MID-AMERICA TRANSPORTATION CENTER

Report # MATC-MS&T: 135-2

Final Report
WBS: 25-1121-0005-135-2

Crash Prediction and Avoidance by Identifying and Evaluating Risk Factors from Onboard Cameras

Ruwen Qin, PhD
Associate Professor
Department of Engineering Management & Systems Engineering
Missouri University of Science & Technology

Zhaozheng Yin, PhD
SUNY Empire Innovation Associate Professor
AI Institute
Department of Biomedical Informatics
Department of Computer Science
Department of Applied Mathematics & Statistics (Affiliated)
State University of New York at Stony Broook

Muhammad Monjurul Karim, PhD Student

Yu Li, PhD Student

Zuhui Wang, PhD Student
Department of Computer Science
State University of New York at Stony Brook

2020

A Cooperative Research Project sponsored by
U.S. Department of Transportation- Office of the Assistant Secretary for Research and Technology

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 in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Crash Prediction and Avoidance by Identifying and Evaluating Risk Factors from Onboard Cameras

Ruwen Qin, Ph.D.
Associate Professor
Department of Engineering Management & Systems Engineering
Missouri University of Science and Technology

Zhaozheng Yin, Ph.D.
SUNY Empire Innovation Associate Professor
AI Institute
Department of Biomedical Informatics
Department of Computer Science
Department of Applied Mathematics & Statistics (Affiliated)
State University of New York at Stony Brook

Muhammad Monjurul Karim, Ph.D. student
Department of Engineering Management & Systems Engineering
Missouri University of Science and Technology

Yu Li, Ph.D. student
Department of Engineering Management & Systems Engineering
Missouri University of Science and Technology

Zuhui Wang, Ph.D. student
Department of Computer Science
State University of New York at Stony Brook

A Report on Research Sponsored by

Mid-America Transportation Center

University of Nebraska–Lincoln

July 2020

Technical Report Documentation Page

1. Report No. 25-1121-0005-135-2	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Crash Prediction and Avoidance by Identifying and Evaluating Risk Factors from Onboard Cameras		5. Report Date September 2020
		6. Performing Organization Code
7. Author(s) Ruwen Qin PhD ORCID: 0000-0003-2656-8705, Zhaozheng Yin PhD ORCID: 0000-0002-9602-6488, Yu Li, Muhammad Monjurul Karim, Zuhui Wang		8. Performing Organization Report No. 25-1121-0005-135-2
9. Performing Organization Name and Address Mid-America Transportation Center Prem S. Paul Research Center at Whittier School 2200 Vine St. Lincoln, NE 68583-0851		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 69A3551747107
12. Sponsoring Agency Name and Address Office of the Assistant Secretary for Research and Technology 1200 New Jersey Ave., SE Washington, D.C. 20590		13. Type of Report and Period Covered January 2019-July 2020
		14. Sponsoring Agency Code MATC TRB RiP No. 91994-41
15. Supplementary Notes		

Motor vehicle crashes are a huge concern of roadway transportation safety, resulting in over 37,000 fatalities and $800 million losses annually. In recent years, the number of road fatalities has been growing. Traditionally used and identifiable risk factor explanations no longer fully account for the causes of a recent increase in road fatalities. Human beings have bounded abilities in vision, cognition, making judgment, and simultaneously handling multiple tasks, particularly in complex, dynamic environments or in response to sudden situations. Therefore, assisting them in cognition of risks and making the right decisions rapidly in a near real-time manner is in need to advance transportation toward a zero fatality rate. This project's motivation is to develop a data-driven, computer-vision (CV) empowered, verifiable system that can predict crashes, and thus improve drivers' ability to avoid them. Pursuing a systematic approach, this project seamlessly integrates data analytics, deep learning, computer vision technology, and a rigorous verification process to achieve the goal. Specifically, this project creates a spatio-temporal attention guidance for CV-based crash risk assessment through analyzing fatal crash report data retrieved from Fatality Analysis Reporting System (FARS). The guidance informs the likelihood of crash and crash types given the time and location information of driving scenes, thus giving the driving scene analysis a clear focus. Then, a system of deep neural networks is developed to perform a driving scene analysis in support of crash risk assessment and prevention. The scene classification result allows for retrieving the relevant guidance for crash risk assessment and prevention. The joint results from the object detection and drivable area segmentation help identify risky pedestrians and vehicles in the surrounding traffic. Evaluation and examples demonstrate the effectiveness of the proposed technologies.

17. ORCID No. of each Researcher Ruwen Qin, 0000-0003-2656-8705 Zhaozheng Yin, 0000-0002-9602-6488	18. Distribution Statement		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 59	22. Price

Table of Contents

Disclaimer vii
Abstract viii
Chapter 1 Introduction 1
1.1 Problem Statement 1
1.2 The Proposed System 2
1.3 Overview of Research Methods 3
1.3.1 Crash Report Data Analysis 3
1.3.2 Deep Neural Networks for Driving Scene Analysis 4
Chapter 2 Crash Report Data Analysis for Creating A Spatio-Temporal Attention Guidance for Vision Based Real-Time Crash Risk Assessment 5
2.1 Introduction 5
2.2 The Literature 7
2.2.1 Applications of Crash Data Analysis 7
2.2.2 Crash Risk Indicators 8
2.2.3 Methods of Exploratory Data Analysis 10
2.3 The Approach 10
2.3.1 Determining Fatal Crash Data of Study 11
2.3.2 Identifying Factors for Characterizing Fatal Crashes 13
2.3.3 Developing the Spatio-Temporal Attention Guidance 17
2.4 Illustrative Examples 29
2.5 Conclusion and Future Work 32
Chapter 3 A System of Vision Sensor Based Deep Neural Networks for Complex Driving Scene Analysis in Support of Crash Risk Assessment and Prevention 34
3.1 Introduction 34
3.2 System for complex driving scene analysis 36
3.2.1 Multi-Net for multi-task classification of driving scenes 36
3.2.2 Object detection with YOLO v3 38
3.2.3 Distance to the nearest vehicle 39
3.2.4 Detecting drivable area and risky pedestrian 40
3.3 Dataset Development 42
3.3.1 Data acquisition 42
3.3.2 Description of the datasets 43
3.4 Experimental Evaluation 44
3.4.1 Training the Multi-Net and DeepLab v3 44
3.4.2 Evaluation of classifier performance 45
3.4.3 Examples of classification results 48
3.4.4 Examples of object detection 49
3.4.5 Inference speed 51
3.5 Conclusion 52
Chapter 4 Recommendations Developed from the Project 54
References 56

List of Figures

Figure 1.1: The schematic diagram of the data-driven, computer vision empowered system for crash risk assessment and crash prevention 2
Figure 2.1 The proposed approach to creating the spatio-temporal attention guidance 11
Figure 2.2 Data decomposition by fatal crash scenarios that are classified in accordance to five variables 13
Figure 2.3 The frequency distribution of fatal crash (2012–2017) on months of year, split by hours of day 14
Figure 2.4 Reduction of seven daily patterns to weekday and weekend Patterns 15
Figure 2.5 Fatal crash distribution on the road configuration tree 17
Figure 2.6 Spatially defined groups of fatal crashes 19
Figure 2.7 The clustering approach and result 21
Figure 2.8 Association rules for fatal crashes in cluster 1 by FHE types 28
Figure 2.9 Association rules for fatal crashes in cluster 1 by MC types 28
Figure 2.10 Association rules for fatal crashes in cluster 2 by FHE types and weekday/weekend 29
Figure 2.11 An illustrative example of driving scene analysis result for fatal crash risk assessment 31
Figure 2.12 More examples of driving scene analysis with a spatio-temporal attention guidance 32
Figure 3.1 The schematic diagram of the driving scene analysis system 37
Figure 3.2 The architecture of the Multi-Net 37
Figure 3.3 Measuring the distance to another vehicle 40
Figure 3.4 Detecting drivable area and risky pedestrians 40
Figure 3.5 The pipeline for mining images for dataset development 42
Figure 3.6 Confusion matrices 48
Figure 3.7 Examples with results generated by the Multi-net 49
Figure 3.8 Example of detecting and locating risky pedestrian and the nearest vehicle 51

List of Tables

Table 2.1 Clustering Results 22
Table 2.2 Moran's I values for spatial-temporal clusters 24
Table 3.1 Summary of the datasets 43
Table 3.2 Performance of classifiers 47

List of Abbreviations

Computer Vision (CV)
Fatal Analysis Reporting System (FARS)
Department of Transportation (DOT)
Crash Data Analysis (CDA)
Driving Scene Analysis Network (DSAN)
Crash Prediction (CP)
Federal Highway Administration (FHWA)
Crash Report Sampling System (CRSS)
Safety Data Initiative (SDI)
National Highway Traffic Safety Administration (NHTSA)
Exploratory Advanced Research (EAR)
Exploratory Data Analysis (EDA)
Manner of Collision (MC)
First Harmful Event (FHE)
Functional System (FUNC_SYS)
Relation to Trafficway (REL_ROA)
Relation to Junction-Specific Location (JUN_INT)
Convolutional Neural Networks (CNNs)
Multi-task deep neural networks (Multi-Net)
Convolutional Neural Network (CNN)

Disclaimer

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

Abstract

Motor vehicle crashes are a huge concern of roadway transportation safety, resulting in over 37,000 fatalities and $800 million losses annually. In recent years, the number of road fatalities has been growing. Traditionally used and identifiable risk factor explanations no longer fully account for the causes of a recent increase in road fatalities. Human beings have bounded abilities in vision, cognition, making judgment, and simultaneously handling multiple tasks, particularly in complex, dynamic environments or in response to sudden situations. Therefore, assisting them in cognition of risks and making the right decisions rapidly in a near real-time manner is in need to advance transportation toward a zero fatality rate. This project's motivation is to develop a data-driven, computer-vision (CV) empowered, verifiable system that can predict crashes, and thus improve drivers' ability to avoid them. Pursuing a systematic approach, this project seamlessly integrates data analytics, deep learning, computer vision technology, and a rigorous verification process to achieve the goal. Specifically, this project creates a spatio-temporal attention guidance for CV-based crash risk assessment through analyzing fatal crash report data retrieved from Fatality Analysis Reporting System (FARS). The guidance informs the likelihood of crash and crash types given the time and location information of driving scenes, thus giving the driving scene analysis a clear focus. Then, a system of deep neural networks is developed to perform a driving scene analysis in support of crash risk assessment and prevention. The scene classification result allows for retrieving the relevant guidance for crash risk assessment and prevention. The joint results from the object detection and drivable area segmentation help identify risky pedestrians and vehicles in the surrounding traffic. Evaluation and examples demonstrate the effectiveness of the proposed technologies.

Chapter 1 Introduction

1.1 Problem Statement

In 2017, about 37,151 fatalities resulted from motor vehicle traffic crashes (1). Crashes cost the Nation over $800 billion annually in lives lost or injuries, lost productivity, and property damage. Traditionally used and identifiable risk factors no longer fully explain the causes of the recent increase in traffic fatalities. Identifying and understanding changes and emerging risk factors that contribute to the increase in transportation-related fatalities are greatly needed by policy makers, providers, operators of transportation systems, and the public (2).

Human factors are found to play a significant contributing role in 94 percent of traffic crashes (3). Human beings have bounded abilities in vision, cognition, making judgment, and simultaneously handling multiple tasks, particularly in complex, dynamic environments or in response to sudden situations. Improving their ability to appropriately use or operate the transportation system through assisting them in cognition of risks and rapidly making the right decisions in a near real-time manner is in particular need in order to advance transportation toward a zero fatality rate (4).

Pursuing a data-driven systematic safety approach has been a strategic objective of the U.S. Department of Transportation (DOT), to mitigate risks and encourage infrastructure and behavior change (3). Yet the development of this approach is technically challenging (5). Data sources are abundant and growing, and existing data analytical processes are diverse. Near real-time systems for enhancing risk recognition and mitigation are powered by sophisticated technologies. Comprehensive risk evaluation and prediction are complex.

The integration of data analytics, technology for safety enhancement, and risk prediction and avoidance require systems thinking. We propose the development of a data-driven,

computer-vision empowered, verifiable safety enhancement system for drivers, which addresses the identified technical challenges and meets the new needs for crash prediction and avoidance.

1.2 The Proposed System

The proposed system is illustrated in figure 1.1. The system is composed of three integrated modules. The Crash Data Analysis (CDA) module of the system identifies and characterizes fatal crash scenes by analyzing the big data of crash reports using data exploration and visualization. The Driving Scene Analysis Network (DSAN) is a system of Deep Neural Networks which analyzes the live video stream captured by an onboard camera. Detection and classification results of DSAN and high-risk scenes identified by CDA are sent to the Crash Prediction (CP) module to determine if crash risks are present.

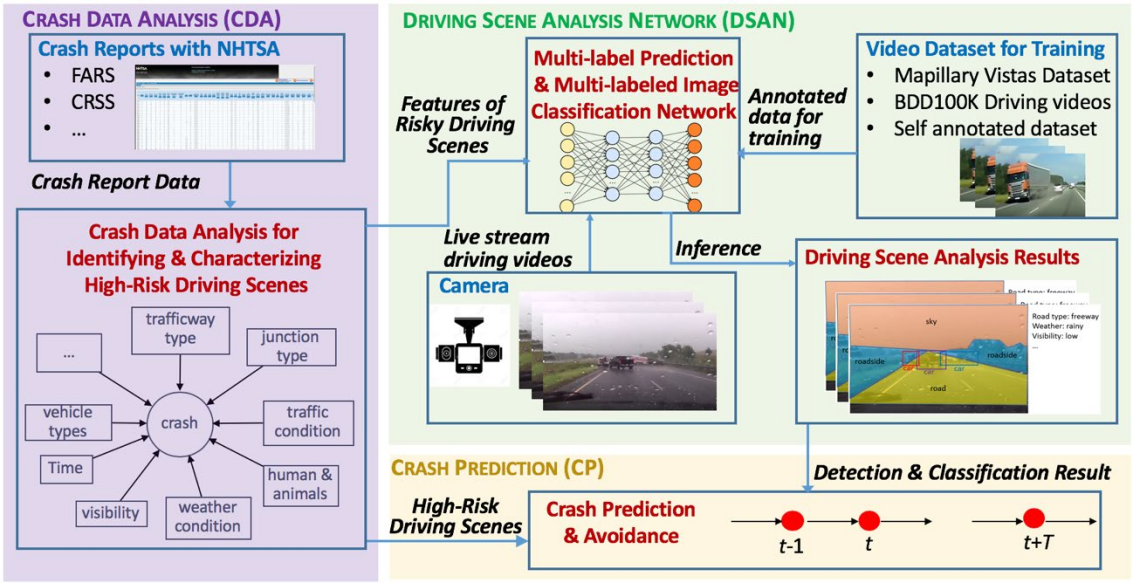


Figure 1.1 The schematic diagram of the data-driven, computer vision empowered system for crash risk assessment and crash prevention

1.3 Overview of Research Methods

Two major efforts are required to create the proposed system: a crash report data analysis to create a spatio-temporal attention guidance for computer vision-based real-time crash risk assessment and the development of vision sensor based neural networks for complex driving scene analysis in support of crash risk assessment and crash prevention. They are briefly introduced below and delineated in the next two chapters.

1.3.1 Crash Report Data Analysis

With the rapid development of sensor and artificial intelligence technologies, the CV-based crash risk assessment and prevention is receiving growing attentions. Current CV algorithms lack an attention guidance for driving scene analysis; for example, they may detect every single object from stream videos. Consequently, they are computationally expensive, and the delivered driving scene analysis result may increase the cognitive load of human drivers or cause misunderstanding. To effectively assist drivers and autonomous driving systems in analyzing driving scenes, this project proposes to create a spatio-temporal attention guidance for CV-based, real-time fatal crash risk assessment through an analysis of 5–year (2013–17) fatal crash report data retrieved from the FARS. Firstly, the project justifies the need for scenario-based attention and identifies five variables for classifying crash scenarios and decomposing the data accordingly. Then, it illustrates the development of an attention guidance for the common scenario of fatal crashes. Time and location related variables were identified or created from an exploratory data analysis for featuring driving scenes. These feature variables are used to cluster fatal crash data. The spatio-temporal attention guidance was learnt through association rule mining from those clusters, which inform the likelihood of a fatal crash and crash types given the

time and location information of any driving scene. At the end, examples are presented to illustrate the improved results of driving scene analysis or crash risk assessment.

1.3.2 Deep Neural Networks for Driving Scene Analysis

To assist human drivers and autonomous vehicles in assessing crash risks, driving scene analysis using dash cameras mounted in vehicles and deep learning algorithms is of paramount importance. Although these technologies are increasingly available today, comprehensive analysis of driving scenes in support of crash risk assessment and crash prevention still remains an unsolved challenge. This is mainly due to the lack of annotated large datasets for analyzing crash risk indicators and identifying the crash likelihood, and the lack of an effective method to extract lots of required information from complex driving scenes. To fill the gap, this project developed a scene analysis system composed of classifiers, an object detector, and a region of interest segmentor. The Multi-Net of the system includes two multi-task neural networks that perform scene classification to provide four labels for each scene, including the crash likelihood, road function, weather, and time of day. The DeepLab v3 and YOLO v3 are combined to detect and locate risky pedestrians and the nearest vehicles in the scene. All the identified information can provide the situational awareness to autonomous vehicles or human drivers for identifying crash risk from the surrounding traffic. To address the scarcity of annotated datasets for studying traffic crashes, two completely new datasets have been developed by this project and made available to the public, which were proved to be effective in training the proposed deep neural networks for scene classification. The project further evaluated the performance of the developed Multi-Net in classifying driving scenes and illustrated comprehensive scene analysis results with examples. Results demonstrate the effectiveness of the developed system and datasets for driving scene analysis, and their usefulness in support of crash risk assessment and crash prevention.

Chapter 2 Crash Report Data Analysis for Creating A Spatio-Temporal Attention Guidance for Vision Based Real-Time Crash Risk Assessment

2.1 Introduction

Enhancing traffic safety is a top priority of the US Department of Transportation (DOT). In 2018, there were 36,560 deaths in motor vehicle crashes (6). Federal Highway Administration (FHWA) aims to reduce motor vehicle fatalities and serious injuries across the transportation system (7). Various safety enhancement programs consider a multi-pronged approach that encompasses countermeasures of enforcement, education, emergency response, and engineering (named 4E) (8). 4E's countermeasures are usually developed for specific risk problems, such as pedestrian involved crashes (e.g. (9) (10)) and work zone related crashes (e.g. (11)). Those countermeasures are implemented by state DOTs, local entities, and other transportation agencies.

With the rapid development of sensor and artificial intelligence technologies, real-time crash risk detection and prevention for both autonomous and human driving vehicles are attracting growing interests (12) (13) (14) . For example, a computer vision (CV)-based deep learning model can be trained to analyze stream data captured by the camera mounted on a vehicle to estimate the chance of crash occurrence (15) (16). Real-time risk assessment and prevention techniques like such complement 4E's countermeasures and may advance the roadway safety to a new paradigm. However, so many objects and elements are present in a driving scene. Attempting to detect everything from driving scenes and analyzing all of these are not only computationally expensive but can be misleading. For example, a pedestrian in a non-crossing area at night probably is more risky than one who is crossing the road at an intersection controlled by traffic signals. One possible solution is to create an attention guidance, which

suggests regions of interests and objects that deserve attention for the real-time driving scene analysis.

Crash report systems, such as the Fatality Analysis Reporting System (FARS) (6) and Crash Report Sampling System (CRSS) (17), contain comprehensive historical data of crashes, which can be used to identify representative crash scenarios. Data analysis and visualization for safety enhancement have received tremendous attentions. The Safety Data Initiative (SDI) of DOT (18) launched a challenge (19) to invite participants to develop innovative ways of visualizing crash data for gaining insights of safety enhancement. National Highway Traffic Safety Administration (NHTSA) completed the visualization of their traffic safety fact sheets on speeding (20) and pedestrians (21) using Tableau. The Exploratory Advanced Research (EAR) Program of FHWA is supporting projects (e.g., TRIP (22)) to make massive amounts of transportation data accessible (23). Crash risks are often inferred from risk indicators that are either factors contributing to crashes (e.g., bad weather) or variables on which crash related measurements may vary (e.g., the volume of crashes varies with the time of day). While being widely used in crash data analysis, basic descriptive statistics and graphic techniques are not efficient. Researchers have started using data mining methods to discover the relation of multiple variables from big data and predict future trends to support decision-making.

This paper is motivated to explore an advanced data analytic approach to developing attention guidance for CV-based, real-time crash risk assessment. Based on our best knowledge, no work has been done to timely meet this need. A comprehensive attention guidance requires a systematic development effort that is impossible to be completely reported in one research article. This paper chose to focus on fatal crashes in “common scenario”, defined as one most commonly seen by human drivers and autonomous driving systems, due to the following

reasons. Firstly, fatal crashes are the most concerned crash type with an average economic cost of $1.7 million per death in 2018 (24). Preventing fatal crashes is an urgent need. Secondly, fatal crash risks have varied features or causes in different crash scenarios, making it necessary to assess crash risks by scenarios. Thirdly, CV-based driving sense analysis is able to classify driving scenes by crash scenarios and retrieve the corresponding attention guidance. Last but not the least, methods developed in this study are applicable to the crash risk assessment of other scenarios.

The remainder of this chapter is organized as follows. The next section summarizes related studies. Then, the proposed approach to developing the attention guidance is explained in Section 2.3. Section 2.4 illustrates examples that implement the proposed attention guidance in driving scene analysis. Conclusions drawn from this study and recommended future work are summarized at the end, in Section 2.5.

2.2 The Literature

The literature relevant to this study is in three streams: applications of crash data analysis, crash risk indicators, and methods of exploratory data analysis. Details of the literature are discussed below.

2.2.1 Applications of Crash Data Analysis

Crash report data are widely analyzed for developing 4E's safety strategies. For example, Abdel-Aty *et al*. (25) studied crashes caused by visibility obstruction such as fog and smoke. Findings from this work recommend locations for installing solar and battery-powered systems to provide warnings to drivers. Law enforcement for safety improvement must be well justified by evidences, such as results from crash data analysis (e.g., (11) (26) (27)). For example, Harb *et al*. (27) developed logistic regression models to unveil traits of freeway work zone crashes and

found that the inattentiveness and hostile driving habits of drivers are over-represented in work zones. Therefore, additional enforcement was recommended such as doubling the fine for traffic violations at work zones. Pour-Rouholamin and Zhou (26) also suggested programs of educating senior people to use pedestrian crosswalks and contrasting clothing at night. Sharma *et al*. (28) showed that the survival rate of injured persons involved in crashes largely depends on how quickly emergency responders arrive at the crash scene. Accordingly, they developed an automatic accident detection system that uses the sensor data collected from smartphones.

The 4E's strategies for road safety enhancement are developed for addressing a variety of specific safety concerns. These include, but are not limited to, pedestrian involved crashes (10) (26), work zone crashes (11) (27) (29), large truck involved crashes (30), the effectiveness of speed limit (31), the weather impact (25) (32), the seat-belt law (33), and teenage and adult drivers (34).

2.2.2 Crash Risk Indicators

Risk indicators investigated by crash data analysis are in multiple categories, including time related, location related, driving environment related, human factors, and so on. Fatal crash cases are found to distribute non-homogeneously in the temporal space. Thus, the crash risk can be inferred from time-related factors such as hour, week, month, and season. Pour-Roubolamin and Zhou (26) found the volume of severe injury cases from 8:00pm to 5:59am is higher than other hours. Chong et al. (10) split the 24 hours of day into four segments and sorted them in the increasing order of pedestrian involved fatal cases. The dashboard of pedestrian crash data on the DOT website (21) showed that the volume of pedestrian involved crashes in winter (November, December and January) is higher than in other seasons. More crashes happened between 5:00pm

to 7:59pm during winter time, but after 8:00pm in other seasons. Compared to weekdays, more crashes occurred after midnight on Saturday and Sunday.

Crashes are also non-homogeneously distributed in the spatial space defined by factors such as road functional system, junction or intersection types, and land use. Work of (21) showed that pedestrian involved crashes are the highest on other principal arterial roads. Rovšek *et al*. (35) found that injury severity in the circumstance of inappropriate speed is most likely to be fatal on freeway, regional or local roads. Statistical comparisons in (11) indicate work zone activities are more risky on the interstate than on other road types. Meng and Weng (29) identified that rear-end crash risk is higher on the expressway than on other arterial. Pedestrian fatalities at non-intersection areas are almost three times of those at intersections (21). Pedestrian fatalities at urban areas are three times more than those in rural areas. Rural areas have more work zone related crashes and pedestrian involved crashes than in urban areas (27) (36).

Driving environment conditions are not related to human factors but they influence drivers' ability to operate vehicles, such as visibility, weather, and the road condition. Factors related to visibility are widely explored (e.g., (26) (27) (35)). Darkness, poor lighting condition, and no lighting condition are contributing factors for the severity of pedestrian involved crashes (26). Pedestrian fatalities that happened in the dark condition were three times more than those in the daylight condition (21). Bad weather (e.g., fog, snow, and severe wind) is also a reason for crashes (27). Abdel-Aty *et al*. (25) analyzed crashes in fog, smoke fog, and smoke conditions, and identified multiple scenarios those crashes are more likely to occur, such as morning hours of December to February on undivided roads. Zhai *et al*. (9) found that people tend to be less patient and more likely to violate traffic rules in undesirable weather conditions. Injury severity is impacted more by a bad weather condition than a poor lighting or road surface condition (35).

Hamdar *et al*. (32) found that drivers become more aggressive in dealing with challenging roadway and weather conditions. Some other factors pertain to drivers themselves, such as the use of drugs or alcohol, not wearing seat belt and helmet (11) (35), speeding, emotion, and cell browse (10) (11) (26) (31) (35) (37).

2.2.3 Methods of Exploratory Data Analysis

Data exploration along with visualization is an analytic approach to summarizing and understanding data. Typical graphical techniques used in Exploratory Data Analysis (EDA) for crashes include scatter charts, line charts, bar or column charts, bubble charts, and heat maps (10) (25) (27) (31) (38) (39). Data mining techniques such as classification trees and association rules are also proposed as advanced methods for discovering new knowledge, particularly from big crash data. Rovšek, *et al*. (35) used a classification tree to find key risk factors impacting the severity of traffic injuries on Slovenian roads. Montella *et al*. (36) innovatively converted results of classification trees into association rules. Each terminal node of the tree is treated as the consequent of a rule and all the splits of parent nodes are the antecedent.

Spatial and temporal patterns of crashes are useful decision-making tools. Moran's I is a measurement to evaluate the existence of a pattern of clustering by calculating the spatial autocorrelation (40). Pearson's r is a coefficient that measures the correlation between two variables. Lee (41) integrated Pearson's r and Moran's I to develop a bivariate spatial association measure, named L.

2.3 The Approach

The proposed approach to creating the spatio-temporal attention mechanism is a three-step process described in figure 2.1. Details of each step are discussed below.

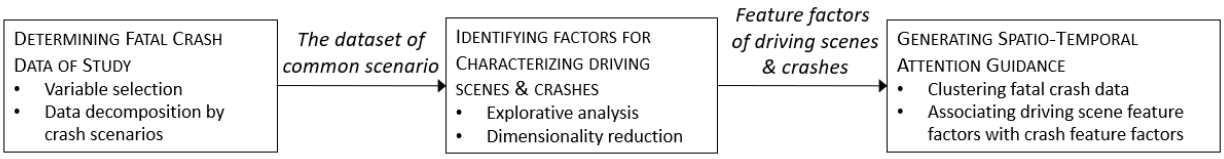


Figure 2.1 The proposed approach to creating the spatio-temporal attention guidance

2.3.1 Determining Fatal Crash Data of Study

The data used in this study is the 2013-2017 fatal crash data of the United States retrieved from FARS (6). There are 162,104 fatal crashes during this five-year period. FARS has 30 variables (42) that generally characterize fatal crashes.

2.3.1.1 Variable Selection

Thirteen variables in the following four categories were chosen by this study, which can be detected either by CV-techniques (e.g., the light condition) or provided by systems equipped in vehicles (e.g., the time of day).

- Time related: Hour, Day of Week, Month;
- Location related: Functional System, Relation to Trafficway, Relation to Junction-Specific Location, Type of Intersection, Related to Junction-Within Interchange Area;
- Environment related: Light Condition, Atmospheric Condition;
- Special crash types: Work Zone Related, School Bus Related, Related Factor-Crash Level.

Two additional variables were chosen to feature crashes, which are “Manner of Collision” (MC) and “First Harmful Event” (FHE).

Each driving scene captured by the camera mounted on a vehicle will be processed by an image analysis deep learning algorithm (43) to generate thirteen feature labels. With the labels,

the attention guidance to be developed in this paper will provide the likelihood of a fatal crash and possible crash types.

2.3.1.2 Data Decomposition by Crash Scenarios

A fatal crash scenario is defined as a set of fatal crashes with certain similarities. For example, all the work zone related fatal crashes can be seen as a crash scenario. By creating a finite number of mutually exclusive and collectively exhaustive fatal crash scenarios, a driving scene is mapped to one and only one fatal crash scenario. Crashes of the same scenario have one or several strong linkages and, thus, are likely to have some common features. Therefore, a hypothesis in this study is that individuals crash scenarios have their unique risk patterns that are more distinct than risk patterns of the population.

Five categorical variables were identified as suitable for classifying fatal crash scenarios, which are “School Bus Related”, “Work Zone Related”, “Within Interchange Area”, “Related Factor-Crash Level”, and “Atmospheric Conditions”. Each of these five variables is relevant to only a small portion of fatal crashes. That is, if these five variables are coded as binary variables, most values of the binary variables are zeros. The data sparsity of them provide an opportunity of data decomposition. A hierarchical tree method was proposed in this study to decompose the fatal crash data into mutually exclusive and collectively inclusive subsets with each corresponding to a crash scenario, as figure 2.2 illustrates. The study ranked the five variables for classifying fatal crash scenarios in the increasing order of non-zero elements and decomposed the dataset in that sequence, as figure 2.2 illustrates. The number of school bus related fatal crashes is only 521 (0.32% of the 162,104 cases), smaller than the number of any other type of crashes. Therefore, the first step of data decomposition was made according to the variable “School Bus Related”. After that, crashes not related to school bus were further split according to

the variable “Work Zone Related”. This process was continuing until all of the five variables have been used for data decomposition. Figure 2.2 shows that the biggest subset is the “common scenario” composed of fatal crashes not related to school bus, work zone, any crash factors and any bad weather types, and not occurred within an interchange area. This subset contains 128,149 fatal crashes, equal to 89.93% of total cases during the five-year period. The common scenario dataset was analyzed in the remainder of this paper with respect to the remaining eight variables for featuring driving scenes and the two variables for featuring crashes.

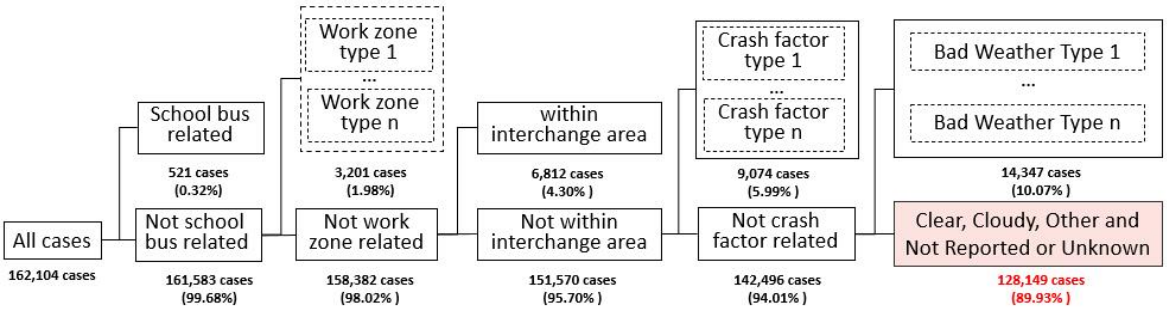


Figure 2.2 Data decomposition by fatal crash scenarios that are classified in accordance to five variables

2.3.2 Identifying Factors for Characterizing Fatal Crashes

Among the eight feature variables for characterizing the driving scenes, three variables are time related, four are location related, and one is environment related. The environment related factor is the light condition, which is verified to strongly correlate with hours of day. Therefore, this study performed a basic exploratory analysis of this dataset to develop an initial understanding of it on the time and location dimensions, respectively.

2.3.2.1 Fatal Crash Distribution on Time Related Variables

The count of fatal crashes by month and hour is displayed in figure 2.3. It is observed that the distribution of crashes by month and hour are not even. The volume of fatal crashes is the lowest in February, increases afterward, reaches the highest level during fall, and gradually dropped during winter. Afternoon and evening hours (1:00pm-23:59pm) have more fatal crashes than other hours of day. Moreover, the hours with a peak volume of fatal crashes demonstrate a seasonality pattern. The volume reaches the peak during 6:00-7:59PM in February, 7:00-8:59PM in March, 8:00-9:59PM from April to September, 7:00-8:59PM in October, and 5:00-6:59PM during winter.

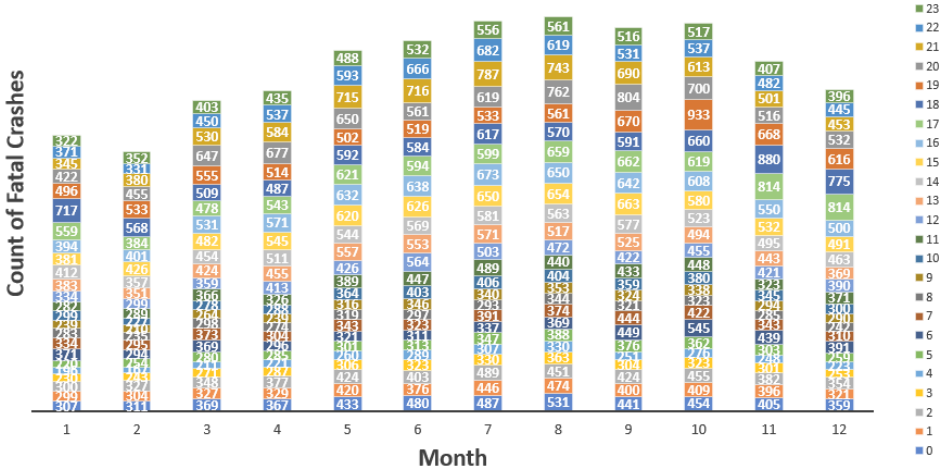
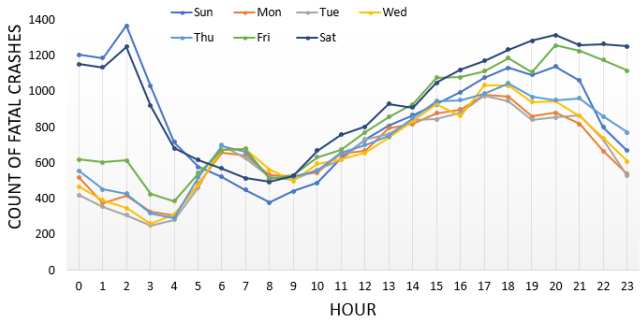
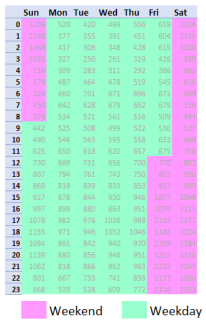


Figure 2.3 The frequency distribution of fatal crash (2012–2017) on months of year, split by hours of day

Figure 2.4(a) further displays the trend lines of fatal crashes during a day, split by day of week. This figure shows that the trend lines of Monday, Tuesday, Wednesday, and Thursday are close to each other, suggesting the existence of a weekday pattern. The trend line of Saturday is clearly separated from the weekday pattern, representing a weekend pattern. The trend lines of

Friday and Sunday stretch to both patterns. The afternoon to evening portion of Friday is close to the weekend pattern, so is Sunday from midnight to early morning. Observations from figures 2.3 and 2.4(a) suggest that two month-hour 2D distributions of fatal crashes should be created, one for weekday and the other for weekend.





(a) Hourly series for the seven days of a week (b) weekday and weekend split

Figure 2.4 Reduction of seven daily patterns to weekday and weekend Patterns

An optimization problem was formulated in this study to split crash data into two subsets, one for weekday and one for weekend. Let n_i^j represent the number of fatal crashes in hour i of week j, for $i = 0, 1, \dots, 23$ and $j \in \{Mo, Tu, We, Th, Fr, Sa, Su\}$. Let x_i^D and x_i^E respectively stand for the average number of fatal crashes at hour i of weekday and weekend. They are initialized as

$$x_i^D = \frac{n_i^{Mo} + n_i^{Tu} + n_i^{We} + n_i^{Th}}{4}, \quad x_i^E = n_i^{Sa}. \tag{2.1}$$

The hourly series of fatal crashes on Friday is split into two segments; that is, $\{n_i^{Fr} | i = 0, \dots, 23\} = \{n_i^{Fr} \big| i = 0, \dots, p\} \cup \{n_i^{Fr} | i = p+1, \dots, 23\}$. The first segment of Friday belongs to

weekday and the second segment is weekend. The place to split the time series needs to be chosen so that the sum of squared errors due to merging the two segments of Friday data into the weekend and weekend time series, respectively, is minimized:

$$\min_{p\in\{0,\dots,23\}} ss^{Fr}(p) = \sum_{i=0}^{p}(n_i^{Fr} - x_i^{D})^2 + \sum_{i=p+1}^{23}(n_i^{Fr} - x_i^{E})^2 . \quad (2.2)$$

Similarly, the hourly series of fatal crashes of Sunday is split as two segments, with the first q hours belonging to weekend and remaining hours belonging to weekend.

The counts of crashes, $\{n_i^j | \forall i,j\}$, in Eqs. (2.1–2.2) were estimated using the 128,149 fatal crashes. The optimal values of q and p were found to be 11 and 8, respectively. The result is displayed in Figure 2.4(b). That is, 9:00am of Sunday to 11:59am of Friday is the time frame for the weekday crash pattern and 12:00pm of Friday to 8:59am of Sunday is the time frame for the weekend pattern.

2.3.2.2 Fatal Crash Distribution on Location Related Variables

The location of a driving scene can be described by the type and configuration of the road. "Functional System (FUNC_SYS)" is the variable indicating the road type and the road configuration is described by the combination of another three location related variables: "Relation to Trafficway (REL_ROA)", "Relation to Junction-Specific Location (JUN_INT)", and "Type of Intersection". Due to their hierarchical relationship, this study merged "Type of Intersection" into JUN_INT by replacing the attribute "Intersection or related" of JUN_INT with specific intersection types such as "Four-Way Intersection", "T-Intersection", "Y-

Intersection", and so on. In the remainder of the paper, the road configuration is defined by REL_ROA and the revised JUN_INT.

Figure 2.5 uses a tree structure to show the distribution of fatal crashes by road configurations. The figure shows that fatal crashes are unevenly distributed on different road configurations. 86.8% fatal crashes occurred at the following five road configurations: on the roadway and non-junction areas (32.8%), on the roadside and non-junction areas (28.8%), on the roadway and four-way intersections (14.7%), on the roadway and T-intersections (7.0%), and one the roadway and driveway access or related (3.5%).

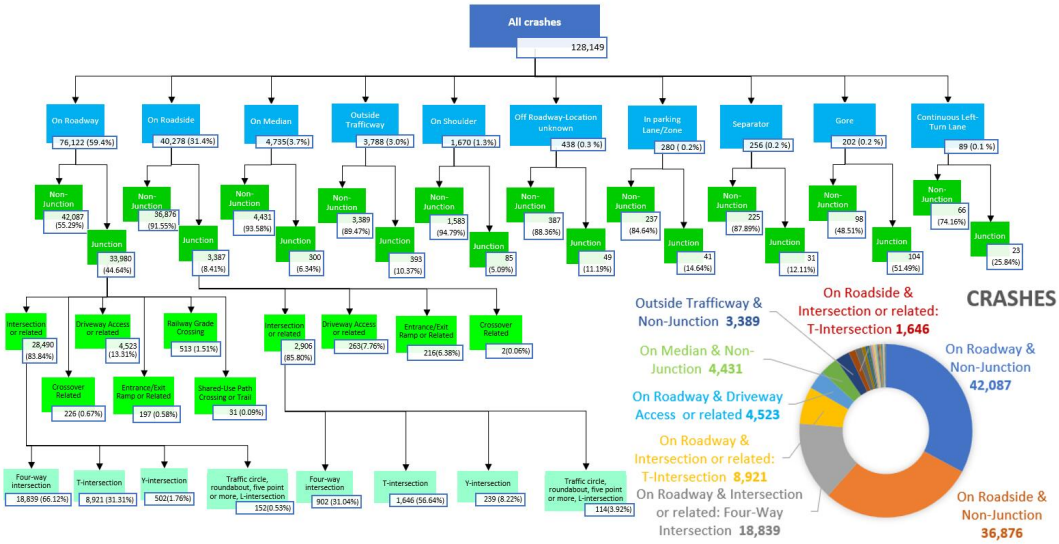


Figure 2.5 Fatal crash distribution on the road configuration tree

2.3.3 Developing the Spatio-Temporal Attention Guidance

The spatio-temporal attention guidance is learnt from the crash data patterns discovered in this section. Given the location and time, the attention guidance describes the likelihood of a fatal crash and the crash types.

2.3.3.1 Spatially Defined Groups of Fatal Crashes and Their Temporal Distribution Patterns

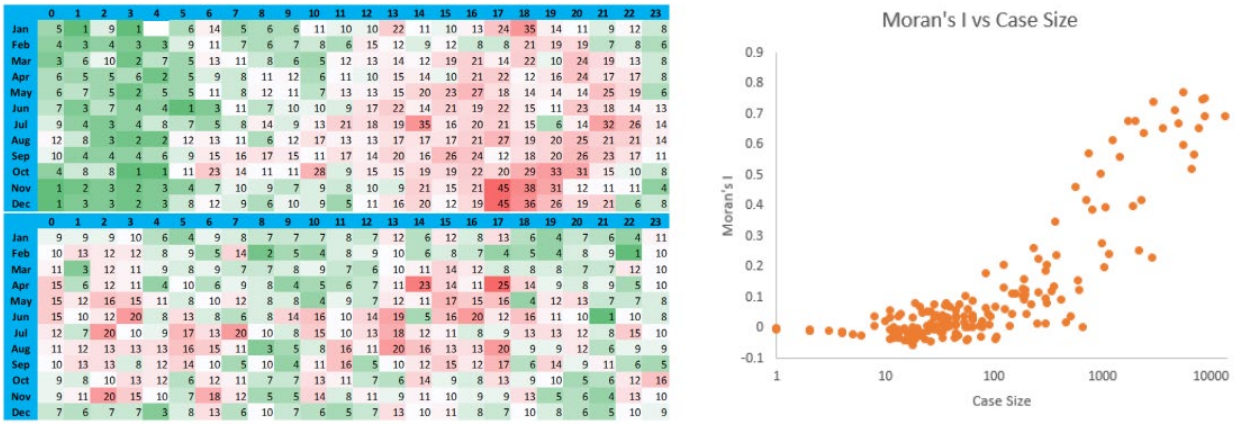
The common scenario dataset is broken down into 184 mutually exclusive and collectively exhaustive groups that each contains fatal crashes with the same road type and configuration. For example, the spatially defined group of fatal crashes occurred on "Local Road, On Roadside, and Driveway Access or Related" includes those with FUNC_SYS=7, REL_ROA=4, and JUN_INT=4.

Then, the month-hour 2D temporal distribution of fatal crashes is created for each of the spatially defined groups. The 2D distribution counts the number of fatal crashes in each unit of the month-hour grid, and it uses colors to visualize the variation of cases on this 12×24 grid – the darker the red (and green) color is, the larger (and less) the volume of cases in that unit of the temporal grid. Figure 2.6(a) illustrates two examples. The one on the top is the group "Principal Arterial – Other, on Roadway, and T-Intersection" with 3,609 fatal crashes. The one at the bottom is the group "Interstate, On Roadside, and Non-Junction" with 2,858 cases. These two distributions are very different. The distribution on the top has a pattern with a red spot on the right and a green spot on the left. The distribution at the bottom is more random, without a clear pattern. Moran's I was calculated for each spatially defined group to measure the strength of its temporal pattern. Moran's I in this paper is the autocorrelation of the 288 month-hour units of the temporal grid:

$$I = \frac{288}{\sum_{i,j=1}^{288} w_{ij}} \frac{\sum_{i,j=1}^{288} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{288}(y_i - \bar{y})^2} \tag{2.3}$$

where i and j are linear indices of any two units on the grid. y_i and y_j are the numbers of fatal crashes in units i and j, respectively. $\bar{y}$ is the average number of fatal crashes per unit. w_{ij} is a binary coefficient that takes one when units i and j are neighbors and zero otherwise. Each unit has at least three and up to eight neighbors. The value of Moran's I is from -1 to 1, where -1 means perfect dispersion, 0 is perfect randomness, and 1 indicates perfect clustering.

In figure 2.6(a), the Moran's I value of the distribution on the top is 0.65 and the value of the bottom one is 0.23. Figure 2.6(b) further displays the scatter plot of the 184 spatially defined groups with their Moran's I value and the group size. This figure indicates groups of smaller sizes (e.g., no greater than 100 fatal crashes) are likely to have a random temporal distribution. Therefore, this study further analyzes the similarity of the 68 groups that each contains more than 100 fatal crashes. There are 123,782 fatal crashes in those 68 groups, which count for 96.6% out of the total 128,149 crashes in the common scenario dataset. Groups of size 100 or less contain 4,367 cases, only 4.4% of the data in the common scenario dataset.



(a) Examples of 2D distribution on the month-hour grid (b) Moran's I vs group size

Figure 2.6 Spatially defined groups of fatal crashes

In figure 2.6(a), the Moran's I value of the distribution on the top is 0.65 and the value of the bottom one is 0.23. Figure 2.6(b) further displays the scatter plot of the 184 spatially defined groups with their Moran's I value and the group size. This figure indicates groups of smaller sizes (e.g., no greater than 100 fatal crashes) are likely to have a random temporal distribution. Therefore, this study further analyzes the similarity of the 68 groups that each contains more than 100 fatal crashes. There are 123,782 fatal crashes in those 68 groups, which count for 96.6% out of the total 128,149 crashes in the common scenario dataset. Groups of size 100 or less contain 4,367 cases, only 4.4% of the data in the common scenario dataset.

2.3.3.2 Clustering Spatially Defined Groups by Temporal Pattern Similarity

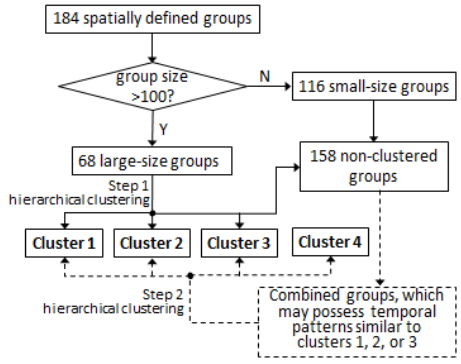
The 184 spatially defined groups were clustered in a two-step approach, as figure 2.7(a) shows. During the first step, a down-to-top hierarchical clustering algorithm below was used to cluster the 68 groups of size greater than 100. The algorithm begins from the bottom of the hierarchical tree, which has 68 single-group clusters (i.e., each cluster contains only one spatially defined group). In each step of iteration, the Pearson's r is calculated for every pair of clusters and, accordingly, the most similar clusters are merged as a new cluster. The number of clusters is reduced by at least one in each iteration. The process is automatically terminated when all groups are merged as a single cluster, reaching the top of the tree. But this study observed the growth and formation of three meaningful clusters during the iterative process of clustering and terminated the iteration process. The three groups are summarized below with details listed in Table 2.1:

Cluster 1: 2 groups that are on roads with a higher speed limit, on roadway, and non-junction areas.

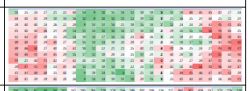
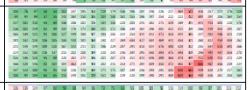
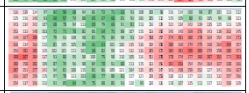
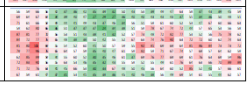
Cluster 2: 18 groups on roads with a slower speed limit, on roadway, and non-junction or some types of junction.

Cluster 3: 6 groups on the slower speed limit road, on roadside, and non-junction areas.

When the first step of the clustering process was terminated, 28 out of 68 groups (85.3% fatal crashes in the 68 groups) were clustered into the three groups above, and 42 groups still remained as single-group clusters.



(a) The clustering approach

Cluster	Description	Details				Pattern on Month and Hour
		FUNC_SYS	REL_ROAD	JUN_INT	No. Cases	
1	The higher speed limit road; on roadway and non-junction.	1,2	1	1	7420	
2	The slower speed limit road; on roadway.	3,4,5,6,7, 96,98,99	1	all except for 98&99	65192	
3	The slower speed limit road; on roadside or outside trafficway; non-junction.	3,4,5,6,7, 96,98,99	4,5	1	32955	
4	Others	-	-	-	16684	

(b) Four clusters

Figure 2.7 The clustering approach and result

Table 2.1 Clustering Results

(a) Details of the four clusters

Cluster	Description	Details			
		FUNC_SYS	REL_ROAD	JUN_INT	No. Cases
1	The higher speed limit road; on roadway and non-junction.	1	1	1	5521
		2	1	1	1899
2	The slower speed limit road; on roadway.	3	1	1	13582
		4	1	1	8782
		7	1	1	5474
		3	1	22	8387
		4	1	22	4674
		3	1	23	3613
		5	1	1	5055
		4	1	23	2399
		7	1	22	2898
		7	1	23	1427
		6	1	1	1069
		96,98,99	1	1	705
		3	1	4	1722
		5	1	22	2021
		4	1	4	1238
		7	1	4	736
		5	1	23	950
		5	1	4	560
		3,4,5,6,7,96, 98,99	1	All Junctions	2784*
3	The slower speed limit road; on roadside or outside trafficway; non-junction.	3	4	1	6529
		4	4	1	6934
		5	4	1	7664
		7	4	1	8726
		6	4	1	2295
		96,98,99	4	1	807
		3,4,5,6,7,96, 98,99	5	1	3114*
4	Others	-	-	-	16684

Note: Groups with* were added to clusters in step 2.

(b) Code and definitions

Variable	Code	Definition
FUNC_SYS (Functional System)	1	Interstate
	2	Principal Arterial – Other Freeways and Expressways
	3	Principal Arterial – Other
	4	Minor Arterial
	5	Major Collector
	6	Minor Collector
	7	Local
	96	Trafficway Not in State Inventory
	98	Not Reported
	99	Unknown
REL_ROA (Relation to Trafficway)	1	On Roadway
	2	On Shoulder
	3	On Median
	4	On Roadside
	5	Outside Trafficway
	6	Off Roadway – Location Unknown
	7	In Parking Lane/Zone
	8	Gore
	10	Separator
	11	Continuous Left-Turn Lane
	98	Not Reported
	99	Unknown
JUN_INT (Relation to Junction - Specific Location and Type of Intersection)	1	Non-Junction
	4	Driveway Access or related
	5	Entrance/Exit Ramp or Related
	6	Railway Grade Crossing
	7	Crossover Related
	16	Shared-Use Path Crossing or Trail
	17	Acceleration/Deceleration Lane
	18	Through Roadway
	19	Other Location Within Interchange Area
	22	Intersection or related: Four-Way Intersection
	23	Intersection or related:T-Intersection
	24	Intersection or related: Y-Intersection
	25	Intersection or related: Traffic Circle
	26	Intersection or related: Roundabout
	27	Intersection or related: Five-Point, or More
	30	Intersection or related: L-Intersection
	98	Not Reported
	99	Unknown

The second step of the clustering analysis was focused on the remaining 22,582 fatal crashes in the 158 non-clustered groups, including the 116 groups of size no greater than 100 and the 42 groups left from the first step of clustering. Some small size groups, after being combined into a bigger one, may possess a temporal pattern similar to one of the three clusters initially created from step 1. From observing the three clusters, this study identified a few candidates and successfully merged 2784 cases of "roads with a lower speed limit, on the roadway, and other junction types in addition to those included in cluster 2" with cluster 2. Then, 3114 cases of "roads with a lower speed limit, outside trafficway, and non-junction" were merged with cluster

3. The hierarchical clustering in the second step was stopped after two iterations. The remaining 16,684 fatal crashes were simply treated as cluster 4. These four clusters are depicted in figure 2.7(b).

The 2D temporal distribution of each cluster may be further split into the weekday and weekend distributions (by following the method in Section 2.4.2.1) if the two distributions both have recognizable patterns and the two patterns are different. Table 2.2 lists the Moran's I values for the whole-week, weekday and weekend distributions, respectively, for each cluster. The weekday temporal distribution of cluster 1 becomes more random compared to its whole-week temporal distribution, which suggests to keep the whole-week temporal distribution for cluster 1. The same conclusion was derived with respect to cluster 4. The weekday and weekend distributions of both clusters 2 and 3 have strong temporal patterns. Furthermore, the Pearson's r between the weekday and the weekend distribution of cluster 3 is 0.34, indicating these two patterns are quite different. The Pearson's r between the weekday and weekend distributions of cluster 2 is 0.67, which means certain degrees of similarity and dissimilarity between these two patterns both present. Therefore, this study decided to separate the weekday and weekend patterns for cluster 2. In total, six spatio-temporal clusters, whose Moran's I values are highlighted in table 2.2, were further analyzed for creating the attention guidance for fatal crashes. They are cluster 1, cluster 2-weekday, cluster 2-weekend, cluster 3-weekday, cluster 3-weekend and cluster 4.

Table 2.2 Moran's I values for spatial-temporal clusters

Spatial Cluster	Temporal Distribution		
	Whole-week	Weekday	Weekend
1	**0.64**	0.41	0.66
2	0.83	**0.80**	**0.84**
3	0.84	**0.81**	**0.81**
4	**0.63**	0.54	0.68

2.3.3.3 Learning the Spatial-Temporal Attention Mechanism through Association Rule Mining

A spatio-temporal attention guidance for each of the six clusters highlighted in table 2.2 was learnt through association rule mining. The attention guidance, composed of a set of association rules, informs the human driver or the autonomous driving system about "if a fatal crash would occur at the location and the time of the driving scene, what types of fatal crash (in terms of the First Harmful Event (FHE) and Manner of Collision (MC)) are likely to occur?"

For each of the six highlighted clusters in table 2.2, let x_{ij} denote the number of fatal crashes occurred in month i and hour j, and $x_{i,j,k}$ be the portion with type k FHE. A rule telling the association of fatal crash of type k FHE with the location and time of the crash is

$$x_{i,j} \Rightarrow x_{i,j,k}. \tag{2.4}$$

The study computed the supports of $x_{i,j}$ and $x_{i,j,k}$ respectively, as well as the confidence and lift of this rule to quantify the strength of this rule.

The support of $x_{i,j}$, denoted as $S_{i,j}$, calculates the fraction of fatal crashes in this cluster occurred in month i and hour j:

$$\text{Support}(x_{i,j}) \coloneqq S_{i,j} = \frac{x_{i,j}}{\sum_{i,j} x_{i,j}}. \tag{2.5}$$

The support $S_{i,j}$ helps determine if the rules relevant to $x_{i,j}$ are worth considering for further analysis. This study excluded rules with $S_{i,j} < 0.003472$ (=1/288) for further consideration.

Similarly, the support of $x_{i,j,k}$ calculates the fraction of fatal crashes with type k FHE in this cluster:

$$\text{Support}(x_{i,j,k}) \coloneqq S_k = \frac{\sum_{i,j} x_{i,j,k}}{\sum_{i,j} x_{i,j}}. \tag{2.6}$$

The study excluded rules with $S_k < 5\%$ to give attention to a consequence with at least 5% chance of occurrence.

The confidence of the rule in equation 2.4, denoted by $C_{i,j,k}$, calculates the likelihood a fatal crash would involve type k FHE if it occurs in month i and hour j:

$$\text{Confidence}(x_{i,j} \Rightarrow x_{i,j,k}) \coloneqq C_{i,j,k} = \frac{x_{i,j,k}}{x_{i,j}}. \tag{2.7}$$

The confidence value indicates how reliable the rule is.

The lift of the rule in equation 2.), designated as $L_{i,j,k}$, measures the rise of the probability of type k FHE with the time information being presenting over the same probability but without knowing the time:

$$\text{Lift}\left(x_{i,j} \Rightarrow x_{i,j,k}\right) \coloneqq L_{i,j,k} = \frac{C_{i,j,k}}{S_k}. \tag{2.8}$$

A lift value greater than 1 indicates a high association between antecedent of the rule (i.e., the specific time) and the consequence of the rule (the specific type of FHE). This study considered rules with a lift value greater than 1. The higher the lift value, the larger the chance that fatal crashes occur in month i and hour j would have the type k FHE. MC related rules were discovered and evaluated using the same method, and the discussion is skipped in the paper.

The spatio-temporal attention guidance for each cluster is visualized. The paper discusses three examples below and the other results and the developed rule query tool can be found at (44). Figure 2.8 shows the attention guidance for cluster 1 with respect to FHE. It includes three month-hour grids with colored dots on the grids. Each dot represents an association rule for fatal crashes in locations of cluster 1 and at the time specified by the dot's position on the grid. The color of the dot indicates the FHE type, the color scale measures the confidence of the rule, and the size of the dot indicates the lift of the rule. The attention guidance is used in the following approach. For example, if a vehicle is driving at a location belonging to cluster 1 and in hour 0 of month 5, there are two rules the driver needs to pay attention. The first rule is the green dot at the intersection of month 5 and hour 0 of the top right grid, which tells the following. If a fatal crash would occur, the chance that the FHE is pedestrian is 53.1% (i.e., the confidence of the rule) and the probability of having the pedestrian FHE at this specific time is 2.2 times of the

probability of having pedestrian FHE (i.e., the lift of the rule). The second rule is the pink dot at the intersection of month 5 and hour 0 of the bottom left grid, which is related to the rollover/overturn FHE. This rule is weaker than the first rule in that the confidence is 9.4% and the lift is 1.2.

The three grids with dots in figure 2.8 are also three temporal patterns of fatal crashes split by FHE types. Some useful information are learnt from comparing the three patterns. For example, fatal crashes of cluster 1 with the pedestrian FHE mainly occurred at night, from 6:00pm to 6:59am. Fatal crashes of cluster 1 with the rollover/overturn FHE mainly occur from April to November. The overall confidence of rules with the motor vehicle in transport FHE is the greatest, followed by those with the pedestrian FHE and then rollover/overturn FHE. Rules with large lift are mainly those with rollover/overturn FHE. Similarly, figure 2.9 visualizes the MC related rules split by MC types. The overall confidence of rules related to the front-to-rear MC is higher than that related to other MC types; but the overall lift of these rules are lower. Rules related to front-to-front MC have larger variations in both lift and confidence, reflecting its heterogeneity in the temporal space. The same for rules related to the same direction sideswipe MC and the angle MC. Moreover, many fatal crashes with the front-to-front MC are from midnight to early morning.

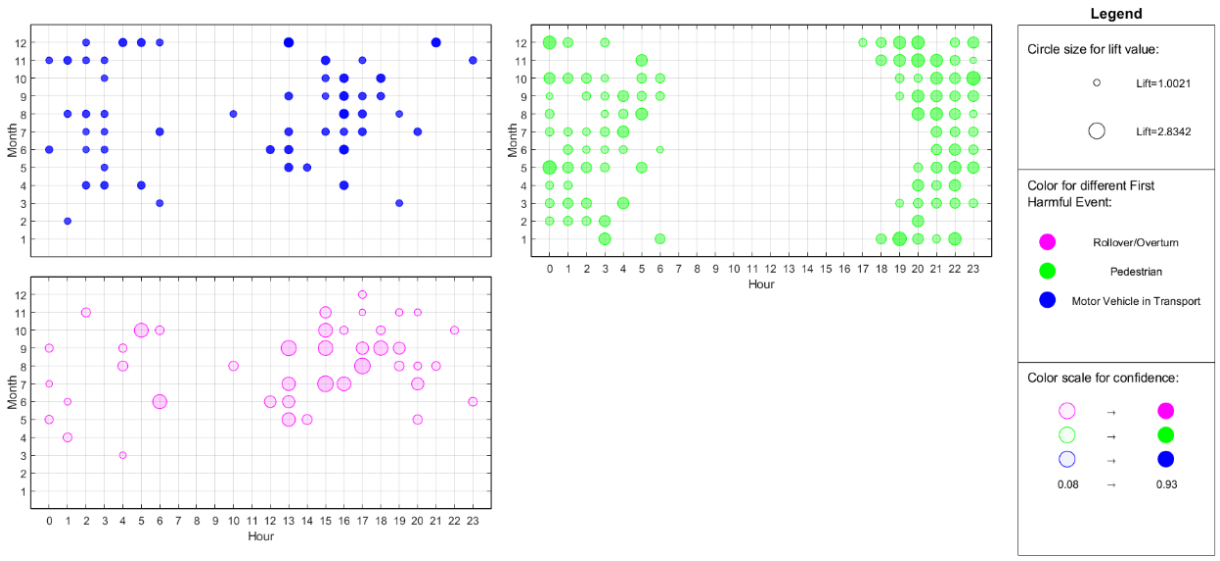


Figure 2.8 Association rules for fatal crashes in cluster 1 by FHE types

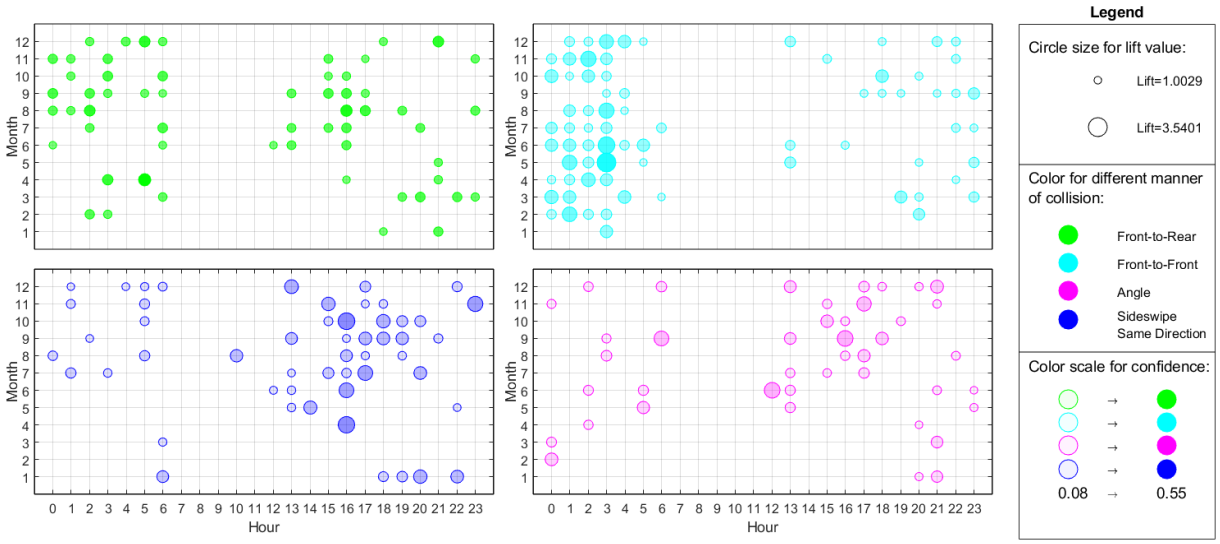


Figure 2.9 Association rules for fatal crashes in cluster 1 by MC types

Rules for the fatal crashes of cluster 2 are displayed in figure 2.10, which are split by FHE types and by weekday/weekend. Differences between clusters 1 and 2 are clearly seen by comparing figure 2.10 to figure 2.8. Fatal crashes with the motor vehicle in transport FHE and those with pedestrian FHE have a very small overlap on the temporal space. Generally speaking,

drivers should pay a special attention to pedestrians at night (5:00pm-2:59am) during the weekend, and at both night time (5:00pm-11:59pm) and morning rush hours (5:00am-7:59am) during weekdays. Drivers should also caution against fatal crashes with the motor vehicles in transport FHE in the afternoon of weekend (12:00pm-8:59pm) and during the daytime of weekday (7:00am-7:00pm). During the weekend, drivers should also be careful about fatal crashes with rollover/overturn FHE during 12:00pm to 2:59am although rules of such fatal crashes are small in the confidence.

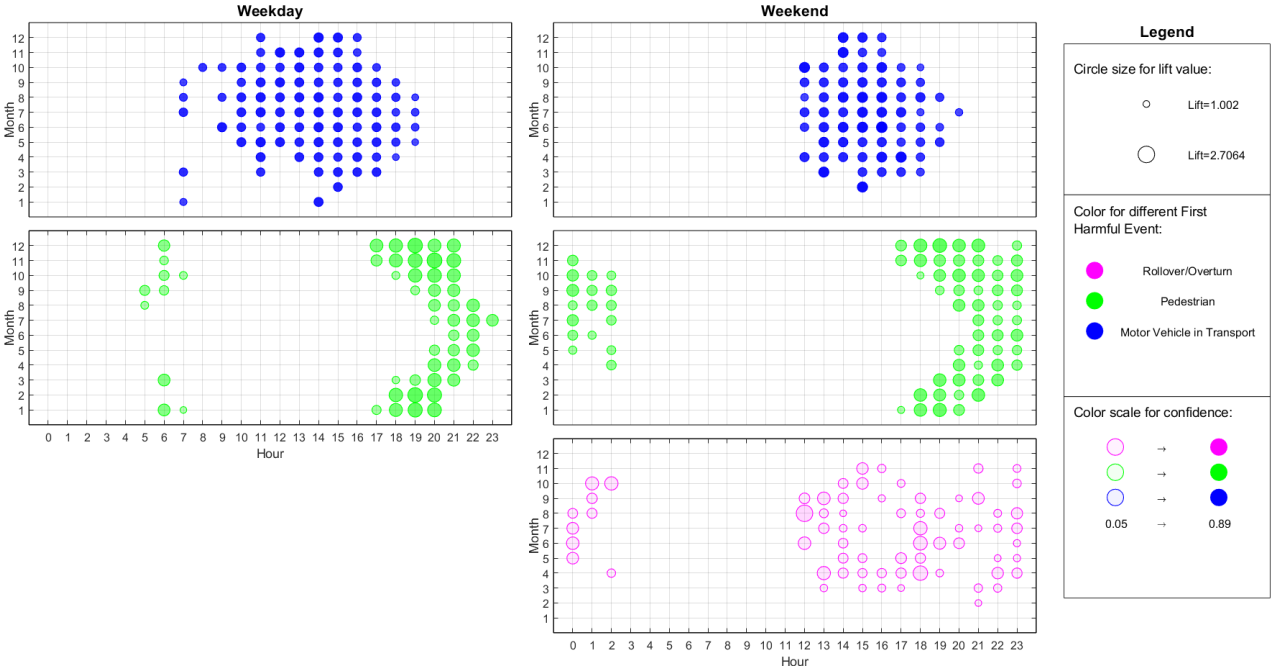


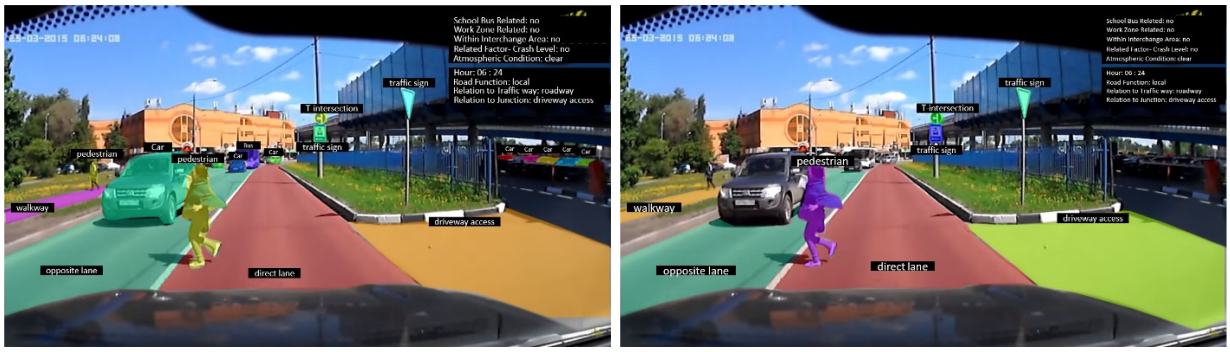
Figure 2.10 Association rules for fatal crashes in cluster 2 by FHE types and weekday/weekend

2.4 Illustrative Examples

Figure 2.11 shows an example that compares the results of driving scene analysis for fatal crash risk assessment without and with the spatio-temporal attention guidance. Figure 2.11(a) was labeled with every detected object, including vehicles staying in the parking garage, and a

pedestrian on the walkway along the opposite traffic, and vehicles far away from the driver's vehicle. To obtain this result may increase the cognitive load of the driver and can be misleading.

Figure 2.11(b) contains the road scene analysis result with the spatio-temporal attention guidance. Firstly, the CV algorithm provides the labels of variables for classifying crash scenarios (i.e. “School Bus Related”, “Work Zone Related”, “Within Interchange Area”, “Related Factor-Crash Level”, and “Atmospheric Conditions”), which are all zero. This indicates the driving scene belongs to the common scenario of fatal crashes. Then, the CV algorithm further classifies the driving scene according to the three location-related variables (i.e., “FUN_SYS”, “REL_ROAD”, and “JUN_INT”). The multi-label classification result determines the driving scene belongs to cluster 2. Given at a location belonging to cluster 2, the chance of involving a fatal crash at 6:24pm of September is 3.7%. There are two rules telling the possible types of the fatal crash in case it occurred. The first rule is related to pedestrian FHE with a confidence 0.4244 and a lift 1.7114. This means the chance of being a pedestrian involved fatal crash is 42.44% and this chance is 1.7114 times of the probability to see a pedestrian involved fatal crash at locations of cluster 2. The second rule is related to the front-to-rear MC with a confidence 0.0988 and a lift 1.3668. This rule tells drivers that any nearby vehicles in front deserve a special attention. The CV algorithm uses these two rules as guidance to detect objects that may be involved in a crash. The CV algorithm detects a pedestrian who is on the roadway and close to the driver's vehicle. Accordingly, this pedestrian is marked due to the risk of pedestrian involved fatal crash. The CV algorithm does not detect any vehicle in front of the driver and determines the front-to-rear collision caused fatal crash is less likely to occur.



(a) without an attention guidance (b) with an attention guidance

Figure 2.11 An illustrative example of driving scene analysis result for fatal crash risk assessment

Additional examples of driving scene analysis with a spatio-temporal attention guidance are illustrated in Figure 2.12. Examples in the first row are at locations of cluster 1, and the rest are at locations of cluster 2. Examples in the second row are in the daytime and those at night are in the third row. No pedestrian or vehicle is highlighted in images (a) and (f) because those driving scenes are found to be relatively safe. Pedestrians are highlighted in images (d), (g), and (h) due to a risk of pedestrian involved fatal crash. The nearby vehicle in images (b), (c), (e), and (i) are highlighted because a set of rules are found related to various vehicle involved fatal crashes such as rollover/overturn, motor vehicle in transport, front-to-rear collision, angle collision, and the same direction sideswipe. In images (g) and (i) nearby vehicles in the same driving direction are highlighted due to the risk of a front-to-rear fatal crash.

Figure 2.12 More examples of driving scene analysis with a spatio-temporal attention guidance

2.5 Conclusion and Future Work

This paper presented an approach to creating a spatio-temporal attention guidance for CV-based, real-time crash risk assessment. Five variables were identified for classifying crash scenarios and, accordingly, crash data were decomposed to generate scenario-based attention guidance. This study was focused on the common scenario of fatal crashes. Four clusters of driving scenes were discovered according to their locations, and up to two 2D temporal distributions were created for each cluster. Furthermore, correlations between the spatio-temporal information of fatal crashes with crash types were identified with association rule mining. Examples were illustrated, which implemented the developed spatio-temporal attention guidance for CV-based driving scene analysis for crash risk assessment. With the attention guidance, the driving scene analysis has clear focuses, thus reducing the cognitive load of human

drivers as well as the computational cost for CV algorithms. The analysis result is more meaningful and helps reduce the misleading information for users.

Important topics of future work have been identified from the paper. The study of this paper was focused on the common scenario of fatal crashes. Although the approach to creating the spatio-temporal attention guidance for other scenarios of fatal crashes are similar, knowledge of these scenarios is important for fatal risk mitigation, but it has not been thoroughly explored. Moreover, association rules are discontinuous and scatter at some spots of the temporal space. This is a difficulty in implementing the attention guidance. Therefore, the visualization of association rules in this paper can be improved to address this issue. The last but not the least, the attention guidance needs to be updated to capture the long-term evolution of crash characteristics after new crash data are released. This paper has built a solid foundation for exploring the above-mentioned future work.

.

Chapter 3 A System of Vision Sensor Based Deep Neural Networks for Complex Driving Scene Analysis in Support of Crash Risk Assessment and Prevention

3.1 Introduction

Driving scene analysis is an emergent topic of both driver assistance and autonomous driving technology. It provides a driver or an autonomous vehicle the situational awareness of the surrounding traffic. For the purpose of crash risk assessment and crash prevention, driving scenes are complex because multiple elements jointly contribute or relate to crashes. These include, but are not limited to, abnormality, nearby vehicles and pedestrians, road type and configuration, weather, and visibility. Therefore, a comprehensive analysis of driving scenes is a prerequisite for a reliable prediction of the immediate future risk condition. Multiple tasks are involved to deliver a comprehensive result of scene analysis, including object detection, instance segmentation, and classifications of scenario, road type, weather, visibility, and so on. A straightforward strategy is to have a system of different tools with each being dedicated to one or a few sub-tasks of driving scene analysis.

Increased availability of dash cameras for vehicles are adding to the visual sensing capability for driving scene analysis. Meanwhile, deep learning methods such as Convolutional Neural Networks (CNNs) have achieved tremendous success in various computer vision tasks such as image classification (45), object detection (46), and instance segmentation (47). CNN has many potentials for driving scene analysis. Yet, driving scene analysis in support of crash risk assessment and crash prevention has not been a thoroughly solved problem, probably due to the lack of effective methods of complex scene classification, a guidance for assessing crash risks from scene analysis results, and publicly available datasets that capture risk indicators and crash likelihood from complex driving scenes.

To tackle these challenges, various sub-tasks for driving scene analysis were developed, such as vehicle detection (48), pedestrian detection (49), lane finding (50), traffic sign detection (51), weather recognition (52), and infrastructure and traffic recognition (53). A few exceptions are noticed, which can describe traffic scenes with multi-label classification (54). However, the work of Chen et al. (54) is solely an image classification problem that completely relies on image-centric features for scene classification. Scene classification by crash likelihood and crash risk indicators is essential for crash prevention, but no work has thoroughly studied this problem. Researchers have also been trying to develop large-scale datasets (55) (56) (57) and deep learning models (58) (59) that well support classification tasks for understanding natural scenes. Some labeled traffic-scene datasets have been developed too (60) (61) (62) (63) to address the need for traffic scene understanding. These datasets are mainly focused on environmental perception and semantic segmentation. Above-mentioned efforts are still insufficient to deliver a satisfying driving scene analysis result to support crash risk assessment, because lacking an ability to capture deep features required to explain crash risks, such as crash likelihood or road function. Being able to identify the pre-crash scenario would allow for taking an immediate action of crash prevention. Crashes have varied features across different road types. Classifying driving scenes by road function (i.e., interstate, collector, local, and so on) can provide important clues for assessing crash risks (64).

To fill some of the identified gaps, this paper developed a system for complex driving scene analysis in support of crash risk assessment and crash prevention. First, this paper developed two new annotated image datasets to support the crash scene classification (no crash, pre-crash, crash) and road function classification, respectively. Second, it built a system of new multi-task neural networks named Multi-Net to generate multiple labels for classifying driving

scenes. Third, the Multi-Net was integrated with an object detector Yolo v3 (59) and an instance segmentation tool DeepLab v3 (65) to characterize vehicles and pedestrians that may cause a crash.

The remainder of this paper is organized as the following. The next section delineates the method to develop the proposed system, followed by the description of the developed datasets. The following section presents an evaluation of the developed system by illustrating various examples of scene analysis results. The conclusion and future work are summarized at the end, in the last section.

3.2 System for complex driving scene analysis

The proposed system for driving scene analysis is illustrated in figure 3.1. Driving scenes, in the form of RGB images captured by dash cameras mounted on vehicles, will be processed to identify the following information of each frame: the drivable area of the road, vehicles, risky pedestrians, crash likelihood, road function, weather, and time of day. This information is critical for assessing crash risks (64). The deep learning models for analyzing driving scenes include the multi-task deep neural networks (Multi-Net), an object detector YOLO v3 (59), and an image segmentation algorithm DeepLab v3 (65). The Multi-Net performs multi-task image classifications to provide the labels of crash likelihood, road function, weather, and time of day for driving scenes. YOLO v3 detects objects in video frames, such as pedestrians and vehicles. The distance to the nearest vehicle is calculated in every frame. The DeepLab v3 segments drivable area. Risky pedestrians who are within the drivable area are also detected.

3.2.1 Multi-Net for multi-task classification of driving scenes

A parallel neural network model named Multi-Net, shown in figure 3.2, was developed in this study to recognize the crash likelihood and another three crash risk indicators-road function,

weather, and time of day. The Multi-Net comprises of two parallel multi-task networks and each network is further split into two branches. Thus, the Multi-Net has four branches in total, which respectively perform four classification tasks to determine the class of the four variables of driving scenes. For example, the crash likelihood has three possible classes: pre-crash, crash, and no crash. The classes of road function, weather, and time of the day are displayed in figure 3.2.

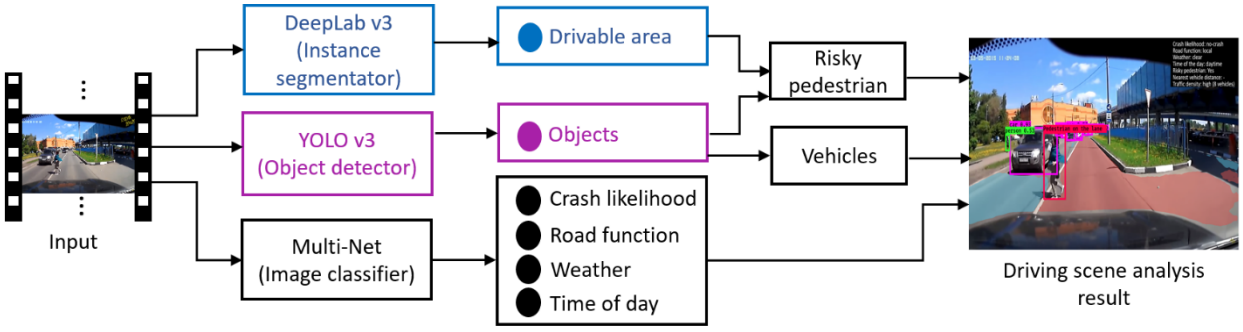


Figure 3.1 The schematic diagram of the driving scene analysis system

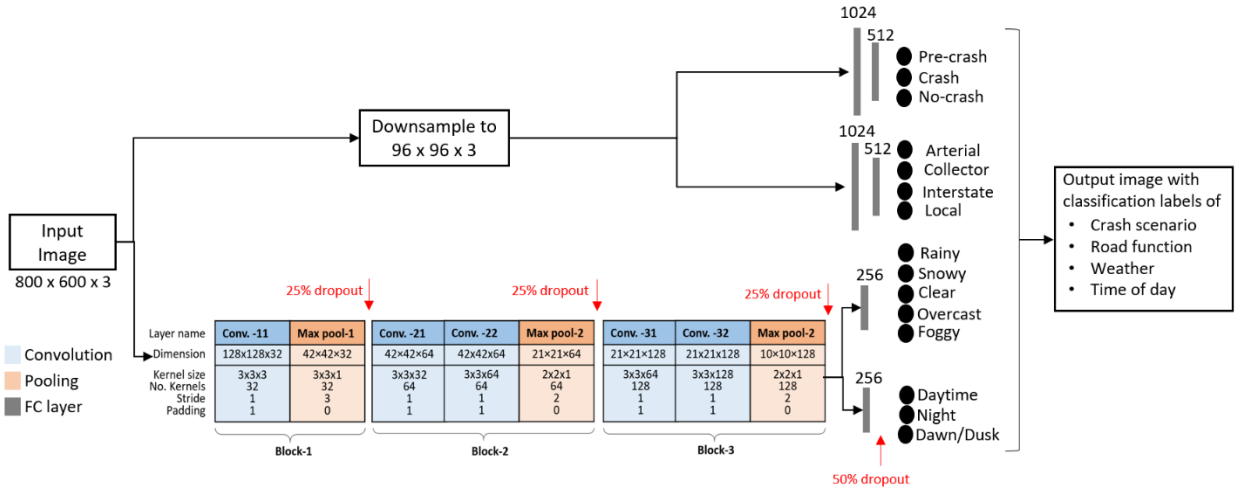


Figure 3.2 The architecture of the Multi-Net

Input RGB images to the Multi-Net are of size w × h × c, where w, h, and c are the width, height, and the number of channels of the images. The first network downsamples each

input image to $96 \times 96 \times 3$, flats the downsampled image, and then feeds it to two separate multi-level perceptron branches for classifying crash likelihood and road function, respectively. Each perceptron branch has two hidden layers. The first layer has 1024 neurons and the following layer has 512 neurons. Then, the final layers of the two branches respectively generate the class scores for crash likelihood and road function.

The second network is a Convolutional Neural Network (CNN) for classifying both weather and time of day. Since almost all the pixels of an image contain information of these two variables, a CNN is more capable than a perceptron in capturing the information, particularly the spatial dependency in an image, through the convolution and pooling operations. Firstly, input images to this CNN are downsampled to $128 \times 128 \times 3$. Then each downsampled image is processed by three blocks of convolutional layers and pooling layer. The kernels, stride, and padding of convolution and pooling operations are delineated in figure 3.2. The convolutional layers are the feature extractor of the CNN. A feature map of size $10 \times 10 \times 128$ is created from the last convolution layer. The feature map is sent to two branches of a fully connected layer with 256 neurons, which respectively perform the classifications of weather and time of day to generate class scores.

3.2.2 Object detection with YOLO v3

YOLO v3 is an object detector that can provide real-time object detection. It is a single-stage detection network that takes an image as the input and detects objects in the image by predicting their bounding box coordinates, the confidence of the bounding boxes, and class probabilities. Compared to its ancestors, YOLO v3 has a better performance in detecting small objects. Interested readers can refer to Redmon et al. (59) for details.

3.2.3 Distance to the nearest vehicle

The distance to the nearest vehicle along the traveling path is an important measurement for crash risk assessment. This study divided any input image into a 4×4 and concentrated on the area containing the travel path of the user's vehicle to reduce computational cost. Figure 3.3(a) illustrates an example wherein the traveling path of the user's vehicle is most likely found in the portion highlighted in yellow. Only vehicles in this highlighted region of interest are considered for the distance calculation.

Assume the region of interest contains *N* objects indexed by *i*. Let h_i be the height of vehicle *i* in the image and *H* be the true height. Let the focal length of the dash camera be *F* and d_i be the distance from the camera to vehicle *i*. Accordingly to the principle of similar triangles, $d_i/H = F/h_i$. This study assumes that *F* is 2.5 inch, *H* is 7 foot for a Van, 6 foot for an SUV, and 4.7 foot for a car. h_i is measured in pixels, and pixels per inch is 100. Therefore, the distance to vehicle *i* is

$$d_i = HF/h_i = 1250/h_i. \tag{3.1}$$

and the nearest vehicle is identified as

$$\underset{i}{\operatorname{argmax}}(1250/h_i) \tag{3.2}$$

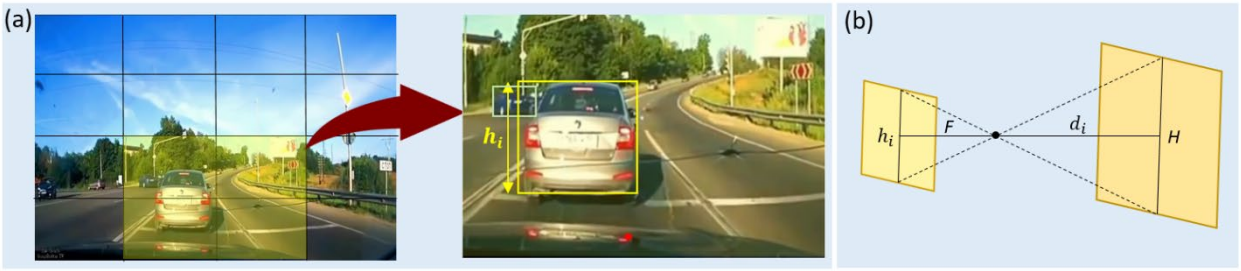


Figure 3.3 Measuring the distance to another vehicle

3.2.4 Detecting drivable area and risky pedestrian

The ability to segment different areas related to trafficway (such as on the trafficway, shoulder, roadside, median, and so on) is useful for crash risk assessment because crash risks differ in those areas. For example, pedestrians near the user's vehicle, particularly those on the same lane as the user, may be risky and require attention. Therefore, this study created the capability of detecting risky pedestrians. DeepLab v3 trained with the BDD100k dataset (63) was used in this study to segment the drivable area that includes two classes of lanes, the direct lane where the user is currently driving and the alternative lanes where the user can go by lane changing. In Figure 3.4, lanes highlighted in red color are direct lanes and those in blue are alternative lanes.

Figure 3.4 Detecting drivable area and risky pedestrians

DeepLab v3 cannot segment the area where objects like pedestrians and vehicles are present. Therefore, DeepLab by itself cannot identify the relationship of pedestrians to the driving lane. This study proposed an algorithm of risky pedestrian detection below. A rectangle bounding box $\mathrm{b_L}$ is extrapolated using the most outside coordinates of the segmented direct lane. If the YOLO v3 object detector detects *P* pedestrians in the driving scene, indexed by p, it returns all possible bounding boxes $\mathrm{b_p}$'s. If the two bounding boxes b_L and b_p have an overlap, pedestrian p is a risky one and a notification of risk is generated and the distance to that pedestrian is similarly estimated using the same method introduced in this paper to estimate the distance to the nearest vehicle. In figure 3.4, a yellow bounding box represents the extrapolated bounding box of the direct lane, and the red one represents the bounding box for the pedestrian. In both frames of Figure 3.4, these two boxes overlap each other, indicating the pedestrian is in the direct lane.

Algorithm 3.1 Risky pedestrian detection

$\mathrm{S_L} \leftarrow \{(x_1, \mathrm{y_1}), (x_2, y_2), \ldots, (x_n, y_n)\}$; //$\mathrm{S_L}$ is the segmented polygon for direct lane
$\mathrm{b_L} \leftarrow (\mathrm{x_{min}}, \mathrm{y_{min}}), (\mathrm{x_{min}}, \mathrm{y_{max}}), (\mathrm{x_{max}}, \mathrm{y_{min}}), (\mathrm{x_{max}}, \mathrm{y_{max}})\}$; //$\mathrm{b_L}$ is the minimum rectangle containing $\mathrm{S_L}$
for $\mathrm{p} \leftarrow$ 1 to P **do**
  $\mathrm{b_p} \leftarrow \{(\mathrm{x_1}, \mathrm{y_1}), (\mathrm{x_2}, \mathrm{y_2}), (\mathrm{x_3}, \mathrm{y_3}), (\mathrm{x_4}, \mathrm{y_4})\}_\mathrm{p}$; // bounding box for pedestrian p
  if $\mathrm{b_L} \cap \mathrm{b_p} \neq \emptyset$ **then**
    Generate a risk notification.
  end if
end for

3.3 Dataset Development

3.3.1 Data acquisition

No public datasets are available for training a deep learning neural network to classify crash likelihood with a pre-crash class and to classify the US road functional system. This study addressed this issue by developing two datasets that contain a diverse set of driving scenes with various road types and crash scenarios. The method for dataset development is illustrated in figure 3.5. This study collected crash videos from YouTube and general driving scene videos from HDD (62) using query terms like “crash” and “road function”. Segments containing classes of interests were retrieved, which were further converted into image frames. Consecutive frames of each video segment are similar and, therefore, sampling a portion of frames that are evenly distributed on the timeline would be sufficient to represent the segment. This study sampled one frame and then skipped five frames. The sampled frames were randomly split into training, validation, and testing sets.

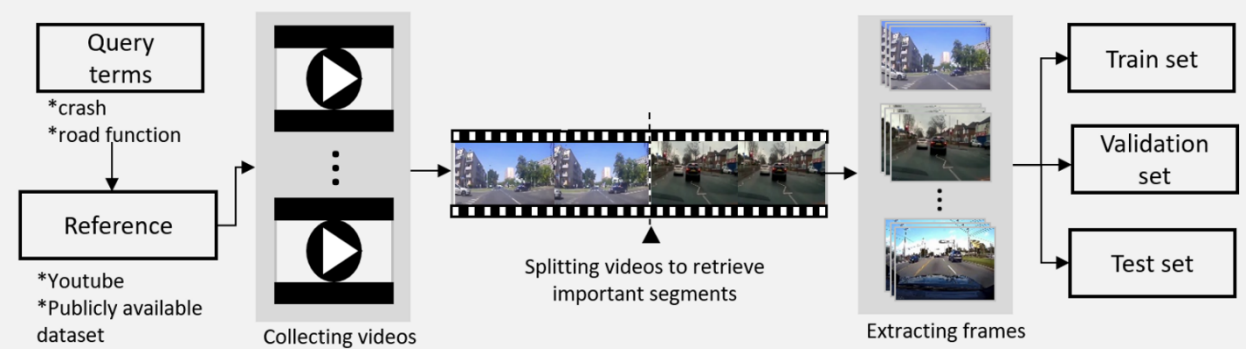


Figure 3.5 The pipeline for mining images for dataset development

The collected datasets were manually labeled with the classes of road function and crash likelihood. Figure 3.2 shows four general classes of road function: 1) arterial, 2) collector, 3)

interstate, and 4) local per the road functional system definition from (66). Crash likelihood has three classes. When a vehicle collides with other objects, then the scenario is defined as a crash. A clip of length two seconds before the crash is defined as pre-crash. Three seconds before the pre-crash is defined as no-crash.

3.3.2 Description of the datasets

TABLE 3.1 below summarizes the datasets, which can be accessed at (67). Each dataset is split into three mutually exclusive subsets: training, validating, and testing sets. In each dataset, samples are approximately evenly distributed among classes.

The crash likelihood classification is challenging because of the large variety of crash scenarios. The crash dataset developed by this study contains 15,900 images.

The classification of road function is less challenging compared to crash likelihood classification because the within-class variation of road function is relatively small. Therefore, a dataset with 6,400 images was developed by this study for classifying road function.

Table 3.1 Summary of the datasets

Task	No. of Classes	Total	Train	Val	Test
Crash likelihood	3	15,900	12,000	3,000	900
Road function	4	6,400	4,000	1,200	1,200
Weather	5	7,900	4,875	1,625	1,400
Time of day	3	7,400	4,875	1,625	900

The dataset for weather classification was created by taking images of clear, foggy, overcast, rainy, and snowy classes from BDD100k. However, there are only 143 images of the

foggy class in BDD100k. Therefore, this study generated synthetic images of the foggy class and combined these with the 143 images from BDD100k to have approximately the same amount of samples for every class of weather. Hence, this dataset has a total of 7,900 images and 6,500 of these are used for training and validation. The same training and validation sets were used for classifying time of day to leverage the annotation of the dataset.

3.4 Experimental Evaluation

3.4.1 Training the Multi-Net and DeepLab v3

This study requires two training procedures, one for training the Multi-Net for multi-task classification of driving scenes and the other for training the DeepLab v3 for segmenting drivable area. The Multi-Net was trained with the dataset developed by this project on a workstation with the following configuration: a 2.90 GHz Intel Xeon W-2102 CPU with 4 CPU cores, 16GB of RAM, and an Nvidia Geforce GTX 1080 Ti GPU with 11 GB memory. The multi-task network for classifying crash likelihood and road function system was trained for 80 epochs, whereas the network for classifying weather and time of day was trained for 30 epochs. Both networks were trained using a learning rate 0.01, a batch size 32, the stochastic gradient descent optimizer, and the categorical cross-entropy loss function. Training the two perceptron branches of the Multi-Net took 32 minutes and 14 minutes, respectively. Training the CNN of the Multi-Net took 90 minutes.

DeepLab v3 was trained with the BDD100k dataset on an Intel ® Xeon ® Gold 6152 CPU @ 2.10 GHz server with 503 GB of RAM and an Nvidia Tesla V100 GPU with 32 GB of memory. This training process took place for 20 epochs with a learning rate of 0.001 and a batch size of 4 images. It took approximately 68 hours to train the DeepLab v3.

3.4.2 Evaluation of classifier performance

Three standard metrics were used for evaluating the performance of the developed Multi-net, which are precision (Pr), recall (Rc), and f1-score (F1) calculated below:

$$\mathrm{Pr} = \frac{\text{No. correct predictions}}{\text{No. predictions}}, \mathrm{Rc} = \frac{\text{No. correct predictions}}{\text{No. ground-truth objects}}, \ \mathrm{F1} = \frac{2}{\mathrm{Pr}^{-1} + \mathrm{Rc}^{-1}} \qquad (3.3)$$

Each classifier of the Multi-Net was evaluated on the testing dataset for it, and the performance measurements are summarized in table 3.2. The right most column shows the number of supporting images for the evaluation. The precision, recall and f1 were calculated for each individual class of tested images. Then, the macro-average performance was calculated for the classifier, which is the average of class level performances. Figure 3.6 further shows the confusion matrices of the four classifiers to provide additional details.

Results in the table confirm that the multi-task network for classifying the crash likelihood and road function has a good capability of learning non-linear complex relationships for very different tasks. The precision and recall achieved by the classifier for crash likelihood are at least 90% for any of the three classes, yielding a macro-average f1 score of 93%. The classifier for road function achieved a similar performance. the precision and recall are 89% or higher for any of the four classes, yielding a macro-average f1 score of 92%.

The classifier for weather achieves a reasonable performance in classifying the foggy, overcast, and rainy classes, with 80% or higher recall. However, the recall values of the clear and snowy classes are 51% and 70%, respectively, indicating that some of these instances are classified as other classes. By looking at the confusion matrix in figure 3.6, 29% testing samples of the clear class are classified as snowy, and 18% as overcast. This is due to the fact that the

clear weather has a pixel distribution similar to the distributions of overcast and snowy weather. 22% testing samples of snowy class are classified as rainy class due to the feature similarity between them. These misclassified instances lower the precision of classifying overcast, rainy, and snowy weathers.

The classifier for time of day achieves an excellent performance in classifying night instances, evidenced by the 98% F1. Yet, this classifier has some difficulty in differentiating dawn/dusk and daytime samples due to the strong similarity in their visual features. The recall for the daytime class is 91% but the precision is 73%. Meanwhile, the recall for the dawn/dusk class is 65%. From the confusion matrix in figure 3.6, it is observed that 34% testing samples of the dawn/dusk class are classified as daytime, thus lowering the precision of daytime class and the recall of dawn/dusk. The precision for classifying the dawn/dusk instances is 87% and those misclassified as the dawn/dusk class are mainly from the daytime class.

Table 3.2 Performance of classifiers

Classifier	Classes	Precision	Recall	F1	Support
Crash	crash	0.95	0.90	0.93	272
	no-crash	0.90	0.98	0.94	322
	pre-crash	0.95	0.90	0.92	306
	macro-crash	0.93	0.93	0.93	900
Road function	arterial	0.93	0.93	0.93	300
	collector	0.94	0.89	0.91	292
	interstate	0.91	0.93	0.92	292
	local	0.91	0.93	0.92	313
	macro-avg	0.92	0.92	0.92	1200
Weather	clear	0.88	0.51	0.65	300
	foggy	0.87	0.85	0.86	300
	overcast	0.7	0.80	0.77	300
	rainy	0.71	0.89	0.79	250
	snowy	0.71	0.89	0.79	250
	macro-avg	0.76	0.75	0.74	1400
Time of day	dawn/dusk	0.87	0.65	0.75	300
	daytime	0.73	0.91	0.81	300
	night	0.98	0.98	0.98	300
	macro-avg	0.86	0.85	0.85	900

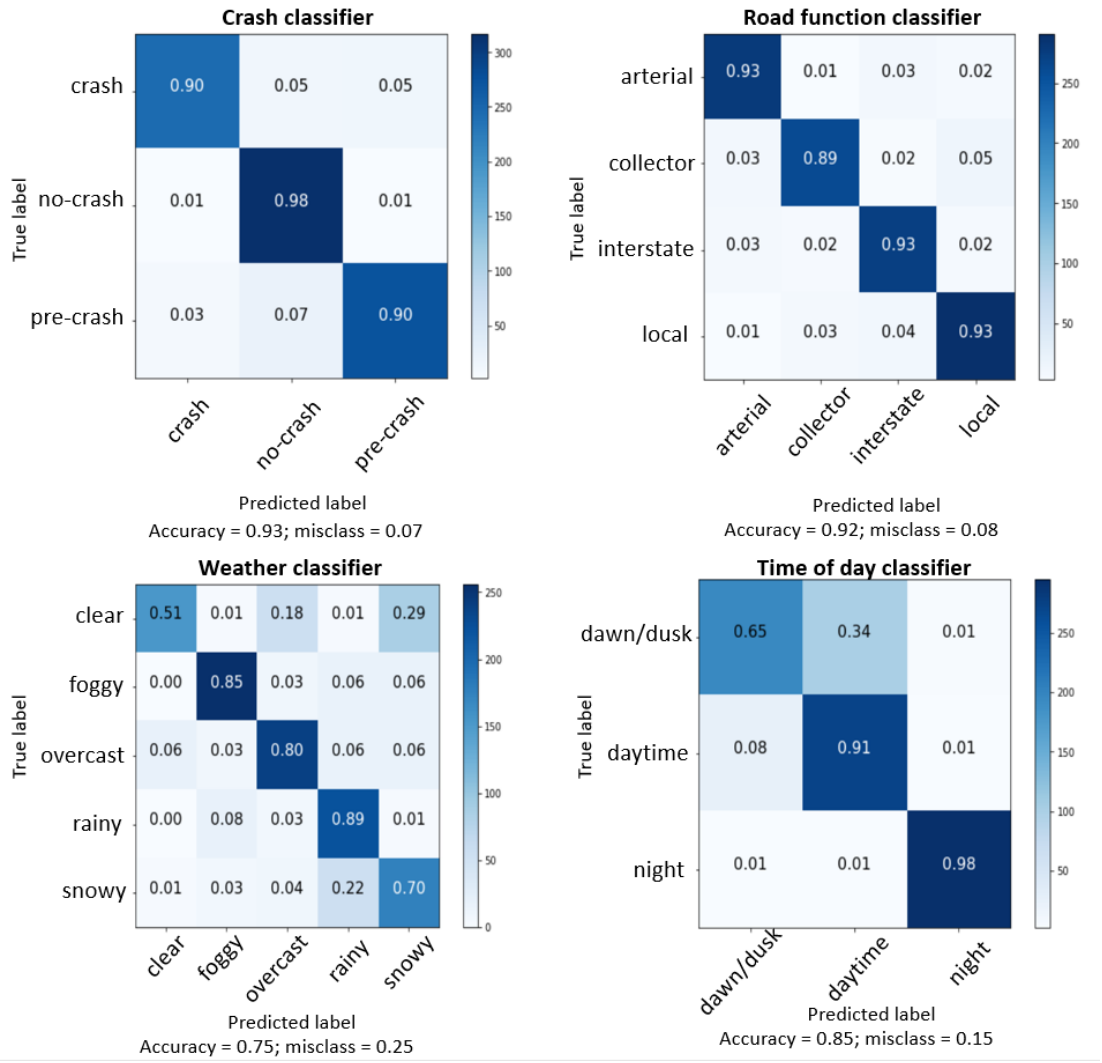


Figure 3.6 Confusion matrices

3.4.3 Examples of classification results

Figure 3.7 illustrates some examples of classification made by the Multi-Net. Four classification labels are generated for each image, shown on the top right of the image. For example, the classification result of (i) says that the driving scene is on a local road during daytime, the weather is clear, and the crash likelihood is pre-crash because the user's vehicle is about to crash into another nearby vehicle. These examples show that the Multi-Net can successfully classify the driving scenes. (h) is classified as a crash scenario when the user's

vehicle collides with that vehicle, and the other seven scenes are classified as no crash. The Multi-Net successfully classified the road function in these examples: scenes (b), (c), and (f) are classified as arterial, (a) is collector, (d) and (e) are interstate, and (g), (h), and (i) are local road. Weather can be classified too. (a), (d), (e), and (g) are classified as snowy, (b), (h), and (i) are clear, (c) is foggy, and (f) is rainy. The time of day is the fourth label of each scene. (b) is classified as night, and the rest are daytime.

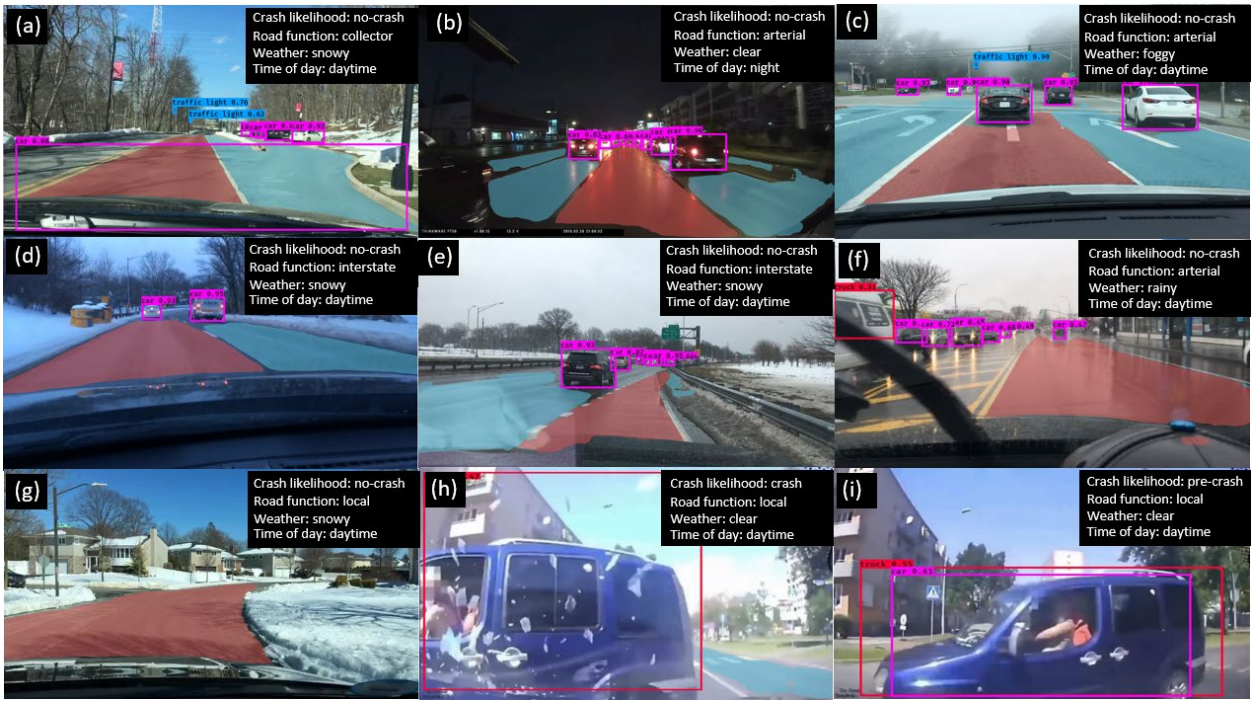


Figure 3.7 Examples with results generated by the Multi-net

3.4.4 Examples of object detection

Beyond the scene classification, the developed system detects and locates objects that may be involved in a crash. The object detector YOLO v3 returns the bounding boxes of detected objects in the image. The segmentation tool DeepLab v3 segments the drivable areas with two colors: red for the direct lane and blue for the alternative lane. Based on those, the

location of pedestrians and vehicles in relation to the drivable area are useful information for assessing crash risks. Figure 3.8 shows some examples with two additional labels. In scene (a) a pedestrian is detected on the direct lane of an interstate highway and the person is identified as a risky pedestrian highlighted with a red bounding box. In (b) two pedestrians were detected, with one on the direct lane and the other outside the drivable area. Therefore, the driving scene analysis system treats the one outside the drivable area as a safe pedestrian and highlights the person with a green bounding box. The pedestrian on the direct lane is highlighted with a red bounding box. Distance to the risky pedestrian is also calculated. For example, the risky pedestrian is 10.23 feet away from the user's vehicle in (a) and 5.07 feet away in (b). The proposed method also calculates the distance to the nearest vehicle. For example, in scene (c) the nearest vehicle is 4.63 feet from the user vehicle, which the user should pay attention to for avoidance of the rear-to-front collision. In scene (d), the nearest vehicle is 14.34 feet away and it is on the alternative lane. Unlike (c), the information of the nearest vehicle in (d) does not suggest any crash risk involving another vehicle.

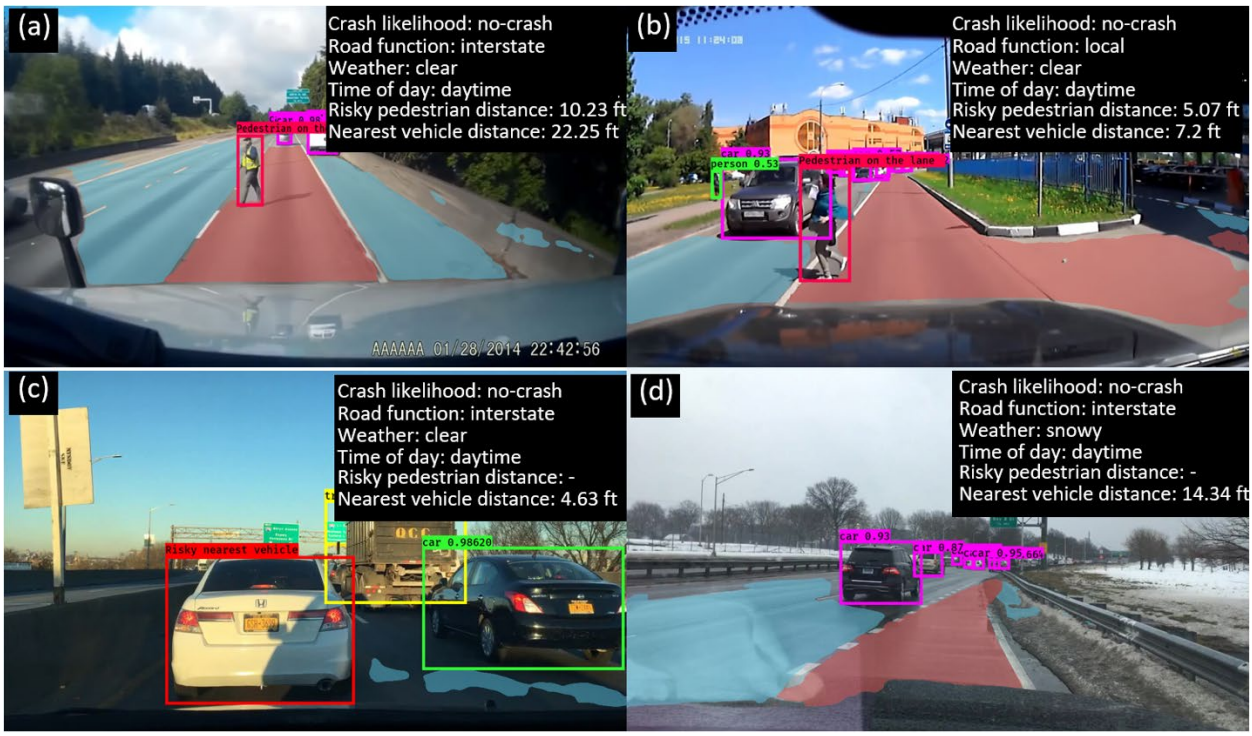


Figure 3.8 Example of detecting and locating risky pedestrian and the nearest vehicle

3.4.5 Inference speed

The inference speed of the proposed system is critical because driving scene analysis needs to be real-time, ideally. This study thus evaluated the efficiency of the proposed system in terms of the inference speed on the workstation that was used to train the Multi-Net. The workstation is Dell Precision 5820, a typical workstation in the market. Loading the trained weights of multiple networks requires a lot of computer memory. To save the memory, the proposed system was tested in two steps. At first, testing images are passed through the DeepLab v3 for the instance segmentation. The segmented images are stored in a disk. DeepLab v3 processed images at an inference speed of 2 frames per second (fps). Then the images were passed through Multi-Net and Yolo v3 for classification and object detection, where images were processed at a speed of 6 fps. A normal dash camera typically captures 30 fps, and this study

samples 5 fps for analysis. That is, the proposed system provides a real-time inference speed without the instance segmentation. The inference speed of the segmentation task needs to be improved in order to achieve real-time scene analysis.

3.5 Conclusion

This paper presented a system for vision sensor based complex driving system analysis in support of crash risk assessment and crash prevention. The system is composed of a Multi-Net model for scene classification, the YOLO v3 object detector, and the DeepLab v3 segmentor. Multi-Net performs scene classification and provides four labels for each scene including the likelihood of a crash, road function, weather, and time of day. The DeepLab v3 and YOLO v3 are combined to detect and locate risky pedestrians and the nearest vehicles in the scene. All the identified information can provide the situational awareness to an autonomous vehicle or a human driver for identifying crash risk from the surrounding traffic. To address the scarcity of annotated image datasets for studying crashes, two completely new datasets were developed, which were proved to be effective in training the deep neural networks with the proposed method by this study.

Vision sensor-based driving scene analysis is still facing a variety of challenges and needs. Firstly, the system can be expanded by adding the ability to classify additional risk indicators. For example, crash risks have varied features at different junction-related locations such as non-junction, four-way intersection, Y-intersection, T-intersection, driveway access, and so on. The ability to classify scenes according to their junction-related feature, as well as the ability to detect other classes of road areas such as shoulder, roadside, and median, will make the crash risk assessment more reliable. Secondly, efforts can be made to further improve the performance of the developed system. The accuracy of weather classification needs to be

improved. Other methods that can substitute the segmentation method for identifying the drivable area need to be developed and integrated to the system in order to increase the inference speed. Besides, a method should be introduced to tackle the memory issue while loading trained weights from multiple networks. The estimation of the traffic density (i.e., the number of vehicles per unit length of driving lane) from dash cameras mounted on vehicles is a challenging research question but it will provide critical information for crash risk prevention. This paper has built a foundation for further exploring these exciting research opportunities.

Chapter 4 Recommendations Developed from the Project

This project analyzed 5-year fatal crash report data in the United States to develope a spatio-temporal attention guidance for computer vision (CV) based real-time risk assessment. Then, a system of vision sensor based deep neural networks was developed for analyzing complex driving scenes in order to assess crash risks. Given the developed attention guidance, the driving scene analysis is focused on suggested elements and their relationship may contribute to traffic crashes. Recommendations that can be made based on this project are the following:

- Sensor and artificial intelligence technologies provide new opportunities for safety enhancement. Yet these high technologies are not designed to address specific needs for roadway safety enhancement. Therefore, data analytics methods, deep learning algorithms, and intelligent computational models need to be defined to adapt to the new tasks of roadway safety. This project is an example of such efforts.
- Data integration is a way of maximizing the usefulness of various sources of information for safety enhancement. Video data have richer information than the traditional reporting method of data collection. With vision sensors mounted in vehicles, video data can be captured and process in near real-time. Yes, compared to the big amount of fatal crash report data, video data of crashes are still limited. Through analyzing the big data in crash report systems, knowledge and patterns of crash risks can be discovered, which can help improve the efficiency and performance of AI algrithms in analyzing video data.
- We envision that more and more autonomous vehicles will operate on the road. This change also has a variety of safety issues that need to be addressed. For example, when autonomous vehicles are performing slow moving operations on the road, the question

becomes how to prevent other vehicles from crashing into the autonomous vehicles? The developed technologies and methods in this project can also be implemented for that purpose.

References

1. Early Estimate of Motor Vehicle Traffic Fatalities in 2017. *National Highway Traffic Safety Administration (NHTSA).* 2018.
2. Solving for Safety Visualization Challenge. *US Department of Transportation (USDOT).* 2018.
3. Strategic Plan for FY 2018-2022. *US Department of Transportation (USDOT).* 2018.
4. Beyond Traffic 2045: Trend and Choices. *US Department of Transportation (USDOT).* 2015.
5. Research, Development and Technology Strategic Plan FY2017-2011. *US Department of Transportation (USDOT).* 2017.
6. FATALITY ANALYSIS REPORTING SYSTEM (FARS) ENCYCLOPEDIA. s.l. : US Department of Transportation National Highway Traffic Safety Administration, 2020.
7. Tang, Eric and Peach, Kara and DeFisher, Joshua and Eccles, Kimberly A and others. *Transportation Safety Planning and the Zero Deaths Vision: A Guide for Metropolitan Planning Organizations and Local Communities.* s.l. : United States. Federal Highway Administration. Office of Safety, 2018.
8. Kuznicki, Scott. *Highway Safety Improvement Program National Scan Tour - Chapter 8– Considering all "4E's".* s.l. : US Department of Transportation Federal Highway Administration, 2016. FHWA-SA-16-024.
9. *Diagnostic analysis of the effects of weather condition on pedestrian crash severity.* Zhai, Xiaoqi and Huang, Helai and Sze, NN and Song, Ziqi and Hon, Kai Kwong. s.l. : Elsevier, 2019, Accident Analysis & Prevention, Vol. 122, pp. 318–324.
10. *Epidemiology cf pedestrian--Motor vehicle fatalities and injuries, 2006–2015.* Chong, Shu-Ling and Chiang, Li-Wei and Allen Jr, John Carson and Fleegler, Eric William and Lee, Lois Kaye. s.l. : Elsevier, 2018, American Journal Of Preventive Medicine, Vol. 55, pp. 98–105. 1.
11. *What role do precrash driver actions play in work zone crashes?: application of hierarchical models to crash data.* Liu, Jun and Khattak, Asad and Zhang, Meng. s.l. : SAGE Publications Sage CA: Los Angeles, CA, 2016, Transportation Research Record, Vol. 2555, pp. 1–11. 1.
12. *Deep learning-based vehicle behaviour prediction for autonomous driving applications: a review.* Mozaffari, Sajjad and Al-Jarrah, Omar Y and Dianati, Mehrdad and Jennings, Paul and Mouzakitis, Alexandros. 2019, arXiv preprint arXiv:1912.11676.
13. *Embedding vision-based advanced driver assistance systems: a ssurvey.* Velez, Gorka and Otaegui, Oihana. s.l. : IET, 2016, IET Intelligent Transport Systems, Vol. 11, pp. 103–112. 3.
14. *Egocentric vision-based future vehicle localization for intelligent driving assistance systems.* Yao, Yu and Xu, Mingze and Choi, Chiho and Crandall, David J and Atkins, Ella M and Dariush, Behzad. s.l. : IEEE, 2019. 2019 International Conference on Robotics and Automation (ICRA). pp. 9711–9717.
15. Crash Prediction and Avoidance by Identifying and Evaluating Risk Factors from Onboard Cameras Supervised Fire and Traffic Accident Scene Classification. s.l. : Mid-America Transportation Center, 2019.
16. *A System of Systems Approach to Optimize a Realtime Risk Situational Awareness System.* Li, Yu and Dagli, Cihan. s.l. : IEEE, 2020. 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE). pp. 17–22.
17. Crash Report Sampling System. s.l. : US Department of Transportation National Highway Traffic Safety Administration, 2020.
18. Safety Data Initiative. s.l. : US Department of Transportation, 2020.

19. Solving for Safety Submissions. *US Department of Transportation.* 2020.
20. Traffic Fatalities in Crashes Involving Speed. s.l. : US Department of Transportation, 2016.
21. Pedestrian Traffic Fatalities Visualization, 2011-2017. s.l. : Department of Transportation, 2020.
22. Transportation Research Informatics Platform (TRIP). s.l. : CUBRC Advantage Through Technology, 2020.
23. Data, Analyzing Big. *Knowledge Discovery in Massive Transportation Datasets.* U.S. Department of Transportation Federal Highway Administration. 2018. FHWA-HRT-17-117.
24. Guide to Calculating Costs. s.l. : Injury Facts,National Safety Council, 2020.
25. *A study on crashes related to visibility obstruction due to fog and smoke.* Abdel-Aty, Mohamed and Ekram, Al-Ahad and Huang, Helai and Choi, Keechoo. s.l. : Elsevier, 2011, Accident Analysis & Prevention, Vol. 43, pp. 1730–1737. 5.
26. *Investigating the risk factors associated with pedestrian injury severity in Illinois.* Pour-Rouholamin, Mahdi and Zhou, Huaguo. s.l. : Elsevier, 2016, Journal Of Safety Research, Vol. 57, pp. 9–17.
27. *Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression.* Harb, Rami and Radwan, Essam and Yan, Xuedong and Pande, Anurag and Abdel-Aty, Mohamed. s.l. : American Society of Civil Engineers, 2008, Journal of Transportation Engineering, Vol. 134, pp. 203–214. 5.
28. *S-CarCrash: Real-time crash detection analysis and emergency alert using smartphone.* Sharma, Harit and Reddy, Ravi Kanth and Karthik, Archana. s.l. : IEEE, 2016. 2016 International Conference on Connected Vehicles and Expo (ICCVE). pp. 36–42.
29. *Evaluation of rear-end crash risk at work zone using work zone traffic data.* Meng, Qiang and Weng, Jinxian. s.l. : Elsevier, 2016, Accident Analysis & Prevention, Vol. 92, pp. 43–52.
30. *Differences in passenger car and large truck involved crash frequencies at urban signalized intersections: an exploratory analysis.* Dong, Chunjiao and Clarke, David B and Richards, Stephen H and Huang, Baoshan. s.l. : Elsevier, 2014, Accident Analysis & Prevention, Vol. 62, pp. 87–94.
31. *An exploratory study on the safety effects of speed limit reduction policy in Brisbane and Melbourne CBDs.* Kamruzzaman, Md and Debnath, Ashim Kumar and Bourdaniotis, Vasili. s.l. : ATRF, 2019. Proceedings of the 2019 Australasian Transport Research Forum. pp. 1–12.
32. *Weather and road geometry impact on longitudinal driving behavior: Exploratory analysis using an empirically supported acceleration modeling framework.* Hamdar, Samer H and Qin, Lingqiao and Talebpour, Alireza. s.l. : Elsevier, 2016, Transportation Research Part C: Emerging Technologies, Vol. 67, pp. 193–213.
33. *Effectiveness of enforcement of seatbelt law: an exploratory empirical analysis using aggregate data.* Ahmad, Numan and Ahmed, Anwaar and Shah, Akhtar Ali. s.l. : Statens vag- och transportforskningsinstitut, 2018. 18th International Conference Road Safety on Five Continents (RS5C 2018), Jeju Island, South Korea, May 16-18, 2018.
34. *Exploratory multinomial logit model–based driver injury severity analyses for teenage and adult drivers in intersection-related crashes.* Wu, Qiong and Zhang, Guohui and Ci, Yusheng and Wu, Lina and Tarefder, Rafiqul A and Alcantara, Adelamar Dely. s.l. : Taylor & Francis, 2016, Traffic Injury Prevention, Vol. 17, pp. 413–422. 4.
35. *Identifying the key risk factors of traffic accident injury severity on slovenian roads using a non-parametric classification tree.* Rovsek, Vesna and Batista, Milan and Bogunovic, Branko. s.l. : Taylor & Francis, 2017, Transport, Vol. 32, pp. 272–281. 3.

36. *Data-mining techniques for exploratory analysis of pedestrian crashes.* Montella, Alfonso and Aria, Massimo and D'Ambrosio, Antonio and Mauriello, Filomena. s.l. : SAGE Publications Sage CA: Los Angeles, CA, 2011, Transportation Research Record, Vol. 2237, pp. 107–116. 1.
37. *Driver crash risk factors and prevalence evaluation using naturalistic driving data.* Dingus, Thomas A and Guo, Feng and Lee, Suzie and Antin, Jonathan F and Perez, Miguel and Buchanan-King, Mindy and Hankey, Jonathan. s.l. : National Acad Sciences, 2016, Proceedings of the National Academy of Sciences, Vol. 113, pp. 2636–2641. 10.
38. *Spring forward at your own risk: daylight saving time and fatal vehicle crashes.* Smith, Austin C. 2016, American Economic Journal: Applied Economics, Vol. 8, pp. 65–91. 2.
39. *Spatial epidemiologic analysis of relative collision risk factors among urban bicyclists and pedestrians.* Delmelle, Elizabeth Cahill and Thill, Jean-Claude and Ha, Hoe-Hun. s.l. : Springer, 2012, Transportation, Vol. 39, pp. 433–448. 2.
40. *The interpretation of statistical maps.* Moran, Patrick AP. s.l. : JSTOR, 1948, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 10, pp. 243–251. 2.
41. *Developing a bivariate spatial association measure: an integration of pearson's r and Moran's i.* Lee, Sang-Il. s.l. : Springer, 2001, Journal Of Geographical Systems, Vol. 3, pp. 369–385. 4.
42. Office of Data Acquisition, State Data Reporting Systems Division. *Fatality Analysis Reporting System (FARS) Analytical User's Manual, 1975-2018.* US Department of Transportation National Highway Traffic Safety Administration. 2019.
43. *A system of vision sensor based deep learning networks for multi-label classification of complex driving scenes.* Karim, Muhammad Monjurul and Li, Yu and Qin, Ruwen and Yin, Zhaozheng. 2020, Washington, DC: Transportation Research Board-USA. Working paper.
44. Li, Yu. Association Rules Results of MATC Project. 2020.
45. *Deep convolutional neural networks for image classification: A comprehensive review.* Rawat, Waseem and Wang, Zenghui. s.l. : MIT Press, 2017, Neural Computation, Vol. 29, pp. 2352–2449. 9.
46. *Object detection with deep learning: A review.* Zhao, Zhong-Qiu and Zheng, Peng and Xu, Shou-tao and Wu, Xindong. s.l. : IEEE, 2019, IEEE Transactions on Neural Networks and Learning Systems, Vol. 30, pp. 3212–3232. 11.
47. *A review on deep learning techniques applied to semantic segmentation.* Garcia-Garcia, Alberto and Orts-Escolano, Sergio and Oprea, Sergiu and Villena-Martinez, Victor and Garcia-Rodriguez, Jose. 2017, arXiv preprint arXiv:1704.06857.
48. *Joint monocular 3D vehicle detection and tracking.* Hu, Hou-Ning and Cai, Qi-Zhi and Wang, Dequan and Lin, Ji and Sun, Min and Krahenbuhl, Philipp and Darrell, Trevor and Yu, Fisher. 2019. Proceedings of the IEEE International Conference on Computer vision. 5390–5399.
49. *High-level semantic feature detection: A new perspective for pedestrian detection.* Liu, Wei and Liao, Shengcai and Ren, Weiqiang and Hu, Weidong and Yu, Yinan. 2019. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 5187–5196.
50. *Learning lightweight lane detection cnns by self attention distillation.* Hou, Yuenan and Ma, Zheng and Liu, Chunxiao and Loy, Chen Change. 2019. Proceedings of the IEEE International Conference on Computer Vision. 1013–1021.
51. *Traffic-sign detection and classification in the wild.* Zhu, Zhe and Liang, Dun and Zhang, Songhai and Huang, Xiaolei and Li, Baoli and Hu, Shimin. 2016. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2110–2118.

52. *A CNN--RNN architecture for multi-label weather recognition.* Zhao, Bin and Li, Xuelong and Lu, Xiaoqiang and Wang, Zhigang. s.l. : Elsevier, 2018, Neurocomputing, Vol. 322. 47–57.
53. *Multi-label classification of traffic scenes.* Sikiric, Ivan and Brkic, Karla and Horvatin, Ivan and {\v{S}}egvic, Sini{\v{s}}a. 2014. Proc. Croatian Comput. Vis. Workshop (CCVW). 9–14.
54. *Deep integration: A multi-label architecture for road scene recognition.* Chen, Long and Zhan, Wujing and Tian, Wei and He, Yuhang and Zou, Qin. s.l. : IEEE, 2019, IEEE Transactions on Image Processing, Vol. 28, pp. 4883–4898. 10.
55. *Imagenet: A large-scale hierarchical image database.* Deng, Jia and Dong, Wei and Socher, Richard and Li, Li-Jia and Li, Kai and Fei-Fei, Li. s.l. : IEEE, 2009. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 248–255.
56. *Microsoft coco: Common objects in context.* Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Dollar, Piotr and Zitnick, C Lawrence. s.l. : Springer, 2014. European Conference on Computer Vision. 740–755.
57. *Places: A 10 million image database for scene recognition.* Zhou, Bolei and Lapedriza, Agata and Khosla, Aditya and Oliva, Aude and Torralba, Antonio. s.l. : IEEE, 2017, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, pp. 1452–1464. 6.
58. *Faster r-cnn: Towards real-time object detection with region proposal networks.* Ren, Shaoqing and He, Kaiming and Girshick, Ross and Sun, Jian. 2015. Advances in Neural Information Processing Systems. pp. 91–99.
59. *Yolov3: An incremental improvement.* Redmon, Joseph and Farhadi, Ali. 2018, arXiv preprint arXiv:1804.02767.
60. *Are we ready for autonomous driving? the kitti vision benchmark suite.* Geiger, Andreas and Lenz, Philip and Urtasun, Raquel. s.l. : IEEE, 2012. 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361.
61. *The cityscapes dataset for semantic urban scene understanding.* Cordts, Marius and Omran, Mohamed and Ramos, Sebastian and Rehfeld, Timo and Enzweiler, Markus and Benenson, Rodrigo and Franke, Uwe and Roth, Stefan and Schiele, Bernt. 2016. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3213–3223.
62. *Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning.* Ramanishka, Vasili and Chen, Yi-Ting and Misu, Teruhisa and Saenko, Kate. 2018. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7699–7707.
63. *Bdd100k: A diverse driving video database with scalable annotation tooling.* Yu, Fisher and Xian, Wenqi and Chen, Yingying and Liu, Fangchen and Liao, Mike and Madhavan, Vashisht and Darrell, Trevor. s.l. : CoRR, 2018, arXiv preprint arXiv:1805.04687, Vol. 2, p. 6. 5.
64. Li, Yu and Karim, Muhammad Monjurul and Qin, Ruwen and Wang, Zuhui and Yin, Zhaozheng and Sun, Zeyi. *Crash Report Data Analysis for Creating A Spatio-Temporal Attention Guidance for Vision Based Real-Time Crash Risk Assessments.* s.l. : Working paper, 2020.
65. *Rethinking atrous convolution for semantic image segmentation.* Chen, Liang-Chieh and Papandreou, George and Schroff, Florian and Adam, Hartwig. s.l. : arXiv preprint arXiv:1706.05587, 2017.
66. FHWA. Highway functional classification concepts, criteria and procedures. *US Department of Transportation Washington, DC.* 2013.
67. Karim, Muhammad Monjurul. Visual dataset for driving scene classification. 2020.