CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Final Report #73
November 2023

Facilitating UAV Application for CAV Deployment

Shuya Zong
Yujie Li
Sikai Chen
Shaoshuai Mou
Samuel Labi

PURDUE UNIVERSITY

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Report No. 73
Project Start Date: January 2022
Project End Date: September 2023

November 2023

Facilitating UAV Application for CAV Deployment

Shuya Zong
Graduate Researcher

Yujie Li
Graduate Researcher

Sikai Chen
Visiting Assistant Professor

Shaoshuai Mou
Associate Professor

Samuel Labi
Professor

Purdue University

ACKNOWLEDGEMENTS AND DISCLAIMER

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant No. 69A3551747105 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The researchers are grateful to cost sharing collaborators including the Indiana Department of Transportation (Drs. Barry Partridge and Samy Noureldin), Delft University (Dr. Gonçalo Homem de Almeida Correia), Indian Institute of Technology, Tirupati (Dr. Krishna Prapoorna), King Khalid University (Dr. Saeed Alqadhi), the National University of Singapore (Dr. Ghim Ping Ong) and Nanyang Technological University (Dr. Feng Zhu). The authors are also grateful to the Purdue Black & Gold Autonomous Vehicle Racing Team, Ms. Ajuna John, and Mr. Jelin Sabu, for their support and assistance in various ways. The institutional support provided by Center for Innovation in Control, Optimization, and Networks (ICON), and the Autonomous and Connected Systems (ACS) initiatives at Purdue University's College of Engineering, is also acknowledged. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Suggested APA Format Citation:

Zong, S., Li, Y., Chen, S., Mou, S., Labi, S. (2023). Facilitating UAV application for CAV deployment, CCAT Report #73, The Center for Connected and Automated Transportation, Purdue University, West Lafayette, IN.
Cover Image: Dollar Gill, unsplash.com/photos/clear-glass-ppanel-XHKuDcyAbKo

Contacts

For more information:

Samuel Labi, Ph.D.
550 Stadium Mall Drive
HAMP G167B
Phone: (765) 494-5926
Email: labi@purdue.edu

CCAT
University of Michigan Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48152

uumtri-ccat@umich.edu
(734) 763-2498
www.ccat.umtri.umich.edu

Technical Report Documentation Page

1. Report No. 73	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Facilitating UAV application for CAV deployment		**5. Report Date** November 2023
		6. Performing Organization Code N/A
7. Author(s) Shuya Zong, Yujie Li, Sikai Chen, Shaoshuai Mou, Samuel Labi		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Center for Connected and Automated Transportation, Purdue University, 550 Stadium Mall Drive, W. Lafayette, IN 47907; and Univ. of Michigan Ann Arbor, 2901 Baxter Rd, Ann Arbor, MI 48109		**10. Work Unit No.**
		11. Contract or Grant No. Contract No. 69A3551747105
12. Sponsoring Agency Name and Address U.S. Department of Transportation, Office of the Asst. Secretary for Research & Tech., 1200 New Jersey Ave., SE, Washington, DC 20590		**13. Type of Report and Period Covered:** Final rep., Jan 2022-Sept 2023
		14. Sponsoring Agency Code: OST-R

15. Supplementary Notes

Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.

16. Abstract

Connectivity-equipped UAVs offer communication opportunities within CAV ecosystems. This study was motivated by challenges in UAV-based monitoring: (a) at intersections where road user trajectories could be irregular, leading to traffic conflicts or collisions, and (b) during inclement weather where video images are often corrupted by falling rain streaks thus impairing the integrity of the image. In response to these challenges, the study developed and demonstrated the efficacy of (a) an algorithm for UAVs to predict vehicle trajectories at intersections and (b) a 2-stage self-supervised algorithm for enhancing the quality of UAV-sourced images. Both parts of the study highlight the UAV potential and challenges for advanced traffic management in the prospective era of CAV operations. There are some potential practical benefits of this research. First, an enhanced UAV-CAV data domain can help road agencies to improve the reliability of their traffic safety risk assessments and vehicle trajectory tracking. Further, from perspectives of systems control, the study products could enhance CAV operations by providing reliable information for safe and efficient trajectory planning and control at urban intersections.

17. Key Words Unmanned aerial vehicles, De-raining, Intersection monitoring, Trajectories, Video image quality, V2X	**18. Distribution Statement** No restrictions.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 65	**22. Price**

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

CAV	Connected and Automated Vehicle
CCA	Cooperative Collision Avoidance
CNN	Convolutional Neural Network
DBT	Detection-Based Tracking
DFT	Detection-Free Tracking
FCW	Forward Collision Warning
IIHS	Insurance Institute for Highway Safety
LDW	Lane Departure Warning
MOE	Measure of Effectiveness
MOT	Multi-Object Tracking
MOTA	Multi-Object Tracking Accuracy
N2N	Noise2Noise
PDS	Pedestrian Detection System
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index
TTC	Time-to-Collision
TTI	Travel Time Index
UAV	Unmanned Aerial Vehicles
V2I	Vehicle-to-Infrastructure
V2N	Vehicle-to-Network
V2P	Vehicle-to-Pedestrians
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything

LIST OF COMMONLY USED TERMS

Traffic parameters	Quantitative measures and characteristics that describe the traffic flow.
Clean image/frame	A video frame/image used as the ground truth. A clean image/frame is without any type of noise.
Denoise/derain	Remove noise/rain streaks from an image/video frame.
Ego motion	Motion of the UAV due to wind action.
Ground truth	Information that is known to be real or true, provided by direct observation and measurement.
Rainy scene	A video frame/image with rain noise.
Self-supervised learning	A group of machine learning technics that learns from unlabeled sample data.
Spatial noise	Noise that is not related to the movement of objects, existing in an individual frame of a video sequence.
Spatial-temporal denoise	A video denoising methodology that deals with both spatial and temporal noise.
Temporal noise	Noise existing between frames of a video sequence; often caused by fast-moving objects.

CHAPTER 1 INTRODUCTION

1.1 Study Background

Traffic monitoring using video has recently drawn increasing attention due to significant advances in image and video technologies. Traffic monitoring involves multiple tasks such as road user classification, vehicle counting, and crash risk assessment. The quality and reliability of traffic monitoring videos play a vital role toward reliable site diagnostics for promoting transportation safety. The videos can be used for road-use enforcement, pre-crash evaluation for safety studies, crash risk assessments, and liability evaluations to enhance the safety of infrastructure. The safety of road users is considered a key indicator of the social pillar of sustainable development (Jeon and Amekudzi, 2005; Timmermans and Beroggi, 2000), and public safety has been identified as an important indicator of infrastructure or transportation sustainability (Shen et al., 2011; Feizi et al., 2020).

The social costs of crashes are immense, as families experience great pain and suffering and, in some cases, loss of family income when they lose a loved one through an accident. Road traffic crashes cause 1.3 million deaths annually (World Health Organization, 2018) and over 20 million people suffer non-fatal injuries, with many incurring a disability because of their injury (Center for Disease Control and Prevention, 2020). Traffic crash impacts extend beyond the social to the economic pillar of sustainable development, particularly when safety costs are converted into dollars using unit crash costs (Blincoe et al., 2014). The global economic cost is immense: over 580 billion US dollars annually, costing most countries as much as 3% of their gross domestic product (Peden, 2005).

Of the various elements of highway horizontal design, intersections are considered particularly critical for traffic monitoring as they involve mixed traffic flows (multiple vehicles, pedestrians, motorcycles, or trucks, etc.). According to the Federal Highway Administration (FHWA), over 50 percent of all fatal and injury crashes occur at or near intersections (Federal Highway Administration, 2021). Consequently, transportation agencies including the National Highway Transportation Safety Administration, Federal Highway Administration, and the Institute of Transportation Engineers continue to support the development of safety countermeasures to reduce crash risk at intersections.

The FHWA continues to sponsor the investigation of crash causation factors and the evaluation of alternative intersection designs that facilitate the safe movement of pedestrians and bicyclists. Recently, the FHWA reiterated the encouragement of edge computing platforms to facilitate real-time actions (detection of traffic events and subsequent decision-making) to enhance safe operations at signal-controlled intersections (Federal Highway Administration, 2021).

An unmanned aerial vehicle (UAV), commonly known as a drone, is a small aircraft devoid of human pilot, crew, or passengers. Originally developed for military missions considered inaccessible or too hazardous for humans, UAV applications for non-military use has mushroomed, a trend accelerated by their declining costs and increasing technical efficiency. The civilian applications for UAVs include aerial photography and videography for various purposes including the monitoring of natural or man-made resources or situations (roadway infrastructure condition

or operations, forest management, natural and man-made disasters, law enforcement, and product deliveries.

In the context of connected and automated vehicles (CAVs), connectivity-equipped UAVs offer a valuable but largely untapped dimension of communication opportunities within the CAV ecosystem. Compared to static surveillance cameras typically mounted on fixed infrastructure (traffic signal posts and sign gantries), UAVs can acquire aerial data on the operation conditions (location, direction, speed) of ground vehicles and other road users quickly and cost-effectively. Further, UAVs provide greater efficiency due to their wide visual field of view, versatile camera angles, and pronounced movement flexibility in all three cartesian dimensions. In addition, UAVs can play a critical connectivity role by serving as a hub to facilitate communications among roadway entities (vehicles, infrastructure, and pedestrians).

For these reasons, UAV-CAV monitoring networks can be useful in a gamut of traffic management applications including monitoring and assessment of crash risk, and safety hazard identification and mitigation in real time. This could foster design of safer facilities and promote safe road-user behavior. It is expected that unmanned aerial vehicles (UAVs) will play a vital role in traffic monitoring. As indicated by Figure 1.1, the market size of UAVs is expanding significantly.

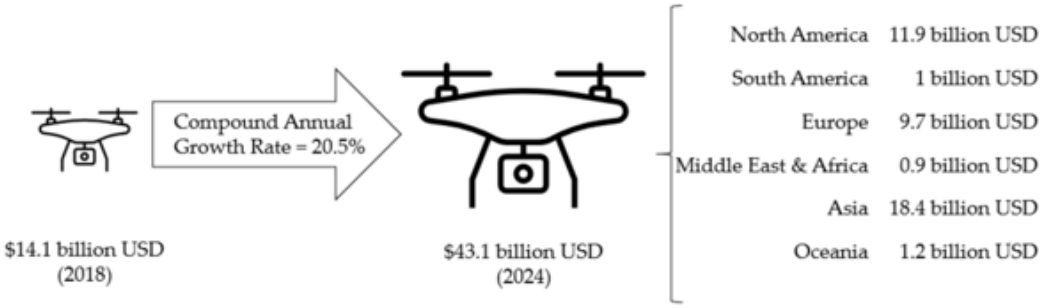


Figure 1.1 The global UAV market size and forecast (2018-2024) (Gupta et al., 2021).

In the application area of traffic monitoring, UAVs have two advantages: First, UAVs are portable, flexible, and robust. In contrast, traditional video data collection by land-based cameras mounted on tall physical structures has several limitations including restrictions on the field of view posed by the height of the camera and camera tilt angle. These could impair accuracy in tracking the trajectories of the road vehicles being monitored. Also, the time-consuming and labor-intensive installation process of mounting cameras on tall buildings prohibits timely and cost-effective ground-based traffic monitoring. UAVs offer a convenient means to address these limitations as it is possible to dispatch them easily and quickly to the site of interest and to adjust their spatial locations and camera positions. For automated vehicles, onboard sensors that use technologies such as cameras and lidar suffer from problems of limited coverage range and occlusion. The vehicle is limited not only qualitatively (in terms of the precision and the richness of the delivered information) but also quantitatively (in terms of the range of its sensors) (Ammoun and Nashashibi, 2009). Onboard sensors often fail to detect persons and objects blocked by trees, vehicles, building corners, and other obstacles, and have difficulty in sensing road users that are not in the same lane/direction as them. Under these conditions, drivers may not easily notice pedestrians and other vulnerable road users in a timely manner. Sensor fusion can address this

problem to a limited extent (Gelbal et al., 2017). In addition, the use of UAVs can help overcome some of these limitations because UAVs offer a global bird-eye-view that helps to generate comprehensive telemetric data on the intersection road user and their movements.

The second advantage of UAVs arises from the realization that it is still challenging to develop a large-scale ground-based vehicle-to-everything (V2X) network at the current time and in the near future. The mobile entities represent nodes in a network and communicate directly with each other or with roadside infrastructure. The resulting information network is termed vehicle-to-infrastructure (V2I), vehicle-to-vehicle (V2V), or vehicle-to-pedestrians (V2P) networks. If the communication is with a data center or information technology network, then the network acquires a vehicle-to-network (V2N) designation. Vehicle-to-everything (V2X) represents a combination of all these communication types. Cooperative V2X communications are intended to support a variety of use cases in risk detection including do-not-pass warning, forward-collision warning, parking discovery, queue-ahead warning, curve speed warning, optimal speed advisory, and other contexts that enhance traffic safety and efficiency (Dong et al., 2021). However, the major drawback of a V2X network is that its effectiveness hinges on the number of vehicles/facilities that are equipped with communication capabilities because non-equipped vehicles are completely "invisible" to other vehicles.

In addition, published research suggests that dedicated short-range communications or DSRC (a major data transfer technology used in V2X networks) is often plagued with issues of reliability, efficiency, and productivity, particularly at high traffic volumes (Kiela et al., 2020). Also, the dynamic nature of certain network typologies, coupled with the network scalability and fortressing against attacks, could lead to security-related issues that reduce the efficacy of complex V2X networks (Ghosal & Conti, 2020). For these reasons, full and effective deployment of V2X systems may not be realized in the very near future. In this regard, UAVs could potentially play a critical role by serving as a point of communication between connected vehicles, infrastructure, and other entities. In such case, the non-UAV entities do not need to be connected.

1.2 Problem Statement

This study was motivated by challenges in UAV-based monitoring (a) at intersections, and (b) during inclement weather, because in such weather conditions, video images are often corrupted by falling rain streaks thus impairing the integrity of the image. Therefore, there is a need for (a) algorithms for UAVs to predict vehicle trajectories at intersections and (b) algorithms for enhancing the quality of UAV-derived images.

Overall, the present study highlights the UAV potential and challenges for advanced traffic management in the prospective era of CAV operations. The practical benefits of the developed products are numerous: a reliable UAV-CAV data domain can help the road agency enhance the reliability of traffic safety risk assessment and vehicle trajectory monitoring. In addition, from perspectives of systems control, the study products can help CAVs by providing to them, reliable information for safe and efficient operational maneuvers such as weaving and lane changing along road corridors and trajectory planning at intersections.

In cognizance of the advantages of UAVs in traffic monitoring, the present study investigates the potential of UAV-supported traffic monitoring framework. The videos captured by UAVs are used to monitor the performance of transportation facilities and assess potential crash risks. The purpose of the proposed framework is to demonstrate the advantages that could foster the application of UAVs to road traffic monitoring.

The next research question pertains to the quality of the image captured by the UAV camera. The traffic monitoring potential of UAVs can be realized only when they provide clean video images where road users and infrastructure facilities are unobstructed and identifiable. In ideal weather and traffic conditions devoid of phenomena that degrade the visual quality of the driving environment, high-quality video images of the roadway environment can be obtained. However certain natural events and traffic conditions may impair the acquisition of good images and jeopardize the effectiveness of UAVs' roadway monitoring. The most common of these conditions is inclement weather, where streaks of falling rain drops severely degrade the visual quality of images of the traffic environment. Typically, rain streaks and raindrops distort the image features, thus impairing the lucidity of the video images. Further, the scattering and blurring effects of rain drops and streaks interfere with neighboring pixels.

These adverse conditions tend to degrade the output of traffic monitoring tasks, such as road user detection and tracking. In addition, rain-damaged images could impair the support of road traffic police monitoring tasks, and the effectiveness of pre- and post-crash information that is communicated to road users (to take proactive actions) and to traffic management centers (to take measures to enhance overall road safety). When videos are taken during such inclement weather, a video-denoising framework is needed to improve the image quality. This report seeks to address this issue.

1.3 Objectives of the Study

Motivated by the great potential of UAVs in a smart city ecosystem, this study investigates the incorporation of UAVs in traffic monitoring at road intersections. The study also examines ways to improve the quality of UAV-captured images and videos. The following specific goals and objectives were identified: (1) Develop and test a framework using UAV to monitor the traffic at intersections; (2) Develop and test a denoising framework to improve the quality of images in UAV videos.

1.4 Study Approach

This report investigates the use of UAVs for effective intersection traffic monitoring in terms of trajectory monitoring (intended ultimately for control of [or, trajectory recommendation to] CAVs) and quality of UAV-sourced images. Therefore, the study developed two frameworks: one for UAV-based intersection traffic monitoring, and the other for denoising videos captured by UAVs during rainy days. The relationship between the two proposed frameworks is shown in Figure 1.2.

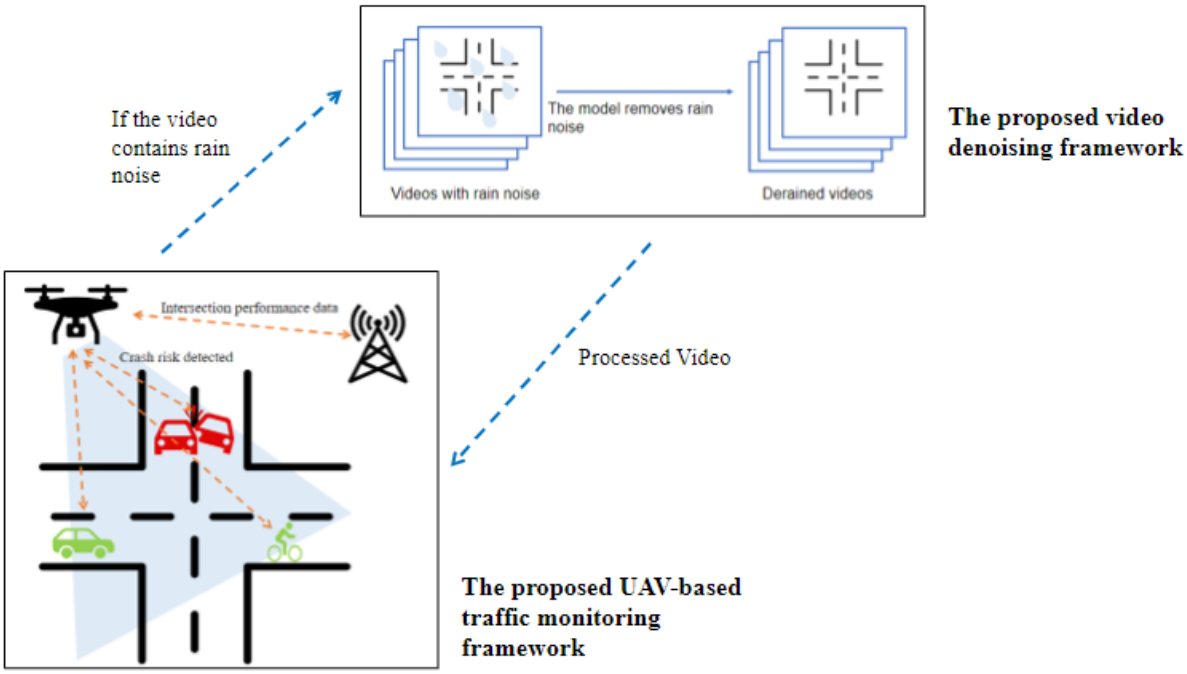


Figure 1.2 Overall study framework

1.5 Organization of this Report

The report is organized as follows: In Chapter 2, a framework using UAV to monitor the traffic at intersections is developed and demonstrated. In Chapter 3, a deep-learning-based denoising framework to remove rain steaks in videos is developed and demonstrated. Chapter 4 provides the conclusions this report and future work. Chapter 5 presents a synopsis of performance indicators, and Chapter 6 discusses the study outcomes and outputs.

CHAPTER 2. USING UAVS FOR VEHICLE TRACKING AND TRAFFIC MONITORING AT INTERSECTIONS

2.1 Introduction

It has been prognosticated that unmanned aerial vehicles (UAVs) will play a vital role in various application or context areas of smart transportation systems (Ke et al., 2017; Xu et al., 2017a; Xu et al., 2017b). This is motivated by the success of UAVs in other sectors and domains including photography, photogrammetry, agriculture, terrain mapping, monitoring, disaster relief and rescue operations, and recreation (Prevot et al., 2016). Due to these applications, the emerging global market for drone-enabled services has been valued at over $12.7 billion (Michał et al., 2016). Also, the UAV industry is predicted to generate at least 100,000 new jobs (Nath, 2020). Outay et al. (2020) stated that 7 million small UAVs have already been deployed in the airspace for commercial use in various sectors including real states, insurance, and agriculture.

In the transportation sector, researchers have investigated various ways in which UAV technology can be applied to enhance transportation operations. For example, drone-based solutions are being developed and tested to increase the general efficiency in transportation, particularly in freight movements (Kure, 2020). Recognizing the immense potential of UAV technology in transportation, the US Congress, in 2012, passed legislation that requires the Federal Aviation Authority (FAA) to integrate small drones into the airspace by 2015 (Outay et al., 2020). That legislation helped further propel UAV research traffic flow analysis (Ke et al., 2017; Zhang et al., 2019), vehicle detection (Raj et al., 2017), and highway infrastructure management (Lovelace, 2015). However, relatively limited attention has been paid to the potential of UAVs in traffic monitoring.

The potential benefits of UAVs in traffic monitoring, particularly safety, are underscored by the fact that in the US, road vehicle accidents are a major cause of unintentional fatality (in 2019, over 33,000 fatal road vehicle crashes occurred). It can be argued that the potential of UAVs to reduce the crashes through enhanced traffic monitoring is due to the flexible nature of UAV operations such that they can facilitate macroscopic and microscopic characterization and analysis of the traffic stream. In addition, UAV connectivity to vehicles, infrastructure, and pedestrians could facilitate intelligent and real-time communications among entities in the road space. Having this capability is useful for safe and efficient operation of connected and automated vehicles (CAV). Due to their accuracy, complexity, range, and availability of the traffic data they generally capture, UAVs have opened up a new chapter in the field of traffic monitoring and management (Ke et al., 2018; Outay et al., 2020). Chen et al. (2017) developed a method for surrogate safety analysis of pedestrian-vehicle conflict at intersections using unmanned aerial vehicle videos.

Against this background, this chapter of the report investigates the potential use of UAVs for effectively monitoring traffic at intersections. The objectives are twofold: (1) propose a framework that uses data obtained from UAV and V2X connectivity to track the movement of road users and assess potential crashes at intersections; and (2) demonstrate the framework using a case study at a specific intersection. The developed model, facilitated using machine-learning

technologies, is intended to enhance the reliable extraction and analysis of trajectory data and measurement of crash risk. The proposed model can help road agencies in their routine tasks of monitoring intersection, recognizing underlying causes of intersection crashes and justifying intersection design for reactive and proactive improvements not only currently but also in the prospective CAV era.

2.2 Related work in the literature

Several research efforts have investigated the applications of UAV in transportation management. Kim and Chervonenkis (2015) studied the detection of emergency and abnormal traffic situations with an UAV artificial vision system with the acknowledgment of the limitations of the efficacy of their algorithm (Kim & Chervonenkis, 2015). Sharma et al. proposed a multi-UAV coordinated vehicular network to analyze driving behavior for improving traffic safety (Sharma et al., 2017). However, their network requires more than two UAVs, which is hard to generalize at present. A few researchers have proposed frameworks for using photographs from drones to reconstruct accident scenes (Liu et al., 2019; Amin et al., 2020). Others compared the use of UAVs to other monitoring platforms for traffic monitoring using manned drones, helicopters, and road patrol vehicles, and carried out multiple criteria analyses to identify the most cost-effective monitoring platform (Mehmood et al., 2018). They found that the UAV has a far lower cost compared to helicopters and is quicker to deploy compared to road patrols, and authors concluded that the UAV is generally the best platform for freeway incident monitoring.

As UAV technology continues to develop, research attention is turning towards the processing of drone images captured at different shooting angles and heights, and improving the quality of reconstructed scenes. For example, researchers have proposed a low-cost method that uses UAV photogrammetry and other techniques to reconstruct traffic accident scenes (Ardestani et al., 2016; Dong et al., 2021; Pérez et al., 2019). Assessment of the quality of a reconstructed image is based on the concepts of peak signal-to-noise ratio and structural similarity (Su et al., 2016). Relatively few research efforts have addressed UAV applications in safety risk assessment. Risk assessment entails a detailed analysis of vehicle trajectories extracted from UAV-based videos. From the trajectories, potential conflicts, high-risk lanes, and risky maneuvers can be identified and crash propensity can be measured. Gu et al. (2019) developed a framework to investigate crash risk at freeway interchange merging areas using data exported from a UAV. The authors used a driver behavior model to identify the factors of risky driving. Other researchers explored UAV applications in smart transportation using concepts including trajectory established from the optical flow model, congestion detection, and driver behavior assessment (Ke et al., 2017, 2018).

The task of accurately extracting trajectories is one of the most challenging aspects of the UAV-based risk assessment. Such difficulty is exacerbated by the heterogeneity that often characterizes the traffic scene. For example, the scene at an intersection may not only be densely crowded but also consist of objects that vary significantly in terms of their features and behaviors. The intersection may have a substantial number of object classes (vehicles, pedestrians, or bicycles) with multiple interactions and behaviors among them. Also, the recognition of specific actions can be challenging. The tasks of manual monitoring and review of large amounts of video data, could be cumbersome and impractical. Therefore, accurate extraction of trajectories from videos is a critical and challenging aspect of video-based applications. In multi-object tracking (MOT),

challenges include occlusion, accurate identification of initialization and termination of tracks, the similarity of appearance of different road users, and interactions among different objects. In recent years, the rapid developments in convolutional neural-network deep-learning based MOT algorithms with high computing speed and accuracy have facilitated the task (Song et al., 2010; Zhou et al., 2020). Most existing MOT research falls into two categories: Detection-Based Tracking (DBT) and Detection-Free Tracking (DFT). DFT performs tracking, matching objects with trajectories and tracking simultaneously. DBT, on the other hand, conducts detection and tracking tasks separately: objects are first detected and then linked to identify the trajectories. Barmpounakis et al. (2019) assessed the reliability of small drones for measuring microscopic traffic parameters; Cao et al. (2014) proposed an ego motion guided particle filter for vehicle tracking in airborne videos; and Rodríguez-Canosa (2012) developed a real-time method to detect and track moving objects using a single-camera UAV. Teutsch and Krüger (2012) and Zhou et al. (2015) discussed the detection, segmentation, and tracking of moving objects in UAV videos. Li et al. (2019) developed an adaptive framework for multi-vehicle ground speed estimation in airborne videos. Other researchers that address UAV based vehicle tracking include Kanistras et al., 2015, Khan et al. (2017), Khan et al (2018), and Kim et al. (2019), and Gomaa et al. (2019).

In recent tracking studies, benchmarks have been established for DBT models (Bose et al., 2007; Song et al., 2010). Bose et al. proposed a framework for detecting and tracking multiple interacting objects with due cognizance of fragmentation (Bose et al., 2007). In their experiments, 89 out of 94 moving objects were correctly tracked and 762 merges and splits were detected. DFT models, on the other hand, are free of pre-trained object detectors but require manual initialization of a fixed number of objects in the first frame (Hu et al., 2012; Zhang & Maaten, 2013). It has been realized by at least one researcher (Zhou et al., 2020) that simultaneous detection and tracking can be carried out using a detection model. DFT models attract significant research attention because they can address disappearing objects or emerging objects in the image frame. DBT is generally more time-consuming compared to DFT because the total time used for the DBT algorithm is the sum of time spent by two components.

Traffic monitoring is expected to provide useful information on traffic conditions to measure the performance of transportation facilities. In the literature, there exist several traffic parameters for performance evaluation at intersections (Sinha and Labi, 2007). The performance measures for intersections could be made to include throughput efficiency, effectiveness, and equity. It is suggested that vehicle arrival patterns, crash probability and volume-to-capacity ratio can be used to measure the efficiency and effectiveness of an intersection (Transportation Research Board, 2010). From the perspective of equity, road user composition can be considered to measure system benefits to the society at large. Of the several traffic monitoring parameters, the assessment of crash risk at road intersections has been identified as a critical task yet to be addressed in the domain of traffic monitoring research (Abrari Vajari et al., 2020; Northmore & Hildebrand, 2019; Shah & Lee, 2021). The American National Standard listed intersections as a highway design context that is due for critical safety evaluation (American National Standard Institute, 2018). The Standard defines an intersection as an area which "(a) contains a crossing or connection of two or more roadways not classified as driveway access and (b) is embraced within the prolongation of the lateral curb lines or, if none, the lateral boundary lines of the roadways". If the distance along a road between two areas meeting the two criteria is less than 33 ft (10 m), then both areas and the connecting roadway are considered as parts of a single intersection.

Due to the availability and quality problems of real crash records, the development of non-crash or surrogate metrics of road safety has piqued the curiosity of researchers (Du et al., 2021; Peng et al., 2004; Wu et al., 2014, Tarko, 2019). Traffic conflict has been one of the most popular surrogate safety measures developed in recent years (Zheng et al., 2014). Traffic conflict has been defined as "a critical traffic situation in which two (or more) road users approach each other in such a manner that a collision is imminent and a realistic probability of personal injury or material damage is present if their course and speed remain unchanged" (van der Horst, 1990). Traffic conflicts are also identified as "an event involving two or more road users, in which the action of one user causes the other user to make an evasive maneuver to avoid a collision" (Parker and Zegeer, 1989). According to the definitions, the proximity of the relevant road users in space and (or) time, or the evasive actions, determines whether a situation constitutes a traffic conflict. To measure the traffic conflict, researchers have developed many proximity-based measures and evasion-based measures as shown in Table 2.1. In recent research, crash risk models that comprehensively consider vehicle motion/location, driver behavior and road geometry information have been proposed (Du et al., 2021; Wu et al., 2014). Although there exist several measures of traffic conflict, to date, none of them have been adopted universally.

In this study, the time-to-collision (TTC) was adopted as the conflict measure for several reasons: first, TTC is effective and efficient. Of the various traffic conflict measures, the most common are based on temporal proximity as they combine spatial proximity and speed (Zheng et al., 2014). Compared to other temporal proximity measures such as post-encroachment time and the time-to-stop line, TTC can be calculated without extra geography information which is difficult to obtain from UAV-captured videos only. Given the values of TTC, traffic conflicts will be effectively recognized once the value of the measure does not exceed a certain specified threshold (Mahmud et al. 2017; Huang et al. 2013). Secondly, the use of evasion-based measures has seen much debate. It is difficult to establish a list of evasive behaviors. Further, it is debatable that or deceleration or braking can serve as reliable indicators of traffic conflict, particularly at intersections where braking behaviors happen frequently. In addition, the capability of a UAV video to adequately characterize an evasive action on the road may be questioned. Therefore, considering the specific scenarios and data set in this study, we decided to use TTC as the measure of traffic conflict. Nevertheless, we recognize that several limitations exist regarding the TTC concept, and we discuss them in the conclusions section of this chapter.

Table 2.1 A synthesis of past research on traffic conflict measures (from Zheng et al., 2014)

Type	Traffic conflict measures
Temporal proximity	Time to collision (Hayward, 1972), post-encroachment time (Cooper, 1984), time to stop line (van der Horst 1990), gap time (Gettman & Head, 2003), time-to-line crossing (Vogel 2003), time to departure (Tarko, 2012), braking time (Lu et al., 2012)
Spatial proximity	The remaining distance to potential collision point (Allen et al. 1978), proportion of stopping distance (Gettman & Head, 2003), lateral distance to departure (Tarko, 2012)
Evasive behavior	Lateral or longitudinal acceleration (Dingus et al., 2006; Bagdadi, 2013), quickness of a driver response to potential traffic event (Dozza & Gonzalez, 2012; Wu & Jovanis, 2012).

The review of literature also showed that the development of the V2X network and cloud computing have enabled Cooperative Collision Avoidance (CCA) and therefore brought the risk detection tasks to a more real-time and proactive level. According to CCA-related studies, CCA systems use (Vehicle-to-Vehicle) communications (Misener et al., 2011; Themann et al., 2015; Tu et al., 2019) or Vehicle-to-Infrastructure (V2I) communications (Misener et al., 2011) to detect the possibility of accidents and to achieve cooperative collision avoidance. Studies have shown that in advanced CCA systems, vulnerable road users (VRU) can be recognized, and warning messages can be sent accordingly. The U.S. Department of Transportation estimates that V2V can potentially mitigate as much as 82% of all crashes in the country that involve unimpaired drivers, thereby prospectively preventing the loss of thousands of lives and billions of dollars (Taleb et al., 2010). Gelbal et al. introduced a pedestrian collision warning and avoidance system for road vehicles based on V2X communication (Gelbal et al., 2017): signals from pedestrians' smartphone apps are used to detect them and their locations using dedicated short-range communications (DSRC). Du et al. incorporated a Model Predictive Control approach in V2V communication systems and proposed a method for autonomous driving vehicles to avoid crashes in a mixed traffic stream that contains aggressive human drivers exhibiting errant lane-changing behavior (Du et al., 2021).

2.3 Methodology

The proposed framework for monitoring intersection traffic consists of three main stages (Figure 2.1). The first stage addresses trajectory extraction and the second stage performs risk assessment. In the first stage, the CenterTrack model (Zhou et al., 2020) trained using UAV-captured traffic videos is applied to obtain the real-time and historical trajectories of each road user. In the second and third stages, the crash risk is measured by time-to-collision (TTC) and the risk associated with each road user is determined. The scales of the frames and speed of each individual road user are first calculated using results from the first stage. The crash risk between each pair of tracked road users is then estimated by calculating the TTC between them. The implementation details and further discussions are provided in subsequent subsections of this section.

Figure 2.1 Overview of the proposed framework

2.3.1 Trajectory tracking

Reliable trajectory tracking is a key basis for the effective generation of profiles. Given an input video sequence, the multi-object tracking (MOT) task needs to be carried out to locate multiple objects, maintain their individual identities, and provide their individual trajectories. In the case study environment, the objects refer to road users (motor vehicles and pedestrians) at the intersection where the volume of objects is typically large. In addition, given the dynamic traffic pattern, an MOT model is required to capture the trajectories of road users quickly and accurately. Recent literature suggests that convolutional neural network (CNN) based multi-object tracking algorithms are promising approaches for doing this (Li et al., 2019). As discussed in the previous sections, CNN-based tracking algorithms fall into two categories: Detection-Based Tracking (DBT) and Detection-Free Tracking (DFT). In this report, it is recognized that traffic monitoring is inherently time-sensitive, and therefore, a DFT algorithm, which is faster than the DBT algorithm, is used. Zhou et al. developed a CenterTrack model, which is a simultaneous detection and tracking algorithm that is simple, fast, and accurate (Zhou et al., 2020) and therefore, is a perfect fit for the case study demonstration. In this study, a need was identified for further enhancement of the CenterTrack model. CenterTrack identifies each object through its center point and then regresses to the object bounding box's height and width. Specifically, it produces a low-resolution heatmap and a size map. In addition to the original output channels in CenterTrack, a new channel is introduced for object classification purposes.

Figure 2.2 presents the structure of the tracking model. At time *t*, an image of the current frame and the previous frame is given, as well as the heatmap of tracked objects from the previous frame. The heatmap is formed by the distribution of the confidence score of object centers. First, the heatmap and frames go through a 7 by 7 convolution module with stride 2 and 128 channels. This is followed by a residual block with stride 2 and 256 channels separately, concatenated to feed into another sequence of convolutional layers. The output from the entire network includes object classification, displacement prediction, height, and width of bounding boxes, and a heatmap for the current frame.

The original loss function of Centertrack consists of three components: focal loss, size, and local location. The focal loss, which is the loss of object detection (L_k), is presented in Equation 2.3(a). In Equation 2.3(a), $Y_{xyc}(= 0,1)$, indicates the ground truth heatmap of annotated objects. $\hat{Y}_{xyc}(= 0,1)$ is the detected heatmap and N is the number of objects. α and β are hyperparameters for focal loss. Compared to cross-entropy loss, the focal loss is an improved version of object detection by assigning greater weight to difficult-to-classify or easily misclassified entities. Therefore, the focal loss is more suitable for detection tasks under complex contexts such as drone-captured intersection images, where the "effective detection region" (i.e., regions occupied by road users) is relatively small compared to the background.

The size prediction is learned by the loss function (L_{size}) in Equation 2.3(b) and is supervised at the center locations. In Equation 2.3(b), $\mathbf{s}_i$ is the bounding box size of the i^{th} object at location P_i and $\hat{S}_{\mathbf{p}_i}$ is the detected size. The offset is calculated as the displacement of object centers and is learned using the loss function (L_{off}) shown as Equation 2.3(c), where:
$\mathbf{p}_i^{(t-1)} - \mathbf{p}_i^{(t)}$ captures the difference in location of the object in the current frame $\mathbf{p}_i^{(t)}$and the previous frame $\mathbf{p}_i^{(t-1)}$ and $\widehat{D}_{\mathbf{p}_i^{(t)}}$ denotes the displacement at time *t* at location P_i learned by the model.

$$L_k = \frac{1}{N}\sum_{xyc}\begin{cases}\left(1-\hat{Y}_{xyc}\right)^{\alpha}\log\left(\hat{Y}_{xyc}\right) & if\ Y_{xyc}=1\\ \left(1-Y_{xyc}\right)^{\beta}\left(\hat{Y}_{xyc}\right)^{\alpha}\log\left(1-\hat{Y}_{xyc}\right) & Otherwise\end{cases}$$ Equation 2.3(a)

$$L_{\text{size}} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{S}_{\mathbf{p}_i}-\mathbf{s}_i\right|$$ Equation 2.3(b)

$$L_{off} = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{D}_{\mathbf{p}_i^{(t)}}-\left(\mathbf{p}_i^{(t-1)}-\mathbf{p}_i^{(t)}\right)\right|$$ Equation 2.3(c)

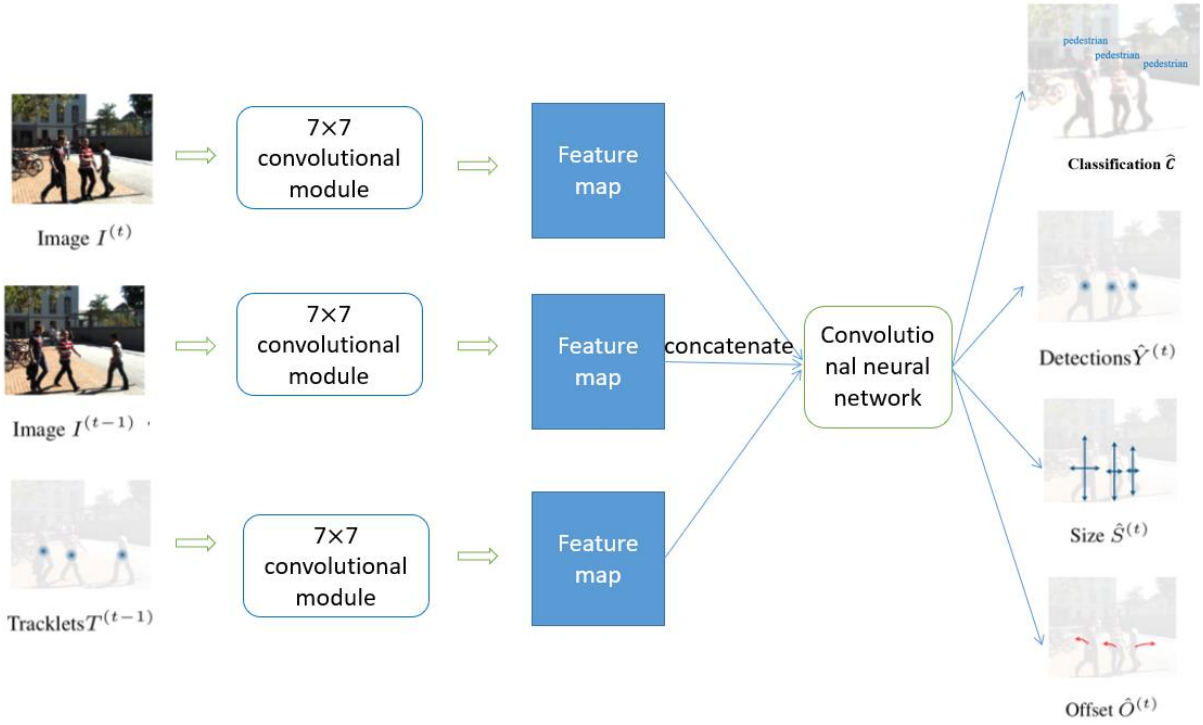


Figure 2.2 Structure of the *CenterTrack* model

2.3.2 Data preparation and traffic conflict assessment

The crash risk of road users can be evaluated using the trajectory extracted at Stage 1 of the methodology. First, the data is prepared to obtain the scale of frames and speed of road users. The trajectory data extracted by the deep-learning model in Stage 1 contains the time step *t* when the frame is captured, coordinates (x_t, y_t) of four corners of the bounding boxes for every road user, categories of every road user, and trajectory vectors for every road user. It is assumed that the length of a typical vehicle is 4 meters, and the width is 1.7 meters. A scale can be obtained by aligning detection boxes of vehicles in the video sequence with the real dimensions of vehicles. The speed of road users is calculated using Equation 2.3(d):

$$v(t) = \frac{\sqrt[2]{(x_t - x_{t-\Delta_t})^2 + (y_t - y_{t-\Delta_t})^2}}{\Delta_t} * scale$$ Equation 2.3(d)

where Δ_t is the video frame frequency and the unit of speed is meter/second, (x_t, y_t) is the location of the object at time *t* and $(x_{t-\Delta_t}, y_{t-\Delta_t})$ is the location of the object at time *t*-Δ_t.

After the data preparation, the information from each road user is assigned a unique ID. The data associated with each ID includes the center of its bounding box, height and width of its bounding box, speed, and the category it belongs to in every frame. A bounding box is a rectangle that surrounds an object, that specifies its position. Figure 2.3 presents an example.

A widely used traffic conflict assessment parameter, the time-to-collision (TTC), is adopted as a measure of traffic conflict. TTC concept was first established in 1972 (Hayward, 1972). The initial definition of TTC is "the time required for two vehicles to collide if they continue at their present speed and on the same path". A lower TTC value corresponds to higher conflict severities and a TTC smaller than 2.5 seconds is typically taken as critical (Nadimi et al., 2020). Hence, as discussed in the section of related work, TTC is generally perceived to be a primary and efficient measure in traffic safety assessment. Table 2.2 presents a summarized set of data (and their notations) used for the risk assessment. In this study, for any two objects (e.g., object 1 and object 2 in Figure 2.4), the TTC is calculated using Equations 2.3(e)–2.3(i).

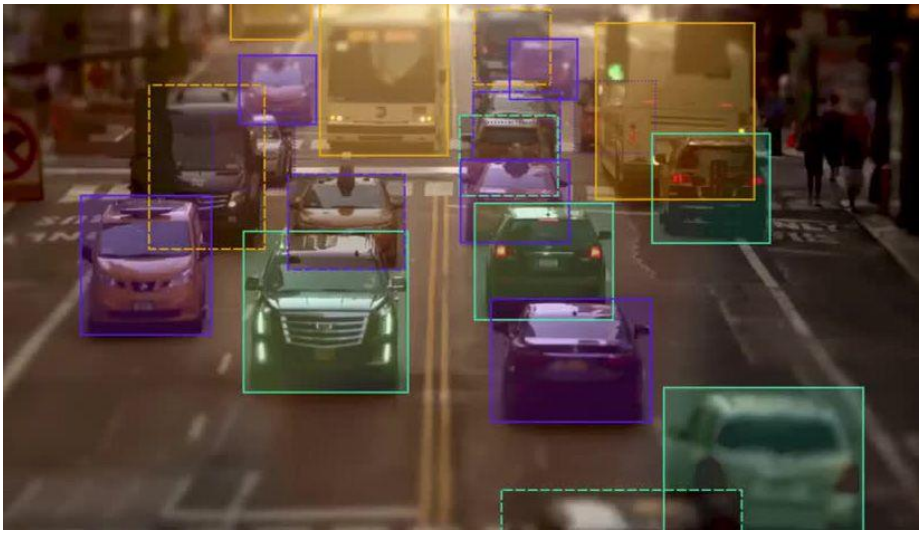


Figure 2.3 An example of bounding boxes (Keymakr, 2021)

Relative speed:

$$\widehat{v_{rela}} = \widehat{v_2} - \widehat{v_1}$$ Equation 2.3(e)

$$|v_{rela}| = \sqrt[2]{{v_1}^2 + {v_2}^2 - 2|v_1||v_2|cos\alpha}$$ Equation 2.3(f)

Distance: $l = \sqrt[2]{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ Equation 2.3(g)

Projected speed: $\hat{v}_{projected} = \hat{v}_{rela} \times cos\theta$ Equation 2.3(h)

TTC: $ttc = \frac{l}{|v_{rela}|}$ Equation 2.3(i)

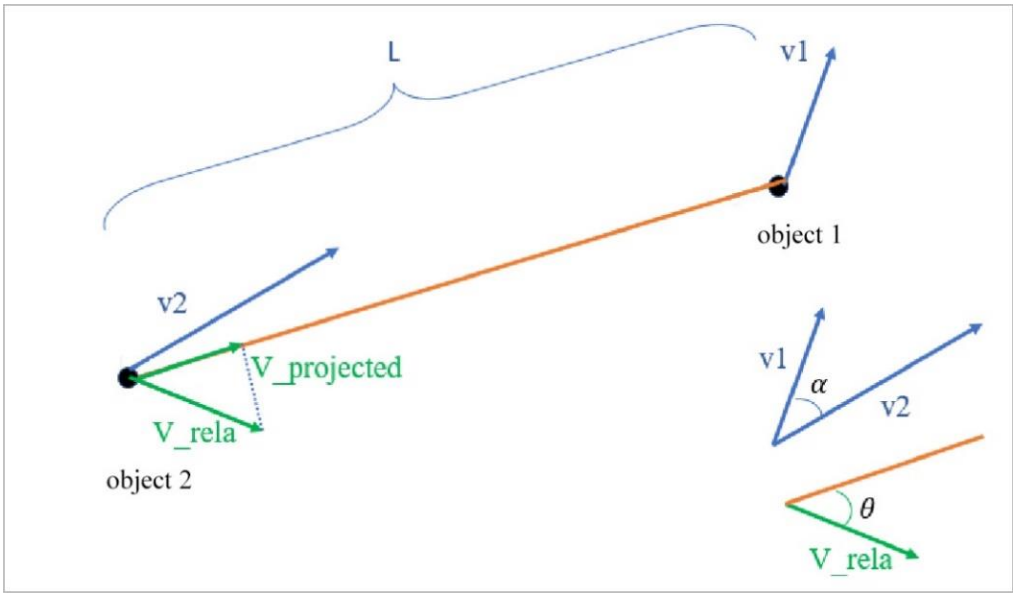


Figure 2.4 TTC calculation

Table 2.2 Data for the risk assessment model

Data	**Notes**
Across the entire video sequence:	
Scale	Match the video to real-world scales
Categories	Log all categories of the different road users
For every individual road user tracked:	
x_t, y_t	Location of the center of the bounding box at time *t*
H_t	Height of the center of the bounding box at time *t*
W_t	Width of the center of the bounding box at time *t*
V_t	The speed at time *t*
C_t	Category of the road user detected at time *t*

From the proposed model, the TTC value of each pair of all tracked road users can be easily achieved and road safety can be assessed at both macroscopic and microscopic levels. From the macroscopic perspective, a risk profile of the studied area at every time step can be established by identifying road users that exhibit the critical TTC. From the microscopic perspective, an individual road user could be informed about their TTC regarding each neighboring entity so that the appropriate maneuver can be undertaken. The case study section presents a detailed demonstration of the assessment.

2.3.3 Performance evaluation

The success of the proposed framework is determined by how many risky TTCs it can correctly detect, which in turn depends on the accuracy of the trajectory tracking task. The TTC ground truth is obtained by feeding the true trajectory data into the Risk Assessment Module. To better fit the UAV scenario, when training the trajectory tracking model, video clips provided by VisDrone (Zhu et al., 2021) which consists of 56 video clips with 24,198 frames captured by UAVs, are used. The trained model is tested on a test set containing 16 video clips with 6,322 frames. In this report, six categories of road users are considered: pedestrian, bicycle, car, van, tricycle, motor. Multi-Object Tracking Accuracy (MOTA) is used to evaluate the tracking results (Li et al., 2016a). MOTA is calculated using Equation 2.3(j) below:

$$\text{MOTA} = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t gt_t} \qquad \text{Equation 2.3(j)}$$

where m_t, fp_t, mme_t and $gt_t$are the number of misses, false positives, mismatches, and ground truth trajectories (road user trajectories), respectively at time *t*. In addition, we measured the tracking accuracy by calculating the root mean displacement error. The displacement is defined as the distance between the center of detected bounding box and the center of ground truth bounding box. As shown in Table 2.3, the tracking algorithm gives 64.89 MOTA on the training set and 63.12 MOTA on the testing set. The root mean displacement error is 0.217 and 0.253 on training set and testing set, respectively.

In using the extracted trajectories to produce risk profiles of the road users, the true positive rate and false positive rate were logged as an evaluation matrix. Using 2.5 seconds (Nadimi et al., 2020) as a threshold, the TTC between each pair of road users was labeled as: safe vs. critical. A true positive means both ground truth and the proposed framework detect the TTC between each pair of road users as critical. A false positive means the proposed framework indicates a TTC as critical while the ground truth shows it is safe. A true negative refers to situations where both ground truth and the proposed framework indicate that the TTC is safe. Similarly, a false negative means that the proposed framework gives a safe TTC while the TTC is critical in the ground truth dataset. It should be noticed that negative TTCs are ignored. As shown in Table 2.4, the model yields a true positive rate of 80% and a false positive rate of 31%. For all the detected TTCs, 78.2% of the model results fall into the ground-truth categories of critical or safe designations.

Table 2.3 Evaluation of the trajectory tracking model

Dataset	MOTA	Root mean displacement error
Training set	64.89	0.217m
Testing Set	63.12	0.352m

Table 2.4 Evaluation of the risk assessment model

Evaluation Metric	Value
Accuracy[1]	78.2%
True Positive Rate	80%
False Positive Rate	31%

[1]. *Accuracy = (number of true positive cases + number of true negative cases)/total number of cases*

2.4 Case study, results, and discussion

To illustrate the analysis framework of the UAV-based traffic monitoring proposed in the sections above, a case study was conducted using drone images captured at an intersection in Tianjin, China. As shown in Figure 2.5, this is a 4-legged intersection. The effectiveness of the proposed framework might be arguable since the traffic signals at the intersection will help to reduce possible conflicts. However, it should be noticed that more than one–third of all the fatalities at intersections occur at signalized intersections, including a large proportion that involve red–light running (Federal Highway Administration, 2022).

Considering the mixed traffic flow at intersections, dangerous road user crossings and violations are common, which yield high chances of traffic crashes. It is recognized that even though each movement direction has a dedicated phase, violations are rife, and therefore, collisions between entities in different directions, are common. For road users such as pedestrians, they are exposed to motorized vehicle flows and are very vulnerable, particularly when crossing by walking outside the bounds of the intersection crosswalk or walking during the signalized intersection's red-light phase. In most large cities, such as Montreal, intersections represent a critical roadway feature, with high concentrations of vehicle–pedestrian crashes and where 60% of pedestrian injuries occur as stated in Morency and Cloutier (2007). Therefore, traffic monitoring at signalized intersections continues to pose a serious problem.

The video data are provided by an open-source dataset (Zhu et al., 2021) that includes intersection videos taken under various conditions including sunny weather, good light, and no electromagnetic interference (which could influence the stability of the video pictures at a vertical angle). The movements and interactions between vehicles in this intersection were captured at a frame frequency of 30fps.

Figure 2.5 Case Study Intersection

2.4.1 Road performance analysis

The UAV-captured data, after being processed by deep-learning networks, offers numerous applications to support road monitoring and management. As proof of concept, this section presents here how these findings could be applied to measure the performance of the study area in terms of the safety and efficiency of traffic movement. Transportation performance measures sometimes referred to as measures of effectiveness (MOEs), are quantitative estimates on the performance of a transportation facility, and include level of service, crash frequency, and travel time (Sinha and Labi, 2007; Vivek et al., 2021).

Proper evaluation of transportation facility performance has always been supported by legislation (U.S. Federal Highway Administration, 2018). This is important because there is a growing demand for information on traffic patterns, to support general transportation administration and management, and in particular, the development and evaluation of road safety policies. In this context, the proposed UAV-supported traffic monitoring framework can be beneficial to road agencies because it can generate large amounts of real-time traffic data for evaluation purposes.

This capability is due to the inherent structure of the deep-learning network used in the framework: the detection results help identify the composition of road users, and the tracking results help measure the speeds and directions of each road user. The data obtained from the proposed framework indicates that in the studied intersection (Figure 2.6), road users consist of motors (36%), vans (1%), bicycles (10.1%), pedestrians (14%), tricycles (2.3%), automobiles (36.6%).

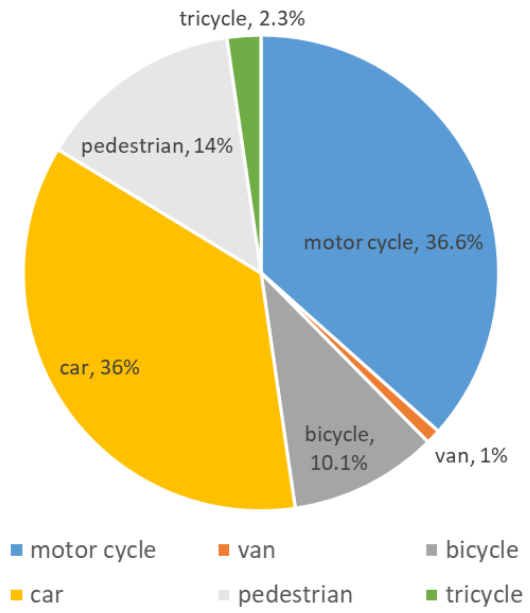


Figure 2.6 Road user composition at the case study intersection (determined from UAV-data)

From the results of the tracking analysis of the UAV data, the speeds of each road user category can be determined (Figure 2.7). The tracking analysis excludes the phase where road users wait for the green signal. Of the road users that pass through the intersection, motor vehicles are those that show the highest speeds and the widest speed range. In contrast, the travel speed of

pedestrians and vans is relatively stable. This information could be used to generate several useful measures of the intersection performance. For example, the Travel Time Index (TTI), which is travel time divided by the free-flow travel time, can be calculated (U.S. Department of Transportation, 2009). A TTI value of 1.00 indicates travel at the free-flow speed, while a TTI value of 2.00 indicates that travel time is twice as long, compared to free-flow conditions. The vehicle speed outcomes can be compared to target or design speeds to assess relative benefit. In analyzing the intersection performance, it is desirable to incorporate more input values such as speed limit from local transportation agencies; however, this is outside the scope of this study.

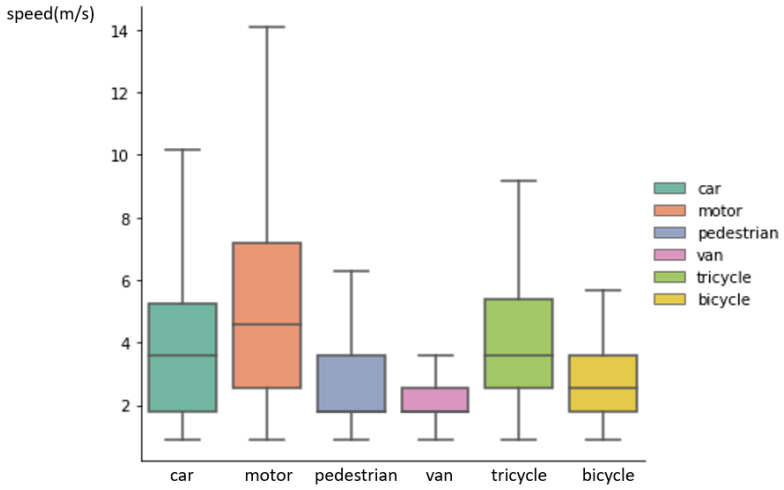


Figure 2.7 Road user travel speeds at the case study intersection (determined using UAV-obtained data)

2.4.2 Risk profiles

At each time step, the trajectory of each road user was tracked using the deep-learning-based model developed in this study. Then, the crash risk for each pair of road users was estimated using the TTC equation provided as Equations 2.3(e)–2.3(i). As discussed in Section 2.3, only positive TTCs are considered in this study and TTCs smaller than 2.5s are labeled as critical. For any road user, if the minimum correlated TTC is critical, the road user is labeled "risky". Figure 2.8 presents a series of consecutive macroscopic risk profiles where risky road users are highlighted by their bounding box; the number indicated at the top of the box is the value of the most critical TTC value correlated to the road user in the box.

As indicated in Figure 2.8, the dynamic variation of the intersection risk profile is captured by the videos. In addition to vehicles, pedestrians and bicycles that are risky are also identified by the proposed framework. The microscopic risk profiles can be obtained by extracting information for an individual road user. Figure 2.9 presents the risk profile of an individual vehicle and highlights the neighbors that have a "critical" level of TTC with respect to the individual vehicle in question. The individual vehicle is indicated by a red circle in the figure. Other road users that are risky are marked with blue boxes. The number indicated above each box is the TTC value between the road user in the box and the studied vehicle.

Time = t_1

Time = t_2

Time = t_3

Time = t_4

Figure 2.8 Macroscope risk profile ("risky" road users are marked with red boxes and the number indicated above each box is their smallest TTC value)

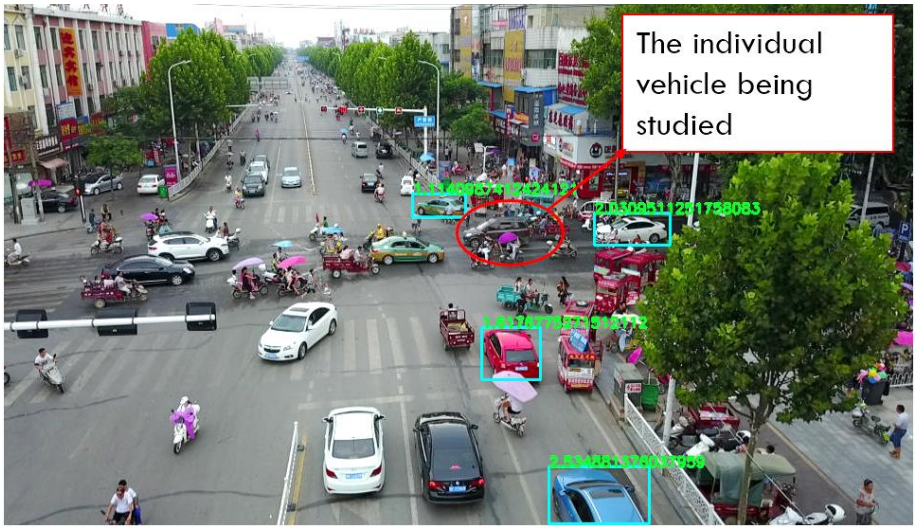


Figure 2.9 The microscope risk profile of an individual vehicle (circled red)

The benefits of such risk profiles are twofold: First, when transmitted (by the UAV) to the individual road user, the microscopic risk profile can help the road user become aware of potential crashes in its surroundings. This is considered particularly important in the prospective era of autonomous vehicles (AVs) because the AVs in-vehicle detectors including cameras and Lidar may fail to identify all potential crashes or hazardous situations due to their narrow detection range and detection challenges. In such situations, the UAV not only serves as a robust source of information relating to a broader view of the surroundings and wider spatial characterization of the environment but also makes available accurate data regarding potential risks. If there exists a centralized control platform, then the platform could convey real-time warning messages to connected risky road users with appropriate crash-avoiding maneuver suggestions to mitigate the crash. Second, the macroscopic risk profile provided by the UAV can provide useful insights to urban planners and transportation managers in their efforts to assess the safety level of an intersection, identify risky road users and analyze the reasons behind potential collisions.

The risk profile patterns can be identified by summarizing data from the study intersection. In the studied intersection, it is observed that 72% of potential crashes occur between vehicles. Of the remaining 28% potential crashes (Figure 2.10), 40% are between pedestrians and vehicles and 30% of collisions are between motors and cars. The road agency overseeing the operations of the intersection may be interested in investigating why pedestrian-vehicle collision risk is so high at certain intersections and therefore, recommend specific initiatives such as providing pedestrian-dedicated facilities to mitigate these problems. Figure 2.11 logs all locations where “risky” road user interactions are prevalent. It can be observed that the most critical potential crashes occurred in the upper right corner of the intersection. This may be due to the large number of bicycles and pedestrians who typically occupy that area, where they share the travel lane with vehicles. As a result, it is difficult for vehicles to undertake safe turning maneuvers. Intersection designers could also use the results of such analysis as a basis to carry out intersection improvements or to revise design policies.

The results of this study are consistent with a national effort to assess safety risks at road sections and intersections. In 2009, the U.S. Federal Highway Administration (FHWA) conducted a program to address pedestrian safety concerns by developing and researching effective tools and countermeasures and by coordinating projects, plans, and discussions with state and local officials and safety advocates (U.S. Department of Transportation, 2009). The initiative has been echoed by efforts at the local level. For example, the Chicago Department of Transportation completed an extensive pedestrian crash analysis to identify specific crash factors and characteristics including when and where pedestrian crashes occurred, who was involved in pedestrian crashes, and the contributing factors related to the pedestrian crash. The Chicago report advocated for the construction of marked crosswalks, in-road state stops for pedestrians’ signs, and pedestrian refuge islands at roads and intersections considered to be risky (Hamilton et al., 2011). Also, the 2012 Chicago Pedestrian Plan identified opportunities and ongoing plans to increase the safety of the city’s pedestrians. Similarly, the City of Austin developed a Pedestrian Safety Action Plan based on a comprehensive analysis of intersections considered dangerous (FHWA, 2021). The risk profiles extracted by UAVs, as demonstrated in this study, would facilitate such programs, and enable them to be more efficient and focused. As depicted in Figure 2.12, the information exchange between UAVs, and connected vehicles and transportation agencies facilitates the dissemination of microscopic and macroscopic risk profiles and helps identify appropriate safety countermeasures.

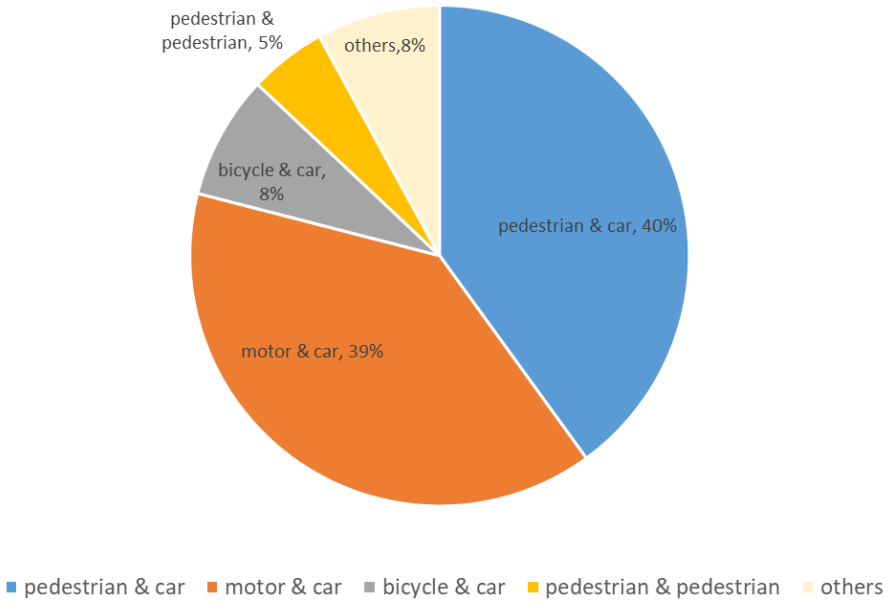


Figure 2.10 Road user pairs associated with potential collisions (excludes car-car pairs)

Figure 2.11 Summary of all locations where potential traffic conflicts could occur. A deeper color indicates a less safe interaction (i.e., a smaller TTC) between road-user pairs

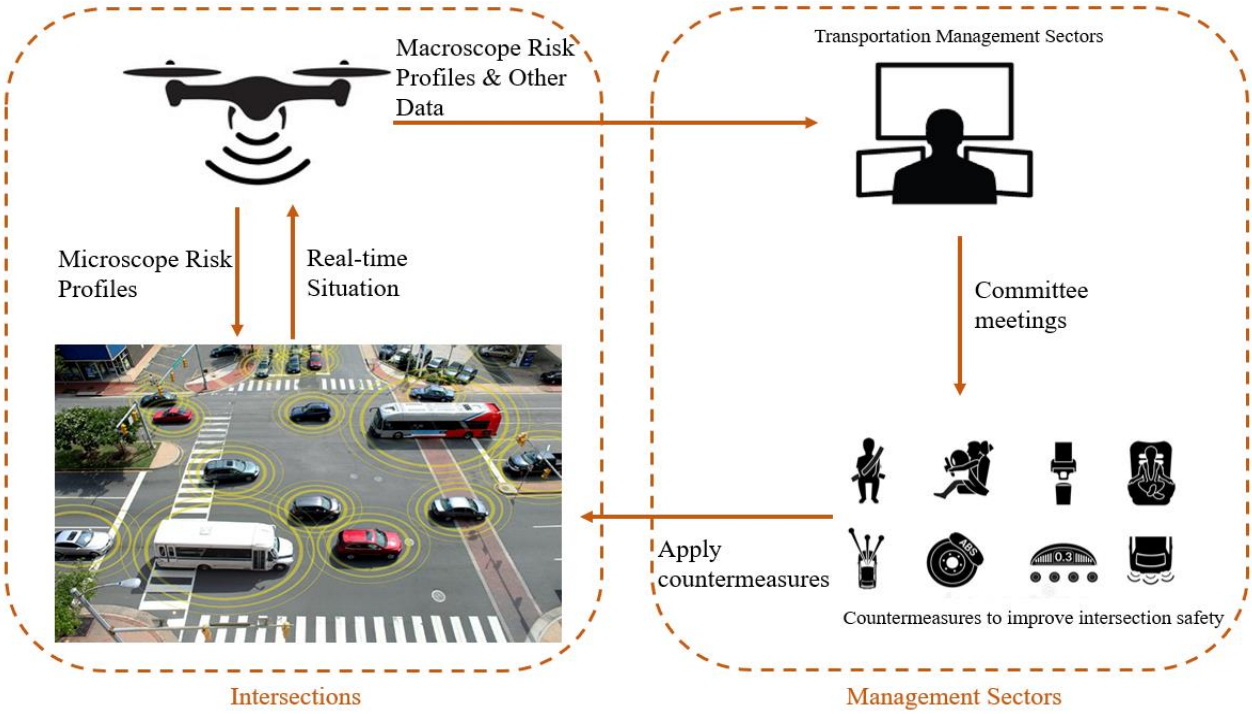


Figure 2.12 Workflow of UAV-supported traffic risk monitoring

2.4.3 Risk prediction

In assessing the risk associated with an intersection, it is also of interest to predict the risk of a vehicle at subsequent time steps. To predict such future crash risk, a Random Forest classifier, a supervised learning algorithm, is deployed. Random forests create decision trees on randomly selected data samples, obtain predictions from each tree and select the best solution through voting. The major advantage of Random Forest is that it provides an indicator of the feature importance which offers insights on features that are influential in distinguishing the data samples. Such information makes the model interpretable and facilitates the identification (before a crash occurs) of vehicles associated with "risky" interactions so that risk could be mitigated in a proactive manner.

It may be noted that only cars are included as studied objects. This is because the moving pattern of car-related pairs (that is, cars and other road users) is different. The features fed into the random forest model include speed, location, safety condition of the studied vehicle and its neighbors, together with TTC and distance between them in the five consecutive previous time steps. The output of the classifier is either "safe" and "risky". 'Safe' means that for the studied vehicle, the smallest predicted TTC at the next time step is greater than 2.5 seconds, and "risky" means the smallest predicted TTC is less than 2.5 seconds. From the Random Forest model, the importance of different features regarding future risk prediction is obtained, which is calculated by a Gini Importance value that sums over the number of splits (across all trees) that include the feature, proportionally to the number of samples it splits. A higher Gini Importance value indicates the feature is more likely to be the essential difference between different categories.

Figure 2.13 presents the top 5 important features and their relative importance. A "dangerous road user" refers to the neighbor with the smallest TTC with respect to the studied vehicle in the last time step. Features with higher importance contribute more when predicting risky vehicles, indicating we can observe these features to predict the future potential risk of a vehicle. The threshold of these features could also be extracted from the Random Forest classifier. The threshold values are not discussed in this study because the threshold values are very specific to the studied area and the time of capturing the video and cannot be generalized. According to the Random Forest classifier, the vehicle speeds in the previous time steps are most related to its future safety condition. The status of the vehicle's neighbors, particularly the location and speed of its most dangerous neighbor, also plays a significant role in deterring a vehicle's future safety condition. In practice, warning messages could be generated based on these results and sent to the vehicles concerned, to remind them to be aware of the imminent danger of traffic conflict. For example, when the speed of a vehicle exceeds the speed threshold found in the classifier, the vehicle could be alerted to reduce speed.

The UAV-based crash predicting and warning system discussed above is promising for supporting and enhancing current crash-avoidance systems. Over the past decade, there has been an upsurge in the availability of crash-warning systems in cars sold in the U.S. (Monticello, 2019). The benefits of pre-crash warning systems have been verified by the Insurance Institute for Highway Safety (IIHS) whose data suggest that collision warning reduces rear-end accidents by 27 percent (Barry, 2019). Wider adoption of collision warning systems could be anticipated considering the rapid advancement of autonomous driving technologies. Currently, the crash-warning systems are mainly powered by on-vehicle ranging sensors (e.g., cameras and radar) and are limited to Forward Collision Warning (FCW), Pedestrian Detection System (PDS) and Lane Departure Warning (LDW). Research has shown that cellular-V2X systems increase communication performance in congested conditions. However, UAV-based V2X systems have attracted scant attention in safety analysis (Vukadinovic et al., 2018). UAV-based collision warning systems can address the inadequacy of onboard sensors and therefore potentially open new directions for AV collision avoidance systems. The design of a UAV-based warning system could be adapted to local intersection data retrieved from videos of local intersection traffic.

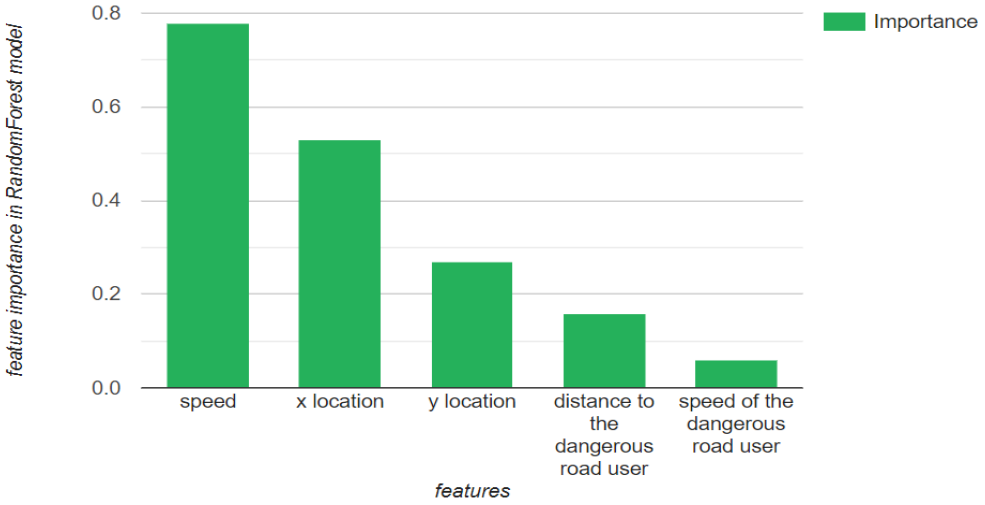


Figure 2.13 Features of top 5 importance in the random forest model

2.5 Conclusions, study limitations and directions for future research

This study presents a methodology to monitor traffic safety at intersections utilizing UAV captured video. The methodology explores the potential crashes between each pair of road users by extracting their trajectories from video images and calculating their time-to-collision values. To develop the trajectories of all detected road users efficiently and accurately, the study used a deep-learning-based multi-object tracking algorithm. The trajectory data were re-scaled for the risk assessment. Then, a method was suggested to compute the crash risk between each pair of road users by calculating the time-to-collision between them. The study shows how the data provided by the UAV (including road user composition, their speed distributions, and TTC values) can help road safety managers to identify conflicts and other problem areas, develop targeted countermeasures, and measure the general performance of intersections and other road facilities. Based on the TTC value, "risky" road users for which the smallest TTC is less than a threshold (2.5s in this study) are identified, and a macroscopic risk profile can be established and presented to the road agency that manages the intersection. An individual road user can acquire their own microscopic risk profile from the UAV so that it can make its safe and informed movement decisions accordingly.

The case study also demonstrates how the proposed framework could assist intersection management in the current era of human driving, and more importantly, in the future era of autonomous driving is demonstrated. The results showed that by investigating consecutive macroscopic risk profiles, the spatial-temporal pattern of risk profiles can be developed. Urban planners and intersection managers may find these results useful in their efforts to improve traffic control, design configuration, and ultimately, safety at intersections. In addition, the study used a Random Forest model to predict the safety condition of a vehicle by utilizing their historical risk profiles, and the results suggest that the travel speed is the most critical factor of a vehicle's future safety condition. The speed of the vehicle's neighbor was also found to be influential. With the proposed model, traffic engineers can be placed in a better position to propose efficient countermeasures to enhance road safety at intersections. Also, the proposed model can provide CAVs information that is helpful for making informed driving decisions and make available data for traffic engineers that may be considering intersection improvements from design or operating policy perspectives.

There are a few limitations of this study, which are indicative of possible future work improvements. First, the VisDrone dataset does not provide the geometry and coordinates of the road infrastructure, therefore, this study did not consider the impact of road geometry on the crash risk. For example, where a vehicle approaches the intersection via a misaligned road segment, the time-to-collision can be influenced by such anomalous geometry of the roadway. Ideally, the TTC calculation should reflect such anomaly. In future studies, this limitation may be addressed by the use of datasets that include more detailed information (such as lane directions, skew angles, and geometry) that captures any irregularities associated with the roadway infrastructure. Secondly, the current research work uses a signalized intersection where the traffic lights control the flow of traffic and assign right-of-way to conflicting movements. The effects of the proposed model would be more prominent when road users are not controlled by traffic signals. Therefore, future work could address unsignalized intersections and other geometric contexts (roundabouts, straight or curved highway segments, and so on). Thirdly, although the TTC alone is used effectively as an indicator for collision risk, one limitation of TTC is that it assumes constant vehicle velocity while

the vehicles at intersections accelerate and decelerate frequently. Future work would address such limitations by taking acceleration and deceleration into account. Finally, the model false-positive is relatively high (31%). This value indicates that model identifies some safe behaviors as risky as well. In the future, the model specificity can be improved by incorporating more training data and refining the structure of the tracking network.

The current chapter assumes the video captured by UAVs is visually clean and readable. However, in real life, the monitoring videos are usually affected by inclement weather. For example, rain streaks, fog, and snow are all possible sources of noise to videos. The recovery of clean traffic monitoring videos is a critical issue in traffic monitoring. In the next chapter, this issue is addressed using an innovative denoising framework.

CHAPTER 3. A PROPOSED SELF-SUPERVISED LEARNING APPROACH FOR TRAFFIC VIDEO DERAINING

3.1 Introduction

The previous chapter demonstrates the effectiveness and advantages of using UAV-captured video in intelligent traffic monitoring at urban intersections. An important prerequisite is that the videos are captured in good weather such that there is no weather-induced noise that impairs the video quality. In ideal weather and traffic conditions devoid of phenomena that degrade the visual quality of the driving environment, high-quality video images of the roadway environment are obtained. Unfortunately, this assumption does not always hold true, and image noise is all too common. For example, certain natural events (rain, snow, fog, mist, and smoke) and traffic conditions impair the acquisition of good images and jeopardize the task of roadway environment monitoring. The most common of these conditions is inclement weather, where falling rain severely degrades the visual quality of the traffic environment. Further, in a video, the motion of fast-moving cars generates motion noise that degrades the visual quality of videos. This chapter discusses how to address cases where videos are affected by noise from the environment, particularly rain noise. Good and clear video images are vital for monitoring the roadway traffic environment for purposes of road-use enforcement, pre-crash evaluation for safety studies, crash risk assessments, liability purposes and roadway characterization in the prospective era of autonomous vehicle operations (Kerkhoff, 1985; Du et al., 2021). The quality, reliability and timeliness of real-time traffic information provided by traffic monitoring videos have a major influence on the efficacy and efficiency of traffic monitoring. In the current era, most urban intersections are equipped with mounted fixed-position cameras. These cameras are often located on the top or side of buildings, road sign gantries, or cables or on traffic signal posts.

The degradation of video images due to natural or anthropogenic conditions has been duly recognized in the research literature. It has been found that these conditions significantly degrade the performance of monitoring equipment that relies on the efficacy of image/video feature extraction techniques (Bay et al., 2006; Junior et al., 2009; Maji et al., 2008; Shehata et al., 2008), including event detection (Roser & Geiger, 2009), image registration (Lucas and Kanade, 1981; Dalal & Triggs, 2005), object tracking (Maji et al., 2008), and scene analysis (Zhang et al., 2017). Therefore, a video noise deraining algorithm to generate clean traffic videos offers significant potential for more efficient traffic environment characterization and ultimately, increased roadway safety. In addition, for highly automated transportation where the safety of driving maneuvers hinges largely on the quality and reliability of the information received from local sensors including videos, video deraining and denoising in general, offers great promise in terms of safety.

Video deraining could be considered as a task of recovering a clean video, V: $\mathrm{V} = \{\mathrm{I}_1, \mathrm{I}_2, \ldots, \mathrm{I}_\mathrm{n}\}$ from a noisy video $\hat{V} = V + eN$
where $\mathrm{eN} = \{\mathrm{N}_1, \mathrm{N}_2, \ldots, \mathrm{N}_\mathrm{n}\}$; and $\{\mathrm{N}_\mathrm{t}\}_{\mathrm{t}=1}^{\mathrm{n}}$ is the rain noise associated with each frame.

Unlike the deraining of single-image features such as a photo, the deraining of multiple image features in media such as videos, should be able to address not only the spatial noise for each frame but also the inter-frame temporal noise caused by motion. Therefore, some image-based video denoising approaches, which rely solely on the denoising of individual frames separately (Zhang et al., 2018), tend to yield unsatisfactory results. Recently, convolutional neural network (CNN)-based video deraining algorithms have attracted much attention due to the quality of their product (Davy et al., 2021; Lehtinen et al., 2018). In this past work, CNNs were trained to learn to utilize temporal information by addressing clues between adjacent frames. The end-to-end CNNs have demonstrated great feasibility and flexibility in video denoising because they can be trained to remove various types of noise from the videos. However, most of the end-to-end models require pairs of noisy-clean images of the same location which may be difficult to obtain. Recently, Lehtinen et al. proposed a self-supervised "Noise2Noise(N2N)" model that yielded satisfactory image-denoising outcomes (Sun et al., 2014). The Noise2Noise model uses noisy images as both input and targets during the training phase and encourages the model to learn the average result from the various noisy pairs.

Inspired by Sun et al.'s approach, this report uses an approach that extends the N2N model to a two-stage model. In the first stage, spatial noise is reduced using a single image based N2N method. In the second stage, a regular spatial-temporal denoising method is applied by adopting the images denoised in first stage, as targets. During the entire training process, the clean videos are not accessible to the model. This novel approach is therefore described as a self-supervised video deraining method.

3.2 Related work

3.2.1 Still-image deraining

Recent years have witnessed significant progress in image deraining algorithms. Current image-deraining approaches fall into two groups: prior-based algorithms (Eigen et al., 2013; Huang et al., 2013; X. Zheng et al., 2013) and data-driven CNN models (Brewer & Liu, 2008; Eigen et al., 2013; Garg & Nayar, 2007). In the prior-based algorithms, priors are proposed to detect rain streaks in images. Elad and Aharon proposed a strategy for detecting rain streaks by checking whether the region exhibits a short duration intensity spike (Elad and Aharon, 2006). In Elad and Aharon's work, it is assumed that image signals hold a sparse decomposition over a redundant dictionary (Elad & Aharon, 2006). Another researcher formulated a correlation model to capture the dynamics of falling rain (Garg & Nayar, 2004). In the proposed model of our study, the average of non-rain temporal neighboring pixels is utilized to estimate rain density after the rain region is detected.

CNN-based deraining algorithms have received significant attention recently due to their successful performance in image deraining (Fu et al., 2017). In DnCNN proposed by Zhang et al. (2017), residual learning and batch normalization were implemented for image denoising. The DnCNN demonstrated its flexibility for tasks including blind Gaussian denoising, JPEG deblocking and image inpainting. Extended from DnCNN, FFDNet was developed to handle spatially variant noise (Zhang et al., 2018). Qian et al. utilized an attentive generative network via adversarial training to recover a clean image from a raindrop-degraded image (Qian et al., 2018). Their model first learns about raindrop regions and their surroundings, and then focuses on such

regions to assess the local consistency for rain removal tasks. Another researcher introduced DerainNet which directly learns the mapping relationship between rain and clean image detail layers from data (Fu et al., 2017).

3.2.2 Video deraining

Compared to still images, videos feature a strong temporal redundancy along motion trajectories. This added information in the temporal dimension provides more information when recovering a pixel from noisy frames. On the other hand, it also creates an extra degree of complexity, which could be difficult to address. The movement of objects introduces motion noise which could be in the form of ghost flickering. Therefore, videos suffer from both spatial noises that exist in each frame and temporal noise.

In this context, motion estimation and compensation have been employed in several video-denoising algorithms to help to improve their performance in temporal consistency (Buades et al., 2016; Maggioni et al., 2012). Niklaus et al. incorporated a pre-computed optical flow as motion information with a frame interpolation CNN (Niklaus & Liu, 2018). Caballero et al. developed a network that estimates the motion by itself for video super-resolution (Caballero et al., 2017). Removal of the multiscale rain streaks in video was accomplished using a multiscale convolutional sparse coding established in prior research (Li et al., 2018).

Furthermore, considering different types of rain, a hybrid rain model is offered to model both rain streaks and occlusions using a dynamic routing residue recurrent network (Liu et al., 2019) However, most of the learning-based models mentioned above learn a noise layer and then remove the noise layer away from the original frames. In the proposed approach of this study, rather than learning an explicit noise layer, the clean frame is learned instead.

3.3 The model

In video denoising frameworks, motion noise (also known as flickering) removal is crucial for improving the visual quality of the video. Flickering due to the high-speed motion of objects or the motion of the camera itself, can be removed using temporal clues that exist in neighborhood frames. Therefore, to recover a reference frame I_n, frames that neighbor I_n are required. Consider a sequence of frames $\{I_{n-k}, I_{n-k+1}, ..., I_{n+k}\}$ in the neighborhood of the reference frame I_n (Wang et al., 2020): the relationship between neighboring frames can be modeled as follows:

$$I_n(x) = I_n(x + \delta_k),\ k \in \{-K,\ -K+1,\ \ldots,\ K\} \qquad \text{Equation 3.3(a)}$$

Where δ refers to the estimated optical flow between two frames. Several state-of-art video denoising models achieve a mapping function F that holds two main effects, namely, motion estimation and prediction of the clean reference frame, which can be represented as:

$$\hat{I}_n(x) = F(\{I_{n+k}(x + \delta_k)\}_{k=-K}^{K}) \qquad \text{Equation 3.3(b)}$$

Where $\hat{I}_n(x)$ is the denoised clean frame.

However, in all models mentioned above, clean-noisy pairs are required and the loss function is therefore expressed as:

$$L_{temp}(\theta) = \frac{1}{2m_t}\sum_{j=1}^{m_t} ||\hat{I}_j - I_j||_2 \quad \text{Equation 3.3(c)}$$

Where $\hat{I}_j$ is the estimated clean frame.

Considering that acquiring such data is labor-demanding and challenging in real-word, in the present work, a model that can be trained without noisy-clean pairs. Recently, a researcher (Wang et al., 2020) has verified, using FIVnet, that the "first image then video" approach adequately addresses the problem of blurred boundaries associated with moving objects and provides a state-of-art denoising result. In the FIVnet, intra-frame noise is firstly reduced separately and then frames are wrapped together to remove inter-frame noises. Inspired by this concept in FIVnet, this report proposes a self-supervised video deraining model which consists of two stages: the first stage takes care of intra-frame noise while the second stage handles inter-frame noise. The deraining process can be expressed as follows:

$$\hat{I}_n(x) = \Phi(\{\varphi(I_{n+k}(x))\}_{k=-K}^{K}) \quad \text{Equation 3.3(d)}$$

Where: φ is a spatial deraining block and Φ is a spatial-temporal denoising block. Figure 3.1 presents the architecture of the model. In the spatial deraining module, the *Noise2Noise* model is adopted which removes rain noise without requiring true labels. With regard to the spatial-temporal denoising module, a regular denoising model named FastDVDnet is adopted to take the derained images from φ as input. Details of each module are discussed in the following sections.

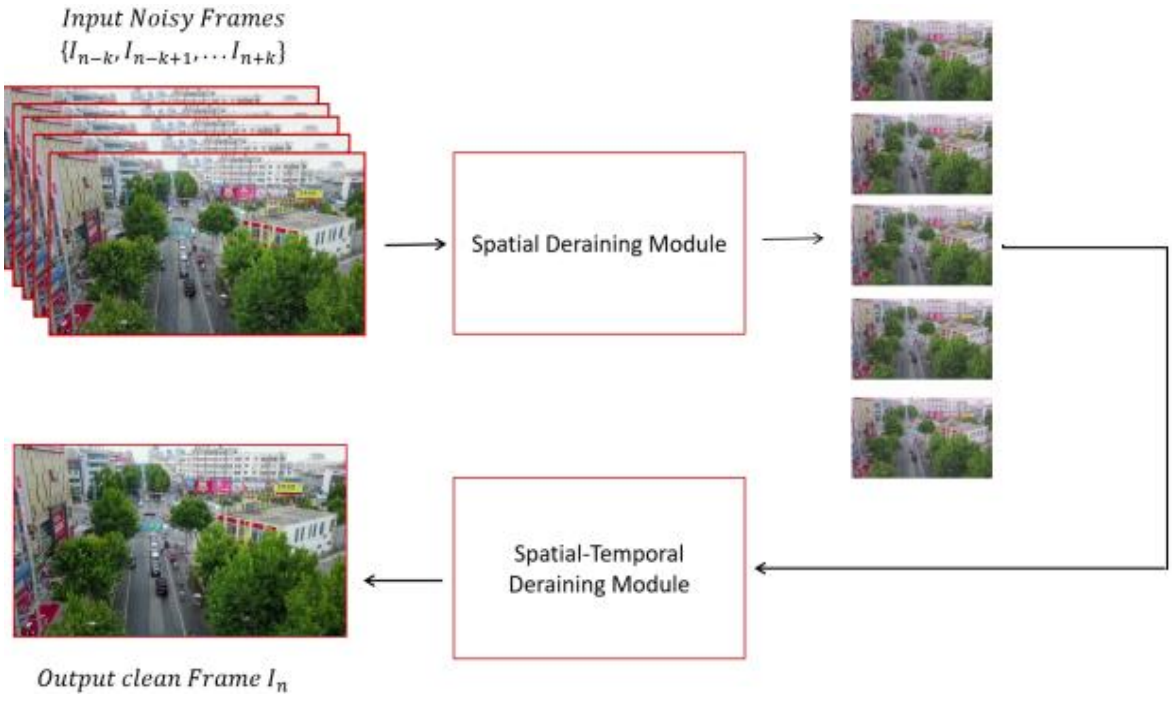


Figure 3.1 Structure of the proposed model.

3.3.1 The spatial deraining module

In the first module of the model, Noise2Noise (Lehtinen et al., 2018) model is implemented to spatially derain 2K+1 frames. The Noise2Noise model provides an approach that can map observations corrupted by rain to clean images by learning mappings between noisy image pairs. The model requires neither an explicit statistical likelihood model of the noise nor a prior image is required. Therefore, unlike most existing noise removal models that remove noise by subtracting noise layers, the Noise2Noise model learns the clean image itself. The theoretical background of this approach is that L2 loss learns to establish an average from the observations. For example, it has been verified that in training the pairs of low-resolution and high-resolution images, L2 learns the average of all plausible explanations, which results in spatial blurriness to network output (Ledig et al., 2017). Therefore, the authors of the Noise2Noise model argued that one could add zero-mean noise to training and target sets without downgrading the network outputs. In past research (Lehtinen et al., 2018), different types of noise were tested for N2N but rain noise was not considered.

In the present study, Noise2Noise is extended to deraining tasks and used as the spatial denoising module as the first part of the model. In the spatial denoising module, the loss function is:

$$L_N = \sum_{k=-K}^{K} ||\varphi(\hat{I}_{n+k}(x)) - \check{\hat{I}}_{n+k}(x)||_2 \qquad \text{Equation 3.3(e)}$$

Where $\hat{I}_{n+k}(x)$ is input noisy frames and $\check{\hat{I}}_{n+k}(x)$ is the noisy frames with rain noise identically and independently distributed with respect to the noise of $\hat{I}_{n+k}(x)$

3.3.2 The spatial-temporal deraining module

In the spatial-temporal denoising stage of the proposed model, FastDVDnet (Tassano et al., 2020) is used to further remove the noise due to object motion. Motion noise removal has always posed a challenging task in video processing. Conventional methods estimate the optical flow and fuse warped frames separately, while efforts have been made to promote the speed of video denoising by incorporating the two elements into one network by TOFlow (Xue et al., 2019). A more efficient network, namely FastDVDnet, is proposed as an alternative for motion estimation. As shown in Figure 3.2, the FastDVDnest architecture contains two stages. The FastDVDnet processes five consecutive frames in the first denoising block and feeds the concatenated features into the second denoising block. Using this model, the explicit estimation of optical flow is skipped, which avoids the distortions and artifacts due to erroneous flow and speeds up the training at the same time. Above all, the expensive computation of warping operations is also eliminated.

In the model, to prevent the FastDVDnet model from learning the identity, different noises are used when training the spatial denoising and the spatial-temporal modules. Instead of rain noise, Poisson noise is added when training FastDVDnet. Therefore, the loss function of the second module is:

$$L_F = \sum_{n=1}^{N} ||\Phi(\hat{I}_n(x)) - \hat{I}_n(x)||_2 \qquad \text{Equation 3.3(f)}$$

Where:

$\hat{I}_n(x)$ is the estimated result from the first spatial denoising block,

Φ represents the FastDVDnet model. Other symbols and subscripts are as defined earlier.

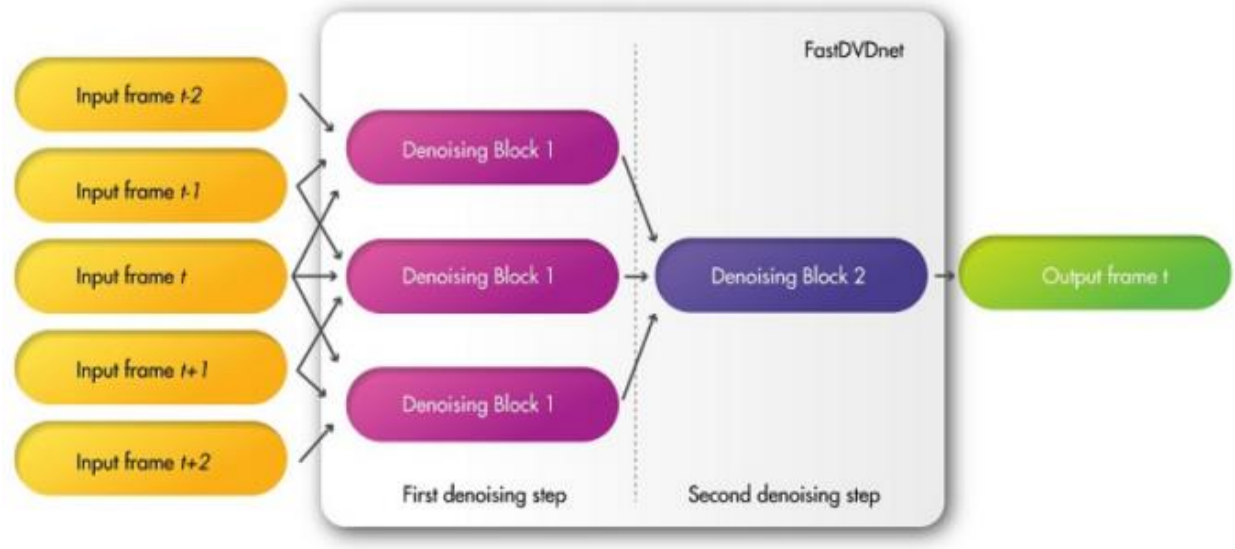


Figure 3.2 Architecture of the FastDVDnet (Tassano et al., 2020)

3.3.3 Rain noise synthesis

Instead of using images with real rain noise, the rain noise in the present study was synthesized for two reasons: first, the spatial draining module requires several rainy images of the same location. However, it is impossible to obtain such data in the real world since the real world is quite dynamic. Even during consecutive rainy days, the configuration (e.g., location of parked of moving vehicles, pedestrians, and color of plants) of a scene is never the same. Secondly, in the absence of clean-rainy image pairs, rainy images are synthesized from their clean versions and used for the analysis because it is needed to compare the deraining result against clean images for evaluation purposes. Rain impairs the visual quality of a scene and affects camera exposure time, depth of field and resolution. In most state-of-art deraining methods, the rain effect is assumed using specific models.

To properly describe rain streaks, it is required to model raindrop size, rain density and rain rotation due to wind. One way to carry out rain modeling is to carry out a linear superposition of the clean background and a layer of line-shaped rain streaks (Li et al., 2018). A rainy scene $\tilde{R}$ can be modeled as:

$$\tilde{R}=R+B \qquad \text{Equation 3.3 (g)}$$

where R is the clean background and B is the rain layer.

In the present study, the concept of transformed Gaussian noise is adopted. This is similar to the approach of PhotoShop software (Flores, 2019) to model the rain effect: the length and width of rain streak is added by noise stretching and the wind effect is modeled by rotating rain drops. Furthermore, rain also tends to introduce a mist effect to the video image. The accumulation and concentration of rain streaks forms an overall fog effect which renders the scene as a blurred image to the human eye. Therefore, a gaussian-blurring step is also implemented to introduce blurring effect to the rainy scene. Figure 3.3 compares a clean image and its rainy version.

(a) Clean image (image during good weather)

(b) Image during rain event

Figure 3.3 Example of a clean image and a corrupted image.

3.4 Experimental Setting

This study used a large-scale and high-quality Visdrone Dataset (Zhu et al., 2021), which is the same dataset as that used in the previous chapter, in the experiments. The dataset provides 400 videos clips formed by 265,228 frames, captured by drone-mounted cameras, covering various real-world scenes. Similar to FastDVDnet, this study realizes the flow map estimation using the DeepFlow algorithm (Weinzaepfel et al., 2013). During the training phase, five consecutive frames are used to recover one central frame and adopted sixty epochs for training. For the spatial deraining module, the current study used UNET (Ronneberger et al., 2015) to extract features from video frames. UNET has been proven to be an effective network to obtain image features in several tasks including classification and detection (Ronneberger et al., 2015). In the initial training of the Noise2Noise model, rain noise with a mean density of 300 and a standard deviation of 10 is added

to the source images while the rain noise with a mean density of 500 and a standard deviation of 20 is added to the target images.

To evaluate the deraining results, average Peak Signal-to-Noise Ratio (PSNR) values and Structural Similarity Index (SSIM) values are calculated and compared between clean frames and derained results for all sequences. For a video sequence, the PSNR and SSIM values are taken as the average score of each frame it holds. Given a reference image f and a test image g (f and g have the same size), the PSNR (dB) between f and g is defined by equations 3.4 (a) and 3.4(b) below.

$$\mathrm{PSNR}(f,g) = 10\log_{10}(255^2/\mathrm{MSE}(f,g)) \qquad \text{Equation 3.4(a)}$$

$$\text{where } \mathrm{MSE}(f,g) = \frac{1}{\mathrm{MN}}\sum_{\mathrm{i}=1}^{\mathrm{M}}\sum_{\mathrm{j}=1}^{\mathrm{N}}(f_{\mathrm{ij}} - g_{\mathrm{ij}})^2 \qquad \text{Equation 3.4(b)}$$

Therefore, a higher value of PSNR indicates a superior deraining result. The Structural Similarity Index (SSIM) is a perceptual metric that quantifies the image quality degradation. It is based on the notion that pixels have strong inter-dependencies particularly when they are in proximity. These dependencies carry important information about the structure of the objects in the visual scene. The SSIM of two images x, y is calculated as follows:

$$SSIM = \frac{(2\mu_x\mu_y+c_1)(2\sigma_x\sigma_y+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \qquad \text{Equation 3.4(c)}$$

Where:
μ_x is the average of x;
μ_y is the average of y;
σ_x is the variance of x;
σ_y is the variance of y;
c_1 and c_2 are the two variables used to stabilize the division;
L is the dynamic range of the pixel values;
$k_1 = 0.01$ and $k_2 = 0.03$ by default.

3.5 Study Results

3.5.1 Results of the deraining task

To demonstrate the efficacy of the proposed model, both quantitative comparison and visual comparison are prepared. Table 3.1 reports the quantitative experimental results. Compared to rainy videos, the proposed model achieves 16% and 10% improvement of PSNR and SSIM, respectively. Figure 3.4 presents a visualization of the model output. The sub-figures in the top line are rainy scenes and the sub-figures at the bottom are the results (derained images). The video sequences present a mixed traffic flow of pedestrians and vehicles. It is easily observed that the result from the proposed model has significantly higher visual quality compared to the noisy images (i.e., the rainy scenes). It would be extremely difficult to efficiently count the number of people or observe the interactions between vehicles from the rain-corrupted video frames. For example, in the heavy rain, it is almost impossible to count the number of pedestrians in the area

bounded by a red circle. However, the existence of pedestrians and the behavior of both pedestrians and vehicles are captured clearly in the derained video frames. Therefore, by providing a superior visual quality, the proposed model can be utilized as a powerful tool for traffic monitoring and management.

Table 3.1 Quantitative results

Method	PSNR	SSIM
Rainy Frames	19	0.76
Derained Results from the proposed model	22	0.89

(a) Video frames with rain noise

(b) Video frames processed by the proposed deraining model

Figure 3.4 (a) noisy frames (b) clean frames from the proposed model

3.5.2 Comparison and discussion

To further explore the performance of the self-supervised video deraining approach, the present study compared the proposed model and TOFlow (Xue et al., 2019) and GMM (Li et al., 2016b), which are existing flow-based denoising model and image-based denoising model, respectively. The image-based model is realized by processing each frame separately. Table 3.2 presents the difference of PSNR and SSIM values for derained results from different models. From the table, it can be stated that even though no prior ground truth is required, the proposed model still reaches comparable denoising performance as the supervised flow-based model TOFlow. Figure 3.5 presents a qualitative assessment of the visual outcomes from the proposed model and the image-based model. It can be noticed that the problem of blur boundary is significant in image-based models in the area highlighted by red circles. This is because the temporal information is considered when removing rain noise. Compared to image-based models, the proposed model exhibits superior temporal coherence.

Table 3.2 Quantitative comparison (rain density is the mean value for Gaussian rain noise)

Method	Rain Density of 500		Rain Density of 300	
	PSNR	SSIM	PSNR	SSIM
The proposed Model	18	0.87	22	0.89
GMM (Y. Li et al., 2016)	17	0.86	21	0.89
TOFlow (Xue et al., 2019)	19	0.90	22	0.92

In the proposed model, the two stages cooperate with (and benefit from) each other. Figure 3.6 and Figure 3.7 present an example of rainy scene and the deraining result extracted from the first stage, respectively. It is observed that the derained image from the first stage is already clean, which significantly reduces the challenges for the second denoising phase. Consequently, the second denoising phase could focus on motion noise removal and improve the output. However, when the original rain noise is heavy, the first deraining module may could not remove all noise. Under this scenario, the following spatial-temporal denoising module could further remove the residual spatial noise. For example, Figure 3.8 presents an example of the result from the first denoising module when the rain density is 800 (the rain density is defined by the level of gaussian noise). In this figure, a few rain streaks remain. However, as indicated in Figure 3.9, after the spatial-temporal denoising phase, the rain streak noise became invisible. Table 3.3 presents the PSNR and SSIM values of images with and without the second stage, and Figure 3.10 presents the image quality difference in a more direct way. The PSNR/SSIM difference value is obtained by subtracting the result of the model without the second stage from the result of the model with the second stage. It can be observed that the second stage further improves video quality based on the first stage, particularly where there exists significant rain noise.

(a) Results from the GMM model, an image-based model

(b) Results from the proposed model

Figure 3.5 Comparison between results from the proposed model and the image deraining model

Table 3.3 Results from the two models

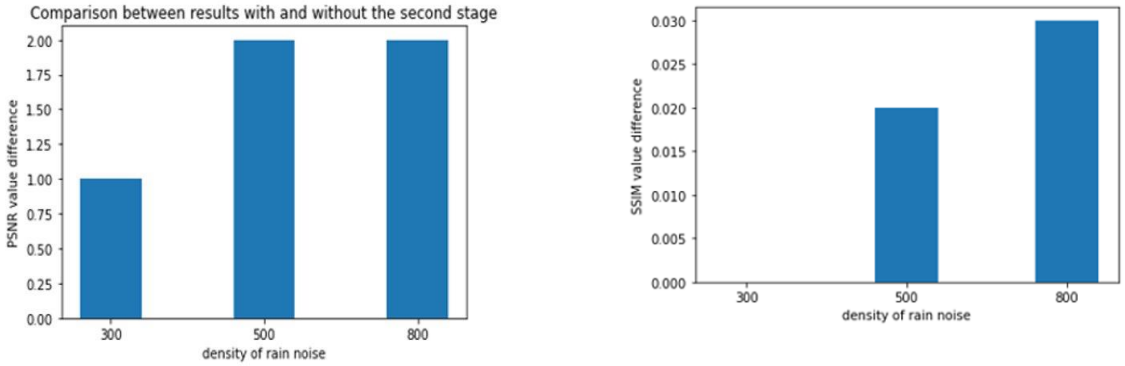
	Results **without** the second stage		Results **with** the second stage	
Rain Density	PSNR	SSIM	PSNR	SSIM
300	21	0.89	22	0.89
500	16	0.85	18	0.87
800	15	0.83	17	0.86

Figure 3.6 An example of a rainy frame

Figure 3.7 Output from spatial deraining block 1

Figure 3.8 Output from the spatial block may be not clean enough

Figure 3.9 Output from the spatial-temporal block



PSNR value difference SSIM value difference

Figure 3.10 Comparison of results with and without the second stage

3.6 Conclusions, study limitations and directions for future research

This report proposes a two-stage model for video deraining to addresses the challenge of rain-induced blur in video images for purposes of enhanced traffic monitoring and management. The model fully considers both rain noise and motion noise. In the spatial deraining stage, rain noise is first reduced from each frame. In the following spatial-temporal deraining stage, the motion noise is removed. A useful feature of the proposed model is that it is self-supervised, and therefore requires no ground-truth labels for video frames. Unlike previous work which was realized based on noisy-clean training pairs, the proposed model is built upon training through noisy-noisy pairs without significantly compromising the model performance. The experiments carried out using the Visdrone dataset showed that although no ground truth is utilized in the proposed model, it demonstrates performance comparable to that of supervised models. The self-supervised approach is achieved by executing the Noise2Noise model in the spatial denoising stage and then connecting it to FastDVDnet which serves as a spatial-temporal denoising module. The spatial-temporal denoising block strives to map the corrupted video frames back to the results generated by spatial block instead of ground-truth clean frames. By removing rain noise from the scenes, the traffic video attains more pleasing visual quality. The clean frames without noise not only demonstrate behaviors of individual vehicles and pedestrians in a clearer image, but also provide visually friendly presentations of moving objects by removing the noised resulting from high-speed movement. This could potentially improve the efficiency of video-supported roadway monitoring.

There are several limitations of this work that give rise to opportunities for further research in this domain. First, the corrupted image used synthesized rain rather than real rain. Although synthesized rain has been widely used in deraining research, limited work has been done to verify whether the synthesized images capture the features of real-world rain. This could be addressed in future research. Secondly, this work considered rain only in streak form. However, rain could also manifest in other shapes including circular raindrops which may cover a part of camera lens or rain mist that tend to span the entire environment space. It is expected that the requisite deraining methods will be different across these forms of rain representation. In the future, the model could be tested on other types of rain representations and its performance assessed vis-à-vis other models. In addition, other self-supervised denoising models (e.g., self2void, self2self) could be investigated for their applicability to video deraining and their efficiency.

CHAPTER 4. OVERALL CONCLUDING REMARKS

UAVs represent potentially a key fundamental infrastructure element of intelligent transportation systems of the future. This report investigated the potential use of UAVs in monitoring the road environment. In addition, recognizing that the quality of UAV videos are often corrupted by rain noise in reality, this report developed a self-supervised framework to remove rain noise from the UAV-acquired video. The contributions of this study are:

- A framework using UAV-captured video to monitor traffic and assess potential crash risks at intersections
- A self-supervised learning approach to remove the rain streaks in videos, enabling better visual quality of traffic monitoring videos
- Identification and validation of the merits of the proposed self-supervised and two-stage deraining model in video image denoising tasks.

The proposed models can help traffic engineers to acquire an UAV's bird-eye perspective to monitor intersection traffic operations, potentially adequately assess safety conditions at intersections, and design intersection infrastructures not only in the current era but also in the prospective era of CAVs. Using the UAV-acquired data, real-time and precise vehicle data could be conveyed to roadway and traffic monitoring centers to facilitate incident response and traffic management. The system could not only monitor existing traffic conditions at intersections but also assess potential traffic conflicts to provide warnings to drivers and other ground entities. Also, the proposed denoising framework provides better visual quality of the videos captured by UAV in inclement weather. This is particularly important at cities with significant rain events and where traffic needs to be monitored daily. With efficient routing advice from UAVs, it is anticipated that the overall safety and efficiency of the entire surface transportation system could be enhanced.

There are a few limitations of this study which are indicative of possible future work improvements. First, full consideration was not given to the efficiency of data transmission. UAVs would need to be connected to vehicles or transportation management centers through wireless communication systems to deliver traffic data in real time. It is challenging to realize real-time transmission of high-quality video data and therefore, this issue of developing reliable data transmission networks should be addressed in future work. Secondly, the report did not consider the aerial stability of UAVs, particularly during high-wind events. The instability of an UAV could impair the quality of the images it captures, and future work could develop machine learning algorithms to acquire a clean image from one distorted by movement of the camera due to wind. Thirdly, only traffic conflicts based on TTC are considered as the safety surrogate measures in this this report. In the future work, more comprehensive surrogate measures will be considered to assess the crash probability with the support more traffic and infrastructure information.

CHAPTER 5 SYNOPSIS OF PERFORMANCE INDICATORS

5.1 USDOT performance indicators I

Two (2) transportation-related courses were offered annually during the study period that was taught by the PI and a teaching assistant who are associated with the research project. One of these was a newly developed course that was inspired and directly associated with this CCAT research. Two (2) graduate students and one (1) post-doctoral researcher (subsequently designated a Visiting Assistant Professor) participated in the research project during the study period. One (1) transportation-related advanced degree program (a doctoral program) utilized the CCAT grant funds from this research project, during the study period to support the graduate students.

5.2 USDOT performance indicators II

Research Performance Indicators: 1 journal article, 3 conference presentations were produced from this project. The research from this advanced research project was disseminated to over 320 people from industry, government, and academia, through 3 conference presentations. These include the 2022 ASCE International Conference on Transportation and Development, the 2022 Purdue Road School, and the 2022 TRB Annual Meeting.
Leadership Development Performance Indicators: This research project generated 3 academic engagements and 2 industry engagements. The PI's held positions in 2 national organizations that address issues related to this research project.
Education and Workforce Development Performance Indicators: The methods, data and/or results from this study were incorporated in the class content of several versions (Fall 2022, Spring 2023, and Fall 2023) of the following courses at Purdue University's undergraduate civil engineering program: (a) CE 299 (Smart Mobility), an optional undergraduate-level course, and (b) CE 398 (Introduction to Civil Engineering Systems), a mandatory undergraduate course. The students in these classes will soon be entering the workforce. Thereby, the research helped enlarge the pool of people trained to develop knowledge and utilize the technologies developed in this research, and prospectively, to put them to use when they enter the workforce.
Collaboration Performance Indicators: There was collaboration with other agencies, and one (1) agency and four (4) academic institutions provided matching funds.

The outputs, outcomes, and impacts are described in Chapter 6.

CHAPTER 6. STUDY OUTCOMES AND OUTPUTS

6.1 Outputs

6.1.1 Publications, conference papers, or presentations

(a) Journal Papers

Zong, S., Chen, S., Alinizzi, M., & Labi, S. (2022). Leveraging UAV Capabilities for Vehicle Tracking and Collision Risk Assessment at Road Intersections. Sustainability 14(7), 4034. https://doi.org/10.3390/su14074034

(b) Conference Presentations

Zong, S., Chen, S., Alinizzi, M., & Labi, S. (2022). Using UAVs for vehicle tracking and collision risk assessment at intersections, 2022 Purdue Road School, March 15-16, West Lafayette, Indiana.

Zong, S., Chen, S., & Labi, S. (2022). Towards Safer Transportation: a self-supervised learning approach for traffic video deraining, 2022 TRB Annual Meeting, Washington, DC.

Li, Y., Chen, S., Labi, S. (2022). Optimal trajectory planning using microscopic traffic estimation from UAV videos. 2022 Purdue Road School, March 15-16, West Lafayette, Indiana.

Li, Y., Chen, S., Labi, S. (2022). Utilizing UAV video for autonomous vehicle trajectory planning, 2022 ASCE International Conference on Transportation and Development (ICTD), May 31-June 3, Seattle, Washington.

Zong, S., Chen, S., Alinizzi, M., & Labi, S. (2022). Leveraging UAV capabilities for vehicle tracking and collision risk assessment at road intersections, 2022 ASCE International Conference on Transportation & Development (ICTD), May 31-June 3, Seattle, Washington.

6.1.2. Other outputs

This study developed a framework for using UAV-captured videos to monitor traffic and assess crash risks at intersections, and a self-supervised learning algorithm for rain noise removal from UAV images. The research produced analytical methods associated with trajectory planning and image de-noising, to:

- help teach relevant concepts in two (2) Purdue undergraduate-level courses: CE 299 (Smart Mobility) and CE 398 (Introduction to Civil Engineering Systems).
- support future research related to the subjects of UAV-CAV network, trajectory planning and operational monitoring for CAVs.

6.2 Outcomes

The outcomes of this project are the prospective changes that an UAV-enabled intelligent transportation ecosystem could bring to the road transportation system, or its regulatory, legislative, or policy framework. These are:

- Increased use of UAVs as mobile sensors in monitoring emerging road technology environments that are driven by internet of things (IoT) and containing CAVs. This could lead to changes in road agency policies and functions regarding road infrastructure monitoring and inspections,
- Development of policies and standards related to the quality of images on road environments, and guidelines for remediation.

6.3 Impacts

The products from this project can affect the performance of road corridors and urban intersections, in terms of reduced delay (and therefore, lower operating costs), environmental benefits (reduced emissions), and community benefits in terms of reduced pedestrian injuries and fatalities. Therefore, the study results can help increase the body of knowledge and technologies in the context of road corridor and intersection monitoring and management. In this context, the study has helped enlarge the pool of people trained to develop knowledge and utilize new technologies and put them to use, and improve the physical, institutional, and information resources that facilitate access of prospective elements of the future workforce to training and new technologies. A list of specific impacts from this research project, are as follows:

- According to the U.S. Federal Highway Administration in 2016, proper evaluation of transportation facility performance has always been supported by legislation. This is important because there is a growing demand for information on traffic patterns, to support general transportation administration and management, and in particular, the development and evaluation of road safety policies. In this context, the proposed UAV-supported traffic monitoring framework can be beneficial to road agencies because it can generate large amounts of useful real-time traffic data.
- The products of this study enhance the capability of UAVs, and their applications to traffic monitoring. By virtue of their exceptional flexibility, extensive monitoring range, aerial perspective, and cost-effectiveness in comparison to conventional traffic sensors, UAVs can be deployed in CAV ecosystems to produce reliable, high-quality images and videos. Therefore, the impacts of the study product on transportation practice are expected to be far reaching. These include enhancements to accident risk assessment and mitigation, traffic flow obstacle detection and removal, post-accident reconstruction, traffic flow parameter estimation, and CAV trajectory planning.
- The demonstrated benefits of the study products, in terms of safety and operational efficiency enhancements, will hopefully help build motivation for manufacturers of UAVs, CAVs, technology companies, and road agencies to invest in infrastructure that promotes connectivity between UAV and CAVs.
- The two graduate students that worked on this project will enter the workforce in 2024 to help support the workforce that will implement and/or improve the algorithms and methods developed in this study.

REFERENCES

Abrari Vajari, M.; Aghabayk, K.; Sadeghian, M.; Shiwakoti, N. A multinomial logit model of motorcycle crash severity at Australian intersections. J. Saf. Res. 2020, 73, 17–24.

Allen, B.L., Shin, B.T., and Cooper, P.J. 1978. Analysis of traffic conflicts and collisions. Report No. TRR 667, Transportation Research Board, Washington, D.C.

American National Standards Institute. (2018, December 18). Manual on Classification of Motor Vehicle Traffic Crashes. (n.d.). Retrieved May 18, 2022, from https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/ansi_d16-2017.pdf

Amin, M. A., Abdullah, S., Abdul Mukti, S. N., Mohd Zaidi, M. H. A., & Tahar, K. N. (2020). Reconstruction of 3D accident scene from multirotor UAV platform. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 43(B2). https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-451-2020

Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., & Zuair, M. (2017). Deep learning approach for car detection in UAV imagery. Remote Sensing, 9(4), 312-322. https://doi.org/10.3390/rs9040312

Ammoun, S., Nashashibi, F. (2009). Real time trajectory prediction for collision risk estimation between vehicles. In Proceedings of the 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 417–422.

Ardestani, S. M., Jin, P. J., Volkmann, O., Gong, J., Zhou, Z., & Feeley, C. (2016). 3D accident site reconstruction using unmanned aerial vehicles (UAV), presented at the Annual Meeting of the Transportation Research Board, Washington DC, United States.

Bagdadi, O. (2013). Assessing safety critical braking events in naturalistic driving studies, Transportation Research Part F: Traffic Psychology & Behavior, 16(1), 117-126.

Barmpounakis, E. N., Vlahogianni, E. I., Golias, J. C., & Babinec, A. (2019). How accurate are small drones for measuring microscopic traffic parameters? Transportation Letters, 11(6). https://doi.org/10.1080/19427867.2017.1354433

Barry, B. K. (2019). Guide to forward collision warning. Consumer Reports. https://www.consumerreports.org/car-safety/guide-to-forward-collision-warning-a8423384882/ Retrieved May 17, 2022.

Bay, H., Tuytelaars, T., & Gool, L. V. (2008). Surf: Speeded up robust features. In European Conference on Computer Vision, 404-417, Springer, Berlin, Heidelberg.

Blincoe, L., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2014). The economic and

societal impact of motor vehicle crashes, Publication Nr. DOT HS 812 013, Transportation Research Board. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013.

Bose, B., Wang, X., and Grimson, E. (2007). Multi-class object tracking algorithm that handles fragmentation and grouping, 2007 IEEE Conference on Computer Vision and Pattern Recognition, 1-8, doi: 10.1109/CVPR.2007.383175.

Brewer, N., & Liu, N. (2008). Using the shape characteristics of rain to identify and remove rain from video. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5342 LNCS. https://doi.org/10.1007/978-3-540-89689-0.

Buades, A., Lisani, J. L., & Miladinović, M. (2016). Patch-based video denoising with optical flow estimation. IEEE Transactions on Image Processing, 25(6), 2573-2586.

Caballero, J., Ledig, C., Aitken, A., Acosta, A., Totz, J., Wang, Z., & Shi, W. (2017). Real-time video super-resolution with spatio-temporal networks and motion compensation. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January. https://doi.org/10.1109/CVPR.2017.304

Cao, X., Gao, C., Lan, J., Yuan, Y., & Yan, P. (2014). Ego motion guided particle filter for vehicle tracking in airborne videos. Neurocomputing, 124, 168-177, https://doi.org/10.1016/j.neucom.2013.07.014

Center for Disease Control and Prevention. (2020). Road Traffic Injuries and Deaths—A Global Problem. www.cdc.gov/injury/features/global-road-safety/index.html, Retrieved May 17, 2022.

Chen, P., Zeng, W., Yu, G., & Wang, Y. (2017). Surrogate safety analysis of pedestrian-vehicle conflict at intersections using unmanned aerial vehicle videos. Journal of Advanced Transportation, Article ID 5202150. https://doi.org/10.1155/2017/5202150

Cooper, P.J. (1984). Experience with traffic conflicts in Canada with emphasis on post encroachment time techniques, International Calibration Study of Traffic Conflicts, NATA ASI Series F5, 75–96.

Dalal, D N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, 886-893.

Davy, A., Ehret, T., Morel, J. M., Arias, P., & Facciolo, G. (2021). Video denoising by combining patch search and CNNs, Journal of Mathematical Imaging & Vision, 63(1), 73-88.

Dingus, T. A., Neale, V. L., Klauer, S. G., Petersen, A. D., & Carroll, R. J. (2006). The development of a naturalistic data collection system to perform critical incident analysis: An

investigation of safety and fatigue issues in long-haul trucking. Accident Analysis & Prevention, 38(6), 1127-1136.

Dong, J., Chen, S., Li, Y., Du, R., Steinfeld, A., & Labi, S. (2021). Space-weighted information fusion using deep reinforcement learning: The context of tactical control of lane-changing autonomous vehicles and connectivity range assessment, Transportation Research Part C: Emerging Technologies, 128, 103192.

Dozza, M., & Gonzalez, N. P. E. (2012). Recognizing safety critical events from naturalistic driving data. Procedia-social & Behavioral Sciences, 48, 505-515.

Du, R., Chen, S., Li, Y., Ha, P. Y. J., Dong, J., Anastasopoulos, P. C., & Labi, S. (2021). A cooperative crash avoidance framework for autonomous vehicle under collision-imminent situations in mixed traffic stream, in 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 1997-2002.

Eigen, D., Krishnan, D., & Fergus, R. (2013). Restoring an image taken through a window covered with dirt or rain. In Proceedings of the IEEE International Conference on Computer Vision, 633-640.

Elad, M., & Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing, 15(12), 3736-3745.

Federal Highway Administration. (2021). Case studies for FHWA pedestrian and bicycle focus states and cities, Tech. Rep. FHWA-SA-21-02, U.S. Dept. of Transportation, Washington, DC. https://safety.fhwa.dot.gov/ped_bike/ped_focus/docs/FHWA_FocusApproach_CaseStudies_508.pdf.

Federal Highway Administration. (2022). About intersection safety. Safety. https://safety.fhwa.dot.gov/intersection/about/#:~:text=Signalized%20Intersection%20Crashes&text=Traffic%20signals%20are%20often%20chosen,that%. Retrieved May 3, 2022.

Feizi, A., Joo, S., Kwigizile, V., & Oh, J. S. (2020). A pervasive framework toward sustainability and smart growth: Assessing multifaceted transportation performance measures for smart cities. Journal of Transport & Health, 19, 100956.

Flores, T. (2019). Gaussian Blurring with Python and OpenCV - Analytics Vidhya. Medium. https://medium.com/analytics-vidhya/gaussian-blurring-with-python-and-opencv-ba8429eb879b Retrieved May 15, 2022.

Fu, X., Huang, J., Ding, X., Liao, Y., & Paisley, J. (2017). Clearing the skies: A deep network architecture for single-image rain removal. IEEE Transactions on Image Processing, 26(6), 2944-2956.

Garg, K., & Nayar, S. K. (2004). Detection and removal of rain from videos, in proceedings of the 2004 IEEE Computer Society Conference on Computer Vision & Pattern Recognition, CVPR 2004, Vol. 1.

Garg, K., & Nayar, S. K. (2007). Vision and rain. International Journal of Computer Vision, 75(1), 3-27.

Gelbal, S. Y., Arslan, S., Wang, H., Aksun-Guvenc, B., & Guvenc, L. (2017). Elastic band based pedestrian collision avoidance using V2X communication. IEEE Intelligent Vehicles Symposium, Proceedings, 270–276. https://doi.org/10.1109/IVS.2017.7995731.

Gettman, D., and Head, L. (2003) . Surrogate safety measures from traffic simulation models. Report No. FHWA-RD-03-050, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Ghosal, A., Conti, M. (2020). Security issues and challenges in V2X: A Survey, Computer NetworksM Volume 169, 107093,

Gomaa, A., Abdelwahab, M. M., & Abo-Zahhad, M. (2019). Real-time algorithm for simultaneous vehicle detection and tracking in aerial view videos. Midwest Symposium on Circuits and Systems, August 2018. https://doi.org/10.1109/MWSCAS.2018.8624022

Gu, X., Abdel-Aty, M., Xiang, Q., Cai, Q., Yuan, J. (2019). Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas, Accident Analysis & Prevention Volume 123, 159-169

Gupta, A., Afrin, T., Scully, E., & Yodo, N. (2021). Advances of UAVs toward future transportation: The state-of-the-art, challenges, and opportunities. Future Transportation, 1(2), 326-350.

Hamilton, L., Wuellner, C., Grove, K., & Considine, J. (2011). 2011 Pedestrian crash analysis, Report Nr. 29, City of Chicago Center for Education and Research in Safety. https://www.chicago.gov/dam/city/depts/cdot/pedestrian/2011PedestrianCrashAnalysisSummaryReport.pdf

Hayward, J. C. (1972). Near miss determination through use of a scale of danger. 51st Annual Meeting of the Highway Research Board, Washington DC.

Hu, W., Li, X., Luo, W., Zhang, X., Maybank, S., & Zhang, Z. (2012). Single and multiple object tracking using log-Euclidean Riemannian subspace and block-division appearance model, IEEE Transactions on Pattern Analysis & Machine Intelligence, 34(12), 2420–2440.

Huang, D. A., Kang, L. W., Wang, Y. C. F., & Lin, C. W. (2013). Self-learning based image decomposition with applications to single image denoising. IEEE Transactions on Multimedia, 16(1). https://doi.org/10.1109/TMM.2013.2284759

Jeon, C.M., Amekudzi, A.A. (2005). Addressing sustainability in transportation systems: definitions, indicators, and metrics, Journal of infrastructure Systems 11 (1), 31-50.

Junior, O. L., Delgado, D., Gonçalves, V., & Nunes, U. (2009). Trainable classifier-fusion schemes: An application to pedestrian detection, in: 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO.

Kanistras, K., Martins, G., Rutherford, M. J., & Valavanis, K. P. (2015). Survey of unmanned aerial vehicles (UAVs) for traffic monitoring, in: Handbook of Unmanned Aerial Vehicles. https://doi.org/10.1007/978-90-481-9707-1_122

Ke, R., Li, Z., Kim, S., Ash, J., Cui, Z., & Wang, Y. (2017). Real-time bidirectional traffic flow parameter estimation from aerial videos, IEEE Transactions on Intelligent Transportation Systems, 18(4). https://doi.org/10.1109/TITS.2016.2595526

Ke, R., Li, Z., Kim, S., Ash, J., Cui, Z., Wang, Y. Real-time bidirectional traffic flow parameter estimation from aerial Videos. IEEE Trans. Intell. Transp. Syst. 2017, 18, 890–901.

Kerkhoff, J. F. (1985). Photographic techniques for accident reconstruction, Science Engineering Medicine, https://trid.trb.org/view/272328, retrieved May 18, 2022.

Keymakr. (2021, May 1). Image Processing Techniques: What Are Bounding Boxes. Keymakr Intro to Machine Learning. Retrieved May 6, 2022, from https://keymakr.com/blog/what-are-boundingboxes/

Khan, M. A., Ectors, W., Bellemans, T., Janssens, D., & Wets, G. (2017). Unmanned aerial vehicle-based traffic analysis: Methodological framework for automated multivehicle trajectory extraction. Transportation Research Record, 2626(1), 25-33. https://doi.org/10.3141/2626-04

Khan, M. A., Ectors, W., Bellemans, T., Janssens, D., & Wets, G. (2018). Unmanned aerial vehicle-based traffic analysis: A case study for shockwave identification and flow parameters estimation at signalized intersections. Remote Sensing, 10(3). https://doi.org/10.3390/rs10030458

Kiela, K., Barzdenas, V., Jurgo, M., Macaitis, V., Rafanavicius, J., Vasjanov, A., Kladovscikov, L., Navickas, R. Review of V2X–IoT Standards and Frameworks for ITS Applications. Appl. Sci. 2020, 10, 4314.

Kim, N.V., Chervonenkis, M.A. Situation control of unmanned aerial vehicles for road traffic monitoring. Mod. Appl. Sci. 2015, 9, 1

Kim, E. J., Park, H. C., Ham, S. W., Kho, S. Y., Kim, D. K., & Hassan, Y. (2019). Extracting vehicle trajectories using unmanned aerial vehicles in congested traffic conditions. Journal of Advanced Transportation, Article ID 9060797. https://doi.org/10.1155/2019/9060797

Kure, M. How drones become a valuable tool for the auto insurance industry. Available online: https://www.forbes.com/sites/sap/2020/01/29/how-drones-become-a-valuable-tool-for-the-auto-insurance-industry/?sh=691002aa1ac9

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T. (2018). Noise2Noise: Learning image restoration without clean data. Proceedings of the 35th International Conference on Machine Learning, in Proceedings of Machine Learning Research 80, 2965-2974 https://proceedings.mlr.press/v80/lehtinen18a.html.

Li, J., Chen, S., Zhang, F., Li, E., Yang, T., & Lu, Z. (2019). An adaptive framework for multi-vehicle ground speed estimation in airborne videos. Remote Sensing, 11(10). https://doi.org/10.3390/rs11101241

Li, J., Ye, D. H., Chung, T., Kolsch, M., Wachs, J., & Bouman, C. (2016a). Multi-target detection and tracking from a single camera in unmanned aerial vehicles (UAVs). IEEE International Conference on Intelligent Robots and Systems, https://doi.org/10.1109/IROS.2016.7759733

Li, Y., Tan, R. T., Guo, X., Lu, J., & Brown, M. S. (2016b). Rain streak removal using layer priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2736-2744.

Li, M., Xie, Q., Zhao, Q., Wei, W., Gu, S., Tao, J., & Meng, D. (2018). Video rain streak removal by multiscale convolutional sparse coding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6644-6653.

Liu, J., Yang, W., Yang, S., & Guo, Z. (2019). D3r-net: Dynamic routing residue recurrent network for video rain removal. IEEE Transactions on Image Processing, 28(2), 699-712.

Lovelace, B. Unmanned Aerial Vehicle Bridge Inspection Demonstration Project. Available online: https://www.lrrb.org/pdf/201540.pdf (accessed on 20 February 2022).

Lu, G., Liu, M., Wang, Y., & Yu, G. (2012). Quantifying the severity of traffic conflict by assuming moving elements as rectangles at intersection. Procedia-Social & Behavioral Sciences, 43, 255-264.

Lucas, B. D., & Kanade, T. (1981). Iterative image registration technique with an application to stereo vision, Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81).

Maggioni, M., Boracchi, G., Foi, A., & Egiazarian, K. (2012). Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms, IEEE Transactions on Image Processing, 21(9). https://doi.org/10.1109/TIP.2012.2199324

Mahmud, S. S., Ferreira, L., Hoque, M. S., & Tavassoli, A. (2017). Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs. IATSS research, 41(4), 153-163.

Maji, S., Berg, A. C., & Maliks, J. (2008). Classification using intersection kernel support vector machines is efficient, 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR. https://doi.org/10.1109/CVPR.2008.4587630

Mehmood, S., Ahmed, S., Kristensen, A. S., & Ahsan, D. (2018, May). Multi criteria decision analysis (MCDA) of unmanned aerial vehicles (UAVS) as a part of standard response to emergencies. In 4th International Conference on Green Computing and Engineering Technologies (p. 31). Gyancity International Publishers.

Michał, M., Adam, W., Jeffery, M. PwC's Global Report on the Commercial Applications of Drone Technology. Available online: https://www.pwc.pl/pl/pdf/clarity-from-above-pwc.pdf (accessed on 20 February 2022).

Misener, J. A., Biswas, S., & Larson, G. (2011). Development of V-to-X systems in North America: The promise, the pitfalls and the prognosis. Computer Networks, 55(14), 3120–3133.

Monticello, B. M. (2019). Car safety systems that could save your life. Consumer Reports. Retrieved May 18, 2022, from https://www.consumerreports.org/automotive-technology/car-safety-systems-that-could-save-your-life/

Morency, P., and Cloutier, M. (2007). From targeted "black spots" to area-wide pedestrian safety, Injury Prevention 12(6):360-4. DOI:10.1136/ip.2006.013326

Nadimi, N., NaserAlavi, S. S., & Asadamraji, M. (2020). Calculating dynamic thresholds for critical time to collision as a safety measure. In Proceedings of the Institution of Civil Engineers-Transport, Thomas Telford Ltd.

Nath, T. How Drones Are Changing the Business World. Available online: https://www.investopedia.com/articles/investing/010615/how-drones-are-changing-business-world.asp (accessed on 20 February 2022).

Niklaus, S., & Liu, F. (2018). Context-aware synthesis for video frame interpolation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2018.00183

Northmore, A., & Hildebrand, E. (2019). Intersection characteristics that influence collision severity and cost. Journal of Safety Research, 70, 49-57.

Outay, F., Mengash, H.A., Adnan, M. (2020). Applications of unmanned aerial vehicle (UAV) in road safety, traffic, and highway infrastructure management: Recent advances and challenges. Transp. Res. Part A Policy Pract., 141, 116–129.

Parker Jr, M. R., & Zegeer, C. V. (1989). Traffic conflict techniques for safety and operations: Observers manual (No. FHWA-IP-88-027, NCP 3A9C0093). United States. Federal Highway Administration, Washington, DC.

Peden, M. (2005). Global collaboration on road traffic injury prevention, International Journal of Injury Control & Safety Promotion, 12(2), 85-91.

Peng, L., Sotelo, M.A., He, Y., Ai, Y., Li, Z. (2004). A Method for Vehicle Collision Risk Assessment through Inferring Driver's Braking Actions in Near-Crash Situations. arXiv 2020, arXiv:2004.13761

Pérez, J. A., Gonçalves, G. R., Rangel, J. M. G., & Ortega, P. F. (2019). Accuracy and effectiveness of orthophotos obtained from low cost UASs video imagery for traffic accident scenes documentation. Advances in Engineering Software, 132, 47–54.

Prevot, T., Rios, J., Kopardekar, P., Robinson, J.E., III, Johnson, M., Jung, J. (2016). The UAS Traffic Management (UTM) Concept of Operations to Safely Enable Low Altitude Flight Operations. In Proceedings of the 16th American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations Conference, Washington, DC.

Qian, R., Tan, R. T., Yang, W., Su, J., & Liu, J. (2018). Attentive generative adversarial network for raindrop removal from a single image. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2018.00263

Raj, C.V., Sree, B.N., Madhavan, R. (2017). Vision Based Accident Vehicle Identification and Scene Investigation. In Proceedings of the 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, India, 14–16 July 2017, pp. 1–5

Rodríguez-Canosa, G. R., Thomas, S., del Cerro, J., Barrientos, A., & MacDonald, B. (2012). A real-time method to detect and track moving objects (DATMO) from unmanned aerial vehicles (UAVs) using a single camera, Remote Sensing, 4(4). https://doi.org/10.3390/rs4041090

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9351. https://doi.org/10.1007/978-3-

Roser, M., & Geiger, A. (2009). Video-based raindrop detection for improved image registration. In 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops (pp. 570-577). IEEE.

Shah, D., & Lee, C. (2021). Analysis of effects of driver's evasive action time on rear-end collision risk using a driving simulator. Journal of Safety Research, 78, 242-250.

Sharma, V. Chen, H.C. Kumar, R. (2017). Driver behavior detection and vehicle rating using multi-UAV coordinated vehicular networks. J. Comput. Syst. Sci. 86, 3–32.

Shehata, M. S., Cai, J., Badawy, W. M., Burr, T. W., Pervez, M. S., Johannesson, R. J., & Radmanesh, A. (2008). Video-based automatic incident detection for smart roads: The outdoor environmental challenges regarding false alarms. IEEE Transactions on Intelligent Transportation Systems, 9(2), 349-360.

Shen, L., Wu, Y., & Zhang, X. (2011). Key assessment indicators for the sustainability of infrastructure projects. Journal of Construction Engineering & Management, 137(6), 441-451.

Sinha, K. C., & Labi, S. (2007). Transportation decision making: Principles of project evaluation & programming. John Wiley & Sons.

Song, B., Jeng, T.-Y., Staudt, E., & Chowdhury, A. K. (2010). A stochastic graph evolution framework for robust multi-target tracking. European Conference on Computer Vision, 605–619.

Su, S., Liu, W., Li, K., Yang, G., Feng, C., Ming, J., Liu, G., Liu, S., & Yin, Z. (2016). Developing an unmanned aerial vehicle-based rapid mapping system for traffic accident investigation. Australian Journal of Forensic Sciences, 48(4), 454–468.

Sun, S. H., Fan, S. P., & Wang, Y. C. F. (2014). Exploiting image structural similarity for single image rain removal. 2014 IEEE International Conference on Image Processing, ICIP 2014. https://doi.org/10.1109/ICIP.2014.7025909

Taleb, T., Benslimane, A., & Letaief, K. (2010). Toward an effective risk-conscious and collaborative vehicular collision avoidance system. IEEE Transactions on Vehicular Technology, 59(3), 1474–1486.

Tarko, A. P. (2019). Measuring road safety with surrogate events. In Measuring Road Safety with Surrogate Events. https://doi.org/10.1016/C2016-0-00255-3

Tarko, A.P. (2012). Use of crash surrogates and exceedance statistics to estimate road safety. Accident Analysis & Prevention, 45(1), 230–240. doi:10.1016/j.aap. 2011.07.008. PMID:22269505.

Tassano, M., Delon, J., & Veit, T. (2020). FastDVDNet: Towards real-time deep video denoising without flow estimation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR42600.2020.00143

Teutsch, M., & Krüger, W. (2012). Detection, segmentation, and tracking of moving objects in UAV videos. Proceedings - 2012 IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2012. https://doi.org/10.1109/AVSS.2012.36

Themann, P., Kotte, J., Raudszus, D., & Eckstein, L. (2015). Impact of positioning uncertainty of vulnerable road users on risk minimization in collision avoidance systems. 2015 IEEE Intelligent Vehicles Symposium (IV), 1201–1206.

Timmermans, J. S., & Beroggi, G. E. G. (2000). Conflict resolution in sustainable infrastructure management. Safety Science, 35(1–3). https://doi.org/10.1016/S0925-7535(00)00030-8

Transportation Research Board. (2010). Highway capacity manual. National Academic of Sciences, Washington, DC.

Tu, Y., Wang, W., Li, Y., Xu, C., Xu, T., & Li, X. (2019). Longitudinal safety impacts of cooperative adaptive cruise control vehicle's degradation. Journal of Safety Research, 69, 177-192.

U.S. Department of Transportation. (2009). Evaluation of the Focused Approach to Pedestrian Safety Program.
https://safety.fhwa.dot.gov/ped_bike/ped_focus/efapsp020509/efapsp020509.pdf

U.S. Federal Highway Administration (FHWA). (2018, May). National Performance Management Measures. https://www.govinfo.gov/content/pkg/FR-2018-05-31/pdf/2018-11652.pdf

van der Horst, A.R.A. (1990). A time-based analysis of road user behavior in normal and critical encounter. Ph.D. Dissertation, Delft University of Technology, Delft, Netherlands.

Vivek, A. K., Khan, T., & Mohapatra, S. S. (2021). Safety and associated parameters influencing performance of railroad grade crossings: A critical review of state of the art. Journal of safety Research, 79, 257-272.

Vogel, K. 2003. Modelling driver behavior – A control theory-based approach. Ph.D. Dissertation, University of Linköping. Katja, V. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1601554&dswid=-5456

Vukadinovic, V., Bakowski, K., Marsch, P., Garcia, I. D., Xu, H., Sybis, M., ... & Thibault, I. (2018). 3GPP C-V2X and IEEE 802.11 p for Vehicle-to-vehicle communications in highway platooning scenarios. Ad Hoc Networks, 74, 17-29.

Wang, C., Zhou, S. K., & Cheng, Z. (2020). First image then video: A two-stage network for spatiotemporal video denoising. http://arxiv.org/abs/2001.00346

Weinzaepfel, P., Revaud, J., Harchaoui, Z., & Schmid, C. (2013). DeepFlow: Large displacement optical flow with deep matching. In Proceedings of the IEEE international Conference on Computer Vision, 1385-1392.

World Health Organization. (2018). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries, retrieved May 18, 2022.

Wu, C., Peng, L., Huang, Z., Zhong, M., Chu, D. A (2014). Method of vehicle motion prediction and collision risk assessment with a simulated vehicular cyber physical system. Transp. Res. Part C Emerg. Technol., 47, 179–191.

Wu, K. F., & Jovanis, P. P. (2012). Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. Accident Analysis & Prevention, 45, 507-516.

Xu, Y., Yu, G., Wang, Y., Wu, X., & Ma, Y. (2017a). Car detection from low-altitude UAV imagery with the faster R-CNN. Journal of Advanced Transportation, Article ID 2823617. https://doi.org/10.1155/2017/2823617

Xu, Y., Yu, G., Wu, X., Wang, Y., & Ma, Y. (2017b). An enhanced viola-jones vehicle detection method from unmanned aerial vehicles imagery. IEEE Transactions on Intelligent Transportation Systems, 18(7). https://doi.org/10.1109/TITS.2016.2617202

Xue, T., Chen, B., Wu, J., Wei, D., & Freeman, W. T. (2019). Video enhancement with task-oriented flow. International Journal of Computer Vision, 127(8), 1106-1125.

Zhang, H., Liptrott, M., Bessis, N., Cheng, J. (2019). Real-time traffic analysis using deep learning techniques and UAV based video. In Proceedings of the 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, Washington, DC, USA

Zhang, K., Zuo, W., & Zhang, L. (2018). FFDNet: Toward a fast and flexible solution for CNN-Based image denoising. IEEE Transactions on Image Processing, 27(9). https://doi.org/10.1109/TIP.2018.2839891

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7). https://doi.org/10.1109/TIP.2017.2662206

Zhang, L., & van der Maaten, L. (2013). Structure preserving object tracking. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 1838-1845.

Zhang, Z., Liu, W., Ma, H., & Liu, X. (2017). Going clear from misty rain in dark channel guided network. International Joint Conference on Artificial Intelligence Workshop on AI for Internet of Things, Melbourne, Australia.

Zheng, L., Ismail, K., & Meng, X. (2014). Traffic conflict techniques for road safety analysis: Open questions and some insights. Canadian Journal of Civil Engineering, 41(7). https://doi.org/10.1139/cjce-2013-0558

Zheng, X., Liao, Y., Guo, W., Fu, X., & Ding, X. (2013). Single-image-based rain and snow removal using multi-guided filter. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8228 LNCS(Part 3). https://doi.org/10.1007/978-3-642-42051-1_33

Zhou, H., Kong, H., Wei, L., Creighton, D., & Nahavandi, S. (2015). Efficient road detection and tracking for unmanned aerial vehicle. IEEE Transactions on Intelligent Transportation Systems, 16(1). https://doi.org/10.1109/TITS.2014.2331353

Zhou, X., Koltun, V., & Krähenbühl, P. (2020). Tracking objects as points. In European Conference on Computer Vision, 474-490, Springer, Cham.

Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., Ling, H. (2021). Detection and tracking meet drones challenge, IEEE Transactions on Pattern Analysis & Machine Intelligence, 44(11), 7380-7399.

APPENDIX

Published Related Work

Paper 1: Zong, S., Chen, S., Alinizzi, M., & Labi, S. (2022). Leveraging UAV Capabilities for Vehicle Tracking and Collision Risk Assessment at Road Intersections. *Sustainability 14(7)*, 4034. https://doi.org/10.3390/su14074034

Abstract

Transportation agencies continue to pursue crash reduction. Initiatives include the design of safer facilities, promotion of safe behaviors, and assessments of collision risk as a precursor to the identification of proactive countermeasures. Collision risk assessment includes reliable prediction of vehicle trajectories. Unfortunately, in using traditional tracking equipment, such prediction can be impaired by occlusion. It has been suggested in recent literature that unmanned aerial vehicles (UAVs) can be deployed to address this issue successfully, given their wide visual field and movement flexibility. This paper presents a methodology that integrates UAVs to track the movement of road users and to assess potential collisions at intersections. The proposed methodology includes an existing deep-learning-based algorithm to identify road users, extract trajectories, and calculate collision risk. The methodology was applied using a case study, and the results show that the methodology can provide beneficial information for the purpose of measuring and analyzing the UAV performance. Based on vehicle movements it observes, the UAV can communicate to each vehicle, its collision risk. That way, the vehicle can make proactive driving decisions. Finally, the proposed framework can serve as a valuable tool for urban road agencies to undertake roadway investments to reduce crash risks.