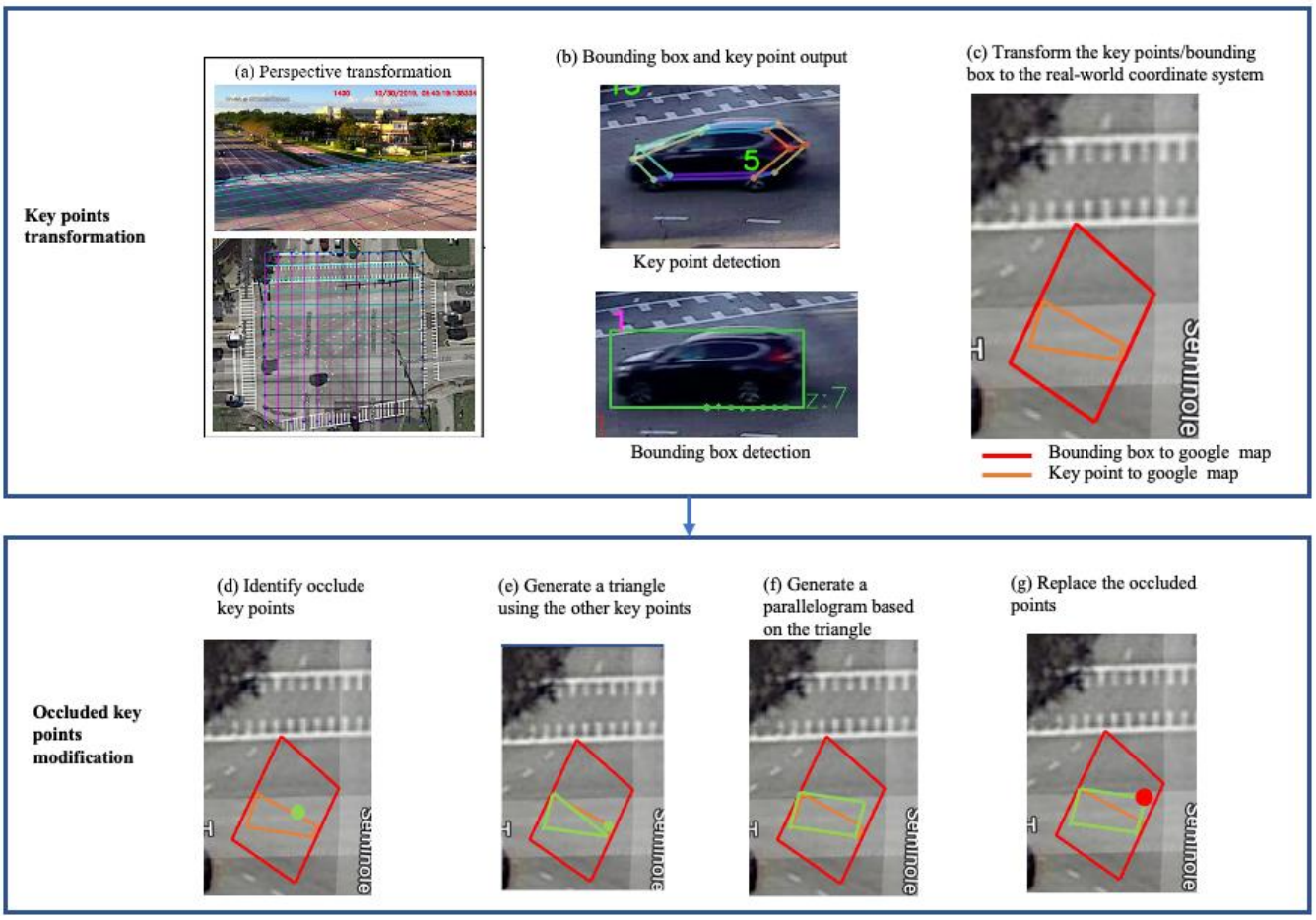
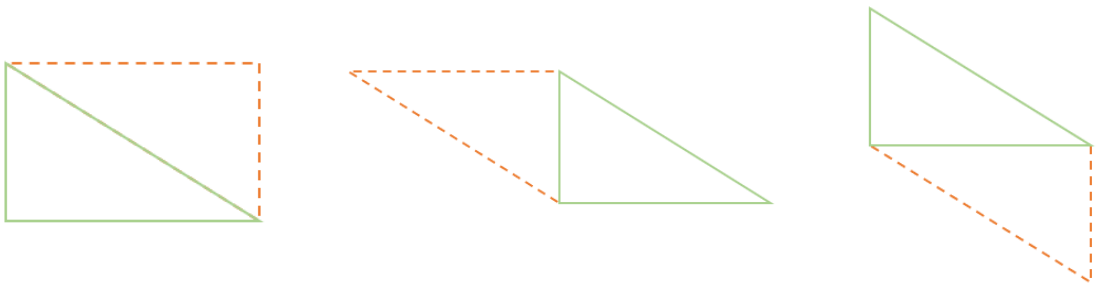


FIGURE 19 OCCLUDED KEY POINT MODIFICATION



(a) **(b)** **(c)**

FIGURE 20 ILLUSTRATION OF PARALLELOGRAMS BASED ON THE TRIANGLE

Step 4: Near Miss/Conflict Identification

PET values are calculated in this study to identify near misses/conflicts between vehicles. FIGURE 21 illustrates the procedures of PET calculation and conflict detection. First, a polygon is generated based on the key points detection and modification for each vehicle at each time frame. For example, for vehicle i, a polygon P1 would be generated based on the key points detection and modification results at time t1. Then, other vehicles' polygons are compared with the polygon to identify the polygons that have intersected with P1. If the polygons are intersected with each other, it indicates those vehicles have occupied the same locations as vehicle i. By comparing the occupying time of the two vehicles, a list of vehicle pairs with potential conflicts could be obtained. Then, we assume that each vehicle could only have one conflict for each frame, the conflict event with minimum PET value would be selected. Thus, redundant data could be filtered out and the near miss/conflict detection results would be generated for further safety diagnostics.

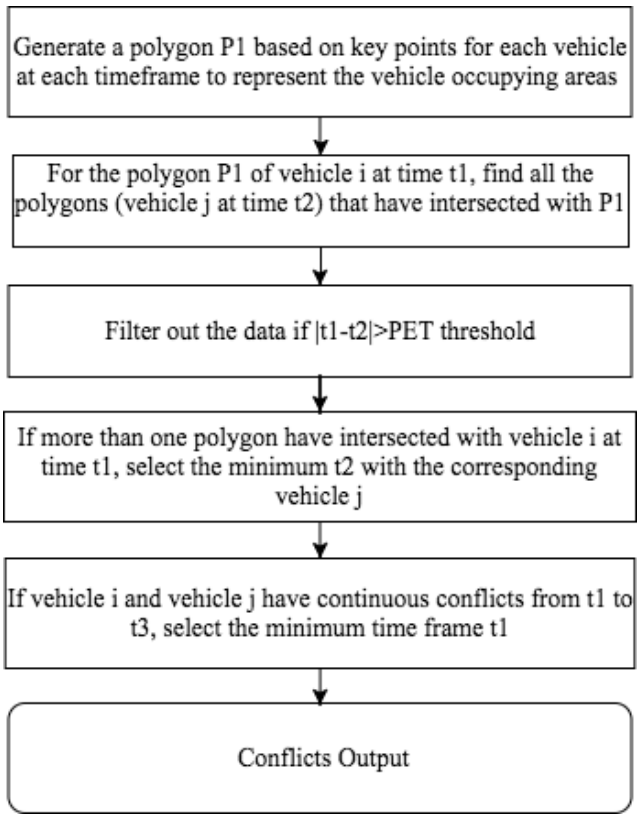


FIGURE 21 CONFLICT IDENTIFICATION

In summary, this study proposed a method to obtain accurate vehicle trajectories from the CCTV videos, with the purpose of traffic safety analysis in the within-intersection area. The method combined the bounding box detection and key points detection algorithms to detect and track vehicles more accurately and efficiently. The key points of vehicles in a 3D reconstruction could be obtained and transformed into the real-world coordinate system based on the perspective transformation. Then, a modification method was proposed to get the correct occluded key points of vehicles. Finally, a framework was developed to identify conflicts based on PET values, which are calculated by using vehicles' trajectories.

3.2.4 System Performance

In order to investigate the performance of the proposed framework, its performance is compared with the following three approaches commonly used in the literature:

(1) Bounding box: vehicle key points detection is not included. Only Mask-RCNN and CSRT tracker are employed to obtain the bounding boxes of the vehicles and the corresponding trajectories;

(2) Key point detection only: Occluded key points modification is not included. The key points by Occlusion-Net were used directly.

(3) Drone data: Drone video process by UCF SST Automated Roadway Conflict Identification System (ARCIS) and manually corrected misdetection data.

Based on the collected data, 514 video images were used to test the proposed performance. In total, 925 vehicles are included in the 514 video images. For each vehicle, the four key points (i.e., right-front headlight, left-front headlight, right-back taillight, and left-back taillight) were collected manually based on the original roadside view as the ground truths. Then, the ground truths are compared with the results from the bounding box detection & tracking, key points detection only, and the proposed framework. Figure 10shows an example of the comparison. In Figure 10(e), the green box indicate the ground truth locations of key points that are collected though drone video Also, a polygon could be generated based on the key points (orange polygon). Meanwhile, the blue line indicates the mask of the vehicle based on the Mask-RCNN detection, while the blue rectangle is the bounding box generated based on the mask. Figure 22 provides examples of the detection and tracking results for both bounding box and key points. Figure 22 (a) and Figure 22 (b) show that the bounding box detection and key points detection could successfully detect vehicles from the camera videos. Figures Figure 22 (c) show example of ARCIS detection output.

Figure 22 (e) and Figure 22 (d) illustrate the comparison of ground truth and results of different detection methods in the roadside and bird's-eye views. As shown in Figure 22 (e), the bounding box (blue polygon) tends to be much larger than the ground truth (green polygon). The occluded key point detection only (upper right vertex of the orange polygon) has a relatively large error. Thus, the results based on the proposed frame (yellow polygon)in Figure 22 (g) shows the best performance in this example.

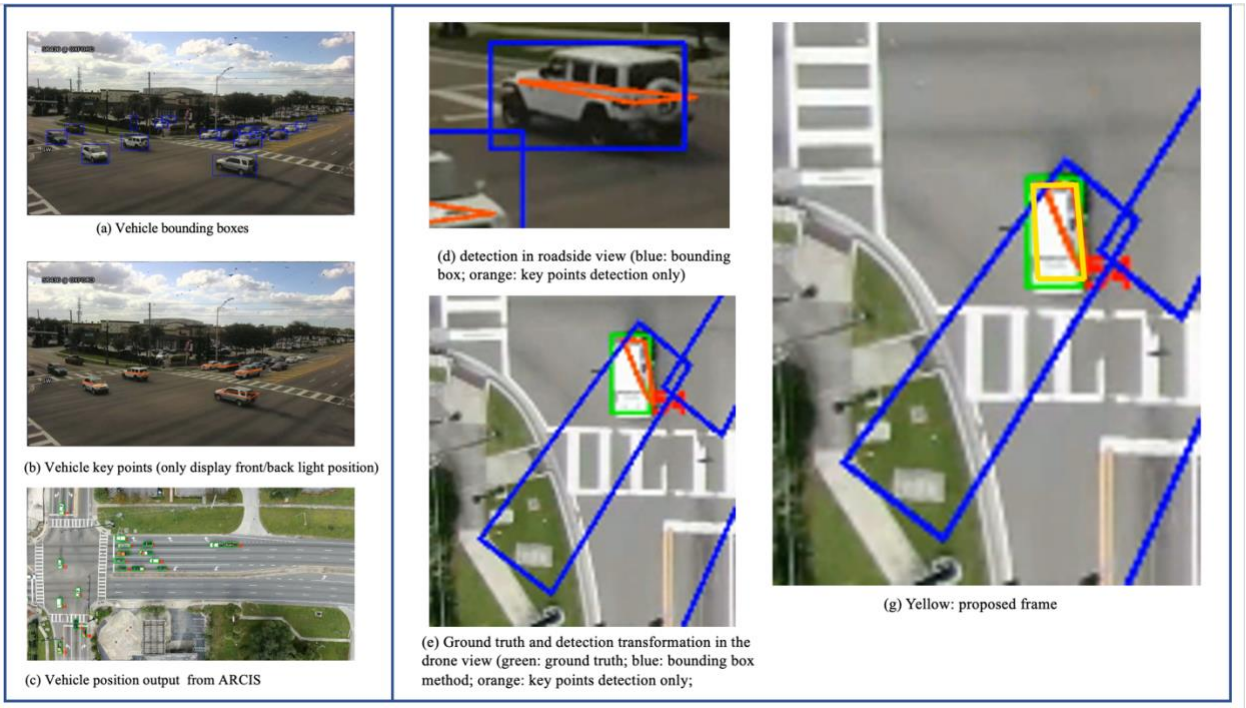


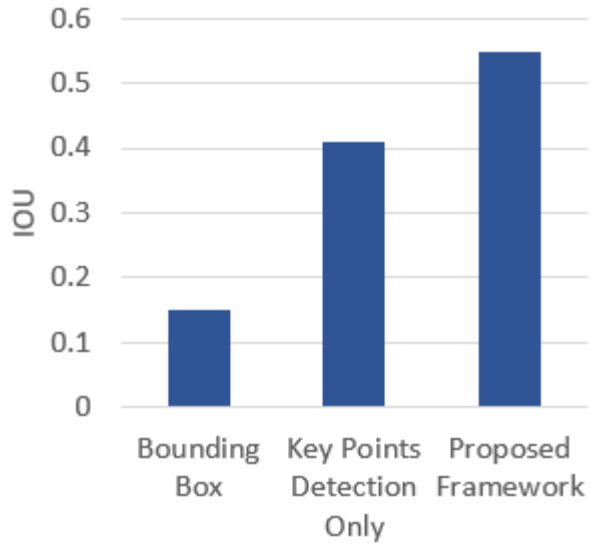
FIGURE 22 GROUND TRUTH DATA COLLECTION FOR LOCATION 2

In this study, the intersection of Union (IOU) is calculated to evaluate the accuracy of the vehicle localization. The IOU value is calculated as follows:

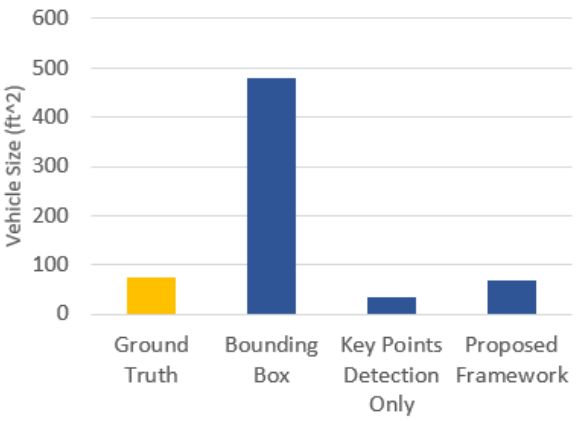
$$IOU = \frac{DetectionResult \cap GroudTruth}{DetectionResult \cup GroudTruth}$$

FIGURE 23 illustrates the comparison between different methods. For the first location, the average IOU for the bounding box is 0.15. Meanwhile, the average IOU of key points detection only and the proposed framework with key point modifications is 0.41 and 0.55, respectively,

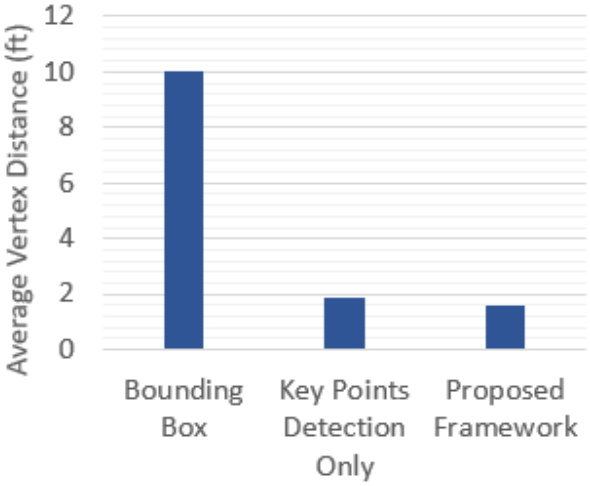
which shows significant improvement compared to the bounding box detection and tracking based on paired t-test (p-value<0.01) (FIGURE 23 (a)). The reason that the bounding box results have significantly lower IOU values is that the areas of bounding boxes tend to be much larger than the ground truths as shown in Figure 22. As shown in FIGURE 23 (b), the average vehicle size of ground truths is 74.57 ft^2, while the average vehicle size is 479.25 ft^2 for bounding boxes. However, due to occlusion point detection errors, the average vehicle size is 35.20 ft^2 based on key points detection only, which is smaller than the ground truths. After the modification based on the proposed framework, the average vehicle size value is 67.47 ft^2, which is closer to the ground truth. The results of the second location are shown in TABLE 2 SYSTEM PERFORMANCE COMPARISON (LOCATION 2). The paired t-test results between the proposed framework and bounding box detection/ key points only detection showed significant improvement with all the p-values smaller than 0.01.



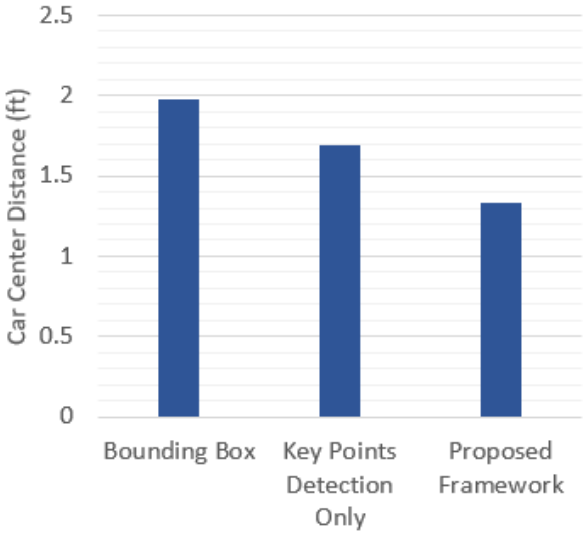
(a) IOU



(b) VEHICLE SIZE



(c) AVERAGE VEHICLE CENTER POINT DISTANCE



(d) AVERAGE VEHICLE KEY POINT DISTANCE

FIGURE 23 SYSTEM PERFORMANCE COMPARISON (LOCATION 1)

TABLE 2 SYSTEM PERFORMANCE COMPARISON (LOCATION 2)

	Ground Truth (ARCIS output UAV Data)	Bounding Box	Key Point Only	Proposed Framework
IOU	1.00	0.14	0.15	0.28
Vehicle Size (ft^2)	104.65	651.93 (MSE= 217319.11)	31.06 (MSE= 5838.93)	74.37 (MSE= 1274.23)
Average vehicle center point distance (ft)	0.00	14.35	2.27	2.23

Moreover, the average car center distance and average key points distance were calculated to evaluate the performance of the different methods. The values are defined as follows:

$$Average\ Car\ Center\ Distance = \frac{\sum \sqrt{(x_c^D - x_c^{GT})^2 + (y_c^D - y_c^{GT})^2}}{n}$$

$$Average\ Vertex\ Distance = \frac{\sum \sqrt{(x_i^D - x_i^{GT})^2 + (y_i^D - y_i^{GT})^2}}{4n}$$

Where (x_c^{GT}, y_c^{GT}) is the center point of the ground truth polygon; (x_c^D, y_c^D) is the center point of the detected polygon; (x_i^D, y_i^D) is the vertex point of the detected polygon. The comparison results are presented in FIGURE 23 (c) and FIGURE 23 (d). It shows that the proposed framework has the smallest average car center distance, which is 1.33 ft. Meanwhile, using key points detection could significantly reduce the average key points distance, and the proposed framework could slightly further reduce the error. The results indicate that the proposed framework could provide

more accurate localization from 2D bird's-eye views, which is expected to influence the results of near miss detection.

3.2.5 Conflict Analysis and Hotspots Identification

3.2.5.1 Comparison of Identified Conflicts

In the previous section, it has been validated that the proposed framework could significantly improve the accuracy of localization from the 2D bird's-eye view. In this section, PET values are calculated in order to identify the near misses/conflicts between vehicles. TABLE 2 SYSTEM PERFORMANCE COMPARISON (LOCATION 2)summarizes the number of near misses/conflicts with different thresholds using different methods. PET values could be utilized to indicate risk levels, and smaller PET values indicate higher risks. FIGURE 24 presents the density of PET values by detection methods. In this study, the PET values smaller than 3 seconds were considered as dangerous situations and utilized for the analysis. It is shown that the conflicts based on the key points detection and the proposed framework follow the normal distribution based on the Shapiro-Wilk test. Also, smaller PET values could be observed based on the bounding box detection method. The CCTV camera used to collect data is located at the bottom-left area in the video image, and the bounding boxes at the upper-right area tend to overlap with each other due to the angle of the view. FIGURE 25 shows an example where the bounding boxes of vehicle #5 and #12 are highly overlapped, which leads to a potential conflict with a small PET value. On the other hand, the key points detection could construct the vehicles' key points in a 3D view. The overlapping problem could be significantly mitigated.

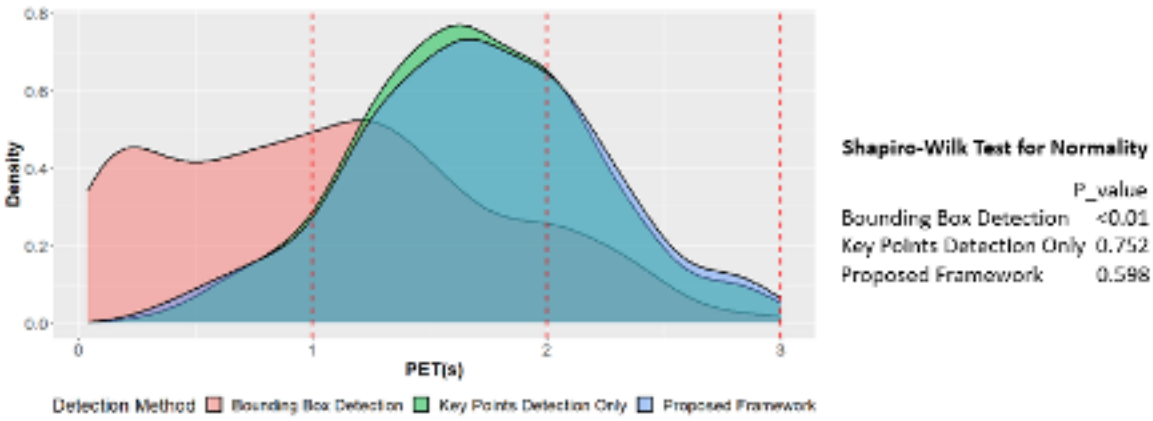


FIGURE 24 DENSITY OF PET VALUES BY DETECTION METHODS

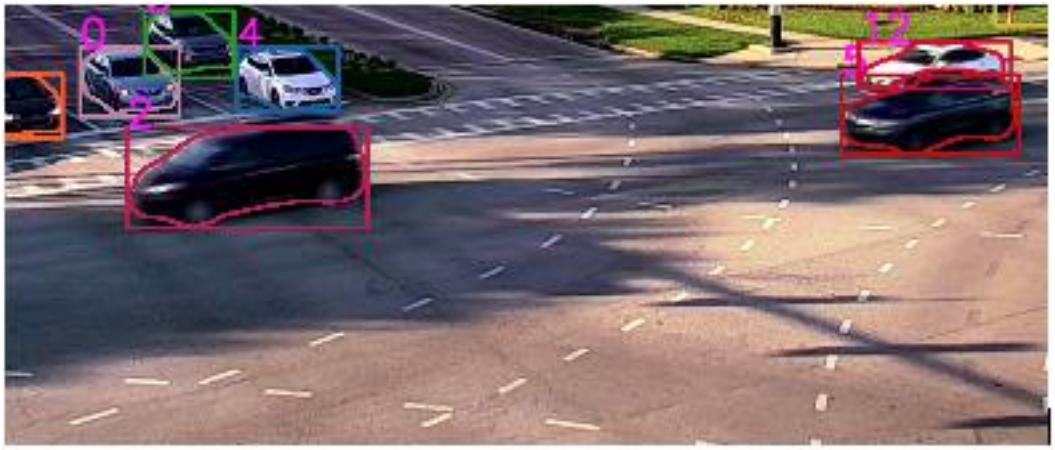


FIGURE 25 EXAMPLE OF VEHICLE OVERLAPPING IN A CCTV CAMERA

The conflicts were aggregated by 3 minutes and 10 groups were obtained for each detection method. FIGURE 26 presents the conflict counts by detection methods and time. Ismail (2011) suggest the PET threshold of 3 seconds to be considered as a severe conflict[11]. Researcher find a conflict with PET threshold of approximate 1 seconds is highly related to crash[11]. Hence, in this paper, we adopted three different PET values: 1 second, 2 seconds, and 3 seconds. The paired Wilcoxon signed rank exact test was conducted to compare the proposed framework with the two counterparts with three the different PET. The results are summarized in **TABLE 3 PAIRED**

WILCOXON SIGNED RANK EXACT TEST FOR PET AT LOCATION 1. A significant difference could be observed between the conflict counts identified by using the bounding box detection and the conflicts based on the proposed framework. Compared to the proposed framework, the bounding detection method at least detected 242% more conflicts. It is suggested the bounding detection is not appropriated for the safety analysis at the within-intersection area since the over detection of conflicts could raise unnecessary concerns and even cause many false alarms if the information of the detected conflicts is sent to the approaching vehicles through the Infrastructure-to-Vehicle (I2V) technology. On the other hand, the difference between the proposed framework and key points detection only is significant when the threshold is 2 and 3 seconds. The proposed framework could identify 2-3% more conflicts than the key points detection only method. The miss detection of conflicts by the key points detection only method should be due to the errors of occluded points. Hence, it could be concluded that the proposed framework could improve conflict detection in the within-intersection area by using CCTVs.

A smaller PET threshold represents a higher conflict severity. The paired Wilcoxon signed rank exact test is conducted for location 1, and the results are shown in TABLE 3 **PAIRED WILCOXON SIGNED RANK EXACT TEST FOR PET AT LOCATION 1**. The bounding box detection method detects way more conflicts than the other two methods at all three thresholds. This is due to the bounding box expand the vehicle size, causing potential overlap between vehicles, which reduces the PET value. When compared to the key points detection only method, the proposed framework showed significant performance improvement for detecting less severe conflicts ($p=0.03$ and $p=0.01$ for PET threshold of 2 seconds and 3 seconds respectively). For the PET threshold of 1 second, although the improvement is not statistically significant, it still detects 1% more severe conflicts than the key points detection only method.

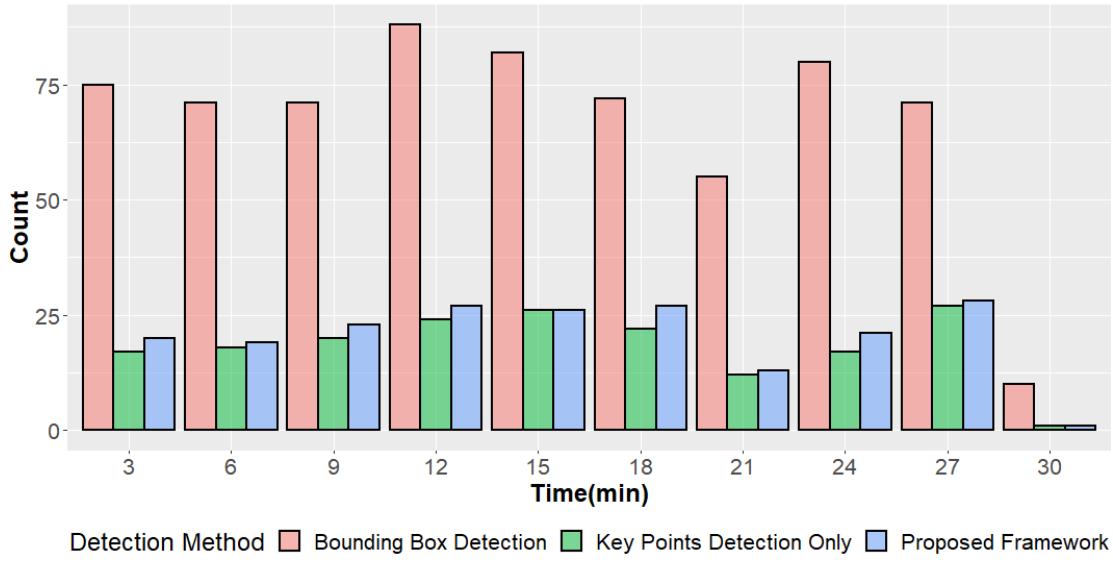


FIGURE 26 CONFLICT COUNTS BY DETECTION METHODS

TABLE 3 PAIRED WILCOXON SIGNED RANK EXACT TEST FOR PET AT LOCATION 1

Threshold	Bounding box detection	Key points detection only	Proposed framework	Comparison between proposed framework and bounding box	Comparison between proposed framework and key points detection only
1 second	310	14	18	1622% (p_value=0.02)	-1% (p_value=0.63)
2 seconds	609	132	143	326% (p_value<0.01)	-2% (p_value=0.03)
3 seconds	694	185	203	242% (p_value<0.01)	-3% (p_value=0.01)

For Location 2, conflicts with PET value smaller than 3 seconds are examined for the three vehicle detection methods. Totally 78 conflicts were identified during the period of video data. The difference in the PET value of each conflict between our proposed algorithm and bounding box/key point detection is calculated, and the results are plotted in FIGURE 27 **DIFFERENCE IN PET BETWEEN PROPOSED VS BOUNDING BOX AND PROPOSED VS KEY POINT**

METHOD. The average difference in PET value between the proposed framework and bounding box detection is 0.808 (p-value<0.01), while the average difference between the proposed framework and the key point detection only method is -0.146 (p-value<0.01). This is because the bounding box detection usually expand the size of the vehicle, while the key point only detection tend to miss a key point and reduce the vehicle size. A bigger vehicle size leads to a smaller PET value given the same vehicle center. Hence, the PET value calculated based the proposed method is greater than the bounding box value and smaller than the key point detection value.

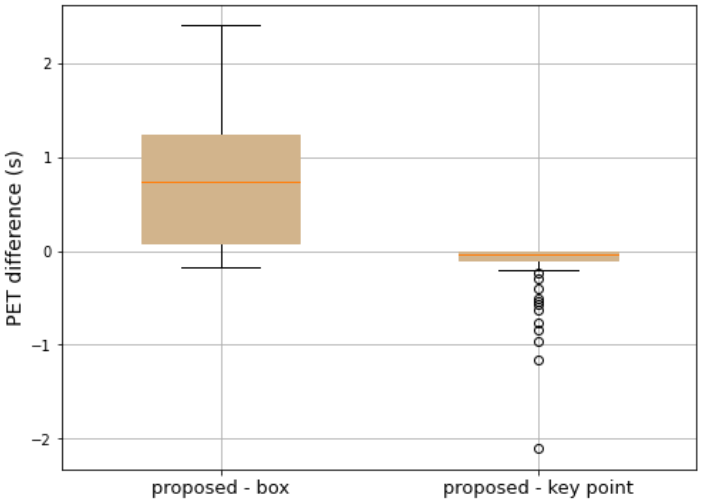


FIGURE 27 DIFFERENCE IN PET BETWEEN PROPOSED VS BOUNDING BOX AND PROPOSED VS KEY POINT METHOD

3.2.5.2 Hotspots Identification

The identified conflicts based on the proposed framework could be used to identify the hotspots, which could help develop the corresponding countermeasures to improve safety. The conflict points in the real-world coordinate system could be determined based on the corresponding vehicles' transformed location in the bird's-eye view. For hotspots analysis, kernel density is the most common method used in the field of traffic safety. While the kernel density could create a smooth and continuous surface of the density of observations for visualization, it has two major

limitations. First, the plot heavily depends on choosing the appropriate bandwidth to best display the data. If the choice of bandwidth is incorrect, it can distort the data by over- or under-smoothing . Second, it could not identify statistically significant hotspots. Hence, it is difficult to tell whether the clusters are random or there are some underlying spatial patterns. Instead, a grid-based hotspot identification method was proposed. Two spatial statistic methods including Moran's I and Getis-Ord G_i^* were used. The study area is divided evenly into grids with uniform side length (FIGURE 28). Then, the identified conflicts are assigned to grids based on the locations.

FIGURE 28 ILLUSTRATION OF GRIDS IN THE STUDY AREA

The grid-based analysis may have the Modifiable Areal Unit Problem (MAUP), which means that the analysis results might be affected by the side length of grids. The Moran's I index is used to measure the spatial autocorrelation by different side lengths from 2 to 20 feet. The length within the most significant spatial autocorrelation will be used for the hotspots identification. Moran's I combines the measure for attribute similarity and the measure of location proximity using an index. In this study, the location proximity weight between two grids is defined as the inverse of the

distance between their centroids. The attribute similarity of conflicts of two grids is determined based on the difference between each value and the global mean value (Wong and Lee, 2005). The Moran's index could be calculated as:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=i}^{n} w_{ij}(x_i - \underline{x})(x_j - \underline{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})(\sum(x_i - \underline{x})^2)}$$

where, w_{ij} is the proximity weight of grid *i* and grid *j*, x_i is the number of conflicts in grid *i*, $\underline{x}$ is the global mean value, and n is the total number of grids by different side length in the study area. The statistical significance for Moran's I is calculated using Z-score methods. Based on the expected values (E[I]) for a random pattern and the variances (VAR[I]), and the standardized Z-score can be calculated as:

$$Z(I) = \frac{I - E(I)}{\sqrt{VAR(I)}}$$

The spatial autocorrelation was tested with different side lengths to find the length with the maximum Z-score, which is shown in TABLE 4. It is indicated that the 10-feet length could reach the highest Z-score value. Then, grids of 10-feet side length are used to identify the hotspots.

TABLE 4 GLOBAL MORAN'S I VALUE BY LENGTH OF GRIDS

Side Length (feet)	Moran's I value	P_value
2	0.018	<0.001
4	0.061	<0.001
6	0.073	<0.001
8	0.081	<0.001
10	0.082	<0.001
12	0.075	<0.001
14	0.056	<0.001
16	0.043	<0.001
18	0.017	0.027
20	-0.005	0.143

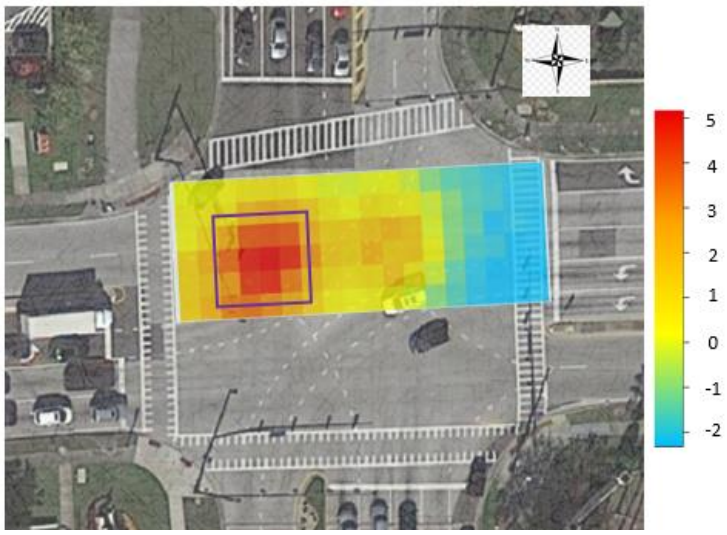
Getis-Ord G_i^* statistic is used to identify the hotspots of conflicts in the within-intersection area. Getis-Ord G_i^* allows the assessment of spatial association in the study area of a particular observation and it could identify statistically significant hotspots (Soltani and Askari, 2017). A high value of Getis-Ord G_i^* statistic represents a cluster of more conflicts (i.e., hotspots), while a low value represents a cluster of fewer conflicts (i.e., cold spots). The exposure is considered for all the grids.The total number of conflicts has been deivded by the volumn of each grid for correct hotpost identification. Getis-Ord G_i^* statistic and its Z-score are mathematically expressed by the following equations:

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} x_j}$$

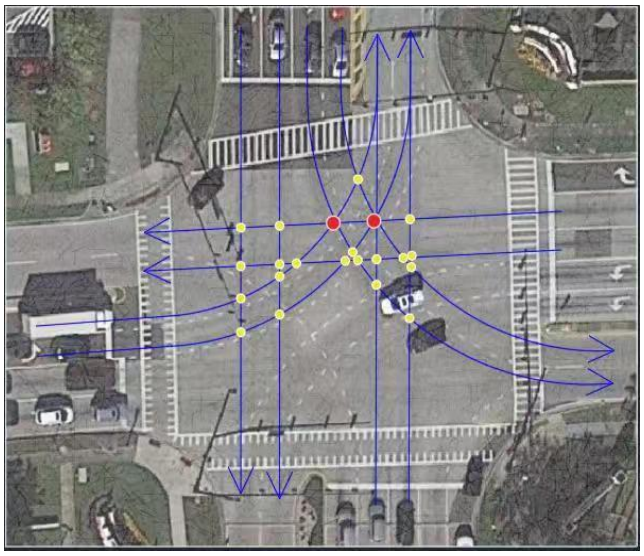
$$Z(G) = \frac{G_i^* - E(G_i^*)}{\sqrt{VAR(G_i^*)}}$$

The results of the analysis of location 1 and location 2 are shown in FIGURE 29(a), and FIGURE 30, respectively. The Z score greater than 2 indicates that the corresponding grid is a hot spot at a statistical significance level of less than 0.05. On the other hand, the Z score less than -2 suggests that the grid is a significantly cold spot at a level of less than 0.05FIGURE 29(b) summarizes the conflict points in the within-intersection area by different vehicle paths. The yellow point indicates a conflict point of two vehicle paths, while the red point represents conflict points of multiple vehicle paths at the same location. Area 1 close to the westbound exit in the study area is identified as the hotspot area. Although Area 2 has more conflict points, it is not identified as the hot areas. The identification of the hotspots at the grid level firstly helps identify specific areas within the intersection prone to have more conflicts, and then the frequency and severity in this area can be further examined to understand the reason of conflicts presence. Based on the outcomes for the identified grid area, the transportation engineers can implement countermeasures, including optimizing signal timing, adding warning signs, add raised medium, and adjusting turning guidelines. For example, the hot zones identified in FIGURE 29(a), the conflicts are mainly caused by the conflict points between southbound through and westbound through, southbound through and eastbound left-turn, and westbound through and eastbound left-turn traffic flows. Hence, the green phase releasing timing between these directions should be closely examined to see whether a sufficient gap for two crossing flows is provided or not. If not, the signal timing should be adjusted to further separate traffic flows from two directions whose conflicts point are in the hotspot grid. Also, the generation of conflicts may be due to another circumstance that the left turn is permissive when is oncoming traffic is on green phase. If this is the case, left turn could be restricted when the opposite through traffic is on green. FIGURE 30presents the conflict hot zones

identified based on different severity levels. FIGURE 30(a) shows that conflicts, regardless of the severity, are prone to be located at the conflict point of northbound through and westbound through, westbound through and eastbound left-turn traffic flows. FIGURE 30 (b) and (c) shows hot zones for severe conflicts are at the intersection entrance of the eastbound through and northbound right turn traffic. Potential cause of the severe conflicts could be some of the vehicles on the rightmost though lane making right turning movement at last moment and decelerate harshly before turning, causing the following vehicle to have little time to react and come too close to the leader. More detailed conflicts analysis should be conducted, and countermeasures could be implemented as the same as discussed above.



(a) HOT ZONE IDENTIFICATION



(b) CONFLICT POINTS BY DIFFERENT VEHICLE PATHS (YELLOW POINT: A CONFLICT POINT BY TWO VEHICLE PATHS; RED POINT: CONFLICT POINTS BY MULTIPLE VEHICLE PATHS)

FIGURE 29 HOTSPOT IDENTIFICATION

(a) PET THRESHOLD OF 3 SECONDS

(b) PET THRESHOLD OF 2 SECONDS

(c) PET THRESHOLD OF 1 SECONDS

FIGURE 30 HOTSPOT IDENTIFICATION WITH DIFFERENT PET THRESHOLD

3.2.6 Conclusions And Future Work

In the within-intersection area, many conflict points are generated by turning vehicles of different movements and drivers are more prone to make mistakes, which could lead to severe crash

outcomes. Closed-Circuit Television (CCTV) cameras, which has been widely installed in the current roadway system, should be a cost-effective sensor to monitor traffic safety in the within-intersection area. This study contributed to propose a framework for safety diagnosis based on videos collected from CCTV cameras named "Near Miss Event Detection System (NMEDS)". The framework combined the Mask-RCNN bounding box detection and Occlusion-Net detection to detect vehicles' key points more accurately and efficiently in a 3D view. Mask-RCNN algorithm is utilized to provide object classification and the initial locations for vehicle tracking. Then, Channel and Spatial Reliability Tracking (CSRT) multi-object tracking algorithm is employed to track the detected objects. Meanwhile, Occlusion-Net is utilized to detect and predict the key points on the vehicles (i.e., right-front headlight, left-front headlight, right-back taillight, left-back taillight) in order to provide precise localization data for vehicles. After transforming the key points to the 2D view (real-world coordinate system), a method was proposed to further modify the occluded points of vehicles. After occluded key points modification, PET values could be calculated to conduct road safety diagnostics by identifying conflicts between vehicles.

An empirical study was conducted at a typical signalized intersection to validate the proposed framework. The proposed framework was compared with two state-of-the-art counterparts in the literature: (1) bonding box detection; (2) key point detection only. Nearly 1000 vehicles were manually labeled from 514 video images as the ground truth from the CCTV camera at the intersection. According to the comparison results, the proposed framework had better performance to detect and track vehicles considering several measures including Intersect of Union (IOU) values, average vehicle size, average vehicle center point distance, and vehicle key point distance. The results indicated that the proposed framework has better performance for vehicle localization, which would be beneficial for road safety analysis, especially when calculating surrogate safety

measures. Then, the conflicts identified by different detection and tracking methods were compared by different PET thresholds. Significant differences could be observed between the proposed framework and other detection methods, which also confirmed the improvement by the proposed framework. By using the conflicts identified by the proposed framework, a grid-based hotspots identification method was proposed by using Moran's and Getis-Ord G_i^* statistical methods. The best side length of girds was determined for the hotspots identification and the significantly dangerous locations were identified in the within-intersection area. It could be concluded that the proposed framework, which combines two detection methods (i.e., bounding box and key points detections) and modifies the occluded key points, could obtain more accurate location and occupying area information of vehicles in the within-intersection area from videos by CCTVs. The improved trajectory data could better identify the conflicts occurring at intersections. The grid-based hotspots identification method could reveal significantly dangerous areas and it has less bias compared with the traditional kernel density method. Given the fact that CCTVs have been installed widely in the existing roadway system, the method proposed in this paper should be a cost-effective method to analyze traffic safety at intersections, especially in a large-scale area.

In the future, the framework could be extended to vulnerable road users (e.g., pedestrians, cyclists, e-scooters) [166].Also, the training datasets from roadside camera images could be prepared to further improve the system performance. Moreover, explore deploying the proposed method into Connected Vehicle Systems could be conducted to provide information to nearby drivers and the traffic management center [64, 167, 168]

3.3 Driving Visual Environment Detection System (D.V.E.D.S)

3.3.1 Introduction

This study aims to explore the effects of drivers' visual environment on speeding crashes based on Google street view (GSV). Deep learning algorithms and computer vision technologies were applied to obtain the clustering and depth information from the GSV images. To quantify drivers' visual environment, several indexes were proposed and calculated. Machine learning algorithms were applied to rank the importance of variables related to visual environment on speeding crashes. The effects of other factors such as traffic volume, speeding proportion, and roadway attributes were also explored. Based on the importance rank, a statistical model was developed to quantify the effects of important factors including the drivers' visual environment. The results validated the visual environment data obtained by the proposed method for the speeding crashes analysis. It was suggested that the visual complexity and proportion of trees in drivers' view have significantly negative effects on speeding crashes. It is expected that this study provided new insights to obtain the detailed information from GSV images for the traffic safety analysis. The findings could also help improve roadway design for the future automated vehicles

3.3.2 Data Collection

The GSV panorama is a 360 degree surrounding image generated from the eight original images captured by multiple cameras by stitching together in sequences. The GSV image could be requested in an HTTP URL form using the GSV image API provided by the Google company. Users can request a static GSV image in a customized direction and angle for the GSV locations. An example of requesting a GSV static image is shown below:

https://maps.googleapis.com/maps/api/streetview?size=640x400&location= 28.78291, -81.2729

& fov=60&heading=0&pitch=0. FIGURE 31 shows the GSV image requested by the above URL. In this example, the output size of GSV image and latitude and longitude of the location was specified. Besides, the heading indicates the compass heading of the camera which ranges from 0 to 360, pitch specifies the up or down angle of the camera relative to the data collection vehicle, and FOV is the horizontal field view of the image. Previous studies suggested the horizontal field view is between 50 *and* 60[169]. Another researcher used FOV of 60 to collect GSVs [90], which was adopted in the current study.

FIGURE 31 AN EXAMPLE OF GSV COLLECTED BY THE URL

3.3.3 Methodology

Studies about information extraction from images has been growing in the field of computer science. Deep learning has been heavily applied and developed to cluster the objects from images. In this study, "Detectron2" from Facebook was used to cluster objects. Detectron2, starting with Maskrcnn-benchmark [5]is Facebook AI research's next generation software system that implements start-of-the-art object detection algorithms[170] . As shown in FIGURE

32 we could detect roads, trees, sky, buildings, etc. from the images. Based on the clustering results, we could know the object type by each pixel in the image. Then, the proportion of pixels by object type in drivers' view could be calculated, such as the proportion of trees, and the proportion of building areas. Besides, the visual complexity level of the driving environment could be calculated as:

$$\text{complexity level} = \frac{-\sum_k (p_k(\ln(p_k)))}{lnN}$$

where k is the category of object, p is the proportion of category k points, N is the number of object categories.

FIGURE 32 ILLUSTRATION OF SEMATIC SEGEMENTATION RESULTS

Meanwhile, the depth information could be obtained from the 2D images. Since the GSV image could be treated as a mono camera, a self-supervised monocular depth estimation proposed by Godard in 2019 was used to obtain the depth information [171].It was suggested that the method

could reach high quality and state-of-the-art results on the KITTI benchmark. FIGURE 33 illustrates the depth information subtracted from the image in black and white colors.

FIGURE 33 ILLUSTRATION OF DEPTH INFORMATION

Through the camera, 3D points are mapped to the image plane (u,v) =f(X,Y,Z). The complete mathematical model that describes this transformation can be written as **p=K[R|t]*P**. In a matrix format, it could be rewritten as:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

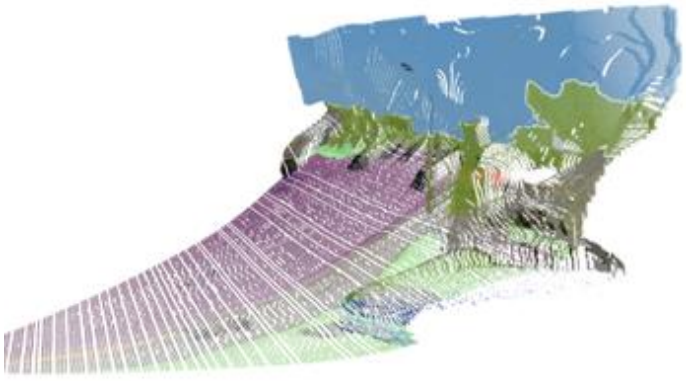
Where

1. p is the projected point on the image plane
2. K is the camera intrinsics matrix
3. [R|t] is the extrinsic parameters describing the relative transformation of the point in the
4. world frame to the camera frame
5. P, [X, Y, Z, 1] represents the 3D point expressed in a predefined world coordinate system

6. in Euclidean space
7. Aspect ratio scaling, s: controls how pixels are scaled in the x and y direction as focal length
8. changes
9. Focal length (fx, fy): measure the position of the image plane wrt to the camera center.

Then, by using an inverse projection process, figures with depth information could be transformed into a 3D point cloud. Then, as shown in FIGURE 34, we could know the exact x, y, z locations in the real world for specific objects. In this study, the proportion of road length with trees was obtained based on the depth information.

3D POINT CLOUD DATA FROM DRIVER'S VIEW



3D POINT CLOUD DATA FROM ROADSIDE VIEW

FIGURE 34 ILLUSTRATION OF 3D POINT CLOUD DATA

3.3.4 Case Study: Explore Effects of Drivers' Visual Environment on Speeding

3.3.4.1 introduction

Speeding is one of the major factors in traffic safety. According to the National Highway Traffic Safety Administration (NHSTA), nearly a third of fatal crashes in the United States have been designated as "speeding-related" in the last decade [20]. On urban arterials, the speed limit violation could significantly increase the severity levels of pedestrian and bicycle crashes. A lot of studies have been conducted to examine the contributing factors for the crash occurrence and speeding behavior. The factors include traffic volume, roadway geometric design, land use, socio-demographic characteristics, and weather, etc. For example, some researcher developed grouped random parameter models to examine the crash occurrence on segments and intersections considering the roadway attributes and the zonal level effects[22]. Another categorized the speeding behavior into three levels by proportions based on the speed camera data. It was found that high speed limits are highly associated with moderate speed limit violations, compared to minor or major speed limit violations[23]. Besides, the study also revealed that a divided median and higher functional class could lead to more major speed limit violations.

Recently, several studies have focused on the effects of the driving environment on drivers' behavior and safety. For example, researcher investigated the effects of road environment visual complexity on travel speed and reaction time by conducting a driving simulator study[172] . It suggested that the visual complexity of the roadside environment is an important contributor to driver workload and performance. Based on a survey study, they revealed the significant effects of the driving environment on speeding and overtaking violations [173]. Researcher developed

statistical models to study the effects of trees on crash frequency in the urban area [174]. The study indicated that tree density could reduce the crashes. However, to the best of the authors' knowledge, the study about the drivers' visual environment on traffic safety is limited. One possible reason is that it is difficult to obtain the data from drivers' view.

This study attempts to propose a novel method to obtain drivers' visual environment from GSV images and explore the effects of the visual environment on speeding crashes. To this end, deep learning models were applied to obtain the cluster and depth information from GSV images. Several indexes were proposed to quantify the visual environment. Then, the effects of the visual environment on speeding crashes were explored by developing both machine learning and statistical models. This paper is organized as follows: the method to process GSV images and obtain the indexes for the visual environment is introduced in Section 2. The method about exploring the effects of the visual environment was also included in this section. Section 3 describes the data used for the analysis including the data related to the visual environment. Section presents the analysis results and discussions. Finally, the last section concludes the findings of this paper.

3.3.4.2 Model for crash analysis

Both machine learning and statistical models have been used for crash count analysis (18-22)[175] [176] [177] [178] [179]. The machine learning could achieve high prediction accuracy and effectively rank the importance of explanatory variables for the crash count. On the other hand, the statistical model could identify the effects of candidate variables in a more interpretable way. Hence, this study took advantage of both approaches by utilizing machine

learning techniques for variable selection and statistical model for an interpretable variable effect evaluation, especially for the factors related to the driving environment

3.3.4.2.1 Statistical model

In the previous studies, the Poisson or Poisson-extended models such as negative binomial (NB) and Poisson lognormal models have been widely used to analyze the count variables. The Poisson model is limited to deal with the over-dispersion issue (i.e., variance exceeds mean) for the count variables. Both NB and Poisson lognormal model could relax the variance assumption and allow for over-dispersed parameter by adding an error term. Recently, it was suggested that the Poisson lognormal model consistently has a better data fit than the negative binomial model for the crash count analysis [56] [180]Hence, the Poisson lognormal model was adopted to analyze the bicycle crash count. The Poisson lognormal model could be specified as:

$$P(y_i) = \frac{EXP(-\lambda_i)\lambda_i^{y_i}}{y_i!}$$

$$\lambda_i = EXP(\alpha \boldsymbol{\beta} \boldsymbol{x_i} + \varepsilon_i)$$

where $P(y_L)$ is the probability of intersection entity i having y_L crashes and λ_L is the Poisson parameter for intersection entity i. $\boldsymbol{\omega_i}$ is the vector of the explanatory variables and Y is a vector of the corresponding parameters. Besides, ε_L is the error term with the distribution of N(0, σ^B). The model parameters were estimated using the Bayesian method by using the free statistical software WinBUGS.

3.3.4.2.2 Machine learning model

Extreme Gradient Boosting (XGBoost) was applied for the analysis in this study. XGBoost is a variant of the gradient boosting framework introduced by Friedman [181], which could provide more accurate results than the regular Gradient Boosting models. It can find the best tree model through a more accurate approximation method. To be specific, it interactively improves learning accuracy by combining an ensemble of weak learners into a single strong model. Besides, as a decision tree algorithm, XGBoost is not affected by multicollinearity [182], which means the model could include variables that reflect the same phenomenon on the study road. The 'XGBoost in the Python interface was used for the analysis. The 'Poisson' object was specified since the crash count was analyzed. The mean absolute error was used to tune the XGBoost to determine the best combination of parameters.

3.3.4.3 Data

The data used in this study were collected from urban arterials in Central Florida. As shown in classifications the urban arterials of nearly 75 miles were included, and around 15,000 GSV images were downloaded and processed to get the indexes about drivers' visual environment. From the images, the proportion of trees, the proportion of buildings, and the complexity level from the drivers' view were collected. In addition, the proportion of road length with trees was calculated based on the cluster and depth information. The corresponding variables were summarized in TABLE 5. The speeding crashes were collected from the Florida Department of Transportation (FDOT). In addition to the driving environment data collected from GSV, other exogenous variables were also collected, which included traffic data, roadway information, lane use attributes, and socio- demographic for each segment. For traffic data, the Average Annual

Daily Traffic (AADT), the proportion of truck traffic, and daily transit frequency were collected from FDOT. Besides, the probe vehicle data INRIX were collected from RITIS by 5 minutes from 2017 to 2019. Based on each segment, the proportion of INRIX speed over the posted speed limit was calculated, which could indicate the general speeding trend. Seven collected roadway attributes are related to speed management strategies in Florida Design Manual (FDM) [183]. They are asphalt pavement, the indicator of narrow lane, average block length, the existence of median island on crossing, number of parking per mile, presence of road diet, and length of the two-way-left-turn lane. Other roadway variables such as lane number, speed limit, median type and width, shoulder type and width were also collected from FDOT. Finally, the land use and socio-demographic variables were also collected in this study.

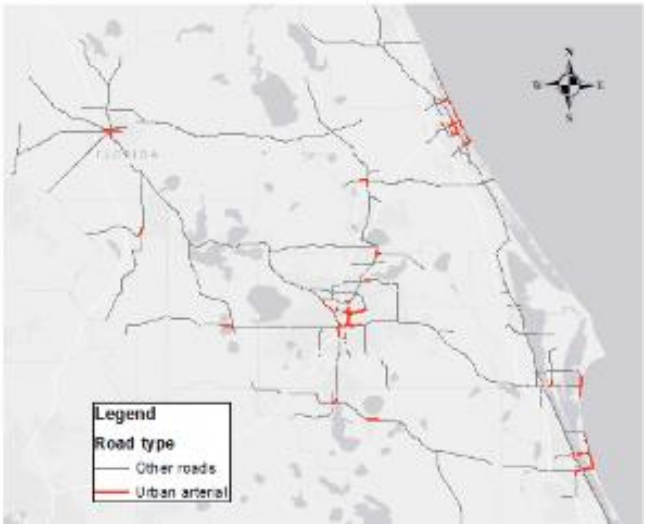


FIGURE 35 STUDY ROADS BY CONTEXT CLASSIFICATIONS

TABLE 5 SUMMARY OF VARIABLES

Variable	Mean	S.D.	Min.	Max.
Speeding Crashes	1.24	1.51	0	8
Length(Mile)	0.52	0.37	0.10	1.80
Traffic Variables				
Average Daily Transit Frequency (Aadt)	29873	95666	9833	66833
Proportion Of Truck Traffic	6.76	3.31	1.77	16.19
Average Daily Transit Frequency	19.97	18.05	0	72
Proportion Of INRIX Speed Data Over The Speed Limit	0.18	0.14	0	0.78
Drivers' Visual Environment Data From GSV Images				
Proportion Of Tree View	0.13	0.06	0.03	0.31
Proportion Of Building View	0.03	0.03	0	0.22
Complexity Level Of Driving View	0.73	0.03	0.64	0.79
Proportion Of Read Length With Trees	0.28	0.06	0.15	
Roadway Variables				
Variables Related To FDM Speed Management Strategies				
Indicator Of Asphalt Pavement (1: Yes; 0: No)	0.84	0.36	0	1
Indicator Of Narrow Lane (Lame Width-12 Feet) (1: Yes; 0: No)	0.36	0.48	0	1
Average Block Length (Mile)	1.60	3.26	0.06	10.33
Existence Of Median Island On Pedestrian Crossing (1: Yes; 0: No)	0.07	0.25	0	1
Log (Number Of Parking Spot Per Mile)	0.27	1.12	0	6
Presence Of Road Diet (1: Yes; 0: No)	0.44	0.5	0	1
Log (Length Of Two-Way-Left-Turn Lane)	0.13	0.23	0	1
Other Roadway Variables				
Number Of Lanes	2.1	0.52	1	4
Speed Limit (Mph)	38.59	5.2	25	55
Pavement Condition	4.17	0.8	0	5
Raised median (1: yes; 0: no)	0.44	0.5	0	1
Median Width (feet)	16.53	9.32	0	55.07
Curb, gutter inside shoulder type (1: yes; 0: no)	0.34	0.47	0	1
Width of inside shoulder (feet)	0.97	1.47	0	9
Curb, gutter outside shoulder type (1: yes; 0: no)	0.48	0.5	0	1
Width of outside shoulder (feet)	3.45	1.88	0	10
Proportion of sidewalk length	0.91	0.25	0	1
Sidewalk width (feet)	5.09	1.35	0	10
Proportion of bike lane length	0.12	0.3	0	1
Proportion of bike slot length	0.01	0.05	0	0.64
Number of signalized intersections per mile	3.11	3.15	0	16.98
Number of access per mile	9.52	5.86	0	28.37
Land use and socio-demographic variables				

Variable	Mean	S.D.	Min.	Max.
Proportion of residential land use	0.04	0.09	0	0.62
Land use mix	0.04	0.09	0	0.62
Proportion of population below poverty	0.06	0.11	0	0.73
Proportion of zero-vehicle household	0.02	0.04	0	0.22
Proportion of commuters by walking or biking	0.02	0.04	0	0.20

3.3.4.4 Results and discussion

3.3.4.4.1 XGBoost model tuning and variable selection

A tuning process was applied to determine the best set for the number of clusters and the number of boosted trees. The mean absolute error was used as the performance measure. The tuning process included values of the number of clusters from 2 to 10 and the number of boosted trees from 20 to 200. The evaluation results are presented in FIGURE 36 MODEL PERFORMANCE BY CLUSTERS AND BOOSTING TREES. The clustering with 4 groups and 100 boosting trees could achieve stable performance, which was used to rank variables' importance.

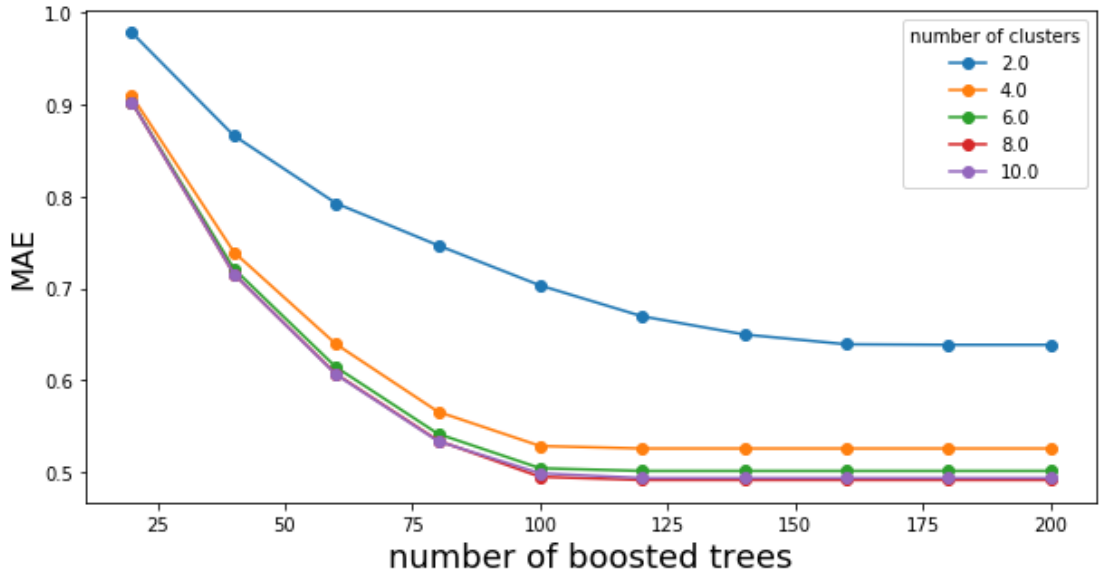


FIGURE 36 MODEL PERFORMANCE BY CLUSTERS AND BOOSTING TREES

The results of variable importance are shown in FIGURE 37 VARIABLE IMPORTANCE. Unsurprisingly, the road length clearly prevails as the most important feature. The traffic volume, speeding proportion, and other road attributes were also identified as important variables. The results revealed the variables about the drivers' visual environment collected from GSV images as important features for speeding crashes. They are the proportion of buildings, the proportion of trees, the complexity level of view, and proportion of road length with trees. The development of the statistical model will be based on the variable importance following a backward stepwise variable selection procedure. It should be noted that the proportion of trees and proportion of road length with trees indicate the attributes of trees. Although both were

included in the XGBoost model, only one of these two variables will be included in the statistical model.

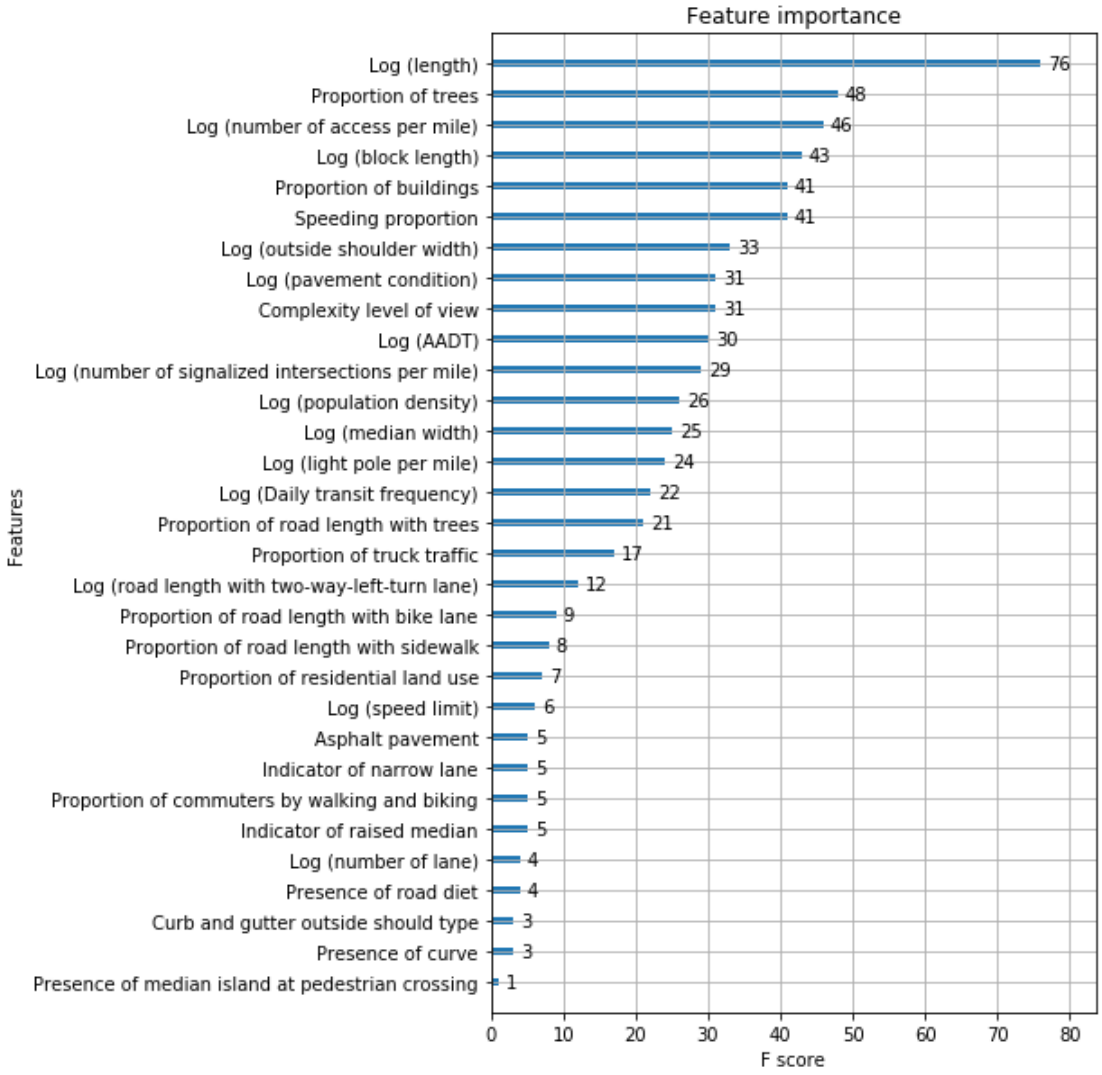


FIGURE 37 VARIABLE IMPORTANCE

3.3.4.4.2 Poisson lognormal modeling results

The Poisson lognormal modeling results are summarized in TABLE 6 POISSON LOGNORMAL MODELING RESULTS. There are nine variables significant for the speeding crashes. The AADT as a traffic exposure has a significantly positive relation with speeding crashes. Meanwhile, the speeding proportion indicates the speeding level of the segment. It is expected that more speeding crashes occur on a segment with a high speeding Two variables

related to drivers' visual environment from the GSV images are found to be significant. It was found that the complexity level from drivers' view is negatively associated with the speeding crashes. The result is consistent with the finding in the literature that drivers select lower vehicle speeds under higher levels of complexity [184]. It suggests that roadside trees could reduce the number of speeding crashes. The previous driving simulator study found that drivers tend to decrease their speeds significantly and move toward the centerline of the road when trees present[185]. Besides, it was found that the crashes could decrease with the increase in tree density ([174]). With the effects of trees on speeding and crash occurrence, it is reasonable to find a negative effect of trees on speeding crashes. Five other road attributes were found significant: pavement condition, road length with two-way-left-turn-lane, outside shoulder width, number of accesses per mile, and segment length. Fewer crashes could be found on the road with good pavement conditions. The two-way-left-turn lane is a speed management strategy. Drivers might reduce the speed if a two-way-left-turn lane presents. The wide outside shoulder could reduce the crashes, which is in line with the previous study (30 [176]). The road access could increase the interactions between road users, resulting in more crashes. The segment length could increase the exposure and result in more speeding crashes.

TABLE 6 POISSON LOGNORMAL MODELING RESULTS

Variable	Estimate	Standard Error	BCI	
			2.50%	97.5%
Intercept	-1.90	1.65	-1.86	1.05
Log(AADT)	0.99	0.20	0.95	1.34
Speeding proportion	1.13	0.49	1.14	2.04
Complexity Level of View	-7.70	1.07	-7.56	-6.00
Proportion of Road Length With Trees	-0.94	0..9	--0.95	-0.20
Log(Pavement Condition)	-1.19	0.50	-1.18	-0.31
Log(Road Length With two-way-left-turn lane)	-0.24	0.10	-0.25	-0.05
Log(outside shoulder width)	-0.61	0.17	-0.60	-0.28
Log(number of access per mile)	0.23	0.08	0.23	0.39
Log(length)	0.73	0.12	0.73	0.97
σ	0.50	0.11	0.50	0.71
DIC		424.49		

3.3.4.5 Conclusions and Future Work

This study applied deep learning models to extract information related to drivers' visual environment from the Google Street View (GSV) images. Around 15,000 GSV images of urban arterials were downloaded. The deep learning algorithm developed by Facebook was used to cluster objects in the images. Based on the clustering results, indexes including the proportion of trees, the proportion of buildings, and the complexity level of the visual environment were calculated by counting the number of pixels. Besides, another deep learning method was applied to get distance information from the images. By combining the clustering information, a 3D point cloud data was generated for each GSV image. The proportion of road length with trees was calculated. The information reflects the environment information from the drivers' view and was used to explore its effects on speeding crashes. XGBoosting analysis was conducted to identify the variable importance and a statistical model was developed to quantify the effects of

information extracted from GSV images on the speed crashes. Other factors including traffic volume, speeding proportions, road attributes, land use, and socio-demographic characteristics were also examined. The XGBoosting regression result revealed that the trees and the complexity level of driver visual view are very important contributing factors for speeding crashes on urban arterials. The statistical model suggested that the proportion of roads with trees and the complexity level could reduce the speeding crashes. The results validated that more insight could be obtained by using deep learning algorithms to extract detailed information from GSV images. Besides, other significant factors for speeding crashes were also revealed, such as traffic volume, pavement conditions, road length with two-way-left-turn lane, outside shoulder width, number of accesses per mile, and segment length. This paper contributed to propose a method to obtain valuable information from GSV images by using deep learning algorithms. The information from drivers' visual view was obtained and used to analyze speeding crashes. The modeling results confirmed the importance of information obtained from GSV images. In this paper, aggregated information from GSV was used for the speeding crash analysis. It would be interesting to use pixel-level data for traffic safety and driving environment analysis. Besides, the current study could be extended to explore safety by using naturalistic driving data.

CHAPTER 4 : CITYSIM: A DRONE-BASED VEHICLE TRAJECTORY DATASET FOR SAFETY ORIENTED RESEARCH AND DIGITAL TWINS

4.1 Introduction

In this dissertation, we introduce the CitySim dataset: a video-based trajectory dataset generated from drone recordings with a focus on traffic safety. This dataset is intended to facilitate safety research by providing traffic trajectories that are rich with conflicts and near misses. Therefore, it was designed with some novel features which are safety research friendly: 1) Accurate vehicle trajectories. CitySim has a five-step procedure to ensure high trajectory accuracy: video stabilization, object filtering, multi-video stitching, enhanced error filtering, and detection and tracking using an integrated algorithm of Mask R-CNN and CRST. The utilized algorithms and data processing methods are further delineated to give the dataset users a clear idea about the quality of the extracted trajectories.

1) Wider vehicle trajectory range. Due to drone flight altitude restrictions and video resolution limitations, the coverage areas of drone cameras are limited. In order to capture wider observation ranges, CitySim utilizes multiple drones hovering over target areas and subsequently stitches the generated videos into one cohesive video. Therefore, CitySim users are able to observe and analyze vehicle interaction behavior from a much longer trajectory range, which accounts for various scenarios and provides comprehensive results.
2) More critical safety events. CitySim selected locations that contain more aggressive and intense vehicle interactions, such as weaving segments. As demonstrated in later parts of this paper, CitySim has more critical safety events compared with other datasets in terms of both event severity level and sample size.

3) Accurate vehicle geometric representation. CitySim provides highly accurate bounding box information for each detected vehicle, which enables a more accurate estimation of the safety levels when compared with only using the vehicle central point information. This paper demonstrates the necessity of using bounding box information to calculate the safety metrics at the intersections in later parts of the paper.
4) Digital twin features. CitySim provides high resolution 3D maps and physical models for each of the collected locations. Additionally, it provides the signal timing information for relevant locations. These assets allow researchers to test and verify their safety research in a high-fidelity virtual environment.

4.2 Dataset Generation

4.2.1 ARCIS

CitySim was generated through five steps (FIGURE 3).The five steps are video stabilization, object filtering, video stitching, enhanced error filtering and detection and tracking. These steps ensure the output trajectory to be accurate as much as possible.

4.3 CitySim Dataset Description

Currently, CitySim contains vehicle trajectories from 1140-minute drone (DJ mavic 2,2 s, 3 or phantom 4) video recordings at 12 locations. The locations include one freeway basic segment, one weaving segment, two merge/diverge segments, five signalized intersections, one stop-controlled intersection and two control-free intersections. The weaving segment and merge/diverge segments are particularly selected to provide sufficient number of critical safety events including cut-in, merge and diverge events. Detailed data collection information can be found in**TABLE 7**, and a bird view of each location is provided in **FIGURE 39**.

CitySim provides the following attributes: seven positions of a vehicle body (measured in pixel and GPS, see Figure 38), speed, heading (measured in global north and image coordinate X-axis), and the lane number of the vehicle,

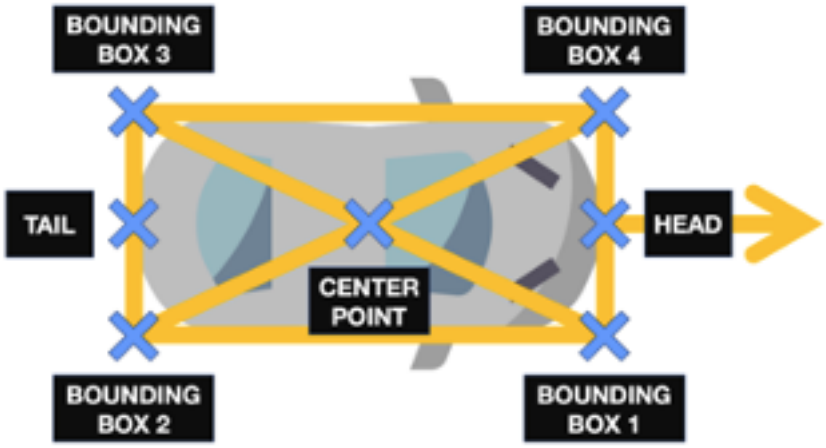


Figure 38 Vehicle body

The dataset can be downloaded from https://github.com/ozheng1993/UCF-SST-CitySim-Dataset, while MATLAB and Python source code to handle the data, create visualizations and extract maneuvers is provided at https://github.com/ozheng1993/UCF-SST-CitySim-Dataset/wiki.

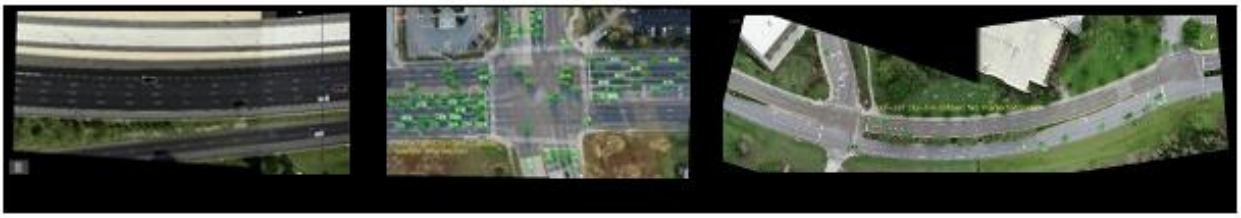
TABLE 7 DATA COLLECTION INFORMATIOM

A: control-free intersection is neither signalized nor stop-controlled.

B: drone height is based on the height of the surface of the take-off

ID	Location	Location Type	FPS	Recording Resolution	Recording Length (min)	Drone Height (m) B
1	Expressway A	Weaving Segment	30	5120 x 2880	120	120
2	Freeway B	Basic Segment	30	5120 x 2880	60	320
3	Freeway C	Merge/Diverge Segment	30	5120 x 2880	60	320
4	I-4 Express Lane Exit	Merge/Diverge Segment	30	3840x2160	60	120
5	University @Alafaya	Signalized Intersection	30	3840x2160	120	120
6	UCF Garage C	Signalized Intersection	30	3840x2160	120	120
7	UCF Garage C V2	Signalized Intersection	30	4096*2160	120	120
8	426@Oxford	Signalized Intersection	30	3840x2160	120	120
9	Hydra Ln@ Gemini (UCF GYM)	Signalized Intersection	30	3840x2160	120	120
10	Aquarius Agora @ Gemini (art Building)	Stop-Control intersection	30	3840x2160	120	120
11	McCulloch @ Seminole	Control-Free Intersection [A]	30	3840x2160	60	120
12	University @ McCulloch	Control-Free Intersection	30	3840x2160	60	120

EXPRESSWAY A FREEWAY B FREEWAY C



I-4 EXPRESS LANE EXIT UNIVERSITY @ ALAFAYA UCF GARAGE C

UCF GARAGE C V2 426 @ OXFORD HYDRA @ GEMINI

AQUARIUS AGORA @ GEMINI MCCULLOCH @ SEMINOLE UNIVERSITY @ MCCULLOCH

FIGURE 39 BIRD VIEW OF THE CITYSIM LOCATIONS

4.4 Toward Safety Research

CitySim is highlighted for its specialties that help with the autonomous vehicle safety research. Here, we compared CitySim with NGSIM and highD, in terms of the following potential safety event:

1) **Freeway cut-in event:** the cutting-in vehicle may cause a significant conflict with the following vehicle on the target lane, and it is one of the contributing factors that cause a rear end crash. In addition, it was observed from the CitySim data that when a vehicle begins a lane change from a source to a target lane, its following vehicle on the source lane often accelerates and generates conflict risk. Therefore, two types of minimum time-to-collision (minTTC) during the cut-in event were calculated (FIGURE 40) the before-minTTC is the minimum TTC between the cut-in vehicle and the following vehicle on the source lane, while the after-minTTC is the minimum TTC between the cut-in vehicle and the following vehicle on the target lane. The cut-in behavior must have following vehicles.

2) **Freeway merge/diverge event:** similar to cut-in behavior, the merge/diverge event often causes a conflict or even a crash to the following vehicle. The corresponding minTTC was calculated as the conflict indicator. The merge/diverge behavior must have following vehicles.

3) **Intersection conflict event:** this type of behavior was measured in minimum post encroachment time (minPET) (*35*) to describe the conflict severity between two vehicles that have an intersected trajectory. The lower the minPET, the server the conflict.

4)

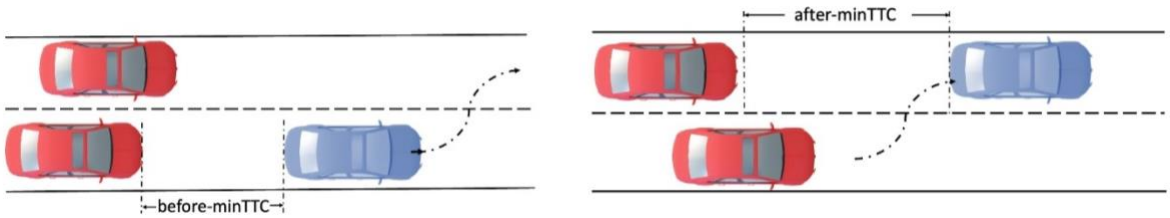


FIGURE 40 TWO TYPES OF MINIMUM TIME-TO-COLLISION DURING THE CUT-IN EVENT.

4.4.1 Freeway/Expressway Cut-in/Merge/Diverge

Regarding the number of potential safety events, CitySim has significantly more cut-in, merge and diverge events compared with the highD and NGSIM. We selected two locations from CitySim, namely Expressway A and Freeway C, and compared them with highD and NGSIM (US-101 & I-80). The NGSIM Lankershim Boulevard Dataset was not included since it contains the urban scene which is not similar to other locations. FIGURE 41shows that two locations from CitySim have 8873 cut-in events, while highD has 5468 cut-in events and NGSIM only has 1363 cut-in events. The two CitySim locations have 1954 merge events and 3910 diverge events. In contrast, the merge and diverge sample sizes of highD and NGSIM are very small. Given that the cut-in scenario, merge scenario and diverge scenario often cause severe conflicts or even crashes, the scenarios are heavily tested for autonomous vehicle safety. As shown in the figure, CitySim can sufficiently provide these types of scenarios for thorough safety testing scenarios for autonomous vehicle.

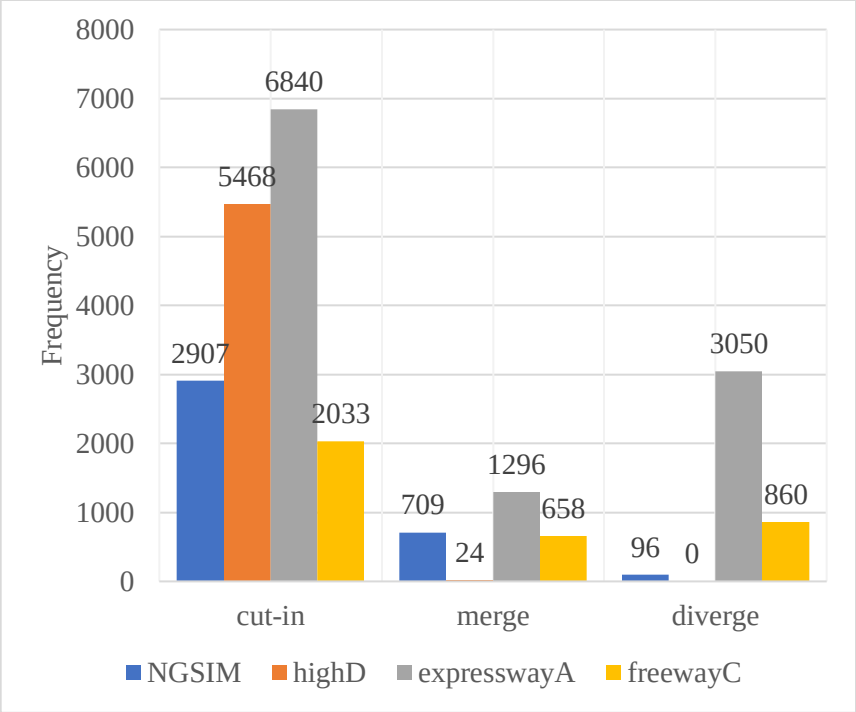
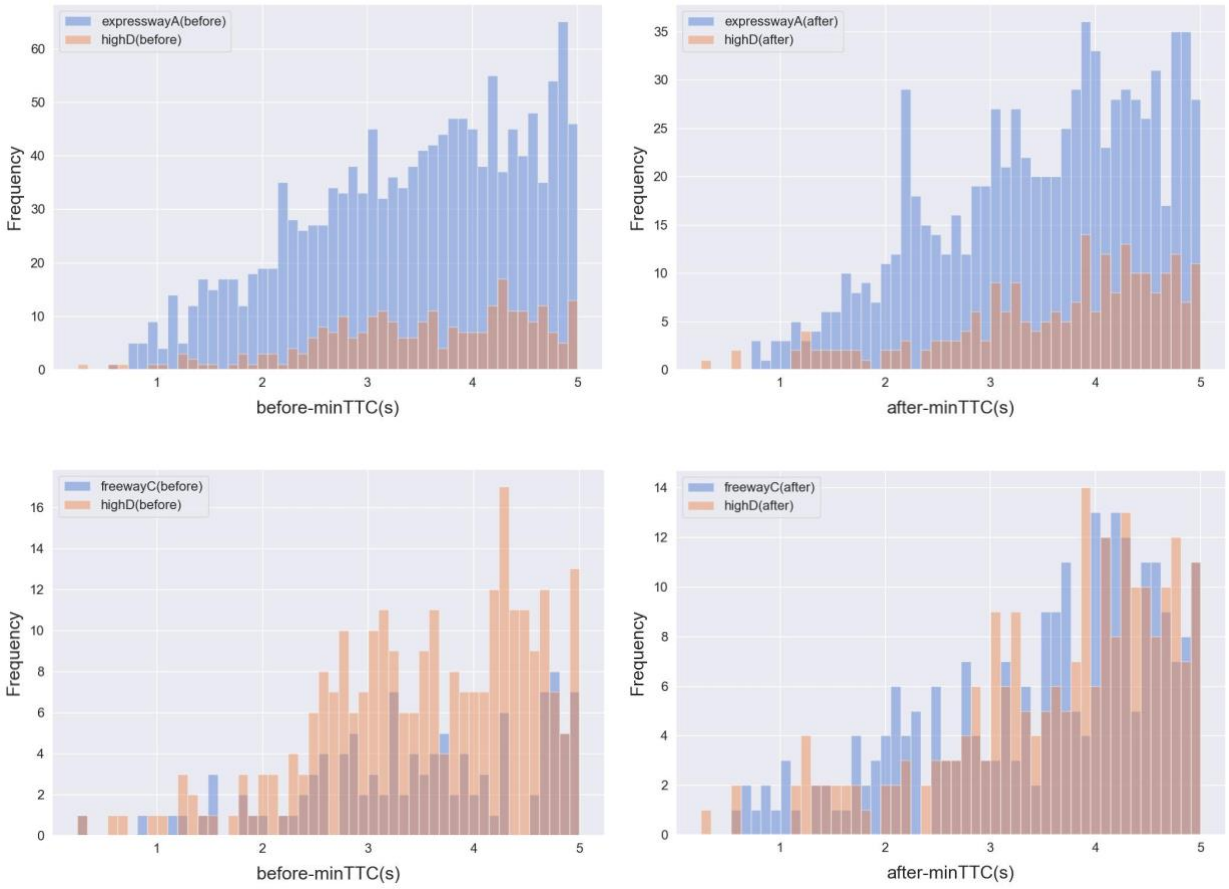
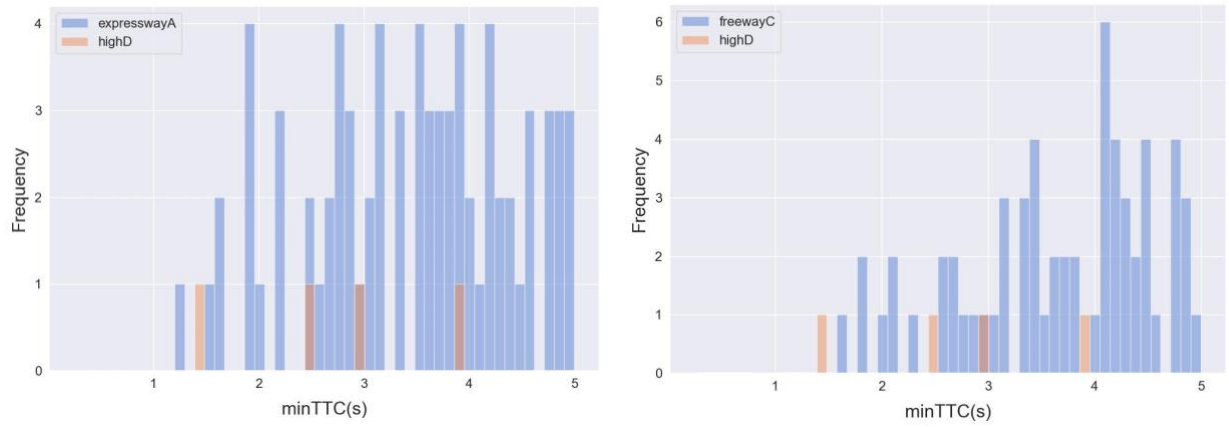


FIGURE 41 THE POTENTIAL SAFETY EVENTS COUNT FOR EACH DATASET

CitySim not only has larger sample size of potential safety events, but also contains much more critical safety events. Critical safety events are defined as a potential safety event with a minTTC/minPET less than a threshold. Therefore, critical safety events are more likely to cause a crash. In this paper, we adopted a threshold of 5.0s and analyzed the distribution of critical safety events. FIGURE 42 shows that, compared with highD, Expressway A has more cut-in events with minTTC less than 5.0s in terms of both types of minTTC, and Freeway C also has more critical cut-in events involving the following vehicle on the target lane. Regarding the merge events, both Expressway A and Freeway C have more critical merge events than highD. The NGSIM was not considered here due to its trajectory error which causes abnormal minTTC. FIGURE 43 shows some example critical safety events observed in CitySim.

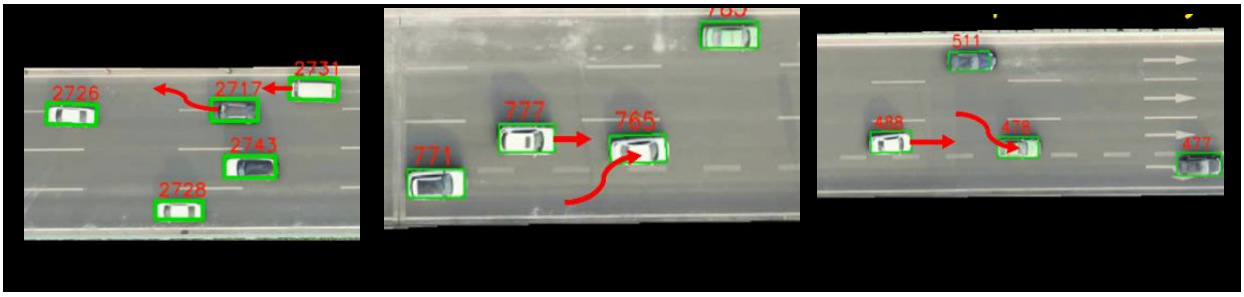


CUT-IN EVENTS (MINTTC<5.0S)

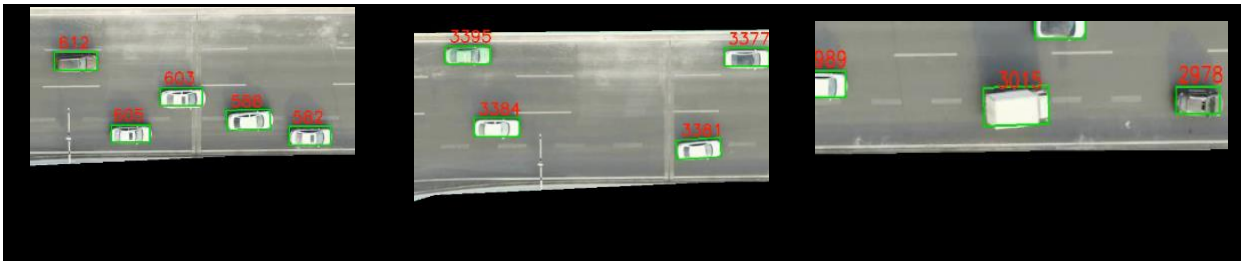


MERGE EVENTS (MINTTC<5.0S)

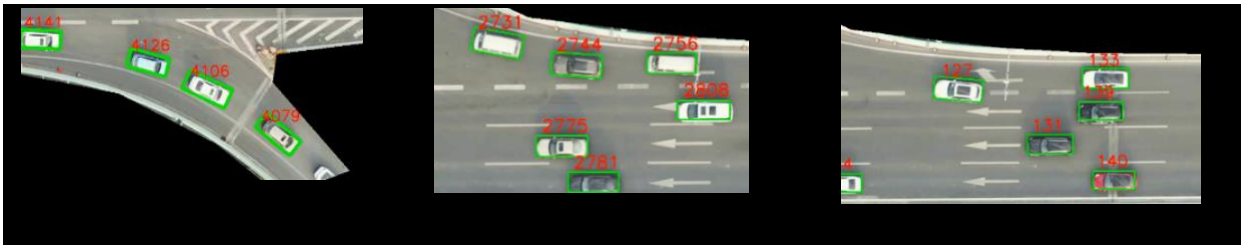
FIGURE 42 CUT-IN AND MERGE EVENT COMPARISON BETWEEN CITYSIM AND HIGHD



(A) CUT-IN EVENTS



(B) MERGE EVENTS



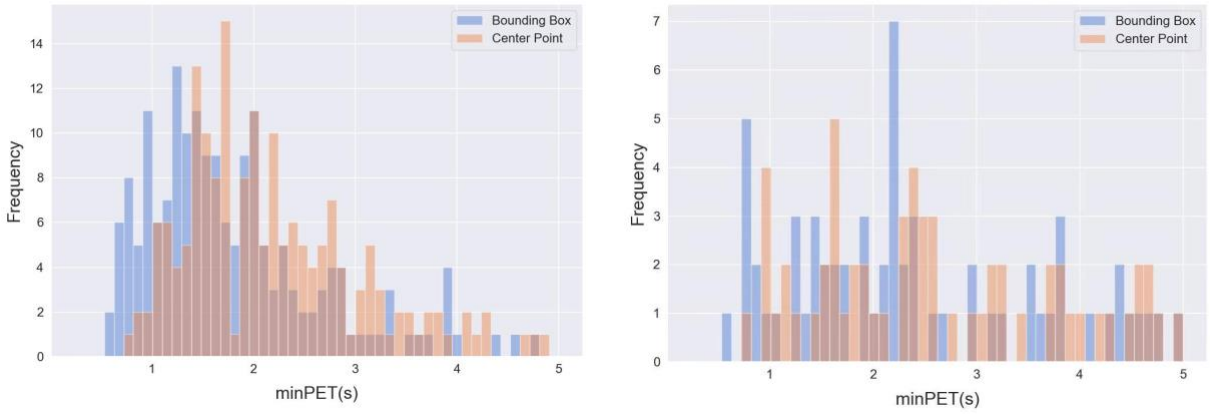
(C) DIVERGE EVENTS

FIGURE 43 CRITICAL SAFETY EVENTS EXTRACTED FROM CITYSIM

4.4.2 Intersection Conflicts

As aforementioned, when safety metrics such as PET/TTC are calculated based on the central points, there might be a bias since the true vehicle body profile is not considered, particularly for the turning trajectories. As shown in FIGURE 44, larger number of critical conflict events are identified by the bounding box measurement; meanwhile, the conflict events are severer than the those identified by the center point. Further, the distribution of the critical conflict events at the intersections are significantly changed (shown in FIGURE 45). This demonstrates the significant

bias caused by the center point for the turning trajectory. Therefore, the bounding box information provided by CitySim is believed to benefit safety evaluation much more than many other datasets.



UNIVERSITY @ ALAFAYA **MCCULLOCH @ SEMINOLE**

FIGURE 44 THE CRITICAL CONFLICT EVENTS AT TWO CITYSIM INTERSECTIONS MEASURED BY DIFFERENT MINPET THRESHOLDS

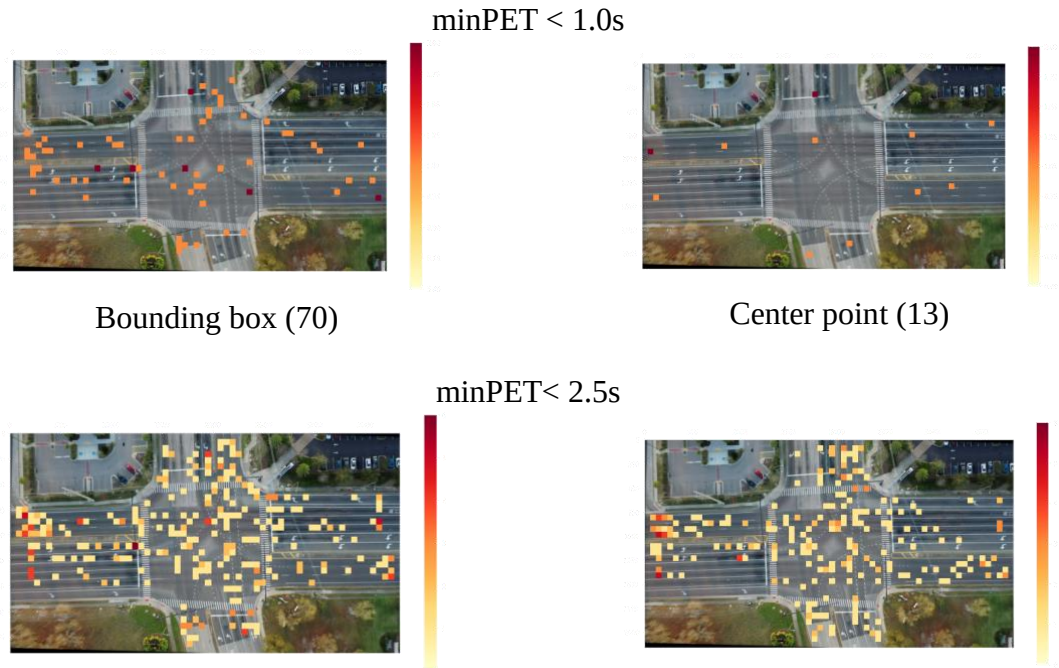


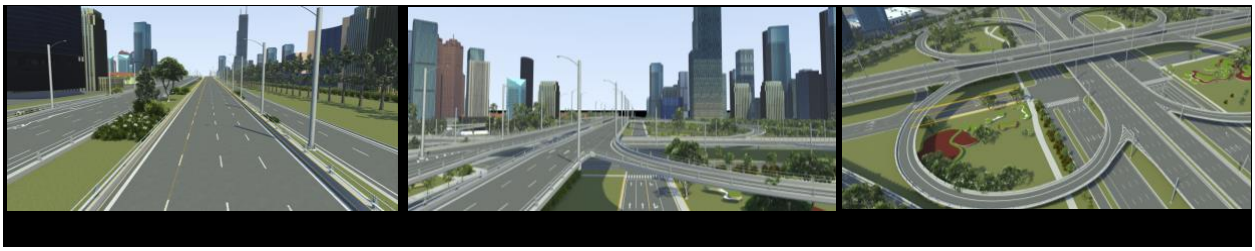
FIGURE 45 THE HEATMAP OF CRITICAL CONFLICT EVENT DISTRIBUTION AT TWO CITYSIM INTERSECTIONS MEASURED BY DIFFERENT MINPET THRESHOLDS

(The number in the parathesis means the identified number of events)

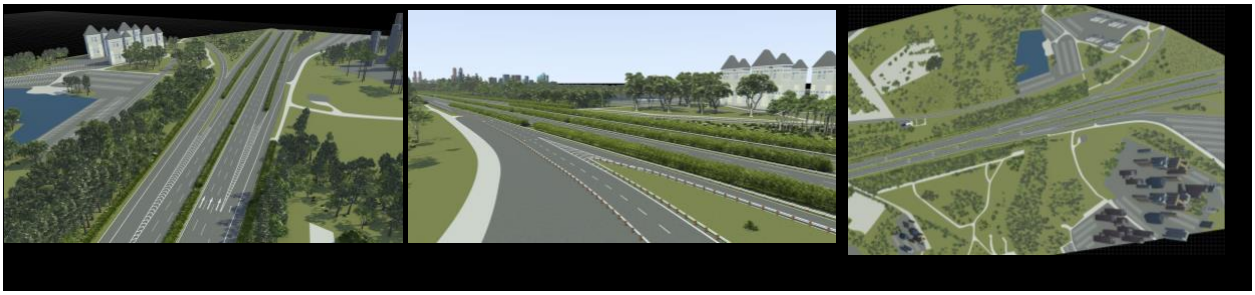
4.5 Towards Digital Twins

CitySim provides 3D base map for each location with an objective to facilitate the digital-twin-based autonomous vehicle safety research. FIGURE 46 shows part of 3D base maps for the CitySim locations. In addition to the digital twin models, CitySim also provides signal timing data related to signalized intersections. The signal data CSV file provides event level data. Each record in the signal data represents a change in one of the signals. The data is represented by eight different digits that corresponds to different signal phases at the intersection.

FIGURE 47 shows a case study of using CitySim to achieve digital-twin-based simulation. CitySim provides accurate vehicle trajectories from its drone data, and the data is used to calibrate microscopic traffic patterns; then, both the traffic patterns and 3D base maps enter a co-simulation platform which integrates the SUMO and Carla. By simulating the virtual testing environment, vehicle dynamics and vehicle sensors, a human-in-loop simulator experiment can be conducted on the co-simulation platform, which fulfills a digital-twin-based experiment that connects the virtual and real world together.



(A) EXPRESSWAY A



(B) FREEWAY C

(C) UNIVERSITY @ ALAFAYA

(D) MCCULLOCH @ SEMINOLE

FIGURE 46 3D MODELS FOR EACH LOCATION OF CITYSIM

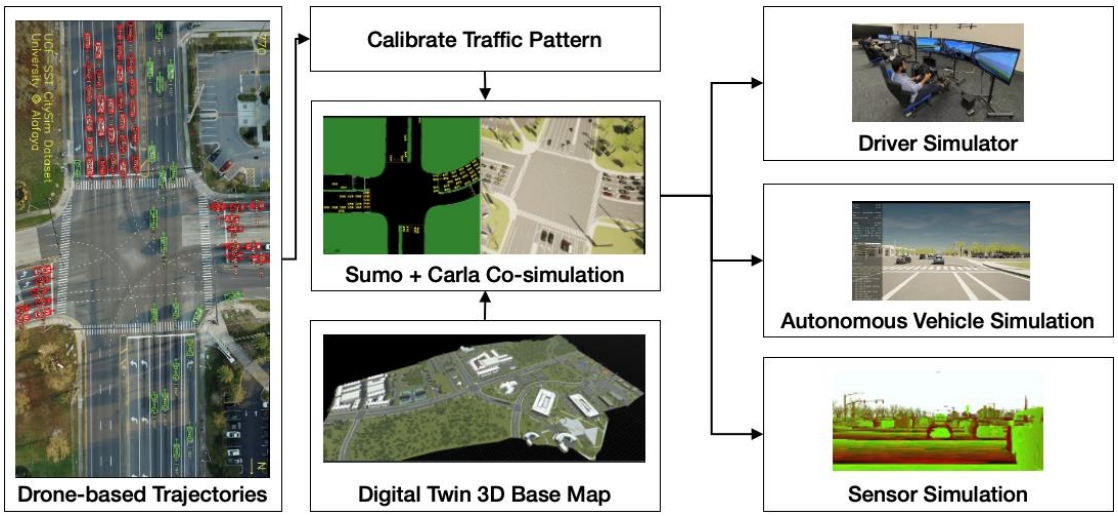


FIGURE 47 A DIGITAL-TWIN-BASED SIMULATION OF INTEGRATING BOTH THE VIRTUAL AND REALITY

4.6 Conclusions And Future Work

This paper introduces a new video-trajectory based dataset, CitySim, which aims to facilitate the autonomous vehicle and conflict based safety research. CitySim has vehicle interaction trajectories extracted from 1140-minutes of video recordings, which have covered variety of locations including freeway basic and weaving segments, expressway segments, signalized intersections, stop-controlled intersections and special intersections without sign/signal control. The biggest advantages of CitySim over other datasets is it has more critical safety events and more severe safety events, which would be a great support for the safety-oriented research. Meanwhile, by providing the bounding box information, CitySim is able to measure the safety level more accurately. In addition, CitySim has the digital twin features including the 3D base maps, and signal timings, which enables a more comprehensive testing environment for the autonomous vehicle safety. In the future, we are planning to increase the size of the dataset and add more digital twin features.

CHAPTER 5: TOWARDS NEXT GENERATION OF PEDESTRIAN AND CONNECTED VEHICLE IN-THE-LOOP RESEARCH: A DIGITAL TWIN CO-SIMULATION FRAMEWORK

5.1 Introduction

The emergence of numerous Intelligent Transportation System (ITS) applications have made an extensive contribution to the transportation system in terms of safety, mobility, and energy consumption. Connected vehicle (CV), as a core component of ITS, has received vast attention since it was proposed in the late 20th century [186, 187][. With the advancement in vehicle connectivity, computing power, and automotive control, the CVs are able to "talk" to other traffic actors (vehicle, infrastructure, pedestrian, etc.) and collaboratively make decisions and perform driving tasks. Apart from the vehicles, pedestrian safety also raised much awareness, as it has the highest fatality rates among road accidents. Many ITS applications have been developed and investigated to enhance pedestrian safety in a CV environment. Due to safety concerns, most of the ITS technologies are developed and tested in a simulation environment, which is based on the assumption that the simulation environment is well-calibrated to represent real-world scenarios. However, this is not always true. Furthermore, there is no interaction between simulated entities and real-world traffic participants, causing a reduction in the validity of testing. Hence, an experimental environment that supports both CV and pedestrian in-the-loop is needed for the research and development of vehicle-to-pedestrian (V2P) ITS applications. Digital Twin, as an emerging technology in the transportation field, provides unprecedented opportunities to support the development of ITS applications and address the above issues. By definition, a Digital Twin is a digital replica of a physical entity in the real world[188]. In transportation, it could be the technology that projects all the traffic participants into a digital road network in a real-time manner. As real vehicle or pedestrian behaviors are reflected in a digital environment, the V2P applications can be implemented virtually. By such means, the

effects of ITS applications in a dangerous scenario can be tested as there is no safety concern in the digital environment. Also, as the digital entities are projections of real traffic participants representing real-world traffic, the experiment results could be as convincing as field experiments. In this work, we propose a Digital Twin framework to support ITS research on CV and pedestrians. A sample architecture under the framework is demonstrated, which incorporates a Sumo-Carla co-simulation platform for CV and a Cave automatic virtual environment (CAVE) for pedestrians. The contribution of this study includes:

- Introduced a CV and pedestrian in-the-loop Digital Twin framework, which consists of physical world, digital world, and the connection in between. The framework describes the features that need to be digitally twined for both CV and pedestrians, including their external and internal states.
- A sample architecture is presented to show a realization method of the Digital Twin framework. The architecture innovatively connects a CV simulation environment (Carla-Sumo co-simulation) and pedestrian simulator (CAVE). The functions of the sample architecture currently include bi-direction data transmission, pedestrian, and driver state capturing, and V2P application display.
- Demonstrated a case study of V2P collision warning under occlusion condition that validates the proposed framework. The case study investigates the effect of V2P warning over baseline (no warning) and autonomous vehicles (equipped with AEB). The advantage of using the CV and pedestrian in-the-loop system is shown through the experiments. The organization of the remaining paragraphs is as follows: Section II reviews the literatures of CV and pedestrian simulation as well as digital twin technology; Section III introduces the proposed Digital Twin framework; Section IV

presents a sample architecture using Carla-Sumo co-simulation and CAVE; Section V is the case study and Section VI is a summary and conclusion

5.2 Digital Twin Framework

The proposed Digital Twin framework consists of the physical world, the digital world, and the data transmission between all the modules. The framework architecture is shown in FIGURE 48. The elements inside the brown rectangle on the left are the digital replica of CV and pedestrians. The physical world is represented in the green rectangle on the right, in which the real drivers and pedestrians are involved. The blue 3 rectangles are the attributes of CV and pedestrian that need to be digitally twined. The grey rectangle at the center containsthe CV and pedestrian-related ITS technologies that will be applied.

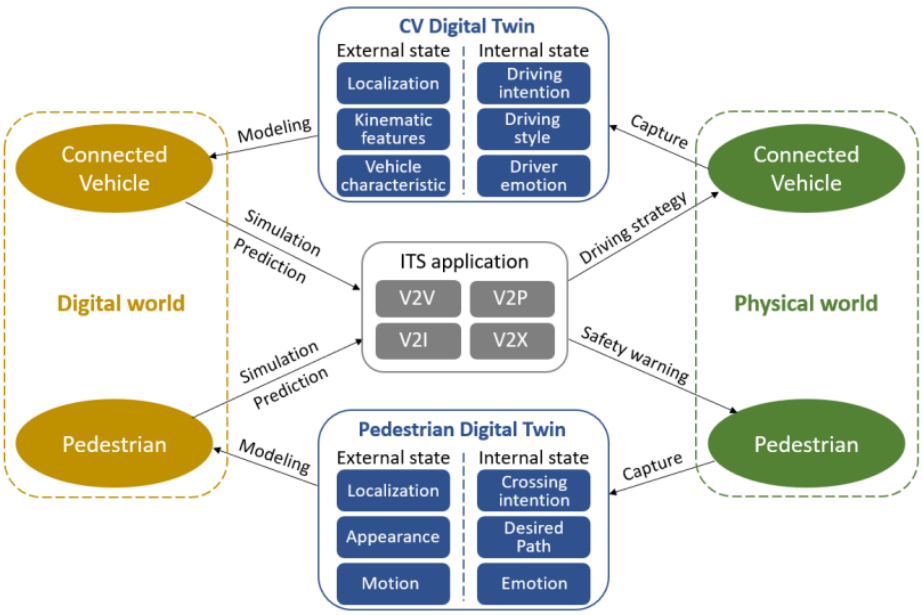


FIGURE 48 DIGITAL TWIN FRAMEWORK FOR CONNECTED VEHICLE, PEDESTRIAN AND TRAFFIC ENVIRONMENT.

5.2.1 Physical world

The Physical world is the space where the real traffic participants (vehicle, driver, pedestrian, signal, and traffic sensors, etc.) interact with the environment. The challenge of digital twinning associated with the physical world is the state detection of the physical entities and the extraction of key features to send to the digital world. The traffic participants could be divided into living entities and non-living entities, and they may need different digital twining techniques. For non-living entities, they only have external states, which are the physical states instead of psychological states, and their states can be fully captured and modeled through sensing or detection. For vehicles, the states include the kinematic features (e.g., position, speed, and acceleration) and ego characteristics (e.g., vehicle model and color). These features can be captured in real-time through onboard sensing (e.g., CAN BUS, GNSS) or external sensor detection (e.g., roadside cameras, radar). The digital twinning of traffic flow can also be challenging. The traditional method like Microwave Vehicle Detection Sensor (MVDS) provides reliable information about traffic flow in terms of volume, speed, and occupancy[189] . However, the MVDS data is still at a macro-level and the micro-level, or vehicle group level information is not captured. The increase in vehicle connectivity makes it easier for precise information collection of a vehicle group. In a mixed traffic environment, which may be a long-term state in the foreseeable future, the connected vehicles could build up a Vehicular ad hoc network (VANET) that collectively sense and share the information of the surrounding traffic. The digital twinning for living entities is more complex than non-living entities, as both external state and internal state need to be detected or modeled. The internal state refers to the psychological status of a traffic participant including intention, emotion, personality, etc. The internal state normally cannot be measured directly and relies on identification or prediction

through external behaviors. For drivers, their external state could be a set of human characteristics (e.g., gender, age) and wellness indicators (e.g., distraction, fatigue) that can be captured through wearable sensors (e.g., EEG sensor, EMG sensor), while the internal state includes driving intention and driving style preference. The internal state needs to be estimated and predicted using the external state as input, and various methods have been proposed including mapping attention field to external environment, RNN-based sequential models, and HMM models. For instance, the driver's gazing intention can be predicted through various methods [24, 190]. For pedestrians, motion is the external state such as walking/running/standing and body gestures, and it could be measured by wearable sensing suites and devices, or through external sensing like video cameras. Pedestrians' internal state mainly includes crossing intention and desired path, and it also needs to be predicted [163].

5.3.2 Digital world

The digital world is a virtual environment that parallels the real-world, where all the traffic participants are the projection of real entities in a real-time manner. Under the proposed Digital Twin framework, the digital world is the place where the vehicle and pedestrian-related ITS applications are deployed and simulated. FIGURE 48 Digital Twin framework for connected vehicle, pedestrian and traffic environment. 4 For connected vehicle digital twinning, the primary task is the projection of the vehicle states from the real world to the digital world. The vehicle's external state, including localization, kinematic features, and vehicle characteristics, are the features that could be directly replicated in the digital world. If considering the CV and drivers as an entity, the vehicle could thus have internal state due to the drivers' psychological attributes. The CVs share information about its ego states and the surrounding traffic, and thus they can conduct ITS applications such as safety warning, collaborative sensing, decision

making, and vehicle control. Assisted by the computing power in the digital world, the effects of these ITS technologies can be simulated. Afterwards, the optimal driving strategy can be derived and feedback to the real-world CVs. The digital pedestrian should also have the external state and internal state. The external state is relatively easy to be represented in the virtual space. It only requires the pedestrians' motion captured from the real-world to be digitally twined such as localization, moving speed and direction, and body gesture. The internal state, referring to pedestrian's psychological condition, needs to be modeled and predicted [191]. As the behavior of the digital avatar is the same as real-world pedestrian, their internal state (e.g., crossing intention, desired path) could be modeled and predicted through probabilistic models or machine learning methods [163, 192]. Once the digital avatar of vehicles and pedestrians are replicated in the digital space, ITS applications could be implemented virtually. For a traffic scenario with both CV and pedestrian presence, various ITS technologies can be applied. Based on the communication subjects, it can be divided into vehicle-to-vehicle (V2V), vehicle-to-pedestrian (V2P), vehicleto-infrastructure (V2I), or vehicle-to-everything (V2X) in general. In terms of functionality, these ITS technologies include driving cooperation, V2V or V2P collision warning, eco-driving guidance, etc. Normally, the digital world has stronger computing power (e.g., using cloud computing), and it can execute fast simulation and prediction to investigate the effects of the ITS technology and to generate optimal driving or crossing suggestions for CV and pedestrian, respectively. Afterwards, the suggestions are forwarded to the real world entities and form a data transmission closed loop.

5.3 Framework Realization

In this section, a sample architecture is presented to demonstrate a possible realization of the proposed Digital Twin framework, as shown in FIGURE 49. The architecture is built based on various simulations and physical environments that incorporates both CV and pedestrians.

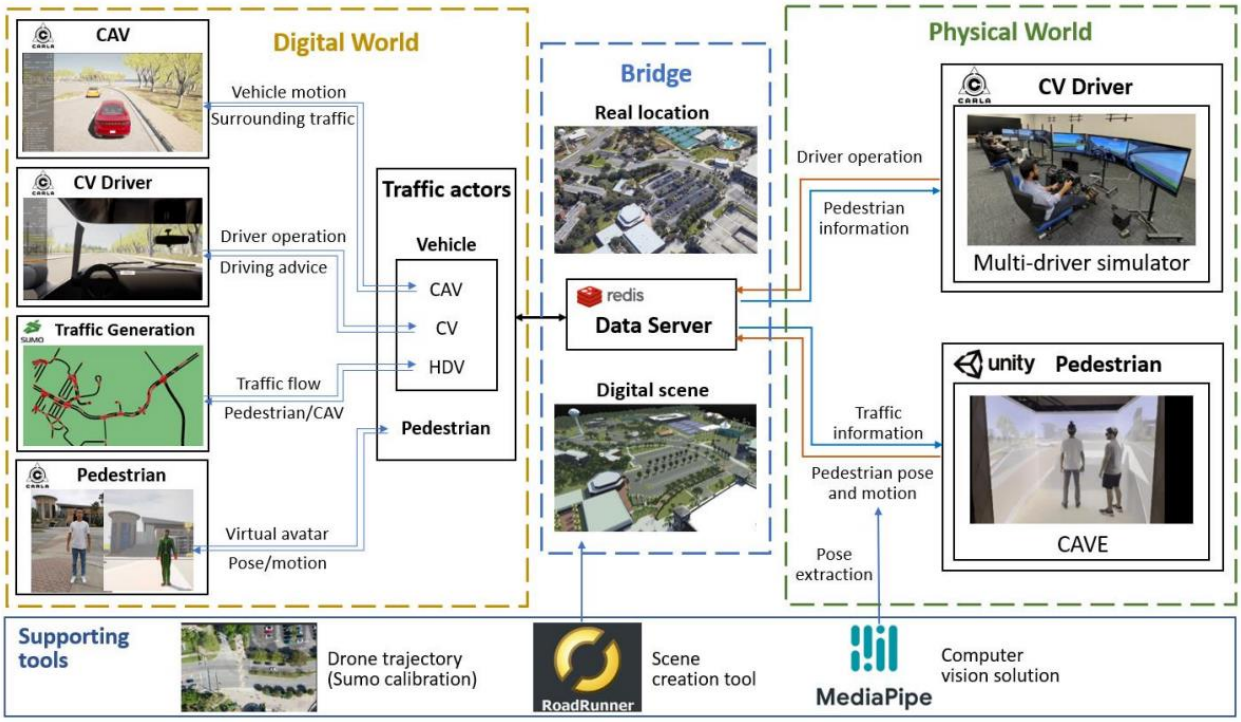


FIGURE 49 SAMPLE ARCHITECTURE OF CV AND PEDESTRIAN IN-THE-LOOP DIGITAL TWIN FRAMEWORK

5.3.1 Connected Vehicle Digital Twin

Since it is technically challenging and cost demanding to use a real vehicle for field tests, driving simulator could be a good alternative to represent a high-fidelity driving environment. Also, for experiments with safety concerns (e.g., a pedestrian crossing in an occlusion condition), simulator is the primary choice. In this sample architecture, a multi-driver co-simulation platform is adopted. The simulator is based on the co-simulation between Carla and Sumo, while multiple participants can drive simultaneously in the Carla virtual environment[193] . FIGURE 50

demonstrates the Carla-Sumo co-simulation based on the location University of Central Florida FIGURE 49Sample architecture of CV and pedestrian in-the-loop digital twin framework 5 (UCF) gym from CitySim dataset[54], which is an open-source vehicle trajectory dataset launched by our Safe&Smart Transportation (SST) team. Carla is an advanced autonomous driving simulator that is built on Unreal 4 game engine, and manual and autonomous driving can be conducted in Carla [125]. For driving simulator experiments, Carla is capable of rendering high-fidelity graphic displays and simulate vehicle dynamics. Sumo is an open-source microscopic traffic simulator [194]. It is powered by vehicle behavior models including driving behavior models and route choice models. After fine model calibration, Sumo can generate traffic flow that represents the real traffic and vehicle behaviors. For model calibration, vehicle trajectory data is adopted as it provides rich data on traffic flow and driving behaviors. The trajectory data used in this project is from the CitySim dataset. The Carla map and sumo road network are exactly matched as they share the same Opendrive road geometry file. The traffic actors are synchronized in both simulators, and Sumo vehicles and Carla vehicles can interact with each other to produce a mixed simulation environment. To power the co-simulation, a PC with Intel® Core™ i7-7800X CPU @ 3.50GHz × 12 and a memory of 64 GB plus NVIDIA GeForce RTX 2080 GPU is used. For the setup of driving simulator, we used triple 40'' Vizio TVs with 1920*1080 resolution, and Fanatec V2.5 steering wheel and hydraulic-supported pedals To reproduce the real-world driving scene, high-fidelity 3D maps of the study location are built using the road scene designer RoadRunner (3D maps are available at CitySim dataset homepage: https://github.com/ozheng1993/UCF-SSTCitySim-Dataset). The maps are built using GIS data to guarantee the quality, including arial images, elevation data, and point cloud data. By such means, the driving behaviors of real drivers at real locations can thus be collected.

Benefiting from using the simulator, the vehicle state can be directly accessed through Carla's Python API, instead of being detected or sensed as in real world. The drivers' state can be captured through detecting or sensing the same methods as real-world incabin technologies such as eye-tracking and fatigue identification. Once the driver and vehicle data are obtained, they are uploaded to the cloud server (Redis in the architecture) and stored. In addition, the connected and automated vehicle (CAV) and human-driven vehicle (HDV) are also included in the sample architecture in FIGURE 49 to be able to simulate a mixed traffic environment.

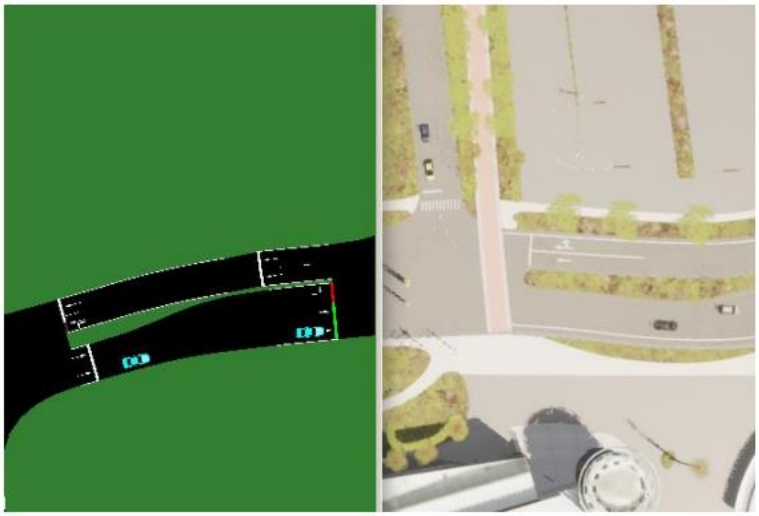


FIGURE 50 CARLA-SUMO CO-SIMULATION (LEFT: SUMO; RIGHT: CARLA)

5.3.2 Pedestrian Digital Twin

Safety is always the primary concern for pedestrian-related studies. When the pedestrian is exposed to potential risk in an experiment, it must be held in simulation instead of field testing. Due to this reason, the objective of pedestrian digital twinning is to create an immersive and close-to-real simulation environment for pedestrians. In this architecture, we used a Cave automatic virtual environment (CAVE) to serve as the testbed for pedestrians. The CAVE equipment used in this study is called "VDen", which is a lightweight, portable version of CAVE. For this architecture, we used a PC with Intel® Core™ i7-10700 3.8GHz CPU, 32GB

DDR4 RAM, NVIDIA GeForce 2080 GPU. The system projects 3D videos on 4 screens (left, front, right and bottom), and provides a 3*3-meter movable physical space. The projectors in this architecture are 4 BenQ ultra-short throw projectors support up to 1920x1200 120Hz 3D output, 60Hz per eye. The users are wearing a 3D glasses and a tracker for localization in the CAVE. Also, an additional tracker is used to track the eye position in order to get the correct perspective of view. Compared to the other VR device Head Mounted Displays (HMDs), CAVE has several evident advantages, especially for pedestrian simulation. First, it feels more natural and comfortable for the user only to wear a pair of lightweight 3D glasses and a tracker instead of a heavy headset. Second, CAVE brings less cybersickness because CAVE displays multiple images on the walls simultaneously instead of generating a new image when the user is moving his head. In addition, when experiencing VR in CAVE, users could naturally be aware of the physical presence of their own and each other's entire bodies, which is crucial for multi-user and social VR scenarios. Users can freely communicate with one another as if they were in the real world. In HMD setups, this is very challenging to achieve. In order to replicate the pedestrian's pose in the digital world, the real body's keypoints are captured and matched to the digital avatar. Keypoint detection is carried out by Google MediaPipe Posekeypoint detector using pre-trained weights from BlazePose GHUM 3D. Unlike most state-of-the-art approaches, which depend on a robust server environment for 6 inference, the MediaPipe achieves real-time pose detection on the 2022 MacBook pro with M1 Pro Max GPU. The MediaPipe outputs are uploaded to a Redis Pub/Sub server in JSON format, which includes 33 human key points' names and the bone transform in 3D space. This JSON file will be pushed into CARLA Client "WalkerBoneControl" class every tick to modify the CARLA walker agent's skeleton. To simulate the pedestrian's walking movement, locomotion method is used. It allows users to move

forward or back by lifting their legs to simulate a real walking pose. The implementation is based on binding two additional lighthouse trackers to the user's legs and monitoring the range of movement of the sensors in the vertical direction to trigger a movement event. The horizontal orientation of the movement of the sensor also determines the direction of motion. The methods of pedestrian pose matching and achieving locomotion are shown in FIGURE 51

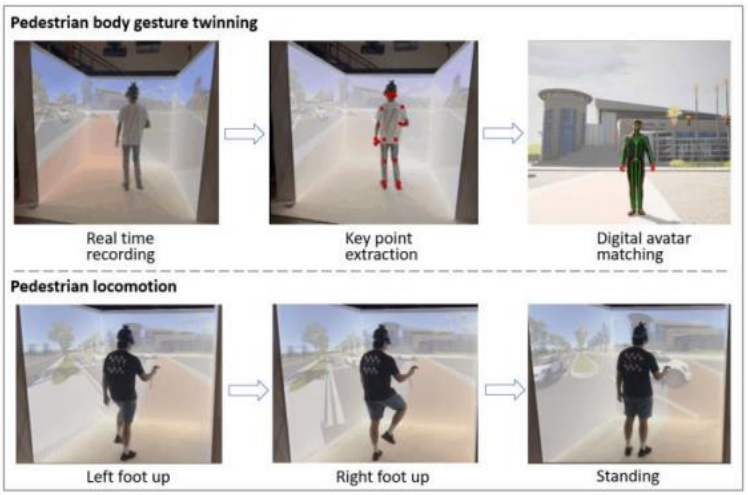


FIGURE 51. PEDESTRIAN DIGITAL TWINING IN CAVE

5.3.3 Closed-loop Data Transmission

Since the presented architecture uses two major platforms: Carla-Sumo (vehicles) and Unity (pedestrian), the two digital spaces need to be synchronized. FIGURE 52 shows the data transmission structure between the two platforms. The data transmitted from Carla-Sumo to Unity includes the vehicle state and driver state; while the data sent from Unity/CAVE to Carla is the pedestrian's location, motion and pose.

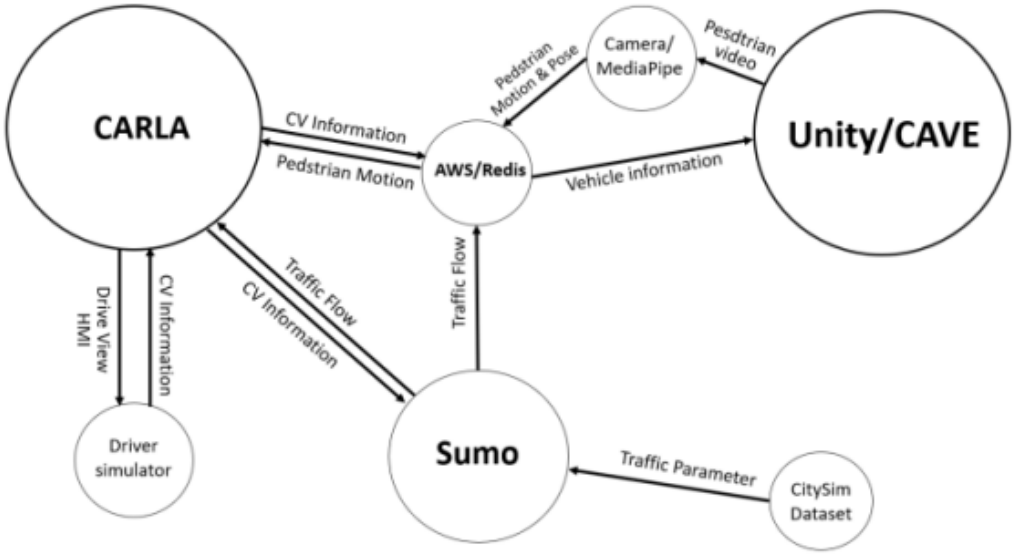


FIGURE 52 DATA TRANSMISSION BETWEEN EACH MODULE

The idea of Digital Twin is to form a closed loop from physical entities to digital avatars and then generate feedback for physical entities. Hence, the proposed architecture follows this concept and keeps a closed-loop data transmission structure. When conducting an experiment, the driver and vehicle information is read from the driving simulator and is uploaded to Redis server. The pedestrians' location, motion and gestures are also tracked in the CAVE and are uploaded to the server simultaneously. At each simulation step, the system accesses the data server to download the vehicle or pedestrian data, and projects the avatar into the digital space. Afterwards, ITS technologies are implemented in the virtual world. For example, V2P collision warning system predicts crashing likelihood and triggers warning messages. Finally, suggestions for CV drivers or pedestrians are generated and feedback to the physical space (e.g., display warning messages on vehicle onboard unit).

5.4 Case study

5.4.1 Investigates The Effects Of V2P Collision Warning System Under Occlusion Conditions

In this section, a case study based on the proposed architecture will be presented that investigates the effects of V2P collision warning system under occlusion conditions. Three experiments are designed to investigate V2P safety of (a) HDV without V2P warning; (b) AV with Automatic emergency braking (AEB); and (c) CV with V2P communication, respectively. In the experiment, HDV and CV are operated by the drivers on the driving simulator and AV is automatically controlled, while the pedestrian is simulated using the CAVE.

5.4.1.1 Experiment Setup

The experiment is derived from a real crash case in front of the gym of UCF (FIGURE 50), where a pedestrian violated the traffic light and was hit by a vehicle under occlusion condition. During school hours, many students are heading to the gym and crossing behaviors with red-light violations are frequently observed, which sometimes leads to conflicts or even crashes. The development of multiple sensing and communication technologies allows the movement of both vehicles and pedestrian to be captured and broadcasted to each other, and V2P collision warning can be activated in a safety-critical scenario to prevent accidents. To test its effectiveness, three experiments are designed to reproduce the crash scene, as shown in FIGURE 53. For all three experiments, the vehicle (HDV/CV driven by the driver, or CAV controlled by Sumo) is driving at 25 mph towards the intersection. When the vehicle's estimated arrival time to the conflict point is 5 seconds, the pedestrian light turns from "Don't walk" into "Walk", and the pedestrian in the CAVE starts to cross the street with a constant speed of 1 m/s.Two buses are placed in front of the zebra crossing, and both driver and pedestrian cannot see each other until closer to the conflict point. For experiment 1, the driver and pedestrian will not receive any warning; in

experiment 2, the AV activates the AEB system, and the vehicle and pedestrian are not receiving warning; while in experiment 3, the collision warning will be displayed through the smartphone (as on board unit OBU emulator) to both CV and pedestrian.

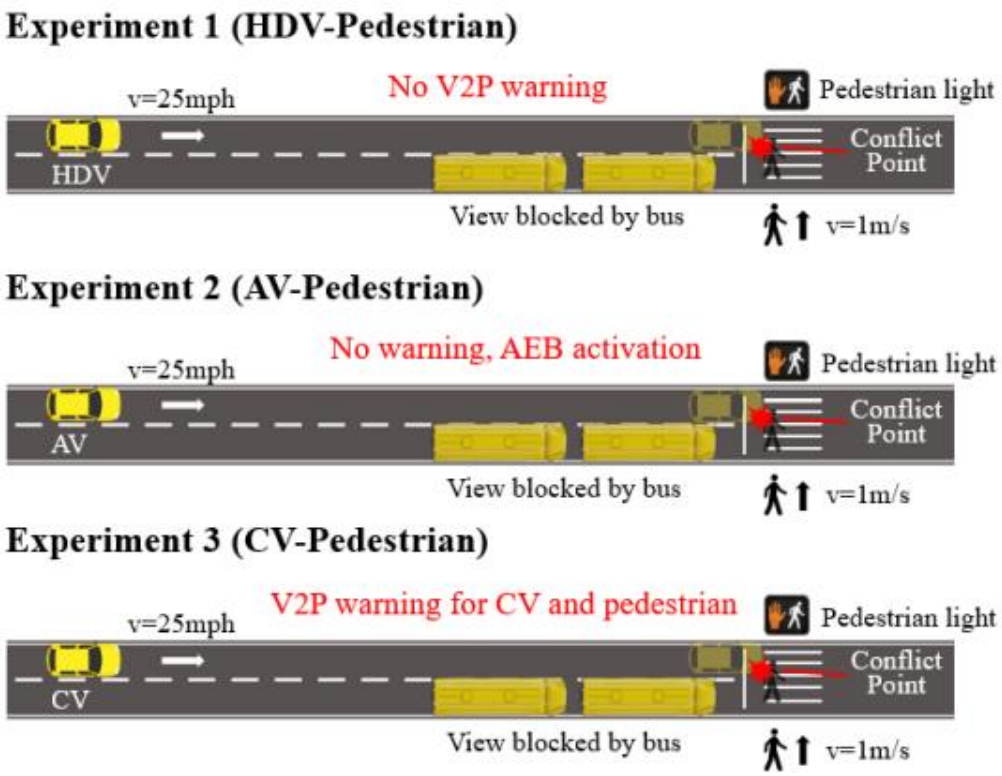


FIGURE 53. EXPERIMENT DESIGN

The collision warning system consistently calculates the time-to-collision (TTC) between vehicle and pedestrian and triggers warning when the TTC is below the predefined threshold. The V2P warning triggering condition is shown in following Equation.

$$Trigger = \begin{cases} yes, \ if \ \ TTC < TTC^{*} \\ no, \ else \end{cases}$$

$$TTC = \left| t_{veh} - t_{ped} \right|$$

$$t_i = d_i / v_i, \ i \in \{veh, ped\}$$

where $TTC*$ is the activation threshold (1.5 second in this study), and $tveh, tped$ are the estimated arrival time to the collision point, which is calculated by the remaining distance to the conflict point divided by current speed (d/v). Once the system is triggered, a warning is sent to the CV driver and displayed through the Onboard Unit (OBU), which is emulated by a smartphone. For AV simulation, the P2V safety is ensured by activating the Automatic emergency braking (AEB) system. A Lidar is attached to the AV in Carla to detect the pedestrian, and once the V2P TTC is smaller than 1.5 seconds, the AEB triggers. This case study adopts the AEB algorithm of standard gradient deceleration on dry road[195, 196], and the braking profile is shown in following Equation.

$$Deceration\ (feet/s^2) = \begin{cases} 0;\ t < 0.25\,s \\ 65.7(t-0.25);\ 0.25 \le t < 0.6\,s \\ 23; t > 0.6\,s \end{cases}$$

where t is the time to the AEB activation time. Six groups (each group includes a driver and a pedestrian) of lab researchers participated in the experiments, and each group experience all three experiments. The participants are requested to behave in a safe manner and try to avoid potential crashes.

5.4.1.2 Experiment Results

The vehicle decelerates once the pedestrian is detected, and the vehicle's distance to the zebra crossing when it completely stops (denote as V2P distance) can reflect the closeness of the V2P conflict. Braking point, which is defined as the vehicle's distance to the zebra crossing when the vehicle starts to brake, is used to measure the braking timing. A larger braking point indicates the driver starts to brake earlier to prevent a crash. Also, the average deceleration of the vehicle during the braking period is examined. The results of the average V2P distance and braking point for the three experiments are shown in TABLE 8

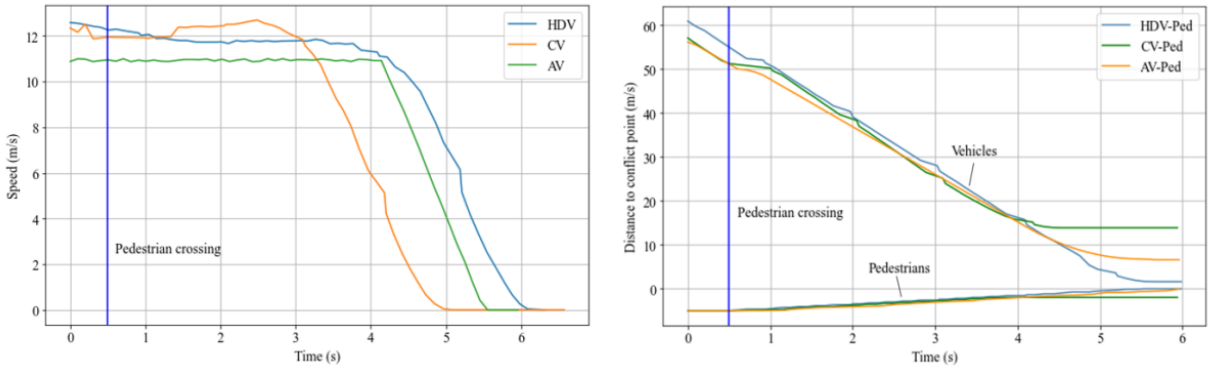
TABLE 8 RESULTS OF V2P DISTANCE AND BRAKING POINT

Experiment	V2P Distance (m)	Braking Point (m)	Average Deceleration (m/s^2)
1. HDV-Pedestrian	3.941	11.458	7.464
2. AV-Pedestrian	5.602	14.612	7.031
3. CV-Pedestrian	17.179	28.325	6.216

5.4.1.3 Conclusions and Future Work

As expected, the experiments of HDV-pedestrian show the smallest V2P distance and braking point, which means the driver starts to brake latest among all experiments and almost hit the pedestrian. The CV-pedestrian experiments have the largest value for both V2P distance and braking point. This indicates that the driver receives V2P warning before seeing the pedestrian and makes braking maneuver in advance to ensure safety. The value of V2P distance and braking point of AV-pedestrian is between the other two experiments. In the occlusion condition, the Lidar cannot detect the pedestrian until the occlusion is relieved, and the AEB activates to avoid a crash. The results show that under occlusion conditions, the HDV gets very close to the pedestrian without the V2P warning. For AV, the AEB system improves the safety slightly. With the collision warning for both CV and pedestrian, the V2P safety has been enhanced by a significant margin. In terms of average deceleration, the result of experiment 1 is the largest with a value of 7.464 m/s^2, indicating that the human drivers braked hard after seeing the pedestrian. The AV also generates an average deceleration over 7 m/s^2, while in experiment 3 the value is significantly smaller. This shows that the V2P warning also enhance comfort while ensuring safety.To further demonstrate the results, a group of experiments are visualized in FIGURE 54. The speed-time plot is presented in FIGURE 54(a), where the speed over time of the HDV, AV and CV are represented in the blue, orange, and green line, respectively. It can be observed that the CV brakes earliest, followed by the AV then HDV. FIGURE 54 (b) shows the space-time plot of the three experiments.

The space is the distance to the conflict point, and the upper lines belong to vehicles and the lines below the x-axis are pedestrians. The plot shows the CV stops far away from the conflict point while the HDV almost reaches the conflict point. In addition, the pedestrians in the HDV-pedestrian and AVpedestrian group arrived at the conflict point, because they are not aware of the presence of vehicles. In the CV-pedestrian experiment, the pedestrian receives the warning and stops to cross the street, as the bottom green line stops to rise after 4 seconds. The differences in pedestrian behavior validate the need for both CV and pedestrian in-the-loop simulation, as conventional CV-related simulation study does not consider pedestrian reaction



(a) Speed-time plot (a) Space-time plot

FIGURE 54 VISUALIZATION OF RESULTS. (A) SPEED-TIME PLOT FOR THREE EXPERIMENTS; (B) SPACE-TIME PLOT FOR BOTH VEHICLES AND PEDESTRIANS, NOTE: THE LINES ABOVE X-AXIS ARE VEHICLES AND THE LINES BELOW ARE PEDESTRIANS

5.4.2 Using virtual simulator to evaluate the automated emergency braking system for avoiding pedestrian crashes at intersections under occlusion

5.4.2.1 Experiment Setup

The AEB system consists of two main parts: one is the sensor model and the other is the braking strategy. The sensor is mainly responsible for the perception of the surrounding environment in which the ego vehicle is moving. As illustrated in the FIGURE 55 sensor scans a segment of a

circle while transmitting detection rays to detect surrounding objects (e.g., a pedestrian). The minimal distance between the ego vehicle and the moving objects could be used to calculate the time to collision (TTC). The TTC could be applied as a metric to activate the brake signal of the AEB system. Once the TTC is below the specific configurable activation threshold, the AEB system is triggered to react to avoid the potential collision. In this study, the brake system described in the previous study [195, 197] has been adopted. The braking system response consists of a brake delay and the build-up time until the full brake, which could be expressed by the following equation.

$$Deceration\ rate\ (unit: feet/s^2) = \begin{cases} 0; t < 0.25 \\ 65.7(t - 0.25); 0.25s \leq t < 0.6s \\ 23; t > 0.6s \end{cases}$$

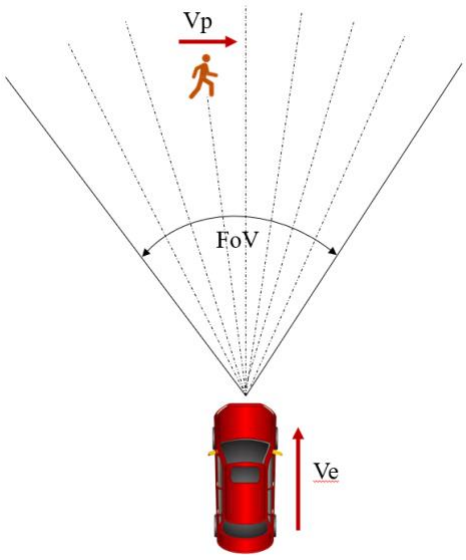
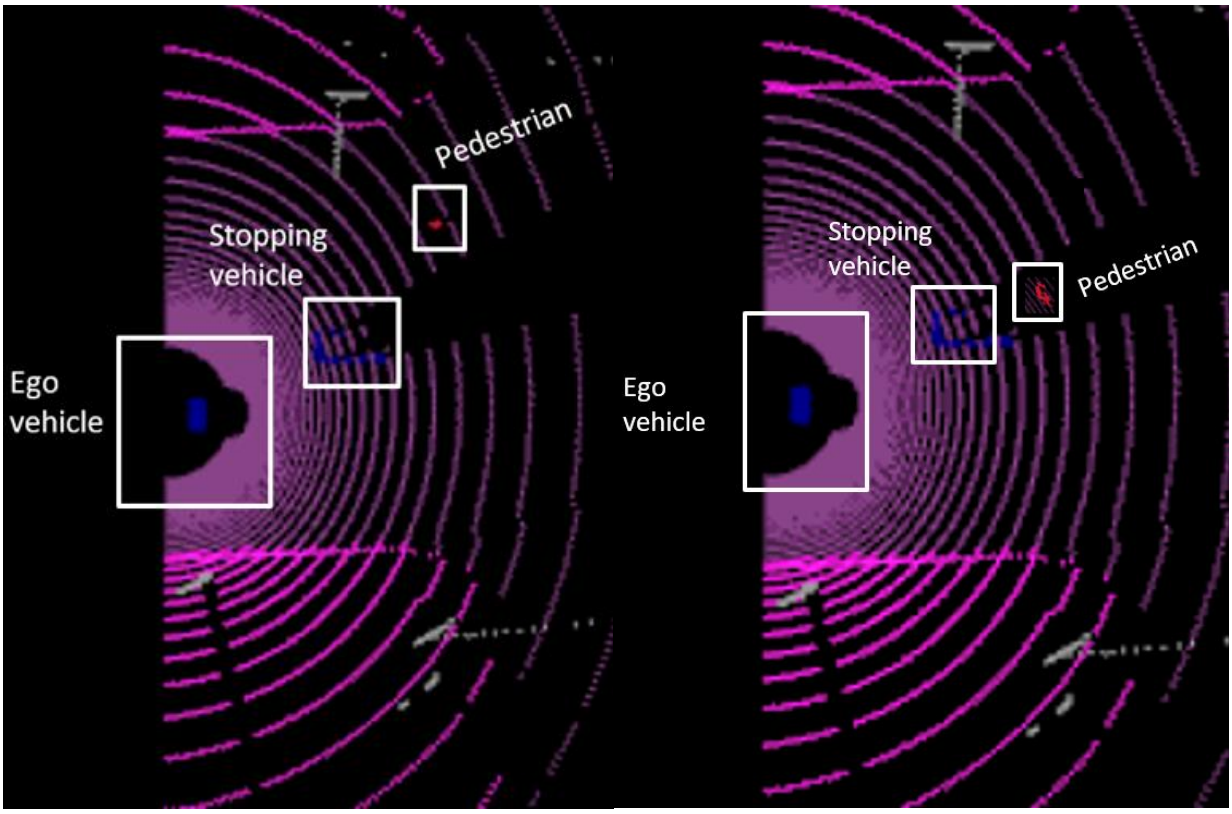


FIGURE 55 ILLUSTRATION OF PEDESTRIAN DETECTION AND FOV OF SENSOR

While the effectiveness of the AEB system on avoiding pedestrian crashes has been validated in the previous studies [198], it remains unclear if the pedestrian is occluded by other vehicles. As shown in FIGURE 56(b), the sensor of the ego vehicle could not detect the pedestrian since the pedestrian is occluded by the stopping vehicle. In that case, there might not be enough time to

activate the brake system to decelerate to avoid collisions. Hence, further investigations are needed to explore the effectiveness of the AEB system under the occlusion conditions.



(a) NOT OCCLUSION CONDITION **(b) OCCLUSION CONDITION**

FIGURE 56 ILLUSTRATION OF OCCLUSION AND NOT OCCLUSION CONDITIONS

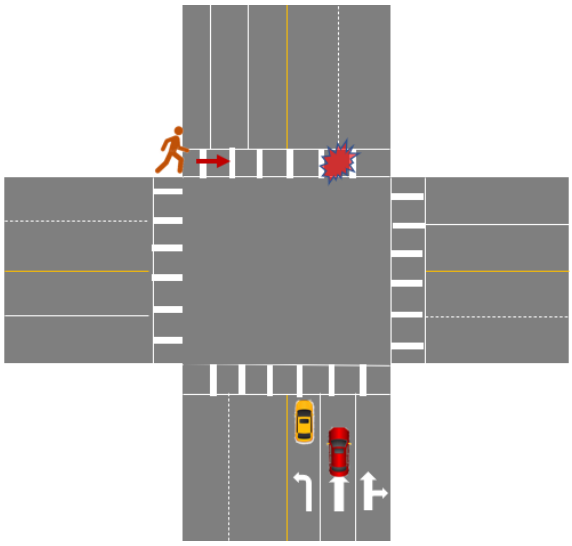
Occlusion scenarios

In this study, three common occlusion scenarios of pedestrians for the through vehicles were investigated based on our previous study about pedestrian crashes [107]. The three scenarios are summarized as follows:

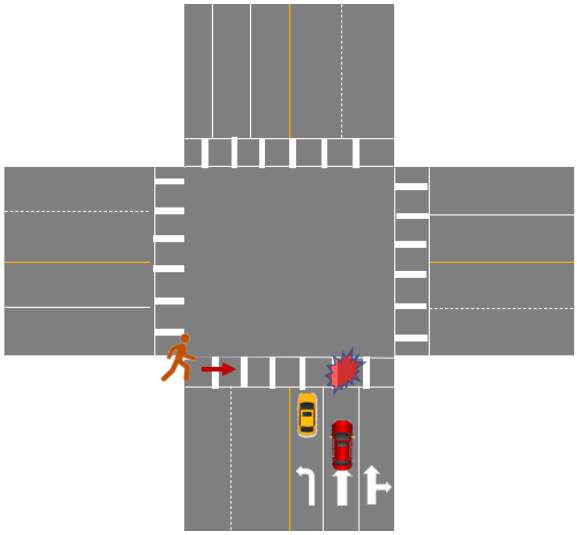
- Scenario 1: the ego vehicle is going through, while a pedestrian is walking on the crosswalk of the ego vehicle's exiting approach and the pedestrian is occluded by a vehicle on the left-turn lane (FIGURE 57(a))

- Scenario 2: the ego vehicle is going through, while a pedestrian is walking on the crosswalk of the ego vehicle's entering approach and the pedestrian is occluded by a vehicle on the left-turn lane (FIGURE 57(b))
- Scenario 3: the ego vehicle is going through, while a pedestrian is walking on the crosswalk of the ego vehicle's entering approach and the pedestrian is occluded by a vehicle on the right-turn lane (FIGURE 57(c))

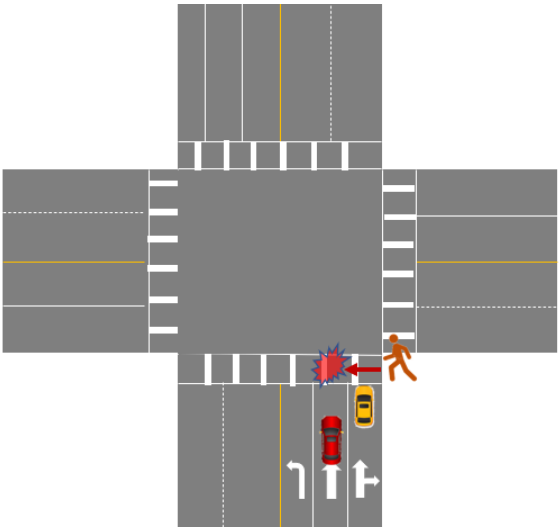
As illustrated in FIGURE 57, the pedestrian could be occluded by the stopping vehicle. If occluded, the pedestrian could not be detected by the sensor of the AEB system of the ego vehicle and the braking system would not be activated even reach the threshold. The potential collision points are also highlighted in the figure. In this study, a typical intersection in the Carla map- Town 3 is selected. The entering approach contains three lanes: one is left lane, one is through only lane, and one is through and right lane. The ego vehicle is on the middle lane and it will drive from the upstream to the intersection. It is assumed that the ego vehicle arrives at the intersection during the green time and it will pass the intersection without a stop. In addition, the intersection has crosswalks on which the pedestrian could walking across the intersection. The pedestrian is assumed to crossing the intersection during the red light without seeing the coming through vehicle. Hence, a collision is highly likely to happen and the effectiveness of the AEB system could even be reduced.



(a) SCENARIO 1



(b) SCENARIO 2



(c) SCENARIO 3

FIGURE 57 ILLUSTRATION OF SIMULATION SCENARIOS

Agents' motion states and system parameters

The development of simulation cases considered the driving speed of the ego vehicle and the crossing speed of the pedestrian in the different study scenarios. As shown in TABLE 9 pedestrian six crossing speeds for the pedestrian are simulated, from 2 feet/s to 12 feet/s, similar to the previous study [198]. To simulate different conditions when the ego vehicle has a collision with

the pedestrian, an offset time was used to describe when the pedestrian starts to cross the intersection based on the arrival time of the ego vehicle to the collision point. According to the previous study [199], six offset time was used in this study. Six ego vehicle's speed categories from 25 mph and 50 mph, which are the typical driving speed on the partial access-control road. Hence, the total combinations considering the pedestrian speed, ego vehicle speed, and the offset time for the pedestrian to cross are 216 for each scenario.

TABLE 9 MOTION STATES OF EGO VEHICLE AND PEDESTRIAN

Parameter	Value	Step size	Counts
Pedestrian initial speed (feet/s)	2-12	2	6
Ego vehicle initial speed (through, mph)	25-50	5	6
Offset time for pedestrian to cross (s)	1-6	1	6
Total	-	-	216

As described above, the maximum deceleration rate to achieve the target speed is 3.28 $feet/s^2$. Besides having to brake due to the curve as described above, the braking process would not consider other interactions with the surroundings once activated. The braking system is activated if the time to collision (TTC) reaches the threshold, and the pedestrian is not occluded by the stopping vehicle. If the pedestrian is occluded by the stopping vehicle, the activation time of AEB could be delayed. Three TTC thresholds to activate AEB were tested from 1 to 3 seconds. Besides, the sensor's field of view (FoV) could affect the time when the pedestrian could be detected and then affect the activation time if under dangerous condition. Hence, five different angles of FoV were tested. Hence, there are 16 AEB control cases (i.e., 15 with AEB control and 1 without AEB

control) included. A total of 10,368 (3 scenarios × 216 motion states ×16 AEB control cases) simulation runs were conducted to evaluate the effects of AEB under occlusion conditions.

TABLE 10 AEB CONTROL CASES

Parameter		Value	Step size	Counts
AEB control	Sensor FoV (angle)	60-180	30	5
	TTC threshold to activate AEB (s)	1-3	1	3
No AEB control		-	-	1
Total		-	-	16

The model of the ego vehicle in this study is a typical passenger car that has a length of 16.5 feet and a width of 6.6 feet. The LiDAR sensor locates at the top front part of the ego car 1.6 feet away from the center of the ego vehicle. The sensor parameters in this study were based on as the following:

- Upper FoV: 15°
- Lower FoV: -25°
- Number of channels: 64
- Detection ranges: 300 feet
- Rotation frequency: 20 HZ
- Points per second: 500,000

Evaluation methods

Different measures were included to evaluate the performance of the AEB system. The output information from the Carla simulation such as agent center position (x, y) in the global coordinate

system, agent speeds, yaw angles of the ego vehicle, and vehicle dimension could be used to compute the measures. First, the collisions could be determined by the geometrical overlap of the agent contours, which could be calculated by each time frame. As shown in FIGURE 58,the position of both the ego vehicle and the pedestrian at the collision time could be identified. By using the geometrical features of the ego vehicle and the pedestrian, the relative impact location to the center of the respective contour edge can be estimated. Besides, the impact speed of the ego vehicle could be obtained for the evaluation as higher speed would result in more severe pedestrian crashes [109]. During the simulation, the information about whether the pedestrian is occluded (i.e., the pedestrian could not be detected by the LiDAR sensor even in the detection range) could be recorded. Hence, the duration of occlusion could be obtained to explore the effects of occlusion on the effectiveness of the AEB system.

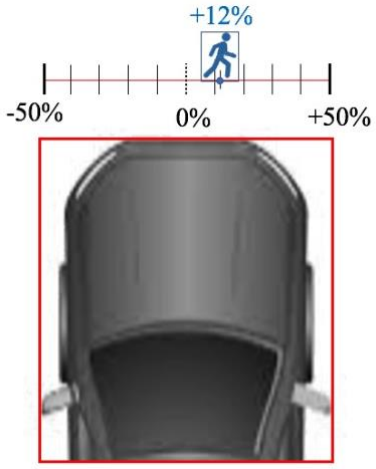


FIGURE 58 COLLISION LOCATION WITH THE RESPECT TO THE CENTER OF THE CORRESPONDING CONTOUR EDGE (IN %)

5.4.2.2 Experiment Results

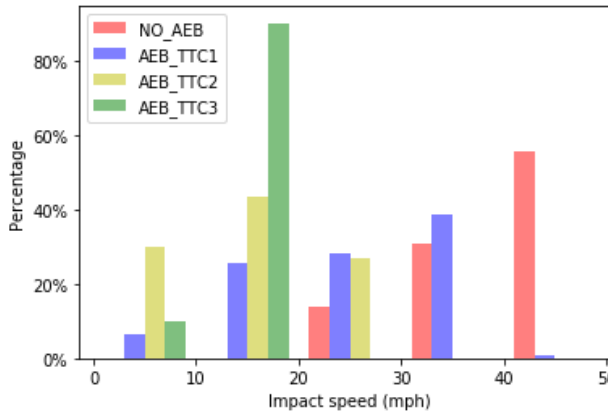
To validate the effectiveness of the AEB under occlusion conditions, the simulation without the AEB braking system was initially simulated as the baseline case. Under this condition, it is assumed that the ego vehicle is not able to react to take any deceleration maneuvers. For each scenario, a total of 226 simulations were conducted to reflect different motion statuses of the ego vehicle and the pedestrian. Corresponding to each case without AEB, fifteen simulations with AEB were conducted considering the different FOVs and AEB activate times. As discussed in Section 2, the activation time is the time when the AEB is activated under the TTC threshold. For each threshold, the number of simulation cases is 1,080 in each occlusion scenario. The percentages of collisions over the simulation cases are summarized in TABLE 11 SUMMARY OF NUMBER OF COLLISIONS UNDER DIFFERENT CONDITIONS Without AEB, more collisions could be found in Scenario 3 (i.e., the pedestrian is occluded by a vehicle at the right side of the ego vehicle). The AEB system could still reduce the number of collisions significantly. With the increase of the activation threshold, more collision could be avoided. If the threshold is 1 second, the ego vehicle would still hit the pedestrian in around 20% of cases.

TABLE 11 SUMMARY OF NUMBER OF COLLISIONS UNDER DIFFERENT CONDITIONS

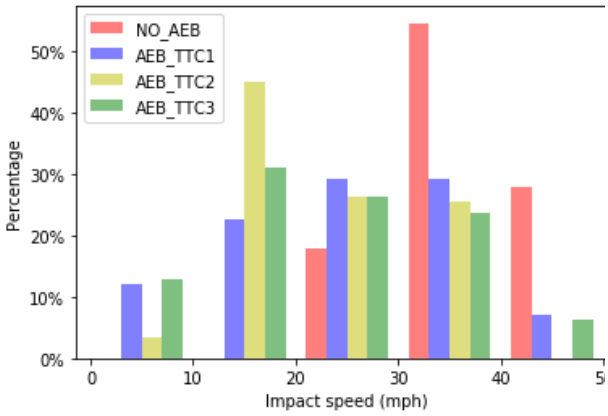
Scenario	Measures	Without AEB (number of cases=216)	With AEB with different activate time (number of cases per threshold=1080)		
			1 second	2 seconds	3 seconds
Scenario 1	Percentage	33.80%	18.61%	3.43%	0.93%
	Reduction percentage	-	15.19%	30.37%	32.87%
Scenario 2	Frequency	28.70%	19.35%	13.80%	10.19%
	Reduction percentage	-	9.35%	14.91%	18.52%
Scenario 3	Frequency	50.56%	19.17%	9.17%	5.19%
	Reduction percentage	-	31.39%	41.39%	45.37%

The percentage of the impact speed once a collision occurs between the ego vehicle and the pedestrian was calculated for different AEB control conditions. The results are presented in FIGURE 59The figure shows that higher impact speeds could be observed under the no AEB conditions, leading to more severe pedestrian crashes. In the cases when the pedestrian is occluded by the vehicle on the left-turn lane, the impact speeds get reduced with the increase of the activate threshold. However, in the cases when the pedestrian is occluded by the left-turn vehicle, the impact speeds tend to be higher with the increase of the activation time. It is because that it is more

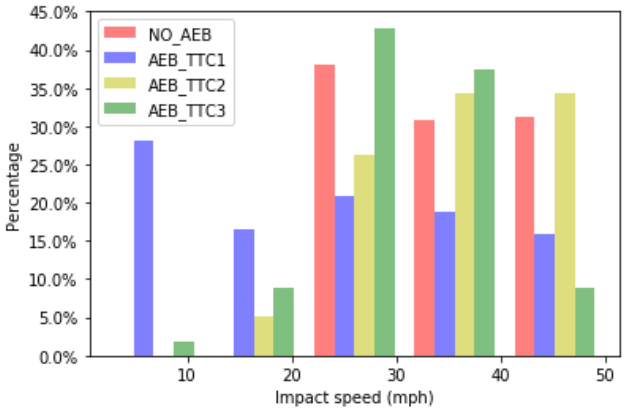
dangerous under this case and the cases which are not avoided by the AEB system with a longer activation time are more critical cases with higher initial speeds.



SCENARIO 1



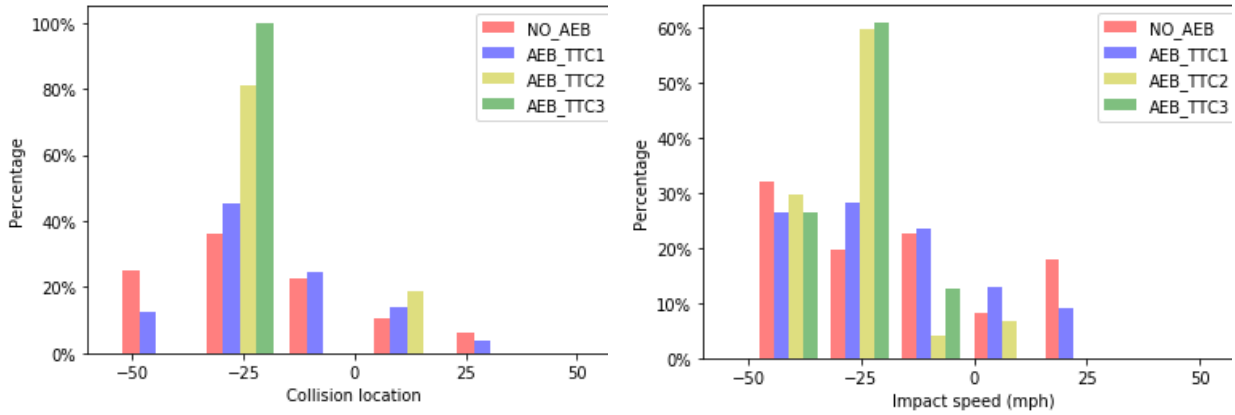
SCENARIO 2



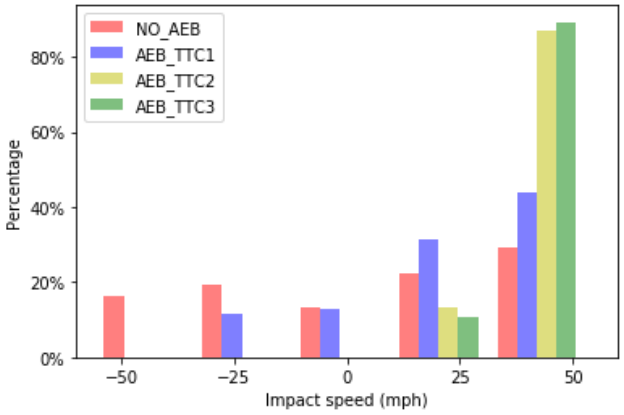
SCENARIO 3

FIGURE 59 DISTRIBUTION OF CRASH SPEED

FIGURE 60 Presents the results about the collision location related to the front center of the ego vehicle. The distributions of collision points among different AEB cases in the same scenario are the same. As expected, more collision points are at the left side of the ego vehicle in Scenarios 1 and 2 since the pedestrian crossed the intersection from the left side of the ego vehicle. Similarly, more collision points are found on the right side in Scenario 3.



SCENARIO 1 **SCENARIO 2**



SCENARIO 3

FIGURE 60 NORMALIZED COLLISION LOCATION TO THE FRONT CENTER OF THE VEHICLE

In the 3 scenarios, there are 73, 62, and 109 cases in which the ego vehicle hit the pedestrian without the AEB system. In the same scenario that the ego vehicle without AEB has a collision with the pedestrian, the AEB could either avoid the collision or reduce the impact speed compared to the condition without AEB. For each scenario, the occlusion time was collected and compared between the 'collision avoided' (i.e., the collision could be avoided by the AEB) and 'collision speed reduced' (i.e., the collision could be avoided by the AEB while the impact speed gets reduced)

conditions. ANOVA test has been conducted to compare the occlusion time under the 2 different effect conditions. As shown in FIGURE 61, the occlusion time is significantly different between the ‘collision avoided’ and ‘collision speed reduced’ conditions. Longer occlusion time could be observed for the ‘collision speed reduced’ conditions. It indicates that the occlusion could make the sensor unable to detect the risky conditions and delay the activation of the AEB braking system. Under the occlusion condition, the effectiveness of AEB could get reduced.

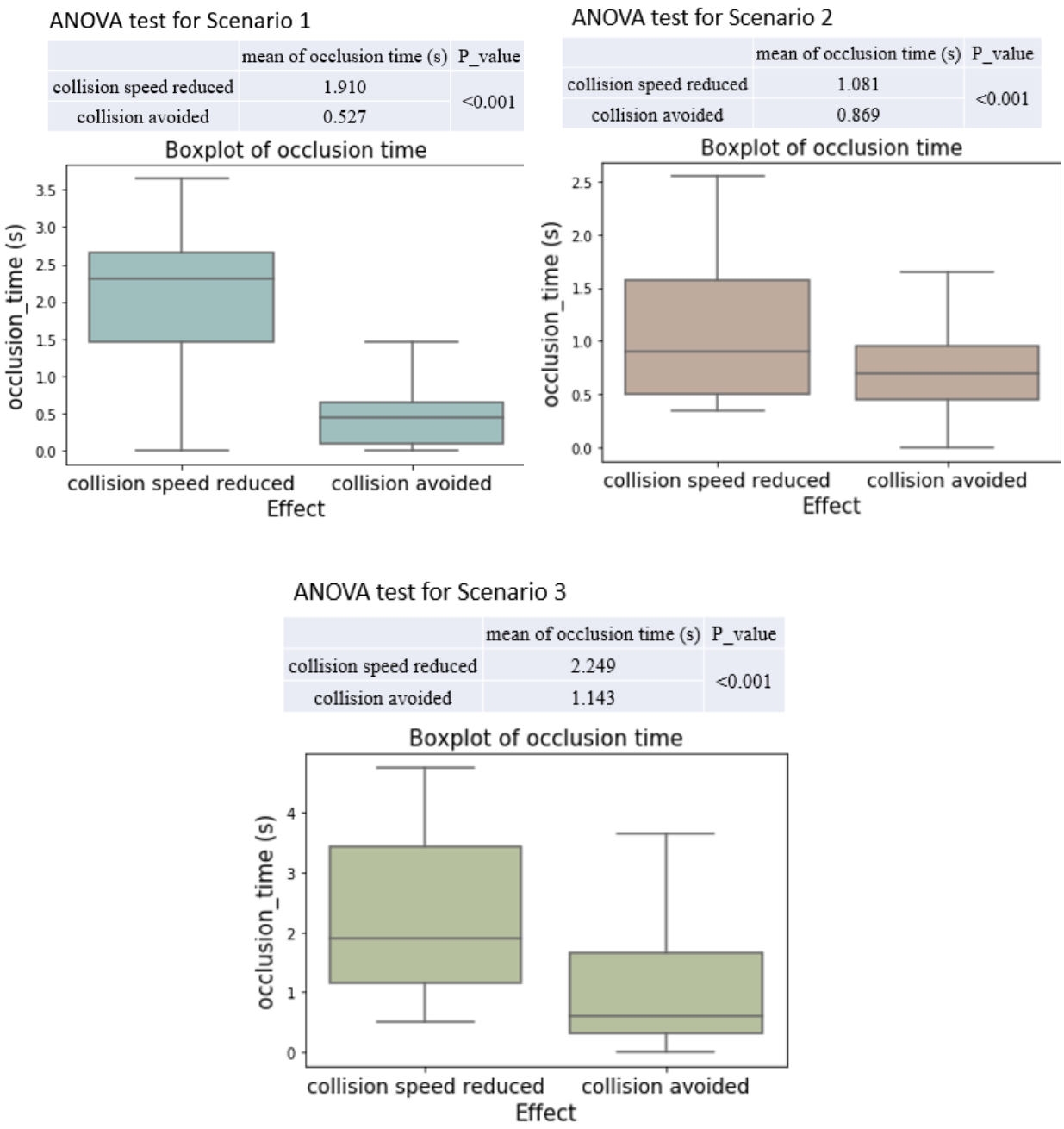


FIGURE 61 COMPARISON OF OCCLUSION TIME FOR DIFFERENT COLLISION AVOIDANCE RESULTS

To further explore the potential effects of different factors on the AEB's effectiveness, a dummy variable is used to indicate if a collision is avoided (indicator=1) or the impact speed for the collision get reduced (indicator=0). A logistic regression model is estimated to quantify the effects

of different factors including the pedestrian initial speed, ego vehicle initial speed, TTC threshold to activate the AEB, occlusion time, FOV, and scenario. The results are summarized in TABLE 12. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are estimators of prediction error, it considers both the goodness-of-fit and the simplicity of a model. Smaller value of AIC or BIC is preferred. All factors except the FOV are significant in the model. A collision could be more likely to get avoided if the pedestrian walks faster since it takes less time for the pedestrian to cross the intersection and the occlusion time is less. The ego vehicle's speed has the opposite effect on the effectiveness of AEB. The higher speed of the ego vehicle makes it difficult to reduce the speed to avoid collisions. It is expected that the ego vehicle could brake earlier with the longer TTC threshold to activate the AEB and avoid the collision. As discussed above, the effectiveness of AEB gets reduced if the pedestrian is occluded longer. Based on the effects of the TTC threshold to activate AEB and occlusion, a longer TTC threshold is needed if a long occlusion time is expected. Compared to Scenario 1, the effectiveness of the AEB system gets reduced in Scenarios 2 and 3 in which the pedestrian is closer to the stopping vehicle and easier to get occluded.

TABLE 12 LOGISTIC REGRESSION MODEL RESULT FOR AEB' EFFECTIVENESS

Variable	Mean	Standard error	Z value	p_value
Intercept	1.081	0.490	-2.206	0.0274
Pedestrian's initial speed	0.920	0.109	-8.422	<0.001
Ego vehicle's initial speed	-0.277	0.017	15.906	<0.001
TTC threshold to activate the AEB	1.582	0.104	-15.145	<0.001
Occlusion time	-0.297	0.071	4.189	<0.001
Scenario (reference=scenario 1)				
Scenario 2	-1.868	0.152	12.321	<0.001
Scenario 3	-1.222	0.246	4.967	<0.001
AIC			2151	
BIC			2191	

5.4.2.3 Conclusions And Future Work

This study introduced an open-source approach by using the CARLA virtual simulator to evaluate the effectiveness of the AEB system under the occlusion conditions. The AEB control algorithm was developed in the virtual simulator. The evaluation was conducted by exploring the collision between a pedestrian crossing at the red light and a through vehicle, which is one of the most dangerous conditions at intersections. Three scenarios in which the pedestrian was occluded by a

stopping vehicle on either the left-turn or right-turn lane were generated for the evaluation. By considering different motion statuses of the ego vehicle and pedestrian, and the AEB controls, a total of 10,368 cases were generated in the simulation platform. Different measures including the percentage of cases with collisions, impact speed of the collision, and collision locations were adopted to extensively evaluate the effectiveness of AEB under the occlusion conditions. The results suggested that the AEB could still effectively avoid collisions between the ego vehicle and the crossing pedestrian. However, the effectiveness of the AEB would get reduced by the occlusion. The longer the pedestrian was occluded by the stopping vehicle, the more the effectiveness of AEB got reduced. The results also suggested that a larger TTC threshold to activate the AEB could improve the effectiveness. In addition, a logistic regression model was developed to explore the effects of other factors. The modeling results suggested that the pedestrian's and ego vehicle's speeds could have significant effects on the effectiveness of the AEB system. Furthermore, different effectiveness of AEB could be found in different occlusion scenarios.

The effectiveness of the AEB system has been evaluated extensively in this study if the pedestrian is occluded by a stopping vehicle. However, there are few limitations in this study. First, the study was built on the assumption that the sensors or motion controllers are 100% reliable and accurate, which omitted the potential impact of sensor errors or controller damages. Second, AEB system may vary from different manufactures by adopting different brake control parameters, and its impact on the AEB performance was not considered in this research. Furthermore, the current study could be extended by considering the scenarios in which the pedestrian is occluded by multiple moving vehicles. Also, since the study confirmed that the effectiveness of the AEB system is reduced under the occlusion conditions, it is important to explore the cooperative perception to reduce the occlusion through the connected vehicle technology.

5.5 Conclusions And Future Work

A Digital Twin framework that incorporates CV and pedestrian is proposed. The framework consists of the physical world, the digital world, and the data transmission between real digital spaces. The attributes of the CV and pedestrian, including their external state and internal state are specified. A sample architecture of the framework is also built. The architecture combines two platforms: Carla-Sumo and Unity for the digital twinning of CV and pedestrian, respectively. The Carla-Sumo platform serves as the driving simulator to simulate a CV environment, while a CAVE powered by Unity is used to provide the simulation environment for the pedestrian. The data transmission between different platforms is defined and the method of forming a closed-loop data transmission structure is discussed. To validate the effectiveness of the proposed framework, a case study that investigates V2P collision warning system under occlusion condition is conducted based on the presented architecture. The results show that the V2P warning enhanced safety compared to HDV and AV. The case study also demonstrates the benefits of the proposed CV and pedestrian inthe-loop framework. The proposed Digital Twin framework is expected to serve as a powerful testbed for future research on ITS technology to enhance CV and pedestrian safety and mobility. The future work of this research is to expand the framework to incorporate other traffic participants such as cyclists, motorists and traffic light, and more complex traffic scenarios. Also, more complex scenarios like V2P safety application considering driver behavior and preference, vehicle connectivity, pedestrian motion prediction, and automated driving algorithms could be investigated using the proposed Digital Twin framework.

CHAPTER 6: CONCLUSIONS

In this dissertation, Chapter 3 presents the development, validation, and integration of three AI-driven computer vision systems: 1) A.R.C.I.S, an automated framework for safety diagnosis utilizing multi-object detection and tracking algorithm for UAV videos. 2)N.M.E.D.S., A new framework with the ability to detect and predict the key points of vehicles and provide more precise vehicle occupying locations for traffic safety analysis. 3)D.V.E.D.S applied deep learning models to extract information related to drivers' visual environment from the Google Street View (GSV) images. Moreover, both A.R.C.I.S., and N.M.E.D.S. used deep learning techniques to process traffic videos collected by UAV or CCTV. Mask region convolutional neural network (R-CNN) is employed to improve the detection of vehicles in UAV videos. The detected vehicles and tracked by a channel and spatial reliability tracking algorithm, and vehicle trajectories are generated based on the tracking algorithm. Missing vehicles can be identified and tracked by identifying stationary vehicles and comparing the intersection of union (I.O.U.) between the detection and tracking results. Rotated bounding rectangles based on the pixel-to-pixel manner masks that are generated by mask R-CNN detection are introduced to obtain precise vehicle size and location data. In addition, the N.M.E.D.S. framework combined the Mask-RCNN bounding box detection and Occlusion-Net detection to detect vehicles' key points more accurately and efficiently ((i.e., right-front headlight, left-front headlight, right-back taillight, left-back taillight)) in a 3D view. Based on the vehicle trajectories from both systems, post-encroachment time (P.E.T.) is calculated for each conflict event at the pixel level. By comparing the P.E.T. values and the threshold, conflicts with the corresponding pixels in which the conflicts can be reported. Various conflict types: rear-end, head-on, sideswipe, and angle, can also be determined. Both systems could reveal significantly dangerous areas and have less bias than traditional

conflict identification with bounding boxes. The third system presents a new method to obtain drive view information (road geometric and vegetation information etc.) from GSV images using deep learning algorithms. Even though GSV image data accuracy is lower than the 3D data from the Digital Twins and LiDAR sensors, its data has a broad coverage at a low cost. The modeling results confirmed the importance of drivers' visual view information obtained by D.V.E.D.S. This system could have various use cases, such as providing road planners and engineers with clear suggestions to select appropriate countermeasures to enhance traffic safety.

Based on the drone video collected and processed by A.R.C.I.S at various locations, Chapter 4 describes the CitySim: a new trajectory dataset extracted from drone videos that aim to facilitate safety research was introduced. CitySim has vehicle interaction trajectories extracted from 1140- minutes of video recordings, which provide a large-scale naturalistic vehicle trajectory that covers a variety of locations, including basic freeway segments, freeway weaving segments, expressway segments, signalized intersections, stop-controlled intersections, and unique intersections without sign/signal control. The advantage of CitySim over other datasets is that it contains more critical safety events in quantity and severity and provides supporting scenarios for safety-oriented research. Furthermore, by providing accurate vehicle geometric representation in the form of rotated bounding boxes, CitySim can measure safety conflicts more accurately. In addition, CitySim provides digital twin features, including the 3D base maps and signal timings, which enables a more comprehensive testing environment for safety research, such as autonomous vehicle safety. To the best of our knowledge, CitySim is the first and largest open trajectory dataset bringing the digital twin concept by providing SUMO and Carla basemap. It leverages the advancement of virtual simulation for traffic safety.

Chapter 5 outlines the study done to investigate and complete the third research objective of this dissertation: the development of the framework of the next generation of pedestrian and connected vehicle in-the-loop digital twin simulation. The framework comprises the physical world, the digital twin world, and the real-time data transmission between those two worlds. The pedestrian and CV's outward and internal states and other characteristics are described. Carla-Sumo and Unity are combined into one architecture for the digital twinning of CV and pedestrians. The pedestrian simulation environment is provided by a CAVE driven by Unity, while the driving simulator for the CV environment is provided by the Carla-Sumo platform. The definition of data transmission between various platforms and the creation a closed-loop data transmission system are discussed. An investigation into a V2P collision warning system under occlusion conditions is done as part of a case study to verify the efficacy of the suggested framework. The findings demonstrate that, in comparison to HDV and AV, the V2P warning improves safety. The case study also illustrates how the suggested CV and pedestrian in the the-loop framework are advantageous. This research is the first to introduce the concept and technical detail of CV and pedestrian in-the-loop for a Digital Twin simulation. Furthermore, future studies on ITS technology to improve CV and pedestrian safety and mobility are anticipated to use the proposed Digital Twin framework as a potent testbed.

The increasing demand for computer vision analytics, digital twin systems, and AI in future smart cities presents both opportunities and challenges. This dissertation explores new traffic safety diagnostic solutions that utilize computer vision and digital twin technology. The study presents the creation, validation, and integration of three AI-driven computer vision systems and a digital twin framework aimed at improving traffic safety diagnostics. Which supported numerous research efforts of its developing team, the University of Central Florida Smart and

Safe Transportation Lab (UCF-SST). Examples of the generated research include crash risk evaluation [200, 201], trajectory prediction [201, 202],and computer vision-based surrogate safety measurement. And many other research paper [51-55, 64, 74, 79, 164, 196, 200-208]based on this dissertation has been published in top journals. Still, there are issues that need to be explored, such as the ability to process real-time data in computer vision systems, the accuracy of detection and tracking models, the precise construction of driver view models, and the scalability of computer vision system. The design and framework of the three computer vision systems will serve as a guide To address these challenges and further advance the field. In addition, we also created the first and largest drone-based vehicle trajectory dataset for safety-oriented research and digital twins, which will increase accessibility for other researchers to delve into digital twin research. Since the release of CitySim beta version in April 2022, the CitySim dataset has attracted international attention from many researchers worldwide. Over 62 top research institutes and over 100 researchers from the United States, China, Europe, the Middle East, and many other places have requested and applied this dataset in their research. We envision that over the next five years, the CitySim dataset will drive the use of trajectory data in various areas of traffic safety research and popularize the adoption of digital twin technology for traffic safety. The CitySim open dataset will provide a foundation for future research on vehicle trajectory digital twins and traffic safety. Our belief is that this dissertation will make a significant contribution to the improvement of global traffic safety.

REFERENCES

1. Chandra, R., et al., *Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms.* IEEE Robotics and Automation Letters, 2020. **5**(3): p. 4882-4890.
2. Lou, Y., et al. *Veri-wild: A large dataset and a new method for vehicle re-identification in the wild*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
3. You, T. and B. Han. *Traffic accident benchmark for causality recognition*. in *European Conference on Computer Vision*. 2020. Springer.
4. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks.* Advances in neural information processing systems, 2012. **25**: p. 1097-1105.
5. He, K.M., et al., *Mask R-CNN.* 2017 Ieee International Conference on Computer Vision (Iccv), 2017: p. 2980-2988.
6. Vaswani, A., et al. *Attention is all you need*. in *Advances in neural information processing systems*. 2017.
7. Wu, Y.N., et al., *Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data.* Transportation Research Part C-Emerging Technologies, 2018. **87**: p. 11-25.
8. Gu, X., et al., *Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas.* Accident Analysis and Prevention, 2019. **123**: p. 159-169.
9. Xing, L., et al., *Examining traffic conflicts of up stream toll plaza area using vehicles' trajectory data.* Accident Analysis and Prevention, 2019. **125**: p. 174-187.
10. Ahmed, M.M. and A. Ghasemzadeh, *The impacts of heavy rain on speed and headway Behaviors: An investigation using the SHRP2 naturalistic driving study data.* Transportation Research Part C-Emerging Technologies, 2018. **91**: p. 371-384.
11. Sayed, T., et al., *Feasibility of Computer Vision-Based Safety Evaluations Case Study of a Signalized Right-Turn Safety Treatment.* Transportation Research Record, 2012(2280): p. 18-27.
12. Zaki, M.H., et al., *Application of computer vision to diagnosis of pedestrian safety issues.* Transportation research record, 2013. **2393**(1): p. 75-84.
13. Fu, T., L. Miranda-Moreno, and N. Saunier, *Pedestrian Crosswalk Safety at Nonsignalized Crossings During Nighttime Use of Thermal Video Data and Surrogate Safety Measures.* Transportation Research Record, 2016(2586): p. 90-99.
14. Puri, A., *A survey of unmanned aerial vehicles (UAV) for traffic surveillance.* Department of computer science and engineering, University of South Florida, 2005: p. 1-29.
15. Puri, A., K. Valavanis, and M. Kontitsis. *Statistical profile generation for traffic monitoring using real-time UAV based video data*. in *2007 Mediterranean Conference on Control & Automation*. 2007. IEEE.
16. Ke, R.M., et al., *Real-Time Traffic Flow Parameter Estimation From UAV Video Based on Ensemble Classifier and Optical Flow.* Ieee Transactions on Intelligent Transportation Systems, 2019. **20**(1): p. 54-64.

17. Kanistras, K., et al. *A survey of unmanned aerial vehicles (UAVs) for traffic monitoring*. in *2013 International Conference on Unmanned Aircraft Systems (ICUAS)*. 2013. IEEE.
18. Khan, M.A., et al., *Unmanned aerial vehicle–based traffic analysis: Methodological framework for automated multivehicle trajectory extraction.* Transportation research record, 2017. **2626**(1): p. 25-33.
19. Coifman, B., et al. *Roadway traffic monitoring from an unmanned aerial vehicle*. in *IEE Proceedings-Intelligent Transport Systems*. 2006. IET.
20. HM, H.C., *Traffic safety facts (No. DOT HS 812 205)*. 2018, National Center for Statistics and Analysis.
21. Kim, J.K., et al., *Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis.* Accident Analysis and Prevention, 2008. **40**(5): p. 1695-1702.
22. Cai, Q., et al., *Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling.* Analytic Methods in Accident Research, 2018. **19**: p. 1-15.
23. Afghari, A.P., M.M. Haque, and S. Washington, *Applying fractional split model to examine the effects of roadway geometric and traffic characteristics on speeding behavior.* Traffic Injury Prevention, 2018. **19**(8): p. 860-866.
24. Jha, S., et al., *The multimodal driver monitoring database: A naturalistic corpus to study driver attention.* IEEE Transactions on Intelligent Transportation Systems, 2021. **23**(8): p. 10736-10752.
25. Zhao, Y., D. Ito, and K. Mizuno, *AEB effectiveness evaluation based on car-to-cyclist accident reconstructions using video of drive recorder.* Traffic injury prevention, 2019. **20**(1): p. 100-106.
26. Zhao, P.B. and C. Lee, *Assessing rear-end collision risk of cars and heavy vehicles on freeways using a surrogate safety measure.* Accident Analysis and Prevention, 2018. **113**: p. 149-158.
27. Naik, B., et al., *Weather impacts on single-vehicle truck crash injury severity.* Journal of safety research, 2016. **58**: p. 57-65.
28. Zhao, S. and A.J. Khattak, *Injury severity in crashes reported in proximity of rail crossings: The role of driver inattention.* Journal of Transportation Safety & Security, 2018. **10**(6): p. 507-524.
29. Administration, U.S.D.o.T.F.H., *Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data.* 2016.
30. Krajewski, R., et al., *The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems.* 2018 21st International Conference on Intelligent Transportation Systems (Itsc), 2018: p. 2118-2125.
31. Bock, J., et al., *The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections.* 2020 Ieee Intelligent Vehicles Symposium (Iv), 2020: p. 1929-1934.
32. Bhattacharyya, R., et al., *Modeling Human Driving Behavior Through Generative Adversarial Imitation Learning.* Ieee Transactions on Intelligent Transportation Systems, 2022.

33. Chen, H., et al., *Driver Behavior Analysis for Advanced Driver Assistance System.* Proceedings of 2018 Ieee 7th Data Driven Control and Learning Systems Conference (Ddcls), 2018: p. 492-497.
34. Huang, Z.Y., J.D. Wu, and C. Lv, *Driving Behavior Modeling Using Naturalistic Human Driving Data With Inverse Reinforcement Learning.* Ieee Transactions on Intelligent Transportation Systems, 2022. **23**(8): p. 10239-10251.
35. Yin, T., et al., *A novel gated recurrent unit network based on SVM and moth-flame optimization algorithm for behavior decision-making of autonomous vehicles.* IEEE access, 2021. **9**: p. 20410-20422.
36. Jin, H., et al., *Gauss mixture hidden Markov model to characterise and model discretionary lane-change behaviours for autonomous vehicles.* Iet Intelligent Transport Systems, 2020. **14**(5): p. 401-411.
37. Benterki, A., et al., *Prediction of Surrounding Vehicles Lane Change Intention Using Machine Learning.* Proceedings of the 2019 10th Ieee International Conference on Intelligent Data Acquisition and Advanced Computing Systems - Technology and Applications (Idaacs), Vol. 2, 2019: p. 839-843.
38. Prabhakaran, N. and M.S. Sudhakar, *Fuzzy curvilinear path optimization using fuzzy regression analysis for mid vehicle collision detection and avoidance system analyzed on NGSIM I-80 dataset (real-road scenarios).* Neural Computing & Applications, 2019. **31**(5): p. 1405-1423.
39. Xu, Z. *Analytical study of the lane-by-lane variation for lane-utility and traffic safety performance assessment using NGSIM data*. in *MATEC Web of Conferences*. 2017. EDP Sciences.
40. Liu, X.B., et al., *Optimizing the safety-efficiency balancing of automated vehicle car-following.* Accident Analysis and Prevention, 2020. **136**.
41. Xing, Y., C. Lv, and D.P. Cao, *Personalized Vehicle Trajectory Prediction Based on Joint Time-Series Modeling for Connected Vehicles.* Ieee Transactions on Vehicular Technology, 2020. **69**(2): p. 1341-1352.
42. Ghaleb, F.A., et al., *Improved vehicle positioning algorithm using enhanced innovation-based adaptive Kalman filter.* Pervasive and Mobile Computing, 2017. **40**: p. 139-155.
43. Lee, D., et al., *Design and field evaluation of cooperative adaptive cruise control with unconnected vehicle in the loop.* Transportation Research Part C-Emerging Technologies, 2021. **132**.
44. Lu, J., O. Grembek, and M. Hansen, *Learning the representation of surrogate safety measures to identify traffic conflict.* Accident Analysis & Prevention, 2022. **174**: p. 106755.
45. Reddy, M.K., et al. *Evaluating Vehicle Pedestrian Interaction at Unsignalised Junction Using Trajectory Data*. in *2022 IEEE Delhi Section Conference (DELCON)*. 2022. IEEE.
46. Wang, C., C. Xu, and Y. Dai, *A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data.* Accident Analysis & Prevention, 2019. **123**: p. 365-373.
47. Chen, P., et al., *Surrogate safety analysis of pedestrian-vehicle conflict at intersections using unmanned aerial vehicle videos.* Journal of advanced transportation, 2017. **2017**.
48. Punzo, V., M.T. Borzacchiello, and B. Ciuffo, *On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program*

data. Transportation Research Part C-Emerging Technologies, 2011. **19**(6): p. 1243-1262.
49. Shi, X., D. Zhao, H. Yao, X. Li, D. K. Hale, and A. Ghiasi, *Video-Based Trajectory Extraction with Deep Learning for High-Granularity Highway Simulation (HIGH-SIM).* Communications in Transportation Research, 2021. **Vol. 1, 2021, p. 100014.**
50. Li, L., et al., *Trajectory data-based traffic flow studies: A revisit.* Transportation Research Part C-Emerging Technologies, 2020. **114**: p. 225-240.
51. Wu, Y.N., et al., *Automated Safety Diagnosis Based on Unmanned Aerial Vehicle Video and Deep Learning Algorithm.* Transportation Research Record, 2020. **2674**(8): p. 350-359.
52. Abdel-Aty, M., et al., *Using closed-circuit television cameras to analyze traffic safety at intersections based on vehicle key points detection.* Accident Analysis & Prevention, 2022. **176**: p. 106794.
53. Cai, Q., et al., *Applying machine learning and google street view to explore effects of drivers' visual environment on traffic safety.* Transportation research part C: emerging technologies, 2022. **135**: p. 103541.
54. Zheng, O., et al., *CitySim: A Drone-Based Vehicle Trajectory Dataset for Safety Oriented Research and Digital Twins.* arXiv preprint arXiv:2208.11036, 2022.
55. Wang, Z., et al., *Towards Next Generation of Pedestrian and Connected Vehicle In-the-loop Research: A Digital Twin Simulation Framework.* arXiv preprint arXiv:2212.05090, 2022.
56. Cai, Q., et al., *Comparative analysis of zonal systems for macro-level crash modeling.* Journal of Safety Research, 2017. **61**: p. 157-166.
57. Cai, Q., et al., *Real-time crash prediction on expressways using deep generative models.* Transportation research part C: emerging technologies, 2020. **117**: p. 102697.
58. Wu, Y., M. Abdel-Aty, and J. Lee, *Crash risk analysis during fog conditions using real-time traffic data.* Accident Analysis & Prevention, 2018. **114**: p. 4-11.
59. Yuan, J., et al., *Real-time crash risk prediction using long short-term memory recurrent neural network.* Transportation research record, 2019. **2673**(4): p. 314-326.
60. Xu, C., et al., *Predicting crash likelihood and severity on freeways with real-time loop detector data.* Accident Analysis & Prevention, 2013. **57**: p. 30-39.
61. Zaki, M.H., T. Sayed, and S.E. Ibrahim, *Comprehensive safety diagnosis using automated video analysis: applications to an urban intersection in Edmonton, Alberta, Canada.* Transportation research record, 2016. **2601**(1): p. 138-152.
62. Roshandel, S., Z. Zheng, and S. Washington, *Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis.* Accident Analysis & Prevention, 2015. **79**: p. 198-211.
63. Lum, H. and J.A. Reagan, *Interactive highway safety design model: Accident predictive module.* Public Roads, 1995. **58**(3): p. 14-16.
64. Wu, Y., et al., *Developing a crash warning system for the bike lane area at intersections with connected vehicle technology.* Transportation research record, 2019. **2673**(4): p. 47-58.
65. Yamazaki, F., W. Liu, and T.T. Vu. *Vehicle extraction and speed detection from digital aerial images*. in *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*. 2008. IEEE.

66. Sayed, T., M.H. Zaki, and J. Autey, *Automated safety diagnosis of vehicle–bicycle interactions using computer vision analysis.* Safety science, 2013. **59**: p. 163-172.
67. Kim, E.J., et al., *Extracting Vehicle Trajectories Using Unmanned Aerial Vehicles in Congested Traffic Conditions.* Journal of Advanced Transportation, 2019.
68. Tang, T., et al., *Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining.* Sensors, 2017. **17**(2): p. 336.
69. Xu, Y., et al., *Car detection from low-altitude UAV imagery with the faster R-CNN.* Journal of Advanced Transportation, 2017. **2017**.
70. Ke, R., et al., *Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow.* IEEE Transactions on Intelligent Transportation Systems, 2018. **20**(1): p. 54-64.
71. Zhao, X., et al., *Automated Traffic Surveillance System with Aerial Camera Arrays Imagery: Macroscopic Data Collection with Vehicle Tracking.* Journal of Computing in Civil Engineering, 2017. **31**(3).
72. Liu, X., et al., *Simulation and evaluation of using unmanned aerial vehicle to detect low-volume road traffic incident*. 2015.
73. Lee, J., et al., *Examining the applicability of small quadcopter drone for traffic surveillance and roadway incident monitoring*. 2015.
74. Xing, L., et al., *Examining traffic conflicts of up stream toll plaza area using vehicles' trajectory data.* Accident Analysis & Prevention, 2019. **125**: p. 174-187.
75. Kim, E.-J., et al., *Extracting vehicle trajectories using unmanned aerial vehicles in congested traffic conditions.* Journal of Advanced Transportation, 2019. **2019**.
76. Chen, X., et al., *High-resolution vehicle trajectory extraction and denoising from aerial videos.* IEEE Transactions on Intelligent Transportation Systems, 2020. **22**(5): p. 3190-3202.
77. Guo, Y., T. Sayed, and M. Essa, *Real-time conflict-based Bayesian Tobit models for safety evaluation of signalized intersections.* Accident Analysis & Prevention, 2020. **144**: p. 105660.
78. Xie, K., et al., *Mining automatically extracted vehicle trajectory data for proactive safety analytics.* Transportation Research Part C-Emerging Technologies, 2019. **106**: p. 61-72.
79. Mahmoud, N., et al., *Vulnerable road users' crash hotspot identification on multi-lane arterial roads using estimated exposure and considering context classification.* Accident Analysis & Prevention, 2021. **159**: p. 106294.
80. Wu, Y., et al., *Automated safety diagnosis based on unmanned aerial vehicle video and deep learning algorithm.* Transportation research record, 2020. **2674**(8): p. 350-359.
81. Kocić, J., N. Jovičić, and V. Drndarević. *Sensors and sensor fusion in autonomous vehicles*. in *2018 26th Telecommunications Forum (TELFOR)*. 2018. IEEE.
82. Reddy, N.D., M. Vo, and S.G. Narasimhan. *Occlusion-net: 2d/3d occluded keypoint localization using graph networks*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
83. Wan, Y.W., Y. Huang, and B. Buckles, *Camera calibration and vehicle tracking: Highway traffic video analytics.* Transportation Research Part C-Emerging Technologies, 2014. **44**: p. 202-213.
84. Mooney, S.J., et al., *Use of Google Street View to assess environmental contributions to pedestrian injury.* American journal of public health, 2016. **106**(3): p. 462-469.

85. Li, X.J., et al., *A novel method for predicting and mapping the occurrence of sun glare using Google Street View.* Transportation Research Part C-Emerging Technologies, 2019. **106**: p. 132-144.
86. Tanprasert, T., et al., *Recognizing traffic black spots from street view images using environment-aware image processing and neural network.* IEEE Access, 2020. **8**: p. 121469-121478.
87. Gong, F.-Y., et al., *Mapping sky, tree, and building view factors of street canyons in a high-density urban environment.* Building and Environment, 2018. **134**: p. 155-167.
88. Middel, A., et al., *Urban form and composition of street canyons: A human-centric big data and deep learning approach.* Landscape and Urban Planning, 2019. **183**: p. 122-132.
89. Richards, D.R. and P.J. Edwards, *Quantifying street tree regulating ecosystem services using Google Street View.* Ecological Indicators, 2017. **77**: p. 31-40.
90. Li, X.J., et al., *Assessing street-level urban greenery using Google Street View and a modified green view index.* Urban Forestry & Urban Greening, 2015. **14**(3): p. 675-685.
91. Peesapati, L.N., M.P. Hunter, and M.O. Rodgers, *Evaluation of postencroachment time as surrogate for opposing left-turn crashes.* Transportation research record, 2013. **2386**(1): p. 42-51.
92. Zhan, W., et al., *Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps.* arXiv preprint arXiv:1910.03088, 2019.
93. Duret, A., C. Buisson, and N. Chiabaut, *Estimating individual speed-spacing relationship and assessing ability of Newell's car-following model to reproduce trajectories.* Transportation research record, 2008. **2088**(1): p. 188-197.
94. Hamdar, S.H., and H. S. Mahmassani, *Driver Car-Following Behavior: From Discrete Event Process to Continuous Set of Episodes.* 2008.
95. Grieves, M., *Digital twin: manufacturing excellence through virtual factory replication.* White paper, 2014. **1**: p. 1-7.
96. Howard, D. *The digital twin: Virtual validation in electronics development and design*. in *2019 Pan Pacific Microelectronics Symposium (Pan Pacific)*. 2019. IEEE.
97. Coraddu, A., et al., *Data-driven ship digital twin for estimating the speed loss caused by the marine fouling.* Ocean Engineering, 2019. **186**: p. 106063.
98. David, J., A. Lobov, and M. Lanz. *Leveraging digital twins for assisted learning of flexible manufacturing systems*. in *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*. 2018. IEEE.
99. Debroy, T., et al., *Building digital twins of 3D printing machines.* Scripta Materialia, 2017. **135**: p. 119-124.
100. Jo, S.-K., et al. *Smart livestock farms using digital twin: Feasibility study*. in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. 2018. IEEE.
101. Knapp, G., et al., *Building blocks for a digital twin of additive manufacturing.* Acta Materialia, 2017. **135**: p. 390-399.
102. Mohammadi, N. and J. Taylor. *Smart city digital twins. Proc*. in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2017.
103. Sivalingam, K., et al. *A review and methodology development for remaining useful life prediction of offshore fixed and floating wind turbine power converter with digital twin*

technology perspective. in *2018 2nd international conference on green energy and applications (ICGEA)*. 2018. IEEE.
104. Pargmann, H., D. Euhausen, and R. Faber. *Intelligent big data processing for wind farm monitoring and analysis based on cloud-technologies and digital twins: A quantitative approach*. in *2018 IEEE 3rd international conference on cloud computing and big data analysis (ICCCBDA)*. 2018. IEEE.
105. NHTSA. *Pedestrian Safety*.
106. Concil, N.S. *Injury Fact*. Available from: https://injuryfacts.nsc.org/motor-vehicle/road-users/pedestrians/.
107. Yue, L., et al., *In-depth approach for identifying crash causation patterns and its implications for pedestrian crash prevention.* Journal of safety research, 2020. **73**: p. 119-132.
108. Zegeer, C.V. and M. Bushell, *Pedestrian crash trends and potential countermeasures from around the world.* Accident Analysis & Prevention, 2012. **44**(1): p. 3-11.
109. Hussain, Q., et al., *The relationship between impact speed and the probability of pedestrian fatality during a vehicle-pedestrian crash: a systematic review and meta-analysis.* Accident Analysis & Prevention, 2019. **129**: p. 241-249.
110. Bertulis, T. and D.M. Dulaski, *Driver approach speed and its impact on driver yielding to pedestrian behavior at unsignalized crosswalks.* Transportation Research Record, 2014. **2464**(1): p. 46-51.
111. Schroeder, B.J. and N.M. Rouphail, *Event-based modeling of driver yielding behavior at unsignalized crosswalks.* Journal of transportation engineering, 2011. **137**(7): p. 455-465.
112. Katz, A., D. Zaidel, and A. Elgrishi, *An experimental study of driver and pedestrian interaction during the crossing conflict.* Human Factors, 1975. **17**(5): p. 514-527.
113. Habibovic, A., et al., *Driver behavior in car-to-pedestrian incidents: An application of the Driving Reliability and Error Analysis Method (DREAM).* Accident Analysis & Prevention, 2013. **50**: p. 554-565.
114. Koh, P., Y. Wong, and P. Chandrasekar, *Safety evaluation of pedestrian behaviour and violations at signalised pedestrian crossings.* Safety science, 2014. **70**: p. 143-152.
115. Rosen, E. *Autonomous emergency braking for vulnerable road users*. in *Proceedings of IRCOBI conference*. 2013.
116. Jeppsson, H. and N. Lubbe, *Simulating automated emergency braking with and without Torricelli vacuum emergency braking for cyclists: effect of brake deceleration and sensor field-of-view on accidents, injuries and fatalities.* Accident Analysis & Prevention, 2020. **142**: p. 105538.
117. Rosén, E., et al., *Pedestrian injury mitigation by autonomous braking.* Accident Analysis & Prevention, 2010. **42**(6): p. 1949-1957.
118. Yue, L., et al., *An augmentation function for active pedestrian safety system based on crash risk evaluation.* IEEE transactions on vehicular technology, 2020. **69**(11): p. 12459-12469.
119. Haus, S.H., R. Sherony, and H.C. Gabler, *Estimated benefit of automated emergency braking systems for vehicle–pedestrian crashes in the United States.* Traffic injury prevention, 2019. **20**(sup1): p. S171-S176.
120. Fridman, L., et al., *MIT advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation.* IEEE Access, 2019. **7**: p. 102021-102038.

121. Shangguan, Q., et al., *An integrated methodology for real-time driving risk status prediction using naturalistic driving data.* Accident Analysis & Prevention, 2021. **156**: p. 106122.
122. Seacrist, T., et al., *Efficacy of automatic emergency braking among risky drivers using counterfactual simulations from the SHRP 2 naturalistic driving study.* Safety science, 2020. **128**: p. 104746.
123. Feng, S., et al., *Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment.* Nature communications, 2021. **12**(1): p. 1-14.
124. Koglbauer, I., et al., *Autonomous emergency braking systems adapted to snowy road conditions improve drivers' perceived safety and trust.* Traffic injury prevention, 2018. **19**(3): p. 332-337.
125. Dosovitskiy, A., et al. *CARLA: An open urban driving simulator*. in *Conference on robot learning*. 2017. PMLR.
126. Yan, X., et al., *Distributionally Consistent Simulation of Naturalistic Driving Environment for Autonomous Vehicle Testing.* arXiv preprint arXiv:2101.02828, 2021.
127. Xu, R., et al., *OpenCDA: An Open Cooperative Driving Automation FrameworkIntegrated with Co-Simulation.* arXiv preprint arXiv:2107.06260, 2021.
128. Lee, T.-K., et al. *Building a V2X Simulation Framework for Future Autonomous Driving*. in *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. 2019. IEEE.
129. Rahman, M.S. and M. Abdel-Aty, *Longitudinal safety evaluation of connected vehicles' platooning on expressways.* Accid Anal Prev, 2018. **117**: p. 381-391.
130. Papadoulis, A., M. Quddus, and M. Imprialou, *Evaluating the safety impact of connected and autonomous vehicles on motorways.* Accident Analysis & Prevention, 2019. **124**: p. 12-22.
131. Xie, Y.C., et al., *Collaborative merging strategy for freeway ramp operations in a connected and autonomous vehicles environment.* Journal of Intelligent Transportation Systems, 2017. **21**(2): p. 136-147.
132. Lee, J. and B. Park, *Development and Evaluation of a Cooperative Vehicle Intersection Control Algorithm Under the Connected Vehicles Environment.* Ieee Transactions on Intelligent Transportation Systems, 2012. **13**(1): p. 81-90.
133. Chen, S.K., et al., *Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles.* Computer-Aided Civil and Infrastructure Engineering, 2021. **36**(7): p. 838-857.
134. Xu, B., et al., *Cooperative method of traffic signal optimization and speed control of connected vehicles at isolated intersections.* IEEE Transactions on Intelligent Transportation Systems, 2018. **20**(4): p. 1390-1403.
135. Jin, Q., et al., *Advanced Intersection Management for Connected Vehicles Using a Multi-Agent Systems Approach.* 2012 Ieee Intelligent Vehicles Symposium (Iv), 2012: p. 932-937.
136. Liao, X., et al., *A game theory based ramp merging strategy for connected and automated vehicles in the mixed traffic: A unity-sumo integrated platform.* arXiv preprint arXiv:2101.11237, 2021.
137. Kang, Y., H. Yin, and C. Berger, *Test Your Self-Driving Algorithm: An Overview of Publicly Available Driving Datasets and Virtual Testing Environments.* Ieee Transactions on Intelligent Vehicles, 2019. **4**(2): p. 171-185.

138. Bai, Z., et al., *Cyber mobility mirror: Deep learning-based real-time 3d object perception and reconstruction using roadside lidar.* arXiv preprint arXiv:2202.13505, 2022.
139. Xu, R., et al. *OpenCDA: an open cooperative driving automation framework integrated with co-simulation*. in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. 2021. IEEE.
140. Yue, L., et al., *Influence of pedestrian-to-vehicle technology on drivers' response and safety benefits considering pre-crash conditions.* Transportation research part F: traffic psychology and behaviour, 2020. **73**: p. 50-65.
141. Yue, L., et al., *Assessment of the safety benefits of vehicles' advanced driver assistance, connectivity and low level automation systems.* Accident Analysis & Prevention, 2018. **117**: p. 55-64.
142. Yang, G., et al., *Connected vehicle real-time traveler information messages for freeway speed harmonization under adverse weather conditions: Trajectory level analysis using driving simulator.* Accident Analysis & Prevention, 2020. **146**: p. 105707.
143. Yue, L., M. Abdel-Aty, and Z. Wang, *Effects of connected and autonomous vehicle merging behavior on mainline human-driven vehicle.* Journal of Intelligent and Connected Vehicles, 2022. **5**(3): p. 36-45.
144. Heesen, M., et al. *Investigation of cooperative driving behaviour during lane change in a multi-driver simulation environment*. in *Human Factors and Ergonomics Society (HFES) Europe Chapter Conference Touluse*. 2012.
145. Zhao, X., et al., *Co-simulation platform for modeling and evaluating connected and automated vehicles and human behavior in mixed traffic.* SAE International Journal of Connected and Automated Vehicles, 2022. **5**(4).
146. Khosravi, S., et al., *Assistive system to improve pedestrians' safety and mobility in a connected vehicle technology environment.* Transportation research record, 2018. **2672**(19): p. 145-156.
147. He, S., J. Li, and T.Z. Qiu, *Vehicle-to-pedestrian communication modeling and collision avoiding method in connected vehicle environment.* Transportation Research Record, 2017. **2621**(1): p. 21-30.
148. Deb, S., et al., *Efficacy of virtual reality in pedestrian safety research.* Applied ergonomics, 2017. **65**: p. 449-460.
149. Lobjois, R. and V. Cavallo, *Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task.* Accident analysis & prevention, 2007. **39**(5): p. 934-943.
150. Jiang, Y., et al., *Acting together: Joint pedestrian road crossing in an immersive virtual environment.* ACM Transactions on Applied Perception (TAP), 2018. **15**(2): p. 1-13.
151. Meir, A., Y. Parmet, and T. Oron-Gilad, *Towards understanding child-pedestrians' hazard perception abilities in a mixed reality dynamic environment.* Transportation research part F: traffic psychology and behaviour, 2013. **20**: p. 90-107.
152. Cruz-Neira, C., D.J. Sandin, and T.A. DeFanti. *Surround-screen projection-based virtual reality: the design and implementation of the CAVE*. in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. 1993.
153. Kreimeier, J., et al. *Initial evaluation of different types of virtual reality locomotion towards a pedestrian simulator for urban and transportation planning*. in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020.

154. Doric, I., et al. *A novel approach for researching crossing behavior and risk acceptance: The pedestrian simulator*. in *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*. 2016.
155. Wang, Z., et al., *Mobility digital twin: Concept, architecture, case study, and future challenges.* IEEE Internet of Things Journal, 2022. **9**(18): p. 17452-17467.
156. Hu, Z., et al., *Review and perspectives on driver digital twin and its enabling technologies for intelligent vehicles.* IEEE Transactions on Intelligent Vehicles, 2022.
157. Lv, Z., et al., *Deep learning for security in digital twins of cooperative intelligent transportation systems.* IEEE transactions on intelligent transportation systems, 2021. **23**(9): p. 16666-16675.
158. Mallick, S. *Object Tracking using OpenCV (C++/Python)*. 2019.
159. Xu, Y., et al., *An enhanced Viola-Jones vehicle detection method from unmanned aerial vehicles imagery.* IEEE Transactions on Intelligent Transportation Systems, 2016. **18**(7): p. 1845-1856.
160. Li, Y., et al., *Deep active learning for object detection.* Information Sciences, 2021. **579**: p. 418-433.
161. Yu, W.P., et al., *Consistency-based Active Learning for Object Detection.* 2022 Ieee/Cvf Conference on Computer Vision and Pattern Recognition Workshops (Cvprw 2022), 2022: p. 3950-3959.
162. Gulli, A. and S. Pal, *Deep learning with Keras*. 2017: Packt Publishing Ltd.
163. Li, Y., et al., *Pedestrian trajectory prediction combining probabilistic reasoning and sequence learning.* IEEE Transactions on Intelligent Vehicles, 2020. **5**(3): p. 461-474.
164. Zheng, O., *Developing a traffic safety diagnostics system for unmanned aerial vehicles usingdeep learning algorithms.* 2019.
165. Zhang, S., et al., *Prediction of pedestrian crossing intentions at intersections based on long short-term memory recurrent neural network.* Transportation research record, 2020. **2674**(4): p. 57-65.
166. Das, S., P.S.R. Kishore, and U. Bhattacharya, *An End-to-End Framework for Unsupervised Pose Estimation of Occluded Pedestrians.* arXiv preprint arXiv:2002.06429, 2020.
167. Abdel-Aty, M., et al., *Safety and operational impact of connected vehicles' lane configuration on freeway facilities with managed lanes.* Accident Analysis & Prevention, 2020. **144**: p. 105616.
168. Wu, Y., et al., *Combined connected vehicles and variable speed limit strategies to reduce rear-end crash risk under fog conditions.* Journal of Intelligent Transportation Systems, 2020. **24**(5): p. 494-513.
169. Yang, J., et al., *Can you see green? Assessing the visibility of urban forests in cities.* Landscape and Urban Planning, 2009. **91**(2): p. 97-104.
170. Yuxin Wu, A.K., Francisco Massa,Wan-Yen Lo,Ross Girshick, *Detectron2*. 2019: https://github.com/facebookresearch/detectron2.
171. Godard, C., et al., *Digging Into Self-Supervised Monocular Depth Estimation.* 2019 Ieee/Cvf International Conference on Computer Vision (Iccv 2019), 2019: p. 3827-3837.
172. Edquist, J., C.M. Rudin-Brown, and M.G. Lenne, *The effects of on-street parking and road environment visual complexity on travel speed and reaction time.* Accident Analysis and Prevention, 2012. **45**: p. 759-765.

173. Atombo, C., et al., *Investigating the motivational factors influencing drivers intentions to unsafe driving behaviours: Speeding and overtaking violations.* Transportation Research Part F-Traffic Psychology and Behaviour, 2016. **43**: p. 104-121.
174. Marshall, W.E., N. Coppola, and Y. Golombek, *Urban clear zones, street trees, and road safety.* Research in Transportation Business and Management, 2018. **29**: p. 136-143.
175. Abdel-Aty, M. and K. Haleem, *Analyzing angle crashes at unsignalized intersections using machine learning techniques.* Accident Analysis and Prevention, 2011. **43**(1): p. 461-470.
176. Cai, Q., et al., *Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data.* Transportation Research Part a-Policy and Practice, 2019. **127**: p. 71-85.
177. Cai, Q., et al., *Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models.* Accident Analysis and Prevention, 2016. **93**: p. 14-22.
178. Rahman, M.S., et al., *Applying machine learning approaches to analyze the vulnerable road-users' crashes at statewide traffic analysis zones.* Journal of Safety Research, 2019. **70**: p. 275-288.
179. Zhang, X., S.T. Waller, and P. Jiang, *An ensemble machine learning-based modeling framework for analysis of traffic crash frequency.* Computer-Aided Civil and Infrastructure Engineering, 2020. **35**(3): p. 258-276.
180. Strauss, J., L.F. Miranda-Moreno, and P. Morency, *Cyclist activity and injury risk analysis at signalized intersections: A Bayesian modelling approach.* Accident Analysis and Prevention, 2013. **59**: p. 9-17.
181. Friedman, J.H., *Stochastic gradient boosting.* Computational statistics & data analysis, 2002. **38**(4): p. 367-378.
182. Parsa, A.B., et al., *Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis.* Accident Analysis & Prevention, 2020. **136**: p. 105405.
183. fdot, *FDOT context classification guide*. FDOT context classification guide: Transportation FDo.
184. Tractinsky, N., E.S. Ram, and D. Shinar, *To call or not to call-That is the question (while driving).* Accident Analysis and Prevention, 2013. **56**: p. 59-70.
185. Calvi, A., *Does Roadside Vegetation Affect Driving Performance? Driving Simulator Study on the Effects of Trees on Drivers' Speed and Lateral Position.* Transportation Research Record, 2015(2518): p. 1-8.
186. Yang, D.G., et al., *Intelligent and connected vehicles: Current status and future perspectives.* Science China-Technological Sciences, 2018. **61**(10): p. 1446-1471.
187. Chen, L., et al., *Milestones in autonomous driving and intelligent vehicles: Survey of surveys.* IEEE Transactions on Intelligent Vehicles, 2022.
188. Glaessgen, E. and D. Stargel. *The digital twin paradigm for future NASA and US Air Force vehicles*. in *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*. 2012.
189. Abdelraouf, A., M. Abdel-Aty, and J. Yuan, *Utilizing attention-based multi-encoder-decoder neural networks for freeway traffic speed prediction.* IEEE Transactions on Intelligent Transportation Systems, 2021. **23**(8): p. 11960-11969.

190. Ramirez, J.M., et al., *Inferring drivers' visual focus attention through head-mounted inertial sensors.* IEEE Access, 2019. **7**: p. 185422-185432.
191. Liu, Y., et al., *Vision-cloud data fusion for ADAS: A lane change prediction case study.* IEEE Transactions on Intelligent Vehicles, 2021. **7**(2): p. 210-220.
192. Zhao, B., et al., *A Novel Prediction Algorithm of Pedestrian Activity Region for Intelligent Vehicle Collision Avoidance System.* IEEE Transactions on Intelligent Vehicles, 2022.
193. Wang, Z., et al., *Investigating the Effects of Human-Machine Interface on Cooperative Driving at Non-Signalized Intersection Using an Integrated Co-Simulation Platform.* Available at SSRN 4230161, 2022.
194. Lopez, P.A., et al. *Microscopic traffic simulation using sumo*. in *2018 21st international conference on intelligent transportation systems (ITSC)*. 2018. IEEE.
195. Gruber, M., et al. *The effect of P-AEB system parameters on the effectiveness for real world pedestrian accidents*. in *Proceedings of the 26th ESV Conference Proceedings*. 2019.
196. Abdel-Aty, M., et al., *Evaluation of automated emergency braking system's avoidance of pedestrian crashes at intersections under occluded conditions within a virtual simulator.* Accident Analysis & Prevention, 2022. **176**: p. 106797.
197. Gruber, M., et al. *The effect of P-AEB system parameters on the effectiveness for real world pedestrian accidents*. in *Proceedings of the 26th ESV Conference Proceedings*. 2019. NHTSA Washington, DC.
198. Schachner, M., et al., *Development and evaluation of potential accident scenarios involving pedestrians and AEB-equipped vehicles to demonstrate the efficiency of an enhanced open-source simulation framework.* Accident Analysis & Prevention, 2020. **148**: p. 105831.
199. Almodfer, R., et al., *Pedestrian Crossing Speed Patterns and running frequency analysis at a non-signalized marked crosswalk: Quantitative and qualitative approaches.* Sustainable cities and society, 2017. **34**: p. 183-192.
200. Zhang, J., et al., *Enhanced index of risk assessment of lane change on expressway weaving segments: A case study of an expressway in China.* Accid Anal Prev, 2023. **180**: p. 106909.
201. Islam, Z., M. Abdel-Aty, A. Goswamy, A. Abdelraouf, and O. Zheng., *Modelling the Relationship Between Post Enchroachment Time and Signal Timings Using UAV Video Data.* 2022.
202. Abdelraouf Amr, M.A.-A., Z. Wang, and O. Zheng, *Trajectory Prediction for Vehicle Conflict Identification at Intersections Using Sequence-to-Sequence Recurrent Neural Networks.* arXiv preprint arXiv:2210.08009, 2022.
203. Yue, L., et al., *Understanding e-bicycle overtaking strategy: insights from inverse reinforcement learning modelling.* Transportmetrica A: Transport Science, 2023: p. 1-27.
204. Song, P., et al., *Addressing unobserved heterogeneity at road user level for the analysis of conflict risk at tunnel toll plaza: A correlated grouped random parameters logit approach with heterogeneity in means.* Analytic methods in accident research, 2022. **36**: p. 100243.
205. Anisha, A.M., et al., *Automated Vehicle to Vehicle Conflict Analysis at Signalized Intersections by Camera and LiDAR Sensor Fusion.* Transportation Research Record, 2022: p. 03611981221128806.

206. Zhang, S., et al., *Pedestrian crossing intention prediction at red-light using pose estimation.* IEEE Transactions on Intelligent Transportation Systems, 2021. **23**(3): p. 2331-2339.
207. Zhang, S., et al., *Modeling pedestrians' near-accident events at signalized intersections using gated recurrent unit (GRU).* Accident Analysis & Prevention, 2020. **148**: p. 105844.
208. Xing, L., et al., *Comparison of different models for evaluating vehicle collision risks at upstream diverging area of toll plaza.* Accident Analysis & Prevention, 2020. **135**: p. 105343.