Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

11-15-2022 3:30 PM

Driver Behavior Analysis Based on Real On-Road Driving Data in the Design of Advanced Driving Assistance Systems

Nima Khairdoost, *The University of Western Ontario*

Supervisor: Steven S. Beauchemin, *The University of Western Ontario*
Co-Supervisor: Michael A. Bauer, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Computer Science

© Nima Khairdoost 2022

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Artificial Intelligence and Robotics Commons

Recommended Citation
Khairdoost, Nima, "Driver Behavior Analysis Based on Real On-Road Driving Data in the Design of Advanced Driving Assistance Systems" (2022). *Electronic Thesis and Dissertation Repository*. 9088.
https://ir.lib.uwo.ca/etd/9088

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

The number of vehicles on the roads increases every day. According to the National Highway Traffic Safety Administration (NHTSA), the overwhelming majority of serious crashes (over 94 percent) are caused by human error. The broad aim of this research is to develop a driver behavior model using real on-road data in the design of Advanced Driving Assistance Systems (ADASs). For several decades, these systems have been a focus of many researchers and vehicle manufacturers in order to increase vehicle and road safety and assist drivers in different driving situations. Some studies have concentrated on drivers as the main actor in most driving circumstances. The way a driver monitors the traffic environment partially indicates the level of driver awareness. As an objective, we carry out a quantitative and qualitative analysis of driver behavior to identify the relationship between a driver's intention and his/her actions. The RoadLAB project developed an instrumented vehicle equipped with On-Board Diagnostic systems (OBD-II), a stereo imaging system, and a non-contact eye tracker system to record some synchronized driving data of the driver cephalo-ocular behavior, the vehicle itself, and traffic environment. We analyze several behavioral features of the drivers to realize the potential relevant relationship between driver behavior and the anticipation of the next driver maneuver as well as to reach a better understanding of driver behavior while in the act of driving. Moreover, we detect and classify road lanes in the urban and suburban areas as they provide contextual information. Our experimental results show that our proposed models reached the F1 score of 84% and the accuracy of 94% for driver maneuver prediction and lane type classification respectively.

Summary for Lay Audience

The large number of vehicle collisions leads to both tremendous human and economic costs. Road traffic injury is the leading cause of death among young people and children aged 5-29 years and makes road fatalities the eighth leading cause of death across all age groups. Evidence has shown that a significant number of vehicle accidents are due to driver error. The broad aim of this research is to develop a driver behavior model using real on-road data in the design of Advanced Driving Assistance Systems (ADASs). In many driving situations, drivers may receive an alert from their passengers to avoid an accident with another vehicle or a pedestrian. This role can be played by an intelligent ADAS by warning the driver or even intervening if ADAS finds it necessary. An intelligent ADAS can understand and benefit from valuable information including the state of the driver's behavior, the vehicle, and the environment to analyze driver behavior in different driving situations as well as to predict driver maneuvers. We analyze several behavioral features of the drivers to realize the potential relevant relationship between driver behavior and the anticipation of the next driver maneuver as well as to reach a better understanding of driver behavior while in the act of driving.

List of Acronyms

2D - two-dimensional
3D - three-dimensional
ABS - Anti-lock Braking System
ACC - Adaptive Cruise Control
ACF - Aggregated Channel Features
ADAS - Advanced Driving Assistance System
AIDS - Acquired Immunodeficiency Syndrome
AIO-HMM - Autoregressive Input-Output Hidden Markov Model
ANN - Artificial Neural Network
AV - Autonomous Vehicle
BN - Bayesian Network
BSD - Blind Spot Detection
CANbus - Controller Area Network bus protocol
CNN - Convolutional Neural Network
DBN - Dynamic Bayesian Network
DBRNN - Deep Bidirectional Recurrent Neural Network
DR - Detection Rate
EBA - Emergency Brake Assist
ECU - Electronic Control Unit
EEG - Electroencephalogram
EOG - Electrooculogram
ESC - Electronic Stability Control system
FC - Fully Connected neural network
FCW - Forward Collision Warning
FPPF - False Positives Per Frame
FPR - False Positive Rate
F-RNN-EL - Fusion-Recurrent Neural Network Exponential Loss
F-RNN-UL - Fusion-Recurrent Neural Network Uniform Loss
GA - Genetic Algorithm
GAN - Generative Adversarial Network
GC - Global Context
GPR - Gaussian Process Regression
GPS - the Global Positioning System
GRU - Gated Recurrent Unit
HA - Highway Assist
HEM - Hard Examples Mining
HG - Hypothesis Generation
HIV - Human Immunodeficiency Virus

HOG - Histogram of Oriented Gradients
HV - Hypothesis Verification
IO-HMM - Input Output Hidden Markov Model
IPM - Inverse Perspective Mapping
IR - Infrared Radiation
LBP - Local Binary Patterns
LC - Lane Centering
LDW - Lane Departure Warning
Lidar - Light Detection And Ranging
LoG - Line of Gaze
LSTM - Long-Short Term Memory
MPE - Mean Prediction Error
NDS - Naturalistic Driving Study
NHTSA - National Highway Traffic Safety Administration
NMS - Non Maximum Suppression
OBD-II - On-Board Diagnostic system
PHOG - Pyramid Histogram of Oriented Gradients
PIA - Percentage of Inside Area
PoG - Point of Gaze
RA - Reinforced Attention
Radar - Radio Detection And Ranging
R-CNN - Region-Based Convolutional Neural Network
R-FCN - Region-based Fully Convolutional Network
RNN - Recurrent Neural Network
ROC - Receiver Operating Characteristics curve
ROI - Region Of Interest
RVM - Relevance Vector Machine
SAE - the Society of Automotive Engineers
SHRP 2 - the second Strategic Highway Research Program
SIFT - Scale Invariant Feature Transforms
S-RNN - Simple Recurrent Neural Network
SURF - Speeded Up Robust Features
SVM - Support Vector Machines
TPR - True Positive Rate
UBI - Usage-Based Insurance
WHO - World Health Organization
YOLO - You Only Look Once

Acknowledgements

This thesis represents the persistent pursuit of research over my Ph.D. program, and would not have been possible without the guidance, support, and love of several individuals:

Foremost, I am eternally grateful to my supervisors Dr. Steven S. Beauchemin and Dr. John Barron. Without them, none of this would have been possible. I want to deeply thank them for all the constant support and invaluable mentorship they provided throughout my Ph.D. journey. And to my co-supervisor, Dr. Michael A. Bauer, inimitable, thank you for providing constructive feedback, encouragement along this pathway, and kindly mentoring me to finish my thesis—it was greatly appreciated!

Moreover, I would like to express my deepest appreciation to my thesis committee: Dr. Ziad Kobti, Dr. Jagath Samarabandu, Dr. Anwar Haque, and Dr. Kaizhong Zhang for their brilliant comments, suggestions, and hard questions.

I would like to thank other members of the RoadLAB team for the encouraging discussions and for all your friendship and help throughout this process.

I would also express my deep gratitude to my parents for their unconditional love and sacrifice. They have encouraged and supported me unwaveringly throughout my life. I would not be who I am today without my parents.

Last but not least, and I can write pages here, my special and deepest thanks go to my very kind wife, Hoda, for her continued and unfailing love, encouragement, and patience during my pursuit of the Ph.D. degree. I would not have finished this journey without her unconditional love and support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Avoiding fatalities and serious impacts caused by road accidents is becoming an increasingly important target for governments as well as car manufacturers around the world. According to the global status report on road safety 2018, launched by World Health Organization (WHO), in December 2018 [40], an estimated 1.35 million people die annually in the world as a result of road traffic accidents, and up to 50 million people are injured. Now, road traffic injury is the leading cause of death among young people and children aged 5-29 years and makes road fatalities the eighth leading cause of death across all age groups surpassing HIV/AIDS, diarrhoeal, and tuberculosis diseases. Undoubtedly, driver error is the main cause of road accidents. In order to overcome this, efforts are being made to develop Advanced Driver Assistance Systems (ADASs) in different aspects. The number of road collisions and their serious impacts can be decreased by equipping vehicles with such advanced safety systems to warn the driver in highly dangerous driving situations or even take control of the vehicle by performing automatic actions.

In this research, we aim to analyze and model driver behavior using real driving data for designing ADASs for on-road vehicles. In fact, co-driver ADASs, first should understand and analyze driver behavior during driving to monitor the driver. Also a kind of ADAS system may aim to predict the most probable next maneuver of the driver and assist the driver or intervene if it finds that necessary. In our work, by employing a deep learning model, we predict driver maneuvers using dynamic vehicle and cephalo-ocular behavioral features. Moreover, we identify driver attention based on the attentional visual filed of the driver and four major traffic object types including vehicle, traffic light, traffic sign and pedestrian. For this, we first need a model to detect and recognize the aforementioned traffic objects based on the attentional visual filed and we develop that. Furthermore, we attempt to discover where the driver is gazing at in a course of driving to reach a better understanding of driver gaze behavior. Also, we detect and classify road lanes in the urban and suburban areas which this provides us more contextual information.

The next section presents a literature survey of related research on driver behavior analysis applications, ADAS systems focusing on the relationship between these systems and the driver's role and driver maneuver prediction. After the survey, an over of the research in this thesis is presented, along with several hypotheses motivating the research, and followed by a brief overview of the instrumented vehicle and data collected. The Chapter concludes with a summary of the main contributions and the thesis organization.

1.1 Literature Survey

In order to assist drivers in driving tasks, a variety of ADAS systems have been developed such as Lane Departure Warning (LDW), Forward Collision Warning (FCW), Adaptive Cruise Control (ACC), Highway Assist (HA), Blind

Spot Detection(BSD), and Emergency Brake Assist (EBA). These technologies can assist drivers to experience comfortable driving as well as help to decrease the number of crashes. Some ADAS systems consider the critical role of the driver as the main element in driving events and utilize information related to the driver. These systems analyze driver behavior to predict the driver's intentions in different driving situations [25], [61], [50]. As mentioned, most collisions are due to driver error and driver distraction leading to a notable number of traffic collisions. For example, the results extracted from the second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study (NDS) indicate 60% to 65% of rear-end events occur because of driver distraction. It is obvious, the recent excessive use of in-vehicle devices, such as navigation systems and cell phones, increases driver distraction and consequently, the risk of an accident. Distracted drivers do not attend to the roads effectively, which means they may not be properly aware of the presence of traffic objects and other obstacles. Hence, analyzing and monitoring driver distraction to decrease the hazardous situation is of great importance in the development of a safety monitoring system.

In many driving situations, drivers may receive an alert from their passengers to avoid an accident with another vehicle or a pedestrian. This role can be played by an intelligent ADAS by warning the driver or even intervening if the ADAS finds it necessary to control the vehicle itself. An intelligent ADAS can understand and benefit from valuable information including the state of the driver's behavior, the vehicle, and the environment to perform its augmentation in different driving situations as well as predict driver maneuvers.

In order to make an intelligent ADAS more efficient and practical, one of the most beneficial research areas is the identification of driver behavioral features and objects eliciting visual responses from drivers. The next subsections are devoted to a review of driver behavior analysis applications, advanced driver assistance systems, and driver maneuvers prediction.

1.1.1 Driver Behavior Analysis Applications

Many studies have been conducted on analyzing driver behavior to achieve different goals such as safety driving, traffic management, commercial purposes, and so on. In [65], an overview of different driver behavior analysis methods has been provided. They categorized the driver behavior analysis applications into three classes including vehicle-oriented applications, management-oriented applications, and driver-oriented applications. These categories are described in more detail in the following along with some of their subcategories.

Vehicle-Oriented Applications

These applications focus mainly on the vehicles to improve the driving task and reduce driver workload by creating advanced systems to assist drivers in different driving situations. These systems interact drivers in a real time manner. This category consists of three main subcategories including "Intelligent Vehicles Systems and Autonomous Vehicles", "Driver Assistance" and "Accidents Detection".

The first subcategory is the recent area of exploration which looks to employ new technologies to automate vehicle tasks [12], [10], [9]. In [7], Google developed its first fully autonomous car prototype, followed by car companies of Tesla, Mercedes and Volkswagen. The applications of this subcategory exploit advanced vehicular control and environmental detection technologies [26] using real-time data such as traffic information and nearby vehicles.

The second subcategory includes applications which aim to assist the driver in different driving tasks such as blind spot detection, parking assistance, etc. Nowadays, these systems are employed by car manufacturers to reduce driver error caused by inattention, distraction, such as emergency braking systems [18], [42] and lane keeping assistance systems [6], [56].

The third subcategory includes systems which detect accidents automatically [8], [41], [11]. The role of these systems is to urgently request emergency assistance services for the injured/unconscious driver who may unable to request it by himself. These systems employ some techniques to investigate various vehicle's factors such as speed, brake, acceleration and sudden stop to detect abnormal incidents which can reveal the vehicle has just crashed.

Management-Oriented Applications

The applications that fall into this category aim to optimize the vehicle use, mainly including fleet management and traffic modeling. These applications focus on the management of infrastructure and resources by monitoring the road conditions and the vehicle. These systems identify road conditions based on the driver maneuvers such as acceleration, braking, and the data related to three-axes accelerations [48], [5]. Consequently, these technologies yield effective planning for managing the traffic and also maintaining the roads. Moreover, transport companies can establish effective fleet management and using such applications they can monitor their vehicles in terms of speed, safety inspection as well as fuel consumption. Also, they can reduce the risks for their drivers and vehicles, decrease their costs and improve the performance of their services [23], [31], [1].

Driver-Oriented Applications

Applications in this category consider the driver as the main element. The major application areas that fall into this category are "Driver Attention Evaluation", "Distraction Detection", "Driving Style Assessment" and "Driver Intent Prediction".

Driver attention evaluation is one of the main research areas in the field of driver behavior analysis. These applications analyze the attention of the driver [51], [54], [59], [62] and somnolence of the driver [28], [13] during driving using information such as facial features, gaze activity, heart rate and so on. In distraction detection systems, the degree of driver focus on the road is identified and these applications look to detect driver distraction considering driver reactions [21], [30]. Other applications in this category can be classified into two classes of the driving style assessment and driver intent prediction. The former aims to categorize the driving mode based on a variety of features collected from the vehicle and the driver's actions such as acceleration, steering, speed, braking and GPS [53], [20], [58]. In other words, the data analysis stage in these systems is to find and assess the correlation between driving style and the input data. Aggressive style and risky style are the two common styles in this area of research. The resulting information is of great importance for automobile insurers who calculate Usage-Based Insurance (UBI) [22], [63]. Using these techniques the insurance costs for each driver can be determined based on the driving score. This approach can increase the affordability of insurance for lower-risk drivers, many of whom are also lower-income people [22]. As for driver intent prediction, these applications aim to anticipate the most probable next maneuver (overtaking, lane change, emergency braking, etc.) of the driver using the methods of automatic prediction of maneuvers

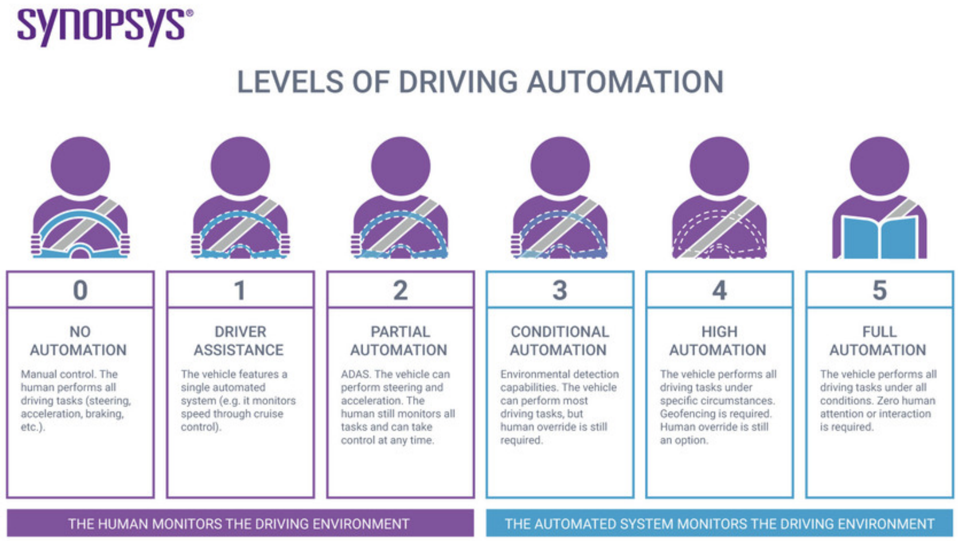


Figure 1.1: *The SAE levels of automation [37]*

[29], [57], [33].

1.1.2 Advanced Driver Assistance Systems (ADASs)

ADASs are designed to increase car and road safety by assisting drivers in dangerous driving situations. ADASs play a critical role to prevent fatalities and injuries by reducing the number of collisions and the serious impacts of those accidents that cannot be avoided. These systems may benefit from various sources of information including the Controller Area Network bus protocol (CANbus) vehicular data, a GPS system, Lidar, Radar, and cameras to perform their tasks. The Society of Automotive Engineers (SAE) has categorized driving automation into five levels [16]. Fig.1.1 illustrates these levels. The following provides an overview of ADASs with consideration of the relationship between these systems and the role of the driver according to the level of automation [16], [37].

Level 0 (No Driving Automation)

The majority of vehicles on the road are manually controlled which means they are in Level 0. These systems monitor the driving environment and provide information to the driver but do not control the vehicle. Several examples of such systems are: *Parking Sensors:* provide an acoustic warning about surrounding obstacles depending on their distances while parking a car. *Lane Departure Warning (LDW):* alarms the driver if the driver accidentally leaves the current lane. *Blind Spot Detection(BSD):* informs the driver if an obstacle exists in the blind spot of the rear-facing mirrors. *Forward Collision Warning (FCW):* provides the driver a warning about an imminent accident with an obstacle ahead. *Night Vision:* by means of IR illuminator and camera, improves driver's perception of the road ahead in the darkness.

Level 1 (Driver Assistance)

Level 1 is the lowest level of automation. These systems perform single functionalities in specific driving situations and also control the vehicle with proper actuators. However, Level 1 and 2 still assign authority to the driver. Examples of Level 1 systems include: *Anti-lock Braking Systems (ABS)* which while braking avoids wheel lock and tire saturation and so provides a reduction in braking distance and better vehicle stability. *Electronic Stability Control systems (ESC)* which can automatically brake a single wheel to better keep the vehicle stable when the system recognizes that it needs to control the steer. *Adaptive Cruise Control (ACC)* which, in addition to keeping the vehicle at the desired speed, can maintain a safe distance from traffic ahead by employing both cutting engine power and actuating the brakes. *Emergency Brake Assist (EBA)* which can automatically apply the brakes if it detects an impending collision. In an urgent situation if the driver is not braking adequately, the system can provide additional braking power to avoid a collision. *Lane Centering (LC)* which, unlike lane departure systems that gives a warning to the driver, maintains the vehicle in the center of the lane by continuously controlling the steer of the vehicle.

Level 2 (Partial Driving Automation)

As mentioned, Level 2 and Level 1 systems leave the authority to the driver but Level 2 systems can perform more complex maneuvers, control both steering and accelerating/decelerating. Tesla Autopilot and Cadillac (General Motors) Super Cruise systems both qualify as Level 2. *Highway Assist (HA)* systems combine ACC, LC, and BSD, for continuously controlling longitudinally and laterally the vehicle. These systems can help reduce driver stress and fatigue and allow drivers to feel safer on highways while driving. *Autonomous Obstacle Avoidance* systems, similar to HA, control the vehicle longitudinally and laterally to avoid an accident with an obstacle. *Autonomous Parking* systems help the driver to find a suitable parking and then assist in parking the car by controlling the steer, speed and avoid collision. These systems still leave the overall authority to the driver.

Level 3 (Conditional Driving Automation)

The leap from Level 2 to Level 3 is substantial from a technological perspective, even if from a human perspective, their functionalities seem quite similar. Level 3 systems perform the maneuvers in the determined scenario, but if the system is unable to execute the task or they detect a self-fault, they require the driver to override. In other words, the driver must be ready to take control of the vehicle although he/she is not required to continuously monitor the driving environment. According to the SAE standard, these systems need redundancies in sensors and decision Electronic Control Units (ECU) to perform their roles. Highway Chauffeur [38] is an example of a Level 3 system. This system is an evolution of HA that autonomously plans when to overtake and accepts full responsibility for the maneuver.

Level 4 (High Driving Automation)

In Level 4 systems, taking control of the vehicle by the driver is not required most of the time. These systems extend the scenarios where they can make decisions, manage situations, and perform all the necessary driving tasks in those situations. For these systems, an integrated intelligence with all-around sources of sensing is required. Automatic Valet Parking [39] is an example of a Level 4 system. In this system, the vehicle takes the responsibility to find a parking spot and to park the car after the driver has left the vehicle. In level 4 systems, communication between the vehicle and the infrastructure is usually needed to improve performance.

Level 5 (Full Driving Automation)

Level 5 is the final automation level so that the vehicles do not require human attention. Level 5 vehicles can even lack interfaces such as steering wheels or acceleration/braking pedals. In fact, the driver is treated as a typical passenger, who just sets a destination and can even sleep while the vehicle is performing all transportation tasks to arrive at the predetermined destination.

1.1.3 Driver Maneuver Prediction

In the ADAS context, the prediction of driver maneuver is one of the principal targets of driver behavior modeling. Driver maneuvers can be considered according to traffic and road infrastructure [2]. Reichart [45] and Tolle [55] categorized driver maneuvers which are mentioned in Table 1.1. These two categories present driving maneuvers on the same level of granularity and only differ to a minor degree. For example, the list of maneuvers provided by Tolle [55], is sufficient to fully cover any trip in city and rural areas as well as on highways. This list does not include unexpected changes in traffic conditions such as the sudden appearance of an obstacle. The other maneuver lists that have been suggested in the literature are similar to the items mentioned above, the differences in the list of maneuvers relating mostly to the aim of the intended application. For instance, the work developed in [35] focuses on maneuvers that occur on highways.

Table 1.1: List of driver maneuvers provided by [45] and [35]

Reichart	**Tölle**
Follow lane	Start
React to obstacle	Follow
Turn at intersection	Approach vehicle
Cross intersection	Overtake vehicle
Turn into street	Cross intersection
Change lane	Change lane
Turn around	Turn at intersection
Drive backwards	Drive backwards
Choose velocity	Park
Follow vehicle	

In order to anticipate driver maneuvers, the temporal aspects of the driving context using multiple sensors are modeled and then the intention of the driver can be inferred. Driver maneuver prediction is still quite a challenging task because the interactions between the sensors are complex and a driver's intentions can not be directly identified. Many internal and environmental factors can influence driver behavior, which ideally should be considered to provide a faithful model [2]. These factors include, but are not limited to:

- emotional features such as stress or anger,
- physical abilities e.g. reaction times,
- environmental conditions, such as lighting, weather,
- cognitive capabilities such as distraction, fatigue, mental load,
- driving skills and driver learning capabilities,
- motivations and goals.

A model that includes all of the aforementioned aspects is highly complicated and not yet feasible in practice. Consequently, driver prediction models that have been presented in the literature deal with subsets of these aspects.

Models for Driver Maneuver Prediction

Driver behavior models can be divided into two classes: cognitive driver models and behaviorist driver models [2].

Cognitive Driver Modeling

Cognitive driver models try to model human behavior based on human information processing. Human aspects such as memory, learning or visual understanding play a critical role in the modeling. Some psychological aspects can be involved in cognitive driver behavior modeling such as reaction time, body strength, distraction, stress, fatigue, etc. [19]. Understanding driver behavior in a cognitive structure is of great importance to find out the driver's motivation for performing an appropriate maneuver. For example, the work presented in [3] utilized cognitive structures to model the driver's situation awareness. In [32], cephalo-ocular behavior of drivers was analyzed in different car/road events including overtaking and crossing an intersection. The authors were able to identify the driver's visual search actions using computer vision, and finally, mapped these events with the driver's behavior. Moreover, a cognitive model was developed in [47] to predict the impact of cellular-phone dialing on driver performance.

Behaviorist Driver Modeling

On the other hand, behaviorist driver models attempt to determine how the driver interacts with his surrounding environment including vehicles, pedestrians, and other traffic objects and also the control elements in the vehicle, such as the steering wheel, the accelerator and brake pedals, and the turn signals. Some examples of maneuvers that have been studied include emergency braking [49], car-following [24], and lane change [14]. In [49], a prediction system was proposed to distinguish merely strong braking behavior from emergency braking. Khodayari *et al.* [24] proposed a car-following model using fuzzy logic technique to predict the driver's car-following behavior. In [14], a method was proposed to model lane changes on curved roads and compare lane changing with lane-keeping scenarios.

In the following, we briefly review some recent methods in the field of driver maneuver prediction which have been developed based on deep learning techniques in recent years.

Some Recent Driver Maneuver Prediction Methods Based on Deep Learning Techniques

Olabiyi *et al.* [36] proposed a method for anticipating driver actions including braking, lane changes and turn anomaly actions. Their prediction system employed Deep Bidirectional Recurrent Neural Network (DBRNN) including multiple Long-Short Term Memory (LSTM) units and/or Gated Recurrent Units (GRU) cells that discovers the spatial-temporal dependencies in temporal data. In [46], the authors presented a new sensory-fusion framework based on deep learning to predict driver maneuvers which utilized a variety of sensory data such as inside and outside camera videos, vehicle speed, GPS and other related information. In order to learn spatial relationships and capture long temporal dependencies, their model took advantage of a combination of dilated CNN and convolutional neural network maxpooling (CNN maxpooling) pairs. In [64], a novel model called Cognitive Fusion-RNN (CFRNN) was proposed to predict driving maneuvers which combined both a cognition-driven model and data-driven model. The CFRNN model included two LSTM units to fuse the data from both inside and outside of the vehicle in a cognitive way and the two LSTM units were regulated by the driver cognition time process. The authors in [34] proposed a method including two parts of processing to anticipate driver maneuvers. In the first part, in addition to the outside features they extracted features using CNN DenseNet121 [44] architecture from the inside frames. The second part mainly included the construction of CNN-LSTM model that is a combination of two standard models of CNN and LSTM. In [33], Mora *et al.* proposed a simplified model to predict the emergency braking intention using a deep learning method and electroencephalogram (EEG) data without transforming the EEG data into gray-scale images. Their method was able to discriminate the events of normal driving and emergency braking using only four electrodes. In [17], a model named Attention-based Global Context Network (AGCNet) was proposed to predict driver maneuvers. This model utilizes multi-modal data, including front view frame data and driver physiological data to perform its task. By proposing the Global Context (GC) block and Channel-wise Attention (CA), AGCNet is capable of generating global context features and choosing valuable ones in a effective way. The AGCNet model coupled with a new Dual attention-based LSTM (DaLSTM) network learns co-occurrence features and predicts driver maneuvers. In [57], a hybrid deep learning based model was proposed to predict lane-changing behavior of the driver. The first level of the hybrid model includes Seq2Seq, a variant of RNN [43], which is mainly employed for temporal data processing to decrease invisible data loss. The second level includes a fully connected neural network (FC) to fuse data and classify lane-changing. The two-level training model enables the Seq2Seq-FC network to deepen the number of network layers while it can avoid gradient dispersion problem.

1.2 Research Overview

The main objective of this research is to analyze and model driver behavior using real driving data in the design of advanced driver assistance systems for on-road vehicles. An intelligent ADAS, as a co-driver, should be able to understand driver behavior in order to be able to identify the most probable next maneuver and assist the driver in different driving situations or intervene if the ADAS finds it necessary. For this, some valuable information from the vehicle, driver, and environment needs to be provided to the ADAS. Therefore, the system would be able to analyze and understand driver behavior in a driving context and monitor it. The system should be able to warn the driver about an unseen obstacle or a traffic object such as a pedestrian, vehicle, or sign or even take control of the vehicle in critical situations. Developing models of understanding and prediction of driver behavior using such data can enable advancement in technologies relating to the vehicle and its passenger's safety and at a higher level, road safety.

1.2.1 Primary Conjecture

ADASs in several different aspects such as detection of drowsiness, distraction, etc. have been studied to help drivers to increase driver safety. It seems the most appropriate method may be an approach that evaluates and monitors driver behavior in order to avoid future hazardous maneuvers [15]. Driver cephalo-ocular behavior and visual attention have been shown to be beneficial in understanding driver behavior and predicting driver maneuvers [60], [61]. Based on observations, as the main conjecture, connecting driver visual behavior and the driver environment (vehicle, pedestrian, etc.) can lead to a better understanding and predictive model of driver behavior.

1.2.2 Hypotheses

In this section, we break down the main conjecture into several hypotheses which can be empirically investigated in the following. We address these hypotheses in Chapters 2, 3, 4, 5 and 6 accordingly.

1. *Driver maneuvers can be partly anticipated using dynamic vehicle and cephalo-ocular behavioral features:* The authors in [61] employed driver behavioral features and vehicle dynamics features to anticipate driver

maneuvers using a traditional Input Output Hidden Markov Model (IO-HMM). They have shown that both extracted features from the cephalo-ocular behavior of drivers and vehicular dynamics are necessary to predict the next driving action with proper accuracy. By employing a deep learning LSTM-based model, we explore this hypothesis with the aim of improving prediction accuracy as well as expanding the types of predicted driver maneuvers in comparison to the previous work [61]. We explore this hypothesis to make the model which takes advantage of three merits that make it a competitive and reliable model in comparison to previous works. Since our model employs LSTM which is capable of keeping long-term dependencies in the temporal data, it can predict driver maneuvers with more performance in comparison to the works employing classifiers which are not suitable for time series data. The second aspect is the fact that our model can predict five maneuver types although there are many works in the literature that predict less than five types. As the last merit, our model utilizes gaze information to perform its task but many previous works ignore such this useful information.

2. *It is possible to detect and recognize all traffic objects inside the attentional visual field of the driver:* The attentional visual area of drivers is a central part of safe driving which is computed as a 2D ellipse in the imaging plane of the stereo system. We verify this hypothesis by the fact that we can find objects in the traffic scene. Therefore, those objects would locate inside the attentional field of the driver, which has been previously obtained by Kowsari *et al.* [27]. This enables us to detect and recognize those objects located inside the visual attentional area of the driver. To explore this hypothesis, we focus on the traffic ob-

jects including vehicles, traffic lights, traffic signs, and pedestrians to be detected and recognized. For this, we first need to develop a framework to perform this task. Prior to the time of the implementation of our framework, there had been a very little attention in previous research focusing on simultaneously detecting traffic objects of different major classes. Hence, one aspect which makes our framework different than others is the fact that in addition to detection of more major classes of traffic objects, we also classify them into their own sub-classes.

3. *It is possible to detect and classify lane types including lane boundary in urban and suburban areas:* Lanes provide contextual information which can be helpful in different applications such as lane keeping assistance systems, driver attention evaluation, driver maneuver prediction and so on. To explore this hypothesis, we employ deep learning methods to detect and classify eight types of lane including road boundary as one type of lane when there is no actual lane marking within urban and suburban areas. Our work is different from similar previous work in several aspects. The majority of previous studies apply their model to highways, where lanes are typically well defined and generally ignore different types of lanes. Other works do classify lanes into types, though assume fewer lane types and ignore road boundaries (no lane markers) whereas our work identifies eight types of lanes and considers the road boundaries.

4. *Driver attention can be estimated based on the driver's visual attentional field and major classes of traffic objects:* It is generally accepted that the driver gaze area is considered the range to extend ± 6.5 degrees [52]. Consequently, a driver cannot attend to the whole driving environment.

In addition, a driver may miss some information because of inappropriate driving habits, driving skills, or even distractions that affect the choice of proper maneuver of the driver. There has been little previous work done to estimate driver attention where multiple classes of traffic objects have been considered. To investigate this hypothesis, we focus on a driver's attention to four kinds of traffic objects (traffic lights, signs, vehicles, pedestrians). We develop an analytical model to estimate a driver's average traffic scene attention based on the attentional area of the driver. Our model is the first model of its kind that along with detection of the four aforementioned traffic object types, takes advantage of the attentional visual field of the driver to perform its task.

5. *It is possible to identify what traffic object the driver is gazing at using the Point of Gaze (PoG) of the driver during driving:* The authors in [27] devised a technique in our laboratory to cross-calibrate the eye-tracker and stereo systems and project the PoGs onto the stereo system imaging plane. In the literature there has been little work to investigate driver's PoG while driving considering multiple object classes, including vehicle, traffic light, traffic sign, and pedestrian simultaneously. We investigate this hypothesis by means of detecting the four aforementioned traffic object types, we aim to discover what traffic object (or elsewhere) the driver is gazing at while in the act of driving using PoG of the driver. As a result, we can estimate a driver's average percentage of the driving time in which the driver has gazed at each aforementioned type of traffic objects in the path of driving.

Investigation of these hypotheses enables us to identify in what type of lanes the driver is driving and also it will increase our knowledge about improved understanding of driver visual behavior by means of estimating average driver attention with respect to traffic objects of the four types as well as estimating average percentage of the driving time in which a driver has gazed at different objects. Moreover, it leads to improved development of predictive modeling of driver behavior in terms of reliability and expanded types of maneuvers in comparison to previous work [61] on real on-road RoadLAB data with the aim of assisting/warning the driver appropriately.

1.2.3 RoadLAB Vehicular Configuration

Our research is based on data gathered in the RoadLAB project. Data was gathered using an experimental vehicle that was equipped with a forward stereoscopic system, OBD-II CANbus, and an eye tracker [4]. (see Fig.1.2.) This configured vehicle was able to record data as follows:

1. The On-Board Diagnostic system (OBD-II) obtained vehicular dynamics data in real-time. These data included steering wheel angle, odometry, accelerator/brake pedal position, and turn indicators.

2. The stereoscopic system mounted on the vehicle's roof recorded the front view of the vehicular driving environment at 30Hz.

3. A non-contact 3D gaze tracker mounted on the dashboard captured several driver cephalo-ocular features, including head/eye motion and gaze information.

This information was collected in real-time as sixteen driving sequences for sixteen drivers, including seven males and nine females. CANbus data was collected via an interface between the on-board computer and the CANbus

Figure 1.2: *RoadLAB vehicular instrumentation configuration.* **a) (left):** *3D infrared gaze tracker;* **b) (right):** *Forward stereoscopic vision system on rooftop*

system of the vehicle to record vehicle odometry information and the driver-related elements such as steering wheel, accelerator/brake pedals, and turn signals. Stereo cameras were employed to collect data on the environment including road markers and traffic signs. FaceLAB, a commercial gaze and head tracking system, was employed to gather eye and head positions. In order to cross-calibrate the stereo system and FaceLAB, a new algorithm was devised in the research RoadLAB group [27].

Each participant drove the instrumented vehicle on a predetermined 28.5km course within the city of London, ON, Canada. (see Fig.1.3.) The course includes downtown, urban and suburban areas of the city. The driver sequences were captured in different weather conditions including sunny (9 driver sequences), partially sunny (4 driver sequences), and partially cloudy (3 driver sequences). Moreover, regarding RoadLAB data, there was ethics approval for the driving experiments and the use of the resulting data for analysis; the data was anonymized.

Figure 1.3: *Map of the predetermined course for drivers, located in London, Ontario, Canada. The path includes urban and suburban driving areas and is approximately 28.5 kilometers long.*

1.3 Contributions

This thesis is an inherent part of the RoadLAB research program, instigated by Professor Steven Beauchemin, and is entirely concerned with vehicular instrumentation for the purpose of the study of driver behavior/intent. Chapters 2, 3, 4, and 5 have been published in recognized peer-reviewed venues. In what follows I describe my contributions with regard to each publication within the thesis:

1. Chapter 2: N. Khairdoost, M. Shirpour, M.A. Bauer, S.S. Beauchemin, *Real-Time Driver Maneuver Prediction Using LSTM.* IEEE Transactions on Intelligent Vehicles, vol. 5, no. 4, pp. 714-724, Dec. 2020.

    - M. Shirpour and I contributed equally in finding appropriate ideas to solve the problem, implementing the algorithms as well as writing the paper. We presented a driver behavior model to predict driver maneuvers using LSTM. For this, we benefited from cephalo-ocular behavior features and dynamic vehicle features to create our LSTM-based model. According to our experimental results, our model outperformed the previous IO-HMM model [45]. It improved the precision from 79.5% to 85.6% and recall from 83.3% to 84.1%. Moreover, we expanded the prediction model to anticipate two more maneuvers (left/right lane changes).

2. Chapter 3: M. Shirpour, N. Khairdoost, M.A. Bauer, S.S. Beauchemin, *Traffic Object Detection and Recognition Based on the Attentional Visual Field of Drivers.* IEEE Transactions on Intelligent Vehicles, 2021.

    - M. Shirpour and I contributed equally in finding appropriate ideas to solve the problem, implementing the algorithms as well as writing the paper. We developed a vision-based model that detects and recognizes simultaneously traffic objects of four major classes including vehicle, traffic light, traffic sign and pedestrian based on the attentional visual field of the drivers. Our framework achieved 91% of detection rate and provided promising results in the object recognition stage.

3. Chapter 4: N. Khairdoost, S.S. Beauchemin, M.A. Bauer, *Road Lane Detection and Classification in Urban and Suburban Areas based on CNNs.* in 16th International Conference on Computer Vision Theory and Applications (VISAPP), Vienna, Austria, 2021.

    - The detection and classification of lanes in urban areas is an important problem. I presented a CNN-based framework to detect and classify lane types in urban and suburban environments. To detect lanes, we used a network that generates lane information in an end-to-end way. In the lane type classification stage, our model categorized the detected lane boundaries into eight classes including road boundary (when there is no actual lane marking) and reached the accuracy of 94% for this stage.

4. Chapter 5: N. Khairdoost, S.S. Beauchemin, M.A. Bauer, *An Analytical Model for Estimating Average Driver Attention Based on the Visual Field.* in 7th International Conference on Signal and Image Processing (ICSIP), Suzhou, China, 2022.

    - For predicting what drivers are paying attention to, it is important to detect relevant objects located inside and outside the attentional visual area of drivers. I provided a new analytical vision-based model including three proposed metrics to estimate average driver attention with respect to several classes of important traffic objects including vehicles, traffic lights, traffic signs, and pedestrians. Our presented model is the first model of its kind that takes advantage of the attentional visual field of the driver to perform its task at any moment while in the act of driving.

1.4 Thesis Organization

The thesis is organized as follows: in Chapter 2, we present a model using LSTM to predict a driver maneuver a few seconds before it occurs. In Chapter 3, we explain our method to detect and recognize traffic objects inside and outside the attentional visual field of the driver. In Chapter 4, we present our

CNN-based method to detect and classify road lanes in urban and suburban areas. In Chapter 5, contributions related to average driver attention estimated based on the attentional visual field of the driver with respect to traffic objects are presented. In Chapter 6, we describe our method to measure the average percentage of the driving time in which a driver has gazed at traffic objects. Finally, Chapter 7 provides conclusions and outlines paths for future research.

Bibliography

[1] Ahmad Aljaafreh, Nabeel Alshabatat, and Munaf S Najim Al-Din. "Driving style recognition using fuzzy logic". In: *2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012)*. IEEE. 2012, pp. 460–463.

[2] C. Bauer. "A driver specific maneuver prediction model based on fuzzy logic". PhD thesis. Freie Universität Berlin, 2012.

[3] M.R.K. Baumann and J.F. Krems. "A comprehension based cognitive model of situation awareness". In: *Digital Human Modeling*. Springer, 2009, pp. 192–201.

[4] S.S. Beauchemin, M. A. Bauer, T. Kowsari, and J. Cho. "Portable and Scalable Vision-Based Vehicular Instrumentation for the Analysis of Driver Intentionality". In: *Instrumentation and Measurement, IEEE Transactions on* 61.2 (2012), pp. 391–401.

[5] Ravi Bhoraskar, Nagamanoj Vankadhara, Bhaskaran Raman, and Purushottam Kulkarni. "Wolverine: Traffic and road condition estimation using smartphone sensors". In: *2012 fourth international conference on communication systems and networks (COMSNETS 2012)*. IEEE. 2012, pp. 1–6.

[6] Yougang Bian, Jieyun Ding, Manjiang Hu, Qing Xu, Jianqiang Wang, and Keqiang Li. "An advanced lane-keeping assistance system with switchable assistance modes". In: *IEEE Transactions on Intelligent Transportation Systems* 21.1 (2019), pp. 385–396.

[7] B. Bilger. *Has the self-driving car at last arrived? The New Yorker (2013).* `http://www.newyorker.com/reporting/2013/11/25/131125fa_fact_bilger?currentPage=all`.

[8] Nimisha Chaturvedi and Pallika Srivastava. "Automatic vehicle accident detection and messaging system using GSM and GPS modem". In: *Int. Res. J. Eng. Technol.(IRJET)* 5.3 (2018), pp. 252–254.

[9] Xiaozhi Chen, Kaustav Kundu, Ziyu Zhang, Huimin Ma, Sanja Fidler, and Raquel Urtasun. "Monocular 3d object detection for autonomous driving". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2147–2156.

[10] Haruna Chiroma, Shafi'i M Abdulhamid, Ibrahim AT Hashem, Kayode S Adewole, Absalom E Ezugwu, Saidu Abubakar, and Liyana Shuib. "Deep learning-based big data analytics for internet of vehicles: taxonomy, challenges, and research directions". In: *Mathematical Problems in Engineering* 2021 (2021).

[11] Jae Gyeong Choi, Chan Woo Kong, Gyeongho Kim, and Sunghoon Lim. "Car crash detection using ensemble deep learning and multimodal data from dashboard cameras". In: *Expert Systems with Applications* 183 (2021), p. 115400.

[12] Michał Czubenko, Zdzisław Kowalczuk, and Andrew Ordys. "Autonomous driver based on an intelligent system of decision-making". In: *Cognitive computation* 7.5 (2015), pp. 569–581.

[13] Md Tanvir Ahammed Dipu, Syeda Sumbul Hossain, Yeasir Arafat, and Fatama Binta Rafiq. "Real-time Driver Drowsiness Detection using Deep Learning". In: *International Journal of Advanced Computer Science and Applications* 12.7 (2021).

[14] Ueruen Dogan, Hannes Edelbrunner, and Ioannis Iossifidis. "Towards a driver model: Preliminary study of lane change behavior". In: *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*. IEEE. 2008, pp. 931–937.

[15] B. Donmez, L.N.g. Boyle, and J.D. Lee. "Safety implications of providing real-time feedback to distracted drivers". In: *Accident Analysis & Prevention* 39.3 (2007), pp. 581–590.

[16] M. Galvani. "History and future of driver assistance". In: *IEEE Instrumentation & Measurement Magazine* 22.1 (2019), pp. 11–16.

[17] Jun Gao, Jiangang Yi, and Yi Lu Murphey. "Attention-based global context network for driving maneuvers prediction". In: *Machine Vision and Applications* 33.4 (2022), pp. 1–11.

[18] G. Griffin, D. Kwiatkowski, and J. Miller. U.S. pat. No. 9248815. Washington, DC: U.S. Patent and Trademark Office. 2016.

[19] S. Hamdar. "Driver Behavior Modeling". In: *Handbook of Intelligent Vehicles*. Springer, 2012, pp. 537–558.

[20] Jin-Hyuk Hong, Ben Margines, and Anind K Dey. "A smartphone-based sensing platform to model aggressive driving behaviors". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 4047–4056.

[21] Md Uzzol Hossain, Md Ataur Rahman, Md Manowarul Islam, Arnisha Akhter, Md Ashraf Uddin, and Bikash Kumar Paul. "Automatic driver distraction detection using deep convolutional neural networks". In: *Intelligent Systems with Applications* 14 (2022), p. 200075.

[22] Siniša Husnjak, Dragan Peraković, Ivan Forenbacher, and Marijan Mumdziev. "Telematics system in usage based motor insurance". In: *Procedia Engineering* 100 (2015), pp. 816–825.

[23] Nidhi Kalra, Raman Kumar Goyal, Anshu Parashar, Jaskirat Singh, and Gagan Singla. "Driving Style Recognition System Using Smartphone Sensors Based on Fuzzy Logic". In: *CMC-COMPUTERS MATERIALS & CONTINUA* 69.2 (2021), pp. 1967–1978.

[24] Alireza Khodayari, Reza Kazemi, Ali Ghaffari, and Reinhard Braunstingl. "Design of an improved fuzzy logic based model for prediction of car following behavior". In: *2011 IEEE International Conference on Mechatronics*. IEEE. 2011, pp. 200–205.

[25] I.H. Kim, J.H. Bong, J. Park, and S. Park. "Prediction of driver's intention of lane change by augmenting sensor information using machine learning techniques". In: *Sensors* 17.6 (2017), p. 1350.

[26] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. "Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust,

and performance". In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9.4 (2015), pp. 269–275.

[27] T. Kowsari, S.S. Beauchemin, M.A. Bauer, D. Laurendeau, and N. Teasdale. "Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems". In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 1245–1250.

[28] Vijay Kumar, Shivam Sharma, et al. "Driver drowsiness detection using modified deep learning architecture". In: *Evolutionary Intelligence* (2022), pp. 1–10.

[29] Stéphanie Lefèvre, Ashwin Carvalho, Yiqi Gao, H Eric Tseng, and Francesco Borrelli. "Driver models for personalised driving assistance". In: *Vehicle System Dynamics* 53.12 (2015), pp. 1705–1720.

[30] Tianchi Liu, Yan Yang, Guang-Bin Huang, Yong Kiang Yeo, and Zhiping Lin. "Driver distraction detection using semi-supervised machine learning". In: *IEEE transactions on intelligent transportation systems* 17.4 (2015), pp. 1108–1120.

[31] Eilham Hakimie bin Jamal Mohd Lokman, Vik Tor Goh, Timothy Tzen Vun Yap, and Hu Ng. "Driving style recognition using machine learning and smartphones". In: *F1000Research* 11.57 (2022), p. 57.

[32] S. Metari, T. Prel F.and Moszkowicz, D. Laurendeau, N. Teasdale, S. Beauchemin, and M. Simoneau. "A computer vision framework for the analysis and interpretation of the cephalo-ocular behavior of drivers". In: *Machine vision and applications* 24.1 (2013), pp. 159–173.

[33] Hermes J Mora and Esteban J Pino. "Simplified Prediction Method for Detecting the Emergency Braking Intention Using EEG and a CNN

Trained with a 2D Matrices Tensor Arrangement". In: *International Journal of Human–Computer Interaction* (2022), pp. 1–14.

[34] Abdellatif Moussaid, Ismail Berrada, Mohamed El Kamili, and Khalid Fardousse. "Predicting Driver Lane Change Maneuvers Using Driver's Face". In: *2019 International Conference on Wireless Networks and Mobile Communications (WINCOM)*. IEEE. 2019, pp. 1–7.

[35] A. Okuno, K. Fujita, and A. Kutami. "Visual navigation of an autonomous on-road vehicle: autonomous cruising on highways". In: *Vision-based vehicle guidance*. Springer, 1992, pp. 222–237.

[36] Oluwatobi Olabiyi, Eric Martinson, Vijay Chintalapudi, and Rui Guo. "Driver Action Prediction Using Deep (Bidirectional) Recurrent Neural Network". In: *arXiv preprint arXiv:1706.02257* (2017).

[37] Online. *Available*. URL: `https://www.synopsys.com/automotive/autonomous-driving-levels.html`.

[38] Online. *Available*. URL: `https://www.media.stellantis.com/em-en/corporate-communications/press/c-roads-italy-project-digital-roads-supporting-connected-level-3-autonomous-driving`.

[39] Online. *Available*. URL: `https://www.bosch-mobility-solutions.com/en/solutions/parking/automated-valet-parking/`.

[40] World Health Organization et al. *Global status report on road safety 2018: Summary*. Tech. rep. World Health Organization, 2018.

[41] Sourav Kumar Panwar, Vivek Solanki, Sachin Gandhi, Sankalp Gupta, and Hitesh Garg. "Vehicle accident detection using IoT and live track-

ing using geo-coordinates". In: *Journal of Physics: Conference Series*. Vol. 1706. 1. IOP Publishing. 2020, p. 012152.

[42] D. Parker, K. Cockings, and M. Cund. U.S. pat. NO. 9682689. Washington, DC: U.S. Patent and Trademark Office. 2017.

[43] Yuchen Qiao, Kazuma Hashimoto, Akiko Eriguchi, Haixia Wang, Dongsheng Wang, Yoshimasa Tsuruoka, and Kenjiro Taura. "Parallelizing and optimizing neural Encoder–Decoder models without padding on multi-core architecture". In: *Future Generation Computer Systems* 108 (2020), pp. 1206–1213.

[44] Sivaramakrishnan Rajaraman, Sameer K Antani, Mahdieh Poostchi, Kamolrat Silamut, Md A Hossain, Richard J Maude, Stefan Jaeger, and George R Thoma. "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images". In: *PeerJ* 6 (2018), e4568.

[45] G. Reichart. *Menschliche zuverlässigkeit beim führen von kraftfahrzeugen*. VDI-Verlag, 2001.

[46] Banafsheh Rekabdar and Christos Mousas. "Dilated convolutional neural network for predicting driver's activity". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2018, pp. 3245–3250.

[47] D.D. Salvucci and K.L. Macuga. "Predicting the effects of cellular-phone dialing on driver performance". In: *Cognitive Systems Research* 3.1 (2002), pp. 95–102.

[48] Shahram Sattar, Songnian Li, and Michael Chapman. "Road surface monitoring using smartphone sensors: A review". In: *Sensors* 18.11 (2018), p. 3845.

[49] J. Schmitt and B. Färber. "Verbesserung von FAS durch Fahrerabsichtserkennung mit Fuzzy Logic". In: *VDI-Berichte* 2015.1919 (2005).

[50] Mohsen Shirpour. "Predictive Model of Driver's Eye Fixation for Maneuver Prediction in the Design of Advanced Driving Assistance Systems". In: (2021).

[51] Mohsen Shirpour, Steven S Beauchemin, and Michael A Bauer. "What Does Visual Gaze Attend to during Driving?" In: *VEHITS*. 2021, pp. 465–470.

[52] Kenji Takagi, Haruki Kawanaka, Md. Shoaib Bhuiyan, and Koji Oguri. "Estimation of a three-dimensional gaze point and the gaze target from the road images". In: *Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on*. IEEE. 2011, pp. 526–531.

[53] Farid Talebloo, Emad A Mohammed, and Behrouz Far. "Deep Learning Approach for Aggressive Driving Behaviour Detection". In: *arXiv preprint arXiv:2111.04794* (2021).

[54] Ashish Tawari, Sayanan Sivaraman, Mohan Manubhai Trivedi, Trevor Shannon, and Mario Tippelhofer. "Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking". In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 115–120.

[55] W. Tölle. *Ein Fahrmanöverkonzept für einen maschinellen Kopiloten*. PhD thesis, Universität Karlsruhe, 1996.

[56] Qun Wang, Weichao Zhuang, Liangmo Wang, and Fei Ju. *Lane keeping assist for an autonomous vehicle based on deep reinforcement learning.* Tech. rep. SAE Technical Paper, 2020.

[57] Cheng Wei, Fei Hui, and Asad J Khattak. "Driver lane-changing behavior prediction based on deep learning". In: *Journal of advanced transportation* 2021 (2021).

[58] Samuel Würtz and Ulrich Göhner. "Driving Style Analysis Using Recurrent Neural Networks with LSTM Cells". In: *Journal of Advances in Information Technology Vol* 11.1 (2020).

[59] Guoliang Yuan, Yafei Wang, Huizhu Yan, and Xianping Fu. "Self-calibrated driver gaze estimation via gaze pattern learning". In: *Knowledge-Based Systems* 235 (2022), p. 107630.

[60] S.J. Zabihi, S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. "Detection and recognition of traffic signs inside the attentional visual field of drivers". In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2017, pp. 583–588.

[61] S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. "Real-time driving manoeuvre prediction using IO-HMM and driver cephalo-ocular behaviour". In: *Intelligent Vehicles Symposium (IV), 2017 IEEE*. IEEE. 2017, pp. 875–880.

[62] Yingji Zhang, Xiaohui Yang, and Zhe Ma. "Driver's Gaze Zone Estimation Method: A Four-channel Convolutional Neural Network Model". In: *2020 2nd International Conference on Big-data Service and Intelligent Computation*. 2020, pp. 20–24.

[63] Wenyi Zheng, Wei Nai, Fangqi Zhang, Weiyang Qin, and Decun Dong. "A novel set of driving style risk evaluation index system for UBI-based differentiated commercial vehicle insurance in China". In: *CICTP 2015*. 2015, pp. 2510–2524.

[64] Dong Zhou, Huimin Ma, and Yuhan Dong. "Driving maneuvers prediction based on cognition-driven and data-driven method". In: *2018 IEEE Visual Communications and Image Processing (VCIP)*. IEEE. 2018, pp. 1–4.

[65] Kawtar Zinebi, Nissrine Souissi, and Kawtar Tikito. "Driver Behavior Analysis Methods: Applications oriented study". In: *Proceedings of the 3rd International Conference on Big Data, Cloud and Application (BDCA 2018)*. 2018.

Chapter 2

Driver Maneuver Prediction

This Chapter is a reformatted version of the following article:

N. Khairdoost, M. Shirpour, M.A. Bauer, S.S. Beauchemin, *Real-Time Driver Maneuver Prediction Using LSTM. IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 4, pp. 714-724, Dec. 2020.

Driver maneuver prediction is of great importance in designing a modern Advanced Driver Assistance System (ADAS). Such predictions can improve driving safety by alerting the driver to the danger of unsafe or risky traffic situations. In this research, we developed a model to predict driver maneuvers, including left/right lane changes, left/right turns and driving straight forward, 3.6 seconds on average before they occur in real time. For this, we propose a deep learning method based on Long Short-Term Memory (LSTM) which utilizes data on the driver's gaze and head position as well as vehicle dynamics data. We applied our approach on real data collected during drives in an urban environment in an instrumented vehicle. In comparison with previous IO-HMM techniques [55] that predicted three maneuvers including left/right turns and driving straight, our prediction model is able to anticipate two more maneuvers (left/right lane changes). In addition to this, our experimental results show that our model, using the identical dataset, improved the F1 score by 4% to 84%.

2.1 Introduction

The number of vehicles on our streets and highways increases every day. This fact makes the analysis of traffic situations increasingly complicated. For example, in the US alone, at least 33,000 people on average die in road accidents every year, with unsuitable maneuvers being reported as the main cause for most of these accidents [8]. Hence, vehicle manufacturers have been developing advanced driver assistance systems (ADASs) to assist the driver in various driving tasks where ADASs are able to avoid up to 40% of vehicle accidents [11]. Examples of ADASs include adaptive cruise control, collision avoidance systems, traffic warning systems, smartphone connectivity, lane departure warning systems, automatic lane centering, blind spot monitoring, etc. Obviously, improving the reliability and robustness of these systems would have a significant impact on decreasing the number of collisions and accident injuries.

An ADAS consists of advanced sensors and camera systems and is activated when some specific predefined conditions are satisfied. In traditional ADAS, a threshold is considered for the inputs and if these inputs are greater than the threshold, the ADAS is activated [21]. Modeling driving behavior of the driver in different traffic scenes, in addition to understanding surrounding environment, makes an ADAS more useful for assisting the driver in controlling the vehicle and avoiding collisions. The goal of this research is to model a driver's behavior so that the ADAS can predict the next driving maneuver a few second before it occurs.

In order to predict driver maneuvers, we need to model the temporal aspects of the driving context and to infer the driver's intention from them. This task is still quite challenging because a driver's decisions are not directly detectable and the interactions between them are complex. The contextual information is also obtained from multiple sensors.

We developed a model to predict driver maneuvers using a Long Short-Term Memory (LSTM) neural network. LSTM is a special type of Recurrent Neural Network (RNN) that is capable of learning long-term dependencies [32, 13]. LSTM includes a memory cell and processes the information flow using its input, forget, and output gates which enables the LSTM model to ignore the non-essential data and keep in its memory only essential information relating to the target. Moreover, an LSTM can effectively resolve the problem of gradient disappearance found in the original RNN approach [50, 54]. LSTMs are successful in many applications such as speech recognition [39], image captioning [22] as well as language translation [46]. In many applications relating to driver behavior, LSTM outperforms the traditional models and standard RNNs [37, 17, 9, 58, 53]. Also, in order to model driver behavior, several previous works studied the significance and superiority of LSTM [37]. Like other algorithms, there are some drawbacks to using the LSTM model. For example, an LSTM model is prone to overfitting, although the dropout can deal with this in deep learning-based models. As well, although LSTM became popular since it solves the disappearing gradient problem, it is unable to eliminate the problem completely. In addition, the number of memory units in the network does not change dynamically, so the memory of the network is eventually limited [45]. However, given the overall advantages of LSTM it seemed to be a good choice to use for the sequence learning problem and in particular the problem of driver maneuver prediction where the driver analyzes the driving environment information only based on several seconds before the current situation [4].

In order to predict driver maneuvers, our LSTM-based model learns the parameters from real driving sequences, including vehicle dynamics, driver's head movements, as well as gaze data. Then the model infers the potential driving maneuvers (namely, left/right turns, left/right lane changes and driving straight forward) by means of generating a probability for each maneuver. In other words, the maneuver with the highest probability is considered as the predicted maneuver.

The rest of this paper is structured as follows. In Section 2.2, we review the literature. In Section 2.3, we explain our vehicle instrumentation. Section 2.4 contains a description of the proposed method. Section 2.5 presents a summary of the datasets used, learning parameters, and the experimental results obtained along with a critical analysis of those results. We discuss several common reasons resulting in incorrect maneuver prediction in Section 2.6. We give conclusions and future research directions in Section 2.7.

2.2 Literature Survey

In general, to anticipate a driver maneuver, a trained model analyzes contextual driving information. This means each driver maneuver is predicted by analyzing data about things such as head movements, GPS, vehicle dynamics, driver gaze, etc. Much research has been done to predict the action of a driver in advance of the driver performing one or more actions [55, 17, 16, 35, 10, 51].

Artificial Neural Networks (ANNs) have a powerful ability to discover implicitly complicated nonlinear relationships among input variables. Hence, ANNs are suitable techniques for pattern recognition and action prediction applications, provided that enough experimental data is available. For driver maneuver prediction, the inputs can be behavioral features, such as acceleration, signaling and braking, and the ANN outputs the predicted maneuver. For instance, Kim *et al.* [21] applied an ANN to measurements from the on-board sensors, such as the steering wheel angle, the yaw rate and the throttle position, to classify road conditions and to predict the driver's intention for a lane change. Leonhardt and Wanielik [27] employed an ANN for lane change prediction. MacAdam and Johnson [31] represented driver steering behavior in path regulation control tasks using elementary neural networks. Mitrovic [34] used neural networks for short-term prediction of lateral and longitudinal vehicle acceleration.

Although traditional ANNs, such as feed-forward neural nets, are powerful machine learning techniques, ANNs are black box learning techniques. They cannot interpret the relationship among the input and output. Moreover, in the standard probabilistic framework, they cannot work with uncertainties. Another disadvantage is that ANNs consider all input data independent of each other, while in many applications, such as driver maneuver prediction, the input data is a sequence of observations taken sequentially in time and, of course, this temporal information is of great importance.

A Bayesian Network (BN) is an acyclic directed graph that constitutes the conditional dependencies among a set of variables, where the directed edges reflect the qualitative relationships between variables and conditional probability distributions are considered as the quantitative relationships. BNs have been employed for driver maneuver recognition such as overtaking, lane changes or left/right turns [15, 18, 33]. Amata *et al.* [1] presented a prediction model for driver behaviors, such as stopping at intersections based on traffic conditions. Tezuka *et al.* [48] used a BN and steering wheel angle data to develop a model to detect lane keeping, normal lane changes and emergency lane changes. Also, BNs have been utilized for intersection safety systems to recognize turning maneuvers at intersections as well as red light crossing [56]. BNs have been used for identifying emergency braking situations [44]. On the one hand, BNs are suitable for applications, like driver maneuver modeling, where considering uncertainties in modeling is essential. On the other hand, considering temporal data using BNs is difficult. Li *et al.* [28] used a novel Dynamic Bayesian Network (DBN) in highway scenarios to predict driver maneuvers. DBNs can model temporal changes, although they cause increased complexity in building and analyzing the network.

Temporal behavior analysis of vehicles surrounding the ADAS vehicle plays an essential role in the safety of the driver. Hence, other methods have been proposed to predict the intention of surrounding vehicles. For example, Kim *et al.* [20] used an LSTM to propose a trajectory prediction technique for analyzing the temporal behavior of surrounding vehicles and their future positions. Also, Khosroshahi *et al.* [19] proposed a framework to classify maneuvers of observed vehicles at four-way intersections using LSTM and 3D trajectory cues. Using LSTM, a method has been introduced by Patel *et al.* [40] to predict lane changes of surrounding vehicles in highway driving. An RNN-based model was presented to interpret the time series data about an observed vehicles at signal-less intersections in order to classify their intentions [57].

For recognition of a driver's intention, many researchers have utilized Hidden Markov Models (HMMs). Kuge *et al.* [25] developed steering behavior models for normal/emergency lane changes, as well as lane keeping using HMMs. Another approach was proposed by Tran *et al.* [49] to predict driver maneuvers, including stop/non-stop, left/right lane changes and left/right turns in both urban and highway driving environments. They employed different input sets to investigate the model performance. He *et al.* [12] developed a double-layer HMM structure to model driving behavior and driving intention in the lower and upper layers, respectively. Amsalu and Homaifar [2] employed a Genetic Algorithm (GA) for optimization, as well as for predicting a driver's intentions when the vehicle approaches an intersection. Aoude *et al.* [3] developed two SVM- and HMM-based approaches to estimate driver behaviors at road intersections. Their results showed that the SVM-based approach often outperformed the HMM-based model. Jain *et al.* [16] proposed a maneuver prediction model based on an Autoregressive Input-Output Hidden Markov Model (AIO-HMM), which jointly exploits the information inside and outside of the vehicle.

Similarly, Zabihi *et al.* [55] developed a maneuver prediction model using an Input-Output Hidden Markov Model (IO-HMM) that learns relevant parameters from natural driving sequences. They combined vehicle dynamics features and two features of driver's cephalo-ocular behavior, including driver gaze direction and head pose for detecting driver intent. We followed the work of Kowsari *et al.* [24] and Zabihi *et al.* [55] for feature extraction. We refer the reader to these publications for more details.

Researchers also focused on driver maneuver prediction at (urban) intersections. Klingelschmitt *et al.* [23] created two separate Bayesian Network and Logistic Regression-based models for a vehicle's driving situation and its behavior respectively. Then, they combined them in a single Bayesian Network to design a model able to predict driver intent. In [42], an indicator-based approach for driver intent prediction was proposed. They combined context information with vehicle data. The authors in [30] proposed a new approach for intersection maneuver prediction that was based on personalized incremental learning. In other words, they continuously improved the model accuracy by incorporating individual driving history. Liebner *et al.* [29] proposed an approach to predict driver intent including straight intersection crossing and right turn with the presence or absence of a preceding vehicle. Their model was based on an explicit parametric model for the longitudinal velocity of preceding vehicles.

Recurrent Neural Networks (RRNs), Long Short-Term Memories (LSTMs) and Convolutional Neural Networks (CNNs) have been utilized in different applications of ADAS and they have shown promising results, such as for driver activity prediction [17, 38]. Jain *et al.* [17] employed a RNN with LSTM units to keep long dependencies over the time. They applied their proposed model on a real dataset to predict driver maneuvers. Olabiyi *et al.* [38] proposed a method for anticipating driver action using a deep bidirectional RNN by discovering the relationships between sensor information and future driver maneuver. For this, they used a fusion of the past and future context. Moreover, deep learning has been employed for other ADAS applications, which has brought significant improvements, such as classifying a vehicle's situation for lane changes as safe/unsafe [43] and detecting a driver's confusion level [14].

In this study, we aim to apply LSTM as a deep learning-based method to our natural driving sequences to predict driver maneuvers some number of seconds before they occur. As a result, this would allow an ADAS to take some actions if deemed dangerous or at least warn the driver. Previously, in [55], a traditional method based on IO-HMM was proposed to anticipate three maneuvers of left/right turns and driving straight forward using our dataset. In addition to the aforementioned maneuvers, our model predicts the maneuvers of left/right lane changes as well. Our model takes advantage of three different aspects of a driving environment in comparison to many previous proposed maneuver prediction methods in the literature. First, since our model employs an LSTM, which is capable of keeping long-term dependencies in the temporal data, it is able to predict driver maneuvers better than works employing classifiers which are not suitable for time series data, such as [27, 41, 26]. The second aspect is related to the number of maneuvers that a maneuver prediction system is able to predict. As mentioned, our model predicts five maneuver types although there has been previous work that has predicted maneuvers [35, 10, 51] they consider fewer maneuvers. Finally, our model utilizes gaze information to perform its task while many previous works ignore such this useful information, such as [51, 21, 49].

2.3 Vehicular Instrumentation

We instrumented (hardware and software) a research vehicle capable of recording driver-initiated vehicular actuation and relating the 3D driver gaze direction with environmental stereo imagery. The instrumented vehicle was used to collect data sequences with 16 drivers on a pre-determined 28.5km course within the city of London, Ontario, Canada. (See Figures 2.1 and 2.2). 3TB of driving sequences were recorded, containing forward stereo imaging and depth, 3D PoG and head pose, and vehicular dynamics obtained with the OBD-II CANbus interface (See Figure 2.3). Data frames are collected at a rate of 30Hz.

Our research vehicle is instrumented to find whether driver maneuvers could be predicted ahead of time. The vehicle is fitted with a non-contact infra-red 3D gaze and head pose tracker working at 60Hz. Its purpose is to record head movements and gaze direction as they happen while driving. Both head pose and gaze are recorded in the reference frame of the tracker

Figure 2.1: **a) (left):** *3D infrared gaze tracker;* **b) (center):** *Forward stereoscopic vision system on rooftop;* **c) (right):** *Driver PoG and LoG expressed in the reference frame of stereoscopic vision system and corresponding depth map.*

Figure 2.2: *Map of predetermined route for drivers, located in London Ontario, Canada. The path length is approximately 28.5 and includes urban and suburban driving areas.*

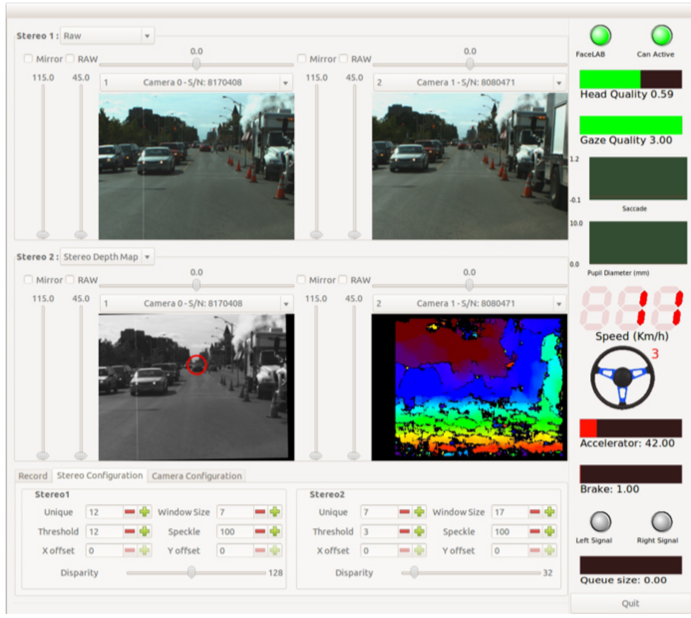


Figure 2.3: *The on-board data recorder interface displaying depth maps, driver PoG, vehicular dynamics, and eye tracker data.*

(See Figure 2.1 a) for a depiction of the tracker). A forward stereoscopic vision system is mounted on the roof of the vehicle to provide dense stereo depth maps at 30 Hz. Depth maps are expressed in the frame of reference of the forward stereo system. Details concerning this instrumentation were described by Beauchemin *et al.* [5].

We devised a cross-calibration technique to transform the 3D driver gaze and head pose, expressed in the tracker coordinates, in the reference frame of the forward stereoscopic vision system. As a result, the 3D Point of Gaze (PoG) and Line of Gaze (LoG) of the driver into the surrounding environment are known in absolute 3D coordinates. The attentional visual area of the average driver is defined as the cone from the eye along the LoG. Here, we briefly describe the procedure we used to determine the attentional visual area, whose contour is defined as an ellipse. We first transform the eye position

$e = (e_x, e_y, e_z)$ and the 3D PoG $g = (g_x, g_y, g_z)$ into the frame of reference of the forward stereo system, and form a cone with apex e that contains the LoG at its center. This cone has an opening of 6.5° with respect to the LoG [47]. Next, we define a plane perpendicular to the LoG that contains the PoG, and compute the intersection this plane makes with the cone, resulting in a 2D circle located in 3D space. The radius of this circle representing the attentional gaze area is obtained as:

$$r = tan(\theta)d(e, g) \tag{2.1}$$

where

$$d(e, g) = \sqrt{((e_x - g_x)^2 + (e_y - g_y)^2 + (e_z - g_z)^2)} \tag{2.2}$$

The circle is reprojected onto the imaging plane of the forward stereo vision system where it becomes a 2D ellipse, as pictured in Figure 2.4. The identification of objects in the scene that elicit an ocular response from the driver can then be identified within this area (Figure 2.5). The cross calibration procedure was devised by Kowsari *et al.* [24]. At the time of its deployment, this was the first publicly known vehicle capable of identifying the 3D PoG of the driver in real-time and in absolute 3D coordinates.

2.4 Proposed Method

In order to anticipate driver maneuvers, we need to jointly model the temporal aspects of the driving context and the driver's intent. For this purpose, we employed LSTM as it has the powerful ability to model time series data with their long-term dependencies.

In general, the aim of driver maneuver prediction is to anticipate the driver's future maneuvers some time before they occur, given information on

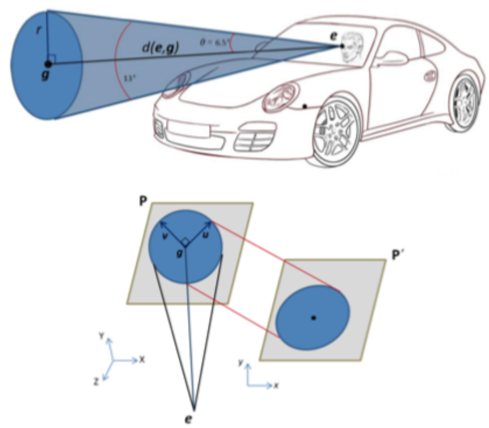


Figure 2.4: *The attentional visual area of driver is defined as the base of the cone located at the depth of sighted features.*

Figure 2.5: *Two projections of the visual attention cone base on the stereo imaging plane.*

driving context. In the model training stage, a set of complete sequences of observations are fed into the model, where at the end of the sequence, an event happens. In our application, the event can be one of five driver maneuvers: a left/right lane change, a left/right turn, or going forward. The model receives an observation at each time slice so as to predict the driver's future maneuver as early as possible. In other words, the model needs to predict the event by only receiving partial observations from a data sequence. To be exact, each time slice consists of the information of a pre-determined number of frames. Hence, by processing the information available up to current time slice, the observation can be represented as a feature vector (described in Section 2.4.2). We discuss our choice for the size of time slices in Section 2.5.2. Finally, for each time slice, the model outputs the SoftMax probability of each maneuver. Then, the maneuver that has the highest probability is proposed as the predicted maneuver, provided that its probability is higher than a preassigned threshold value, otherwise the system makes no prediction. The choice for this threshold value is justified in Section 2.5.3. Algorithm 1 depicts the complete procedure of our prediction model using LSTM. We refer the reader to Zyner *et al.* [57] and Jain *et al.* [17] for more details on this particular technique. Figure 2.6 provides an overview of our proposed method. Below we present an overview of a standard LSTM unit which is illustrated in Figure 2.7.

2.4.1 Long Short-Term Memories (LSTM)

In this work, we focus on driver maneuver prediction using LSTMs [13]. LSTM is a particular form of RNNs which is suitable for time series data. We briefly explain the structure of LSTM. Figure 2.7 shows the internal structure of the LSTM unit. An LSTM is able to keep the information of previous input data

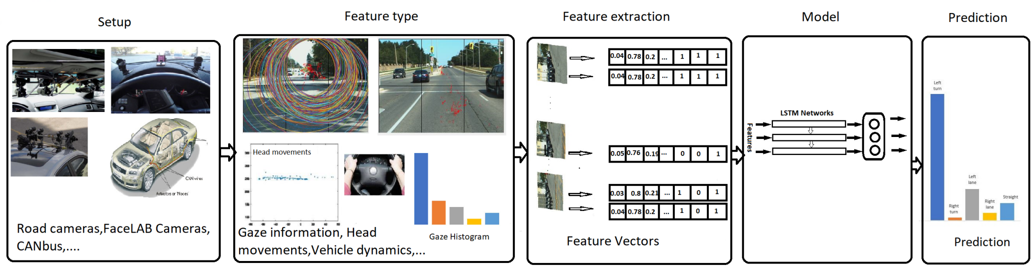


Figure 2.6: *Overview of the proposed approach for predicting driver maneuvers*

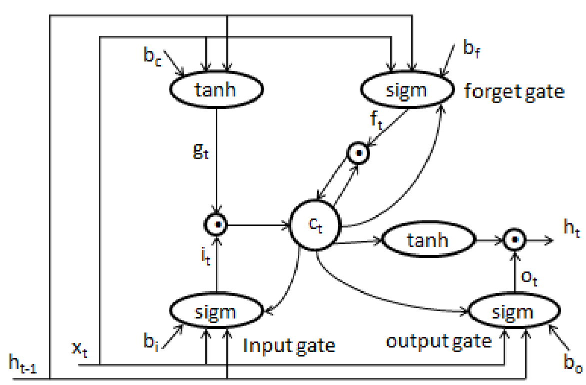


Figure 2.7: *The internal view of an LSTM unit*

in its memory, called a *cell*. Hence, it can overcome the vanishing gradient problem in order to remember long-term dependencies. As mentioned before, LSTMs have been employed in different ADAS applications [20, 19, 17].

We proceed to describe the equations of an LSTM unit [17, 13]. An LSTM unit has a memory cell and three gates, including an input gate i, a forget gate f and an output gate o. At each time step, given the observation x_t, the hidden status from the previous time step h_{t-1}, and the previous cell state c_{t-1}, the unit computes i_t and f_t and then updates c_{t-1} to c_t in order to obtain o_t and h_t. Unlike a RNN, the forget gate in the LSTM unit allows the network to throw away part of memory or learn new information. The following recursive equations encode the mechanism:

$$f_t = sigm(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{2.3}$$

$$i_t = sigm(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{2.4}$$

$$g_t = tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{2.5}$$

$$c_t = c_{t-1} \odot f_t + i_t \odot g_t \tag{2.6}$$

$$o_t = sigm(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{2.7}$$

$$h_t = o_t \odot tanh(c_t), \tag{2.8}$$

where $sigm$, $tanh$ and $\odot$ are the sigmoid function, the hyperbolic tangent function, and the element-wise product, respectively. W and b stand for the weight matrix and bias vector. For multi-class applications, we employ a Soft-Max layer in which the SoftMax function is applied on a linear transformation of h_t. The following notation describes the internal working of a recurrent LSTM unit concisely. In Section 2.4.2, we describe how we reach an observation x (our features).

$$(c_t, h_t) = LSTM(x_t, c_{t-1}, h_{t-1}). \tag{2.9}$$

2.4.2 Features for Driver Maneuver Prediction

We proceed with describing the features that are extracted for maneuver prediction. These features are divided into two major categories called driver cephalo-ocular behavioral features and vehicle dynamics features. These features are aggregated and normalized for each time slice (i.e. after receiving 20 consecutive frames in every 0.67 seconds of driving) and their combination constitutes the feature vector, to be fed into the LSTM model. In what follows, we discuss the extracted features for both categories.

Cephalo-Ocular Behavioral Features

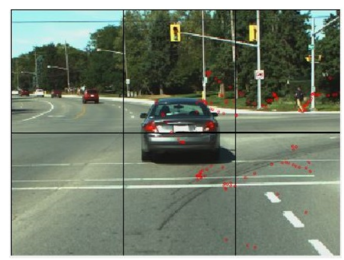
It is generally believed that 3D gaze direction plays a significant role in predicting maneuvers since the driver is observing and focusing on the environment moments before performing a maneuver [36],[55]. Hence, two features of the cephalo-ocular behavior of the driver including 3D Point of Gaze (PoG) in absolute coordinates and also the horizontal head motion have been utilized to predict driver maneuvers. In order to find the 3D PoG of the driver corresponding to its 3D LoG, we used a cross-calibration method proposed by Kowsari *et al.* [24]. This method combines a binocular eye gaze tracker with a binocular scene stereo system and still remains precise for large distances. Once the cross-calibration step is done, the Line of Gaze (LoG) expressed in the coordinates of the eye-tracker is projected onto the imaging plane of the forward stereo system of the instrumented vehicle. Finally, the 3D PoG is identified as the region obtained by intersecting this projected 3D LoG onto the imaging plane of the stereo system with a valid depth estimate.

To extract 3D PoG features, the frame is separated into six non-overlapping equal parts (as shown in Figure 2.8). We create a histogram of 3D PoGs falling into these parts. Figure 2.8 illustrates the PoGs over the last 5 seconds before a maneuver occurs. As can be seen, when drivers are deciding to perform one of the five manuvers they look at different parts of the frame. For more clarification, we discussed the positions of PoGs during a sequence of time slices for a sample of right lane change maneuver. (See Figure 2.9). As shown in Figure 2.9, the driver at first is looking forward, then he decides to check potential obstacles in the right lane before performing the maneuver and then he again looks forward. Finally, he performs the maneuver while is paying attention toward the right lane. We also monitor horizontal driver's head motions and construct a histogram to track that prior to a maneuver.

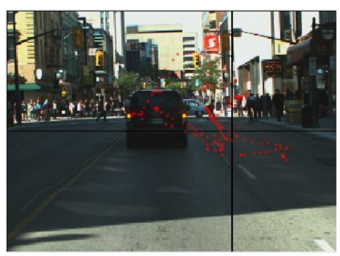
Vehicle Dynamics Features

In 2011, Beauchemin *et al.* [7] instrumented a vehicle with OBD-II CAN-bus. As a matter of fact, all vehicles manufactured after 1996 equipped with on-board diagnostic (OBD-II) systems, which allow physical scan devices by means of vehicle sensors to gather and monitor certain vehicle data on the current status via the OBD-II port. Moreover, since 2008, CANbus protocol (ISO 15765) has been mandatory for OBD-II in all cars sold in the US. As a result, this standardization simplifies examination of the real-time vehicle data (which are generally captured with frequencies between 20 and 200 Hz) for researchers and also car industries to create or improve the performance of the intelligent ADAS (i-ADAS) applications. For example, the captured real-time





(a) ***Left turn*** (b) ***Left lane change*** (c) ***Right turn***




(d) ***Right lane change*** (e) ***Going straight***

Figure 2.8: *Gaze points are shown on the driving frames over the last 5 seconds before a left/right turn, left/right lane change, or going straight maneuver occurs. Frames are divided into six areas.*

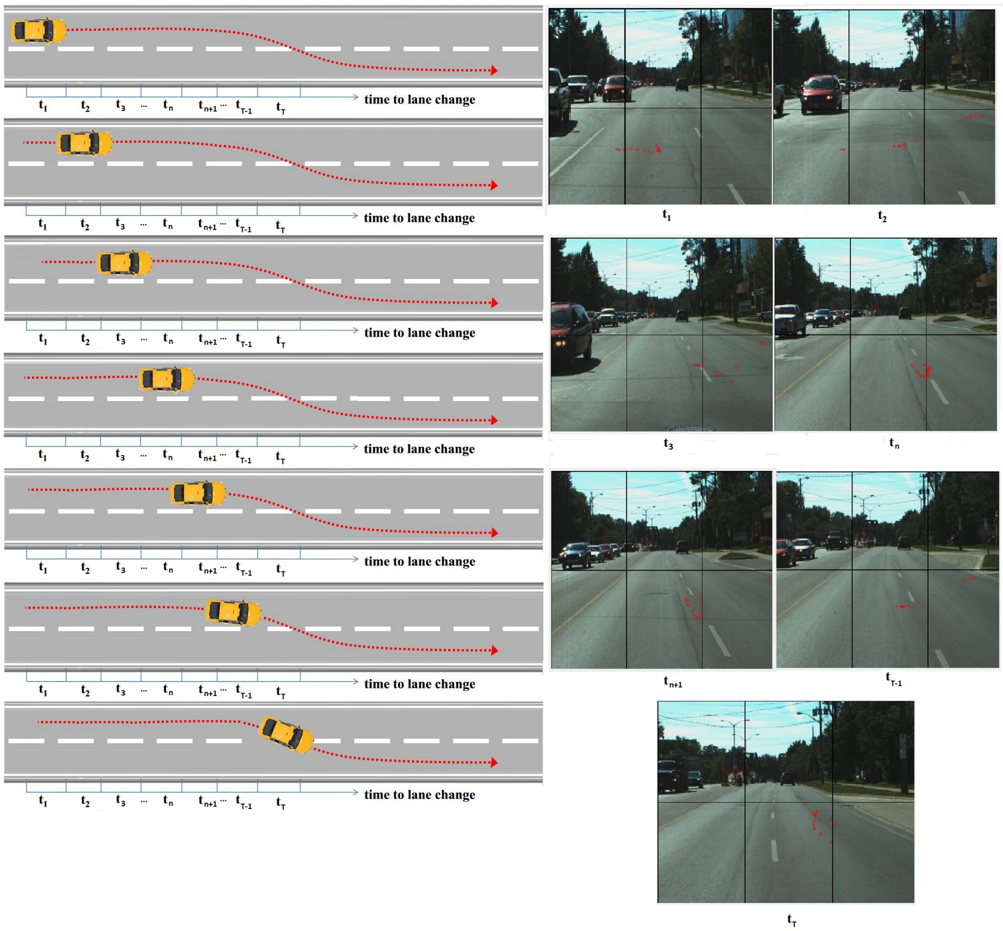


Figure 2.9: *A sequence of time slices belonging to a right lane change event. (t_1): Driver goes straight and looks forward. (t_2 and t_3): Driver decides to initiate an attempt to change lane, and searches visually for potential obstacles in the right lane. (t_n and t_{n+1}): Attention of the driver returns to the current lane and the driver still goes straight. (t_{T-1}): The driver makes the final decision to change lane and looks at the right lane. (t_T): Right lane change event has occurred.*

vehicular data provide the information that is essential for the application of driver maneuver prediction.

Vehicle dynamics-based data include vehicle speed, steering wheel angle, left/right turn signals, brake pedal pressure, gas pedal pressure and the speeds of all wheels. We integrated features to benefit from the sum of them simultaneously. For each time slice, we made a histogram of steering wheel angles and encoded the minimum, average and maximum values of vehicle speed, brake pedal pressure, gas pedal pressure, indicating independent wheel speeds. Finally, for left and right turn signals, we considered a binary feature for each. This feature value is 1 if the turn signal is on, and 0 otherwise.

Algorithm 1 *Driver Maneuver Prediction Using LSTM*

Input: Cephalo-Ocular Behaviour and Vehicle Dynamics Features; Prediction Threshold P_{th}
Output: Predicted Maneuver M; Time-to-Maneuver

```
while t = 1 to T do
    Observe features available up to current time slice
    Max Probability = Calculate and find the maximum of probabilities of
    each maneuver using LSTM model
    if Max Probability > P_th then
        M = Corresponding maneuver with Max Probability
        Time-to-Maneuver = T - t
        break
    end if
end while
Return M, Time-to-Maneuver
```

2.5 Experimental Results

We first give an overview of our maneuver dataset. Then, we explain how we tuned different parameters of the proposed model. Finally, we report our experimental results for maneuver prediction in details.

2.5.1 Dataset

To investigate our proposed model, we applied our approach to driving sequences recorded by the RoadLAB instrumented vehicle in the city of London, Ontario, Canada [6], with the aim of comparing our results with those obtained by Zabihi *et al.* [55], using the same driving sequences as they did. Table 2.1 provides details on the sequences that have been collected by different drivers for our experiments. These driving sequences contain the data, including GPS, 3D driver gaze, head pose, vehicle speed, and the angle of steering wheel, among others. We used a total of 325 events which have been obtained from the aforementioned sequences containing 65 left lane changes, 40 right lane changes, 65 left turns, 75 right turns, and 80 randomly sampled instances of driving straight. Each actual event is considered as one sample, which means our dataset consists of a total of 325 non-overlapping sample events.

Table 2.1: Data description (Each sequence belongs to one driver)

Sequence	**Date of Capture**	**Temperature**	**Weather**
Seq. 8	Sep. 12 2012	27 °C	Sunny
Seq. 9	Sep. 17 2012	24 °C	Partially cloudy
Seq. 10	Sep. 19 2012	8 °C	Sunny
Seq. 11	Sep. 19 2012	12 °C	Sunny
Seq. 13	Sep. 21 2012	19 °C	Partially sunny
Seq. 14	Sep. 24 2012	7 °C	Sunny
Seq. 15	Sep. 24 2012	13 °C	Partially sunny

2.5.2 Learning Parameters

We benefited from 5-fold cross-validation to tune the network parameters and the threshold on probabilities for driver maneuver prediction by searching in ranges for the given different parameters. We selected those set of parameters which give us the highest F1-score on the validation set. Finally, we tested the model on a pre-separated unseen data that consists of a set of randomly selected samples. We performed this strategy several times to estimate the accuracy and generality of the proposed model. We explain more about F1-score and the results in Section 2.5.3. For instance, for the size of the time slice, researchers have reported different number of frames such as 10 [30], 15 [52] and 20 [17] in the literature. We also investigated the performance of the time slice consisting of 10, 15, 20, 25 and 30 consecutive frames and reached better results by employing 20 consecutive frames. Here, we briefly report the other fine-tuned parameters.

Our proposed model consists of 3 hidden LSTM layers. The number of hidden units for the 3 layers was set to 100. We added a dense layer with 5 units for the 5 output classes (including left/right lane changes, left/right turns and driving straight). We employed 0.25, 100 and 10 for the parameters of validation split, epochs and batch size, respectively. The *tanh* activation function for the LSTM layers was used in our experiments. We also used a SoftMax activation function, mean squared error and Adam method for the dense layer, loss function and stochastic optimization, respectively. Dropout is very important to avoid over-fitting, and so we used 0.2, 0.3 and 0.2 for the first, second and third LSTM layers respectively. Moreover, the threshold value in our experiments was set to 0.80 which has been discussed in details in Section 2.5.3. (See Figure 2.11).

2.5.3 Maneuver Prediction Results

In the test step, the model predicts the driver maneuver every 20 frames and we expect the prediction system to anticipate the maneuver using only partial observations of a sequence. Previously, Zabihi *et al.*[55] proposed an IO-HMM-based model to anticipate three maneuvers of left/right turns and driving straight using our real driving dataset. To compare the performance of our model with theirs, as a first experiment, we employed our approach to predict Zabihi's maneuvers only. In the second experiment, in addition to the aforementioned maneuvers, we utilized our method to predict the maneuvers of left/right lane changes. For each time slice (i.e. after receiving 20 frames), the model generates the probability for each maneuver. Obviously, the sum of these probabilities should be 1. Then, the maneuver with the highest probability is chosen as the predicted maneuver only if it is higher than a preset threshold. If the highest probability is less than the threshold (0.8), the system cannot predict the driver maneuver and requires reception of additional features from the next time slice to perform its task. Note that if the maneuver occurs and the system still has not predicted it, the system makes no prediction. We verified the performance of our model by calculating the measures of precision and recall for each maneuver. These measures are defined as follows:

$$P_r = \frac{t_p}{t_p + f_p} \tag{2.10}$$

and

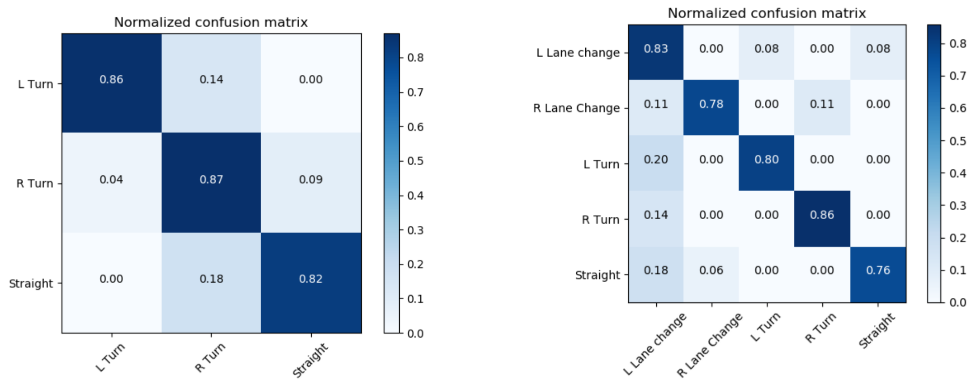
$$R_e = \frac{t_p}{t_p + f_n}, \tag{2.11}$$

where, for each maneuver m, t_p is the number of correctly predicted instances of maneuver m, f_p is the number of incorrectly predicted instances of maneuver

m, and f_n is the number of instances of maneuver m that are wrongly not predicted or the system does not choose any maneuver. In other words, precision is the number of correctly predicted instances of maneuver m divided by the number of instances that were predicted as maneuver m. Recall is the number of instances of correctly predicted maneuver m divided by the total number of instances of maneuver m. We computed the average of precision and recall. We also computed the average of time-to-maneuver, for true predictions (t_p), which indicates the interval between the time of algorithm's prediction and the start of the maneuver. Zabihi *et al.* [55] performed several experiments and reported that utilizing IO-HMM with the data on the driver's gaze and head pose (IO-HMM G+H) made the better model in terms of precision, recall and Time-to-Maneuver.

Table 2.2 compares our results (considering three and five maneuvers) with their best results. As can be seen, our LSTM-based model outperformed their prediction model. To be exact, precision and recall of our model for the three maneuvers are 6.1% and 0.8% respectively higher than those of the previous work by Zabihi *et al.* [55] for these three maneuvers. However, their method can predict the three maneuvers 0.16s earlier on average than ours. The last row in Table 2.2 shows the results of extending our model to predict two more types of maneuvers. In this case, we obviously expect more complexity for the problem and results show that precision, recall and time-to-maneuver have decreased slightly in comparison with our method for predicting only three maneuvers.

Figure 2.10 shows the confusion matrices for our prediction system for three and five maneuvers. In these matrices, a row represents an instance of the actual maneuver class, whereas a column represents an instance of the predicted maneuver class. Consequently, the values of the diagonal elements



(a) ***Model with three maneuvers*** (b) ***Model with five maneuvers***

Figure 2.10: *Confusion matrices of our prediction model*

represent the degree of correctly predicted classes which is greater than or equal to 82% and 76% for the three and five maneuvers respectively.

Table 2.2: Result of different models of driver maneuver prediction on our data set.

	Pr (%)	**Re (%)**	**Time-to-maneuver(s)**
IO-HMM G+H (for three maneuvers)	79.5	83.3	3.8
Our model (for three maneuvers)	85.6	84.1	3.64
Our model (for five maneuvers)	84.2	82.9	3.56

Figure 2.11 compares the changes of the F1-score when we employ our model and the IO-HMM-based model, with different values for the threshold. The F1-score is the harmonic mean of P_r and R_e, where it can reach 1 with perfect precision and recall, and 0 in the worst case. In other words, the prediction threshold is a useful parameter to find a trade-off between the precision

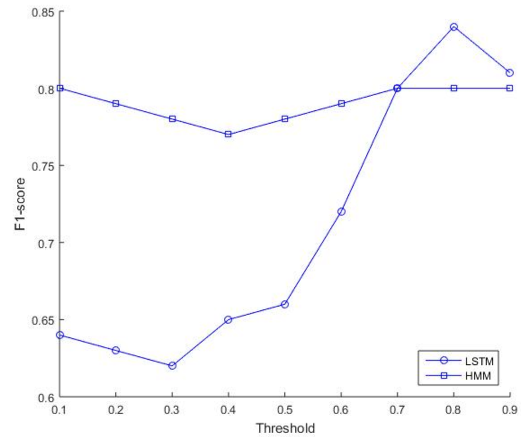


Figure 2.11: *The effect of the threshold on the $F1$ score for IO-HMM and LSTM models.*

and recall of the algorithms. The F1-score is defined as follows:

$$F1 = \frac{2P_rR_e}{P_r + R_e} \tag{2.12}$$

As can be seen, the trend of F1-scores for the IO-HMM model remains roughly stable when the threshold changes. However, when we choose 0.8 for the threshold, the LSTM-based prediction model achieves a significantly higher F1-score in comparison with IO-HMM model. In Table 2.2, we utilized the threshold values which gave us the highest F1-score. Our model predicts maneuvers every 0.67 seconds (20 frames) in 2.8 milliseconds on average on a $3.40GHz$ Core $i7 - 6700$ CPU with Windows 10.

Finally, we briefly mention here the results of several previous works which have also addressed the driver maneuver prediction problem, using their own dataset and features. For instance, Morris *et al.* [36] accomplished a binary classification of lane changes and driving straight maneuvers. They employed a Relevance Vector Machine (RVM; a Bayesian extension to the popular SVM).

In addition, Jain *et al.* [17] evaluated some algorithms for the same purpose (including SVM, Bayesian Network and variants of their deep learning model). The methods listed in Table 2.3 use identical feature vectors, which guarantees a fair comparison[1]. As can be observed, the SVM classification does not model the temporal aspect of the data, and its performance is poor as a result.

Table 2.3: MANEUVER ANTICIPATION RESULTS OF SEVERAL PREVIOUS METHODS.

Method	**Pr (%)**	**Re (%)**	**Time-to-maneuver(s)**
SVM[36]	43.7±2.4	37.7±1.8	1.20
IO-HMM[16]	74.2±1.7	71.2±1.6	3.83
AIO-HMM[16]	77.4±2.3	71.2±1.3	3.53
S-RNN[17]	78.0±1.5	71.1±1.0	3.15
F-RNN-UL[17]	82.2±1.0	75.9±1.5	3.75
F-RNN-EL[17]	84.5±1.0	77.1±1.3	3.58

2.6 Common Reasons for Wrong Maneuver Anticipations

We discuss some major reasons that can generally result in wrong anticipations in the driver maneuver prediction problem. For example, when a driver is interacting with other passengers, head and gaze features are not reliable enough to be taken into account. Also, a driver may be distracted when he/she is watching videos, programming a GPS, using a cell phone, adjusting the radio, smoking and etc. In such situations, wrong anticipation is common as

[1]The methods listed in the Table are: SVM: Support Vector Machine, IO-HMM: Input-Output Hidden Markov Model, AIO-HMM: Auto-Regressive Input Output Hidden Markov Model, S-RNN: Simple Recurrent Neural Network, F-RNN-UL: Fusion-Recurrent Neural Network Uniform Loss, F-RNN-EL: Fusion-Recurrent Neural Network Exponential Loss.

the driver may not be fully focused on the road. Moreover, different drivers have different driving styles. For example, during lane change maneuver, some drivers may merge slowly while others may merge quickly that in this case, the driver has not provided the system with enough data and time to predict maneuver. Hence, in this situation, other features such as speed, acceleration, steering wheel angle can be significant to predict an accurate maneuver. As another example, when drivers rely on their recent perception of traffic scene, they probably do not check blind spots and the surroundings carefully resulting in lack of head information but we may still have valid gaze features. A similar driving situation is when a driver is driving in left/right-turn-only lanes. In this case, the driver might not give us helpful head information as well.

2.7 Conclusion and Future Work

We presented a deep learning-based model to predict driver maneuvers several seconds before they are performed. We employed driver cephalo-ocular behavioral information and vehicle dynamics data as features to train our model. Our experimental results show that our model outperformed the previous IO-HMM model [55]. It improved the precision from 79.5% to 85.6% and recall from 83.3% to 84.1%. Moreover, we expanded the prediction model to anticipate two more maneuvers (left/right lane changes). For predicting the five maneuvers, our model achieved 84.2% and 82.9% for precision and recall respectively. Our model has three features which make it competitive and more reliable in comparison to previous work: it employs an LSTM to utilize long-term temporal dependencies, is able to predict five maneuver types and benefits from using gaze information.

Several limitations do exist and can be addressed for improving the accuracy and generality of the model. Adding more features from the environment, such as the lane in which the driver is located or where the driver is gazing during the driving maneuver, could improve the accuracy of the model. In terms of generality, the tests conducted in this research were based on a limited number of drivers and under specific weather and environmental conditions. Collecting new data under different situations and training the model on a broader set of data could help the generality of the model. Hence, for the commercial use of this model, the mentioned items need to be considered. Lastly, this research area is still challenging and more research is still needed before such models are practical in commercial use. As for future work, we plan to study the extraction of features from video within the attentional visual area of the driver. We believe that utilizing LSTM trained with a combination of these features, with cephalo-ocular behavior and the vehicle dynamics will improve current prediction results.

Bibliography

[1] Hideomi Amata, Chiyomi Miyajima, Takanori Nishino, Norihide Kitaoka, and Kazuya Takeda. "Prediction model of driving behavior based on traffic conditions and driver types". In: *Intelligent Transportation Systems, ITSC'09. 12th International IEEE Conference on.* IEEE. 2009, pp. 1–6.

[2] Seifemichael B Amsalu and Abdollah Homaifar. "Driver behavior modeling near intersections using Hidden Markov Model based on genetic algorithm". In: *Intelligent Transportation Engineering (ICITE), IEEE International Conference on.* IEEE. 2016, pp. 193–200.

[3] Georges S Aoude, Vishnu R Desaraju, Lauren H Stephens, and Jonathan P How. "Behavior classification algorithms at intersections and validation using naturalistic data". In: *Intelligent Vehicles Symposium (IV), 2011 IEEE.* IEEE. 2011, pp. 601–606.

[4] Abdelhadi Azzouni and Guy Pujolle. "A long short-term memory recurrent neural network framework for network traffic matrix prediction". In: *arXiv preprint arXiv:1705.05690* (2017).

[5] S.S. Beauchemin, M.A. Bauer, T. Kowsari, and J. Cho. "Portable and Scalable Vision-Based Vehicular Instrumentation for the Analysis of

Driver Intentionality". In: *IEEE Transactions on Instrumentation and Measurement* 61.2 (2012), pp. 391–401.

[6] Steven Beauchemin, M Bauer, Denis Laurendeau, T Kowsari, J Cho, M Hunter, and O McCarthy. "Roadlab: An in-vehicle laboratory for developing cognitive cars". In: *Proc. 23rd Int. Conf. CAINE.* 2010.

[7] Steven Beauchemin, Michael Bauer, Denis Laurendeau, Taha Kowsari, Ji Cho, Morgan Hunter, Kyle Charbonneau, and Owen McCarthy. "Road-Lab: An In-Vehicle Laboratory for Developing On-Board i-ADAS." In: Jan. 2010, pp. 7–12.

[8] 2012 motor vehicle crashes: "N. Highway Traffic Safety Administration, Washington, D.C." In: *Tech. Rep.* (2013).

[9] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. "Long-term recurrent convolutional networks for visual recognition and description". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015, pp. 2625–2634.

[10] Jun Gao, Jiangang Yi, and Yi Lu Murphey. "Attention-based global context network for driving maneuvers prediction". In: *Machine Vision and Applications* 33.4 (2022), pp. 1–11.

[11] Olaf Gietelink, Jeroen Ploeg, Bart De Schutter, and Michel Verhaegen. "Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations". In: *Vehicle System Dynamics* 44.7 (2006), pp. 569–590.

[12] Lei He, Chang-fu Zong, and Chang Wang. "Driving intention recognition and behaviour prediction based on a double-layer hidden Markov

model". In: *Journal of Zhejiang University SCIENCE C* 13.3 (2012), pp. 208–217.

[13] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[14] Chiori Hori, Shinji Watanabe, Takaaki Hori, Bret A Harsham, JohnR Hershey, Yusuke Koji, Yoichi Fujii, and Yuki Furumoto. "Driver confusion status detection using recurrent neural networks". In: *Multimedia and Expo (ICME), 2016 IEEE International Conference on*. IEEE. 2016, pp. 1–6.

[15] Timothy Huang, Daphne Koller, Jitendra Malik, G Ogasawara, B Rao, Stuart J Russell, and Joseph Weber. "Automatic symbolic traffic scene analysis using belief networks". In: *AAAI*. Vol. 94. 1994, pp. 966–972.

[16] Ashesh Jain, Hema S Koppula, Bharad Raghavan, Shane Soh, and Ashutosh Saxena. "Car that knows before you do: Anticipating maneuvers via learning temporal driving models". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 3182–3190.

[17] Ashesh Jain, Avi Singh, Hema S Koppula, Shane Soh, and Ashutosh Saxena. "Recurrent neural networks for driver activity anticipation via sensory-fusion architecture". In: *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE. 2016, pp. 3118–3125.

[18] Dietmar Kasper, Galia Weidl, Thao Dang, Gabi Breuel, Andreas Tamke, Andreas Wedel, and Wolfgang Rosenstiel. "Object-oriented Bayesian networks for detection of lane change maneuvers". In: *IEEE Intelligent Transportation Systems Magazine* 4.3 (2012), pp. 19–31.

[19] Aida Khosroshahi, Eshed Ohn-Bar, and Mohan Manubhai Trivedi. "Surround vehicles trajectory analysis with recurrent neural networks". In: *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE. 2016, pp. 2267–2272.

[20] ByeoungDo Kim, Chang Mook Kang, Seung Hi Lee, Hyunmin Chae, Jaekyum Kim, Chung Choo Chung, and Jun Won Choi. "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network". In: *arXiv preprint arXiv:1704.07049* (2017).

[21] I.H. Kim, J.H. Bong, J. Park, and S. Park. "Prediction of driver's intention of lane change by augmenting sensor information using machine learning techniques". In: *Sensors* 17.6 (2017), p. 1350.

[22] Ryan Kiros, Ruslan Salakhutdinov, and Richard S Zemel. "Unifying visual-semantic embeddings with multimodal neural language models". In: *arXiv preprint arXiv:1411.2539* (2014).

[23] Stefan Klingelschmitt, Matthias Platho, Horst-Michael Groß, Volker Willert, and Julian Eggert. "Combining behavior and situation information for reliably estimating multiple intentions". In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 388–393.

[24] Taha Kowsari, Steven S Beauchemin, Michael Anthony Bauer, Denis Laurendeau, and Normand Teasdale. "Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems". In: *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*. IEEE. 2014, pp. 1245–1250.

[25] Nobuyuki Kuge, Tomohiro Yamamura, Osamu Shimoyama, and Andrew Liu. “A driver behavior recognition method based on a driver model framework”. In: *SAE transactions* 109.6 (2000), pp. 469–476.

[26] Puneet Kumar, Mathias Perrollaz, Stéphanie Lefevre, and Christian Laugier. “Learning-based approach for online lane change intention prediction”. In: *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2013, pp. 797–802.

[27] Veit Leonhardt and Gerd Wanielik. “Neural network for lane change prediction assessing driving situation, driver behavior and vehicle movement”. In: *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on*. IEEE. 2017, pp. 1–6.

[28] Junxiang Li, Xiaohui Li, Bohan Jiang, and Qi Zhu. “A maneuver-prediction method based on dynamic bayesian network in highway scenarios”. In: *2018 Chinese Control And Decision Conference (CCDC)*. IEEE. 2018.

[29] Martin Liebner, Felix Klanner, Michael Baumann, Christian Ruhhammer, and Christoph Stiller. “Velocity-based driver intent inference at urban intersections in the presence of preceding vehicles”. In: *IEEE Intelligent Transportation Systems Magazine* 5.2 (2013), pp. 10–21.

[30] Viktor Losing, Barbara Hammer, and Heiko Wersing. “Personalized maneuver prediction at intersections”. In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2017, pp. 1–6.

[31] Charles C. MacAdam and Gregory E. Johnson. “Application of elementary neural networks and preview sensors for representing driver

steering control behaviour". In: *Vehicle System Dynamics* 25.1 (1996), pp. 3–30.

[32] Vishal Mahajan, Christos Katrakazas, and Constantinos Antoniou. "Prediction of lane-changing maneuvers with automatic labeling and deep learning". In: *Transportation research record* 2674.7 (2020), pp. 336–347.

[33] Daniel Meyer-Delius, Christian Plagemann, Georg Von Wichert, Wendelin Feiten, Gisbert Lawitzky, and Wolfram Burgard. "A probabilistic relational model for characterizing situations in dynamic multi-agent systems". In: *Data analysis, machine learning and applications.* Springer, 2008, pp. 269–276.

[34] Dejan Mitrovic. "Machine learning for car navigation". In: *Engineering of Intelligent Systems.* Springer, 2001, pp. 670–675.

[35] Hermes J Mora and Esteban J Pino. "Simplified Prediction Method for Detecting the Emergency Braking Intention Using EEG and a CNN Trained with a 2D Matrices Tensor Arrangement". In: *International Journal of Human–Computer Interaction* (2022), pp. 1–14.

[36] B. Morris, A. Doshi, and M. Trivedi. "Lane change intent prediction for driver assistance: On-road design and evaluation". In: *Intelligent Vehicles Symposium (IV), 2011 IEEE.* IEEE. 2011, pp. 895–901.

[37] Jeremy Morton, Tim A Wheeler, and Mykel J Kochenderfer. "Analysis of recurrent neural networks for probabilistic modeling of driver behavior". In: *IEEE Transactions on Intelligent Transportation Systems* 18.5 (2016), pp. 1289–1298.

[38] Oluwatobi Olabiyi, Eric Martinson, Vijay Chintalapudi, and Rui Guo. "Driver Action Prediction Using Deep (Bidirectional) Recurrent Neural Network". In: *arXiv preprint arXiv:1706.02257* (2017).

[39] Jane Oruh, Serestina Viriri, and Adekanmi Adegun. "Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition". In: *IEEE Access* 10 (2022), pp. 30069–30079.

[40] Sajan Patel, Brent Griffin, Kristofer Kusano, and Jason J Corso. "Predicting Future Lane Changes of Other Highway Vehicles using RNN-based Deep Models". In: *arXiv preprint arXiv:1801.04340* (2018).

[41] Jinshuan Peng, Yingshi Guo, Rui Fu, Wei Yuan, and Chang Wang. "Multi-parameter prediction of drivers' lane-changing behaviour with neural network model". In: *Applied ergonomics* 50 (2015), pp. 207–217.

[42] Claas Rodemerk, Hermann Winner, and Robert Kastner. "Predicting the driver's turn intentions at urban intersections using context-based indicators". In: *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2015, pp. 964–969.

[43] Oliver Scheel, Loren Schwarz, Nassir Navab, and Federico Tombari. "Situation Assessment for Planning Lane Changes: Combining Recurrent Models and Prediction". In: *arXiv preprint arXiv:1805.06776* (2018).

[44] Joerg Schneider, Andreas Wilde, and Karl Naab. "Probabilistic approach for modeling and identifying driving situations". In: *Intelligent Vehicles Symposium, 2008 IEEE*. IEEE. 2008, pp. 343–348.

[45] Ralf C Staudemeyer and Eric Rothstein Morris. "Understanding LSTM– a tutorial into long short-term memory recurrent neural networks". In: *arXiv preprint arXiv:1909.09586* (2019).

[46] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. "Sequence to sequence learning with neural networks". In: *Advances in neural information processing systems* 27 (2014).

[47] Kenji Takagi, Haruki Kawanaka, Md. Shoaib Bhuiyan, and Koji Oguri. "Estimation of a three-dimensional gaze point and the gaze target from the road images". In: *Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on*. IEEE. 2011, pp. 526–531.

[48] Shigeki Tezuka, Hitoshi Soma, and Katsuya Tanifuji. "A study of driver behavior inference model at time of lane change using Bayesian networks". In: *Industrial Technology, 2006. ICIT 2006. IEEE International Conference on*. IEEE. 2006, pp. 2308–2313.

[49] Duy Tran, Weihua Sheng, Li Liu, and Meiqin Liu. "A Hidden Markov Model based driver intention prediction system". In: *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2015 IEEE International Conference on*. IEEE. 2015, pp. 115–120.

[50] Shaobo Wang, Pan Zhao, Biao Yu, Weixin Huang, and Huawei Liang. "Vehicle trajectory prediction by knowledge-driven lstm network in urban environments". In: *Journal of Advanced Transportation* 2020 (2020).

[51] Cheng Wei, Fei Hui, and Asad J Khattak. "Driver lane-changing behavior prediction based on deep learning". In: *Journal of advanced transportation* 2021 (2021).

[52] Jürgen Wiest, Matthias Karg, Felix Kunz, Stephan Reuter, Ulrich Kreßel, and Klaus Dietmayer. "A probabilistic maneuver prediction framework

for self-learning vehicles with application to intersections". In: *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2015, pp. 349–355.

[53] Martin Wollmer, Christoph Blaschke, Thomas Schindl, Björn Schuller, Berthold Farber, Stefan Mayer, and Benjamin Trefflich. "Online driver distraction detection using long short-term memory". In: *IEEE Transactions on Intelligent Transportation Systems* 12.2 (2011), pp. 574–582.

[54] Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, and George Toderici. "Beyond short snippets: Deep networks for video classification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 4694–4702.

[55] S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. "Real-time driving manoeuvre prediction using IO-HMM and driver cephalo-ocular behaviour". In: *Intelligent Vehicles Symposium (IV), 2017 IEEE*. IEEE. 2017, pp. 875–880.

[56] Jianwei Zhang and Bernd Roessler. "Situation analysis and adaptive risk assessment for intersection safety systems in advanced assisted driving". In: *Autonome Mobile Systeme 2009*. Springer, 2009, pp. 249–258.

[57] Alex Zyner, Stewart Worrall, and Eduardo Nebot. "A Recurrent Neural Network Solution for Predicting Driver Intention at Unsignalized Intersections". In: *IEEE Robotics and Automation Letters* 3.3 (2018), pp. 1759–1764.

[58] Alex Zyner, Stewart Worrall, James Ward, and Eduardo Nebot. "Long short term memory for driver intent prediction". In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2017, pp. 1484–1489.

Chapter 3

Traffic Object Detection and Recognition Based on the Attentional Visual Field of Drivers

This Chapter is a reformatted version of the following article:

M. Shirpour, N. Khairdoost, M.A. Bauer, S.S. Beauchemin, *Traffic Object Detection and Recognition Based on the Attentional Visual Field of Drivers. IEEE Transactions on Intelligent Vehicles*, Dec. 2021.

Traffic object detection and recognition systems play an essential role in Advanced Driver Assistance Systems (ADASs) and Autonomous Vehicles (AVs). In this research, we focus on four important classes of traffic objects: traffic signs, road vehicles, pedestrians, and traffic lights. We first review the major traditional machine learning and deep learning methods that have been used in the literature to detect and recognize these objects. We provide a vision-based framework that detects and recognizes traffic objects inside and outside the attentional visual area of drivers. This approach uses the driver 3D absolute coordinates of the gaze point obtained by the combined, cross-calibrated use of a front-view stereo imaging system and a non-contact 3D gaze tracker. A combination of multi-scale HOG-SVM and Faster R-CNN-based models is utilized in the detection stage. The recognition stage is performed with a ResNet101 network to verify sets of generated hypotheses. We applied our approach on real data collected during drives in an urban environment with the RoadLAB instrumented vehicle. Our framework achieved 91% of correct object detections and provided promising results in the object recognition stage.

3.1 Introduction

Advanced Driver Assistance Systems (ADASs) have attracted the attention of many researchers and vehicle manufacturers for several decades. Achieving higher performance levels for ADAS requires a robust perception of the driving environment. Hence, vision-based traffic scene perception which refers to the identification of the position of traffic objects such as pedestrians, vehicles, traffic signs, etc is of great importance in designing a modern ADAS. However, in practice, many traffic scene issues, such as occlusions, weather conditions, shadows and distant object identification affect the performance of such systems. Improving the accuracy and adaptability of such methods is still a challenging area of research [89]. In this study, we focus on four essential categories of objects: traffic signs, vehicles, traffic lights, and pedestrians. Correctly detecting and localizing these classes of objects in the context of ADAS is still a difficult challenge. Typically, problems encountered include variations in viewpoints, object shape, size, color, distance from sensors, illumination conditions, and object occlusion [4], [20], [25].

Our contributions include: collecting and labelling a large dataset including images of different objects, and proposing an integrated framework to detect and recognize traffic objects including traffic signs, vehicles, traffic lights, and pedestrians. Our model inherits the advantage of deep neural networks (ResNet and Faster R-CNN) and classical machine learning models (multi-scale HOG-SVM). This framework is the first one of its kind which performs its tasks taking the attentional visual field of the driver into consideration. This is an important aspect of an ADAS, as it allows the ADAS to identify objects seen and not seen by the driver, among other things.

This contribution is organized as follows: In Section 3.2, we review the related literature. Section 3.3 describes the datasets we used and the proposed method. Section 3.4 presents the experimental results obtained along with a critical analysis. Conclusions and future research directions are described in Section 3.5.

3.2 Related Works

3.2.1 Generic Object Detection

Generic object detection algorithms can be divided into two major types of traditional and deep learning-based methods. In this section, we briefly review these generic object detection methods. Several object detection surveys can be found in [113], [114], [90], [118], [98] and [30].

Among the traditional object detectors we find the framework proposed by Viola and Jones which employs searches based on sliding-windows and AdaBoost classifiers [95]. Another popularly used framework is the linear Support Vector Machine (SVM) classifier with such features as Histograms of Oriented Gradients (HOG), Scale Invariant Feature Transforms (SIFT), and Local Binary Patterns (LBP). For example, in [53] and [22], researchers employed SVM and a multi-scale detection framework with HOG features to detect birds and pedestrians respectively. Finally, Aggregated Channel Features (ACF) is as another successful detection framework that has been proposed by [21]. This method also uses sliding-window searches and AdaBoost to detect objects in a multi-scale fashion [70], [64].

Unlike traditional object detection algorithms that benefit from prior knowledge, deep learning-based object detection methods attempt to learn high-level features from a massive amount of data. As a result, they are less sensitive to illumination changes, deformations and geometric transformations [86]. There are two major types of deep learning-based object detection methods: Region-based methods and regression-based methods. The former generates region proposals at first and then classifies them into different object categories while the latter transforms the object detection problem into a regression problem and predicts locations and class probabilities directly from the whole image [113]. The region-based methods mainly include R-CNN [30], Fast R-CNN [29], Faster R-CNN [78], R-FCN [16], SPP-net [38] and Mask R-CNN [36]. On the other hand, the regression-based methods mainly include AttentionNet [107], G-CNN [67], SSD [62], YOLO [75], YOLOv2 [76], YOLOv3 [77], DSOD [82] and DSSD [27].

3.2.2 Traffic Sign Detection and Recognition

Sign detection methods are generally categorized into color-based, shape-based and hybrid approaches [44], [96]. Color-based methods use color information as the main attribute to localize image regions containing traffic signs in the image. Color thresholding segmentation is the more common approach among color-based methods as it reduces the search area by ignoring untargeted regions [18], [54]. These methods are generally sensitive to variations in illumination and the distance to traffic signs [79]. Traffic signs also have specific shapes that can be searched for by shape-based methods. The Hough transform is one of the most common shape-based methods [68], [106], as it is relatively robust against illumination change and image noise. Similarity detection [80] and Distance Transform matching [28] also constitute shape-based methods. Hybrid approaches take advantage of both sign color and shape [19], [74]. Classification stages mostly employ template matching [91], [33], SVM [103], [110], Genetic Algorithm (GA) [50], Artificial Neural Network (ANN) [35], [39], AdaBoost [11], [60] and deep learning-based methods. In recent years, deep learning methods have increasingly attracted a great deal of attention. Convolutional Neural Networks (CNNs) constitute a subset of deep neural network models that have the power to learn robust and discriminative features from raw data. There is a variety of CNN that have been employed for traffic sign recognition such as small-scale CNN [117], multi-scale CNN [81], a committee of CNN [15], multi-column CNN [14], and multi-task CNN [52], CNN-SVM [58], [55], among others. A number of traffic sign datasets have been created in the past decade. However, methods that have been proposed in the literature are mostly based on European datasets. As Traffic signs in North America differ in color and shape, the methods that have been proposed based on European traffic signs are not directly suitable in the North American context [108].

3.2.3 Vehicle Detection

Many traditional vehicle detection approaches comprise a Hypothesis Generation (HG) step followed by a Hypothesis Verification (HV) step. With regards to HG, there are various methods that can be divided into three basic categorizes including knowledge-based, stereo-based, and motion-based [87]. Knowledge-based methods use prior knowledge including shadows [93], symmetry [49], edges [66], color [104], texture [8], corners [5] and vehicle lights [10]. Stereo-based approaches usually exploit the Inverse Perspective Mapping (IPM) [6] or disparity maps [26] to localize vehicles, while motion-based methods detect vehicles with optical flow [63]. HV approaches can be classified into two major categories [87]: template-based and appearance-based. The former employs predefined vehicle patterns and estimates the correlation between templates and candidate image regions [34], while the latter uses machine learning methods such as SVM [92], ANN [31], and AdaBoost [85] to classify hypotheses into vehicle and non-vehicle categories.

Classifiers such as SVM [92], ANN [31], and AdaBoost [85] learn the characteristics of vehicle appearance to draw a decision boundary between vehicle and non-vehicle classes. In HV, a number of local feature descriptors such as HOG [105], PHOG [49], Haar-like [100], Gabor [65], and SURF [59] have shown a remarkable ability in collecting contextual information. Additionally, different vehicle detection approaches that employ deep learning-based methods discussed in Section 3.2.1 have been proposed. For instance, in [23], the authors provided a comparative study on the performance of AlexNet and Faster R-CNN models. Also, in [116], the authors exploited the fine-tuned YOLO [75] for vehicle detection. In [42], vehicles are detected with a simplified Fast R-CNN.

3.2.4 Pedestrian Detection

Many traditional methods for pedestrian detection have been proposed with the majority of them using features such as HOG [17], Haar-Like [72], Viola-Jones [94], and LBP [97], followed by a classification stage using either SVM, ANN, or AdaBoost. Additionally, pedestrian detection methods using deep learning can be categorized as either single-stage or two-stage techniques. RPN+BF [111], and Faster R-CNN [112] are examples where the authors employed a two-stage approach. Examples of single-stage approaches have been proposed: For instance, Lan *et al.* [56] modified YOLOv2 into a single-stage network called YOLO-R for pedestrian detection. Comprehensive surveys on pedestrian detection are provided in [7] and [1].

3.2.5 Traffic Light Detection

Color segmentation is a method often used to reduce the search space in traffic scene images. For example, in [12] and [9], the authors employed HSI and YCbCr color spaces respectively to detect traffic lights. In some studies, a shape-based method such as the circular Hough transform [71] was used after color segmentation to find round traffic lights. Blob detection is another approach to detect traffic lights that analyses the size and aspect ratio of the traffic lights to eliminate regions likely to produce false positives [115]. In [47], saliency maps are employed to detect traffic lights. In [48], GPS data and digital maps are used to identify traffic lights in urban areas. Feature descriptors such as HOG [12], Haar-like [32], and Gabor Wavelets [9] have been extensively used to detect traffic lights. To recognize the state of traffic lights, several methods have been employed mostly including SVM [83], fuzzy algorithms [2] and more recently, deep learning methods. A simple CNN was used by Lee and Park [57] for traffic light classification. Behrendt *et al.* [4] applied YOLOv1 for detection and classification. In [45], YOLO-9000 [76] was applied to the LISA traffic light dataset. The authors in [99] exploited DeepTLR networks for real-time traffic light detection and classification. A novel Faster R-CNN hierarchical architecture was proposed in [73] and trained on a joint traffic light and sign dataset.

Prior to our work, there has been a very little attention in previous research for simultaneously detecting different major classes of traffic objects. Hence, one aspect which makes our work different from others is the fact that in addition to detection of more major classes of traffic objects, we also classify them into their own subcategories.

3.3 Proposed Method

In this section, we describe our proposed method for traffic object detection and recognition based on the attentional visual field of the driver. First, our dataset used in this research is introduced. Following this, we describe the method employed to find the attentional gaze area of the driver in the forward stereo imaging system. Next, in the object detection stage, our trained models and the methods used for enriching our data set are described. We then discuss the Region of Interests (ROIs) integration method we used. Finally, the object recognition stage is presented. Figure 3.1 illustrates our proposed framework.

3.3.1 The RoadLAB Dataset

An essential element of deep learning-based object detection systems is the availability of a large number of sample images. In this section, we present our

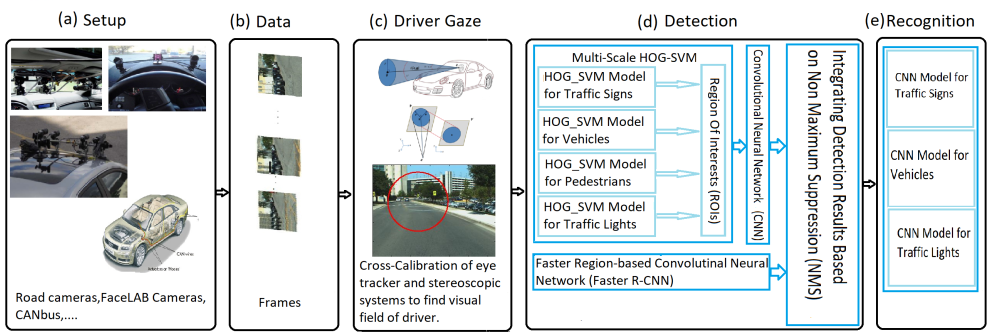


Figure 3.1: **Framework Overview.** *Our framework detects and recognizes traffic objects inside the visual field of driver.* **(from left to right: a)** *The RoadLAB vehicle with forward stereoscopic and eye-tracking systems.* **b)** *Dataset created with the RoadLAB experimental vehicle.* **c)** *Computing the radius of driver's view as attentional gaze cone and locating the re-projected 2D ellipse of the visual field of the driver.* **d)** *We used two different model types in the detection stage of the framework; Model A consists of two steps including multi-scale HOG-SVM followed by applying a CNN, and Model B is a Faster Region-based CNN. Detection results are integrated by an NMS-based algorithm.* **e)** *For the recognition stage, we separately trained three independent models on traffic signs, vehicles, and traffic lights.*

own object dataset from the RoadLAB experimental data sequences [3], [51], [84]. As one of our contributions in this study, in order to train, validate and test our models, we collected 13,546 sample images to detect and recognize traffic objects including traffic signs, vehicles, pedestrians and traffic lights. Our dataset contains 3,225 sample images for the background class in addition to 5,172, 1,984, 1,290 and 1,875 sample images for the object classes of traffic sign, vehicle, pedestrian and traffic light respectively. The vehicle class consists of 3 distinct classes including car, bus and truck. The traffic light class consists of 4 distinct classes including red, yellow, green and not clear. Finally, the traffic sign class includes 19 distinct classes of traffic signs. Additionally, some traffic sign classes include more than one sign type such as "Maximum Speed Limit", "Construction", "Parking", etc. Our samples for traffic signs can be considered as a complete sign dataset including warning signs, regulatory signs, direction signs, and temporary signs.

3.3.2 Driver Gaze Localization

The visual attentional field of the driver consists of a circle in 3D space within the plane that contains the Point of Gaze (PoG), perpendicular to the Line of Gaze (LoG). The radius of the circle is determined by the angular opening of the cone of visual attention as shown in Figure 3.2. The circle generally is projected onto the imaging plane of the stereo sensor as a 2D ellipse. We describe the procedure we employed, as per Kowsari *et al.* [51].

First, both the eye position $e = (e_x, e_y, e_z)$ and the 3D PoG $g = (g_x, g_y, g_z)$ are transformed into the reference frame of the forward stereo sensor. Next, the radius of the circular attentional gaze area is obtained by computing the

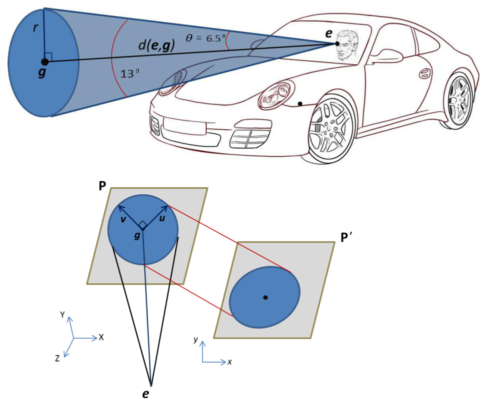


Figure 3.2: **(top):** *Depiction of the driver attentional gaze cone.* **(bottom):** *Re-projection of the 3D attentional circle into the corresponding 2D ellipse on image plane of the forward stereo scene system.*

Euclidean distance between e and g (θ is set to 6.5°: [88]).

$$r = \tan(\theta)\|e - g\|_2 \quad (3.1)$$

We re-project the obtained circle contained in the 3D plane perpendicular to the LoG onto the image plane of the forward stereo imaging sensor where it becomes an ellipse. The coordinates of the ellipse are obtained as:

$$(X, Y, Z) = g + r(\cos\phi\mathbf{u} + \sin\phi\mathbf{v}) \quad (3.2)$$

where $\mathbf{u}$=(u_x, u_y, u_z) and $\mathbf{v}$=(v_x, v_y, v_z) are two orthonormal vectors in the plane orthogonal to the LoG and $\phi \in [0, 2\pi]$. Using perspective projection $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ and applying the intrinsic calibration matrix of the stereo scene system from [51] yields the 2D ellipse on the image plane of the forward stereo sensor. The mathematical details are found in [51] and [109]. Figure 3.3 illustrates several attentional visual areas for several sample frames.

Figure 3.3: *Examples of attentional gaze areas projected onto the forward stereo sensor of the vehicle.*

3.3.3 Object Detection Stage

To detect traffic objects of interest inside and outside of the attentional field of the driver, we employed a framework consisting of two different model types that we proceed to describe:

Model A

The first model consists of two steps that include a multi-scale HOG-SVM followed by the use of a ResNet101 network. The multi-scale HOG-SVM descriptor counts occurrences of gradient orientations in an image region followed by a block-normalization algorithm that results in better invariance to edge contrast and shadows. Since it operates on local cells, it is also relatively invariant to geometric and photometric transformations. In general, the detection algorithm is based on an overlapping sliding window approach. Since the Region of Interest (ROI) contains objects that vary in size, we used a multi-scale method for the object detection problem. We treat the HOG fea-

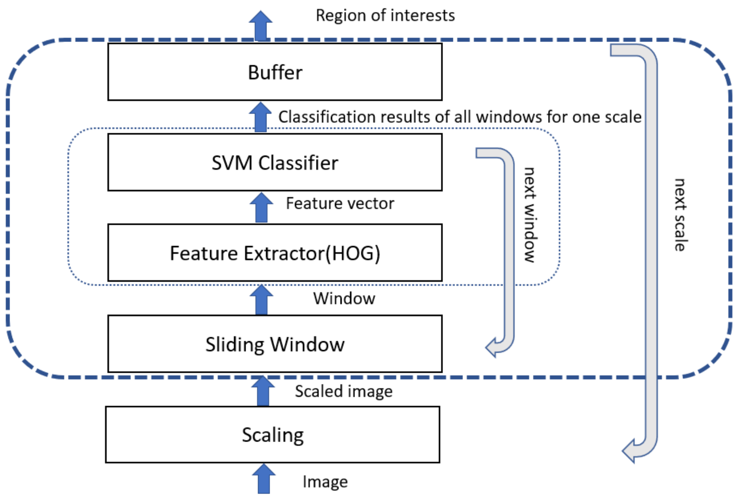


Figure 3.4: *Internal view of a multi-scale HOG-SVM*

tures extracted from each sliding window at each level as independent samples prior to feeding them to the SVM classifier. Figure 3.4 illustrates the internal view of multi-scale HOG-SVM.

We trained four independent multi-scale HOG-SVM models to find ROIs, for our four types of traffic objects (signs, vehicles, pedestrians, and traffic lights). The model moves a sliding window across the images and HOG features are extracted. The model follows this strategy at several imaging scales. Typically, SVM outputs conventional binary decision labels. However, it can also provide a probabilistic confidence score [61] for each sliding window, which we use to threshold on ROIs. With the use of HOG-SVM, we discard the ROIs labelled as background while other candidates are transferred to the next stage of processing.

The remaining ROIs from the HOG-SVM classifier were categorized into five classes: background, traffic sign, vehicle, pedestrian and traffic light. In the second stage we applied ResNet101 [37], which is a popular CNN that has been already trained with more than a million images from the ImageNet

Figure 3.5: *Model A output examples.*

database [69]. Figure 3.5 illustrates sample results obtained with this model. However, during our empirical trials, we noted the multi-scale HOG-SVM had difficulty localizing vehicles occupying a large part of the image (Figure 3.6 illustrates this problem). Hence, we also used a Faster R-CNN model to detect vehicles.

Model B

We trained a Faster R-CNN model on our dataset to localize vehicles. During our empirical trials, we observed that Model B is able to correctly detect vehicles that occupy a large image area, or that are very close to the instrumented vehicle. Conversely, based on our empirical trials as well as our survey of the literature [46], [40] and [43], we found that Faster R-CNN struggled with objects that are low in resolution or small in size. As a result, to detect objects of different sizes, we integrated the results from both Models A and B to take advantage from both. The hypotheses generated in this stage are

Figure 3.6: *Examples of model A missing large vehicle objects.*

directly transferred to an integration stage where detection results are merged. Figure 3.7 displays vehicle detections obtained with Model B.

3.3.4 Data Augmentation

In addition to collecting over 10,000 sample object images, to further enrich our training dataset, we employed a data augmentation technique and a boosting algorithm. Through data augmentation, we made our dataset greater by adding the translated, rotated, scaled, and sheared versions of our original samples resulting in increased performance at the detection stage. To boost the performance of our models, we employed an advanced learning method known as Hard Examples Mining (HEM). HEM refers to examples that are mislabeled by the current version of the model. We trained the SVM, Resnet, and Faster R-CNN models in an iterated procedure on a portion of the training data, and at each iteration, the detector models were applied to a number of unseen images from the training data. Then, we added manually the corrected

Figure 3.7: *Model B output examples.*

mislabeled objects in the preparation of training set for the next iteration. We finally provided the models with additional key samples which made them more robust.

3.3.5 Integrating Detection Results

After completing the detection stage on test images, in order to improve the detection performance, we eliminated redundant detections and merged the remaining ones into a set of integrated results. For this, we used a method that is based on Non Maximum Suppression (NMS) [108], [44]. When multiple bounding boxes overlap, NMS retains the highest-scored bounding box and eliminates any other whose overlap ratio exceeds a preset threshold. We employed Pascal's overlap score [24] to find the overlap ratio a_0 between them. This ratio is obtained as:

$$a_0 = \frac{area(B_1 \cap B_2)}{area(B_1 \cup B_2)} \tag{3.3}$$

where B_1 and B_2 are two overlapping bounding boxes.

The NMS algorithm is not practical in all situations. For instance, consider a situation in which a vehicle is partially occluded by a pedestrian, and both of them are detected. If their overlap ratio is greater than the threshold, NMS wrongly eliminates the lower-scored object. To address this case, we integrated all bounding boxes in three steps. We considered a lower bound and a upper bound threshold for the overlap ratio. In the first step, we employ NMS to merge bounding boxes that belong to the same class. In this step, NMS eliminates the lower-scored bounding boxes whose overlap ratios are between the lower bound and the upper bound thresholds. In the second step, if bounding boxes belong to the same class and their overlap ratio is greater than the upper bound threshold, they are merged into a larger bounding box. In the last step, all remaining bounding boxes are merged without employing NMS to generate the final set of detected hypotheses.

3.3.6 Object Recognition Stage

The output of the detection stage is a list of candidate objects that have been labeled with the class they belong to (traffic sign, vehicle, traffic light, and pedestrian). Except for pedestrian objects, the remaining objects from the list are considered for further analysis at this stage. We separately trained three independent models on traffic signs, vehicles, and traffic lights by using ResNet101 for recognizing the remaining objects. After feeding the candidate object (hypothesis) into its corresponding model, the classifier decides whether the object in the list is either a rejected object or a recognized object and, in this case, the classifier responds with the appropriate class name. More precisely, the traffic light recognizer is able to classify traffic light hypotheses

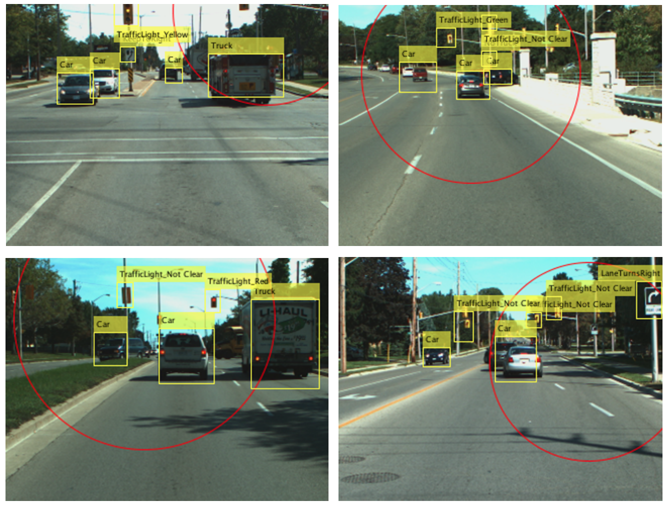


Figure 3.8: *Output samples from the proposed framework superimposed on the attentional visual field of the driver*

into five classes, the vehicle recognizer is able to classify vehicle hypotheses into four classes, while the traffic sign recognizer classifies traffic sign hypotheses into twenty classes. Fig 3.8 shows a sample of results from the proposed framework for four classes of traffic objects.

3.4 Experimental Results

We employed the driving sequences captured with the RoadLAB experimental vehicle [3] and our dataset as described in Section 3.3.1. The proposed method was used to detect and recognize traffic objects inside and outside of the attentional visual area of the driver. Based on the attentional visual field, we can infer whether the driver is likely to have seen the object or not, namely, when the existing object falls inside the driver gaze area. In the following we first provide the parameters which have been used in our experiments and then we report on our experimental results for the proposed detection and recognition stages in detail.

3.4.1 Parameters

To obtain fine-tuned parameters for each classifier model, we used cross-validation experiments on our training dataset. We divided the training data into a basic training set and a validation set. Then, the basic training set was used to train the classifier and subsequently, the validation set was used to evaluate the model. By exploring various ranges for the tuning parameters, we selected the parameter settings that resulted in maximum validation accuracy. Next, the classifier was re-trained on the complete training set using the fine-tuned parameters. Our model achieved 95.1% and 94.2% performance for training and validation sets respectively. Finally, we tested the models on the pre-separated unseen data that consists of a set of randomly selected samples.

We applied a threshold to the score values that each SVM model provided, and ROIs were considered for post-processing only if their SVM score was higher than the threshold value. These score values ranged from 0 (definitely negative) to 1 (definitely positive). We selected the threshold that allowed a maximum of true positives. While some false positives passed this stage, they could mostly be eliminated in the following stage of processing.

Threshold values of 0.50, 0.40, 0.40, and 0.60 were applied to the SVM models for detection of traffic signs, pedestrians, traffic lights, and vehicles respectively. These values provided the best results. We also utilized different augmentation methods to improve the performance of our models. Table 3.1 lists the methods we have used to augment our data.

Table 3.1: Description of data augmentation

Method	Description	Range
Translate	Each image is translated in the horizontal and vertical direction by a distance, in pixels	$(-10, 10)$
Rotate	Each image is rotated by an amount, in degrees	$(-15, 15)$
Scale	Each image is scaled in the horizontal and vertical direction by a factor	$(0.5, 1.5)$
Shear	Each image is sheared along the horizontal or vertical axis by a factor	$(-30, 30)$

3.4.2 Results for the Object Detection Stage

In the following Subsections, we discuss the results we obtained for the object detection stage in detail.

Assessing the Accuracy of the Trained ResNet101 CNN Model

As described in Section 3.3.3, after localizing ROIs by way of multi-scale HOG-SVM, a ResNet101 CNN was trained and used on our dataset to verify and categorize ROIs into our five classes of traffic objects. We computed the confusion matrix from the ResNet101 model on the test data (See Figure 3.9). The model classifies the test data correctly in 94.1% of cases. Notably, 10% of vehicles have been incorrectly classified as background by ResNet101. As a result, we employed a Faster R-CNN-based model to detect vehicles besides Model A.

Assessing the Accuracy the Object Detection Stage

To verify the accuracy of the object detection stage, we report the Detection Rate (DR) and the number of False Positives Per Frame (FPPF), defined as follows:

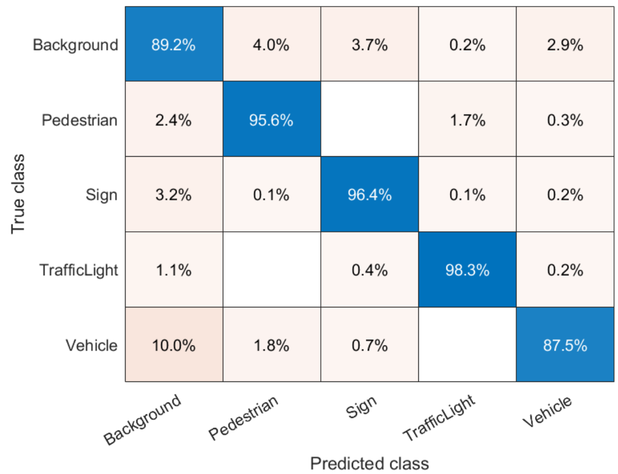


Figure 3.9: *Confusion matrix from trained ResNet101 for labelling of traffic object classes.*

$$\text{DR} = \frac{\text{TP}}{\text{TP+FN}} \tag{3.4}$$

$$\text{FPPF} = \frac{\text{FP}}{\text{F}} \tag{3.5}$$

where TP is the number of correctly detected objects, FN is the number of objects that are wrongly not detected, FP is the number of incorrectly detected objects, and F is the total number of frames.

Moreover, Table 3.2 includes F1-scores for different traffic objects. As can be seen, our model achieved 0.91, 0.90 and 0.06 for DR, F1-score and FPPF respectively. Previously, Zabihi *et al.* [108] detected traffic signs only from the RoadLAB dataset and reported 0.84 for DR and 0.04 for FPPF (last row of Table 3.2). Their model was based on traditional machine learning methods. They employed a linear SVM as a classifier and a HOG as traffic sign features for the detection stage. Our model for traffic sign detection, when compared with the work from Zabihi *et al.* [108], has reached 0.07 more accuracy for DR and shows an increase in FPPF of 0.02. Recent studies have compared several recent object detection models including Faster R-CNN [78], Fast R-CNN [29], YOLO [75], and YOLOv3 [77]. Faster R-CNN has a better performance than R-CNN and Fast R-CNN. However, as mentioned, Faster R-CNN struggles with objects that are small in size. According to [75] and [77], YOLO struggles with small objects as well and on the other hand, YOLOv3 struggles with larger size objects. Using our framework, we were able to detect objects of different sizes. Figure 3.10 illustrates the performance of our detector using a Receiver Operating Characteristics (ROC) curve, comparing the True Positive Rate (TPR) to the False Positive Rate (FPR). In figure 3.10, *Class1*, *Class2*, *Class3*, and *Class4* represent object classes for pedestrians, traffic signs, traffic lights and vehicles, respectively.

Table 3.2: Description of detection results

Description	**DR**	**FPPF**	**F1-score**
Traffic lights	0.93	0.03	0.91
Pedestrians	0.88	0.11	0.87
Traffic signs	0.91	0.06	0.89
Vehicles	0.92	0.04	0.94
Object detection stage, 4 object classes	0.91	0.06	0.90
Previous work [108] for traffic signs	0.84	0.04	-

3.4.3 Trustworthiness Quantification

It is beneficial in any artificial intelligence-based model to know whether the probabilistic description is reliable. Recently, network-level trust quantification has attracted increasing interest from researchers such as [13], [101], [41], where the authors attempted to quantify the overall trustworthiness of deep

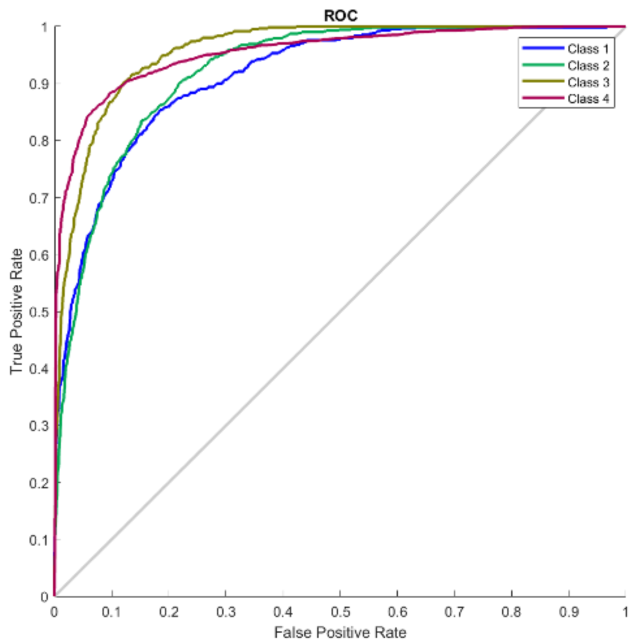


Figure 3.10: *ROC curve obtained from experiments.*

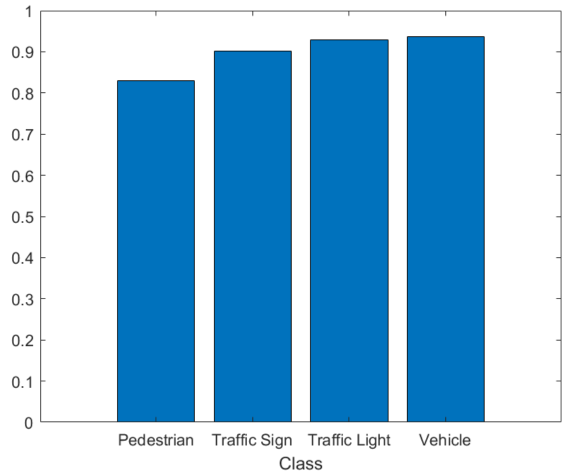


Figure 3.11: *Trustworthiness quantification.*

neural networks. To analyze the trust of the network, they use the concept of trust spectrum to investigate the overall trust across both correctly and incorrectly answered questions [102]. The trust spectrum provides valuable information when trust can break down. The trust spectrum in Figure 3.11 illustrates the overall trust for the four classes, including pedestrian, traffic sign, traffic light, and vehicle. As can be seen, the vehicle class achieved the highest trust while pedestrian class obtained the lowest reliability.

3.4.4 Results for Object Recognition Stage

The object recognition stage is applied to the output of the object detection stage to recognize hypotheses and to provide a classification result. We trained three separate ResNet101 models for classes corresponding to traffic signs, traffic lights, and vehicles using our training dataset. To verify the accuracy of the object recognition stage, we computed the confusion matrix for each class, as displayed in Figures 3.12, 3.13 and 3.14.

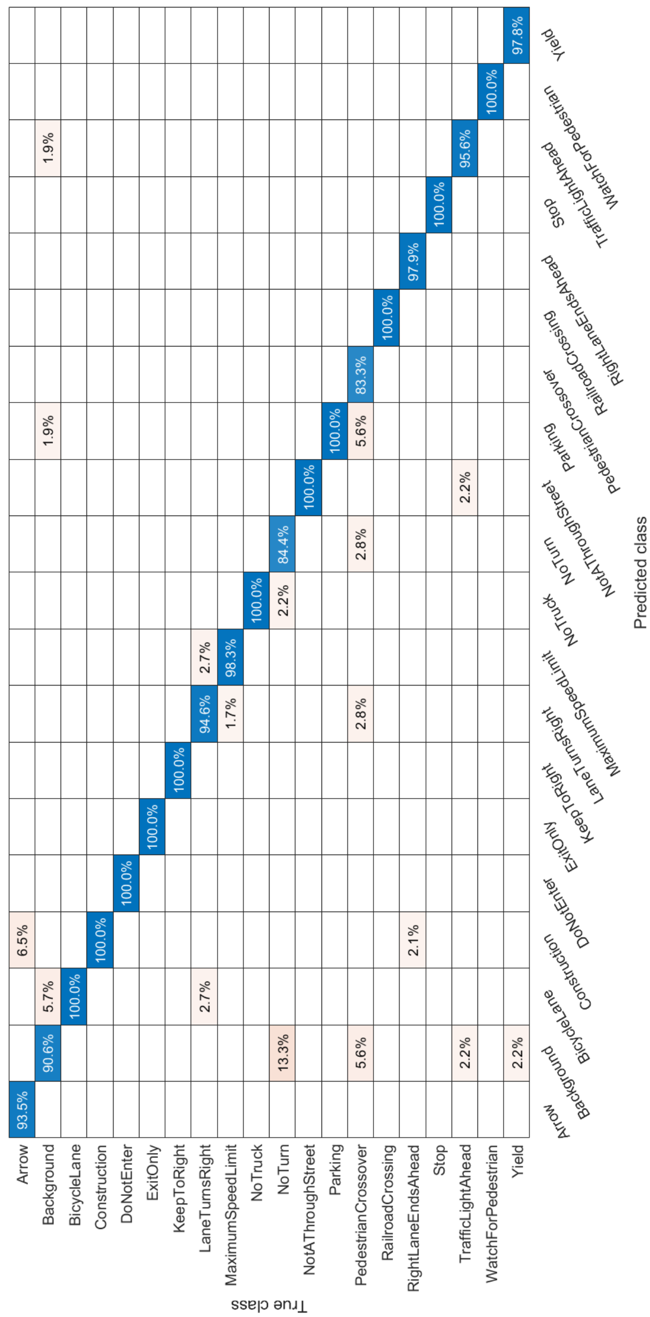


Figure 3.12: *Confusion matrix from trained ResNet101 for traffic sign recognition.*

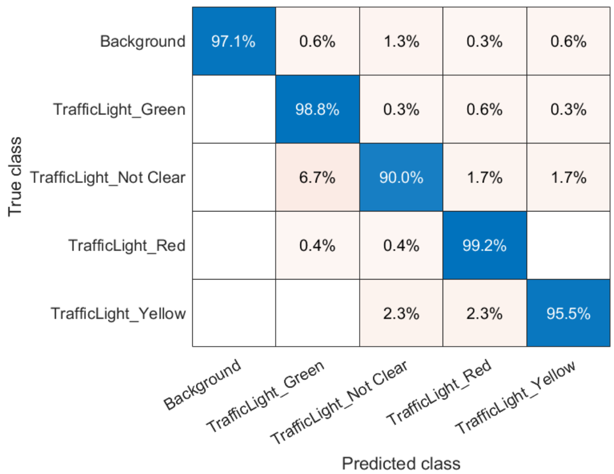


Figure 3.13: *Confusion matrix from trained ResNet101 for traffic light recognition.*

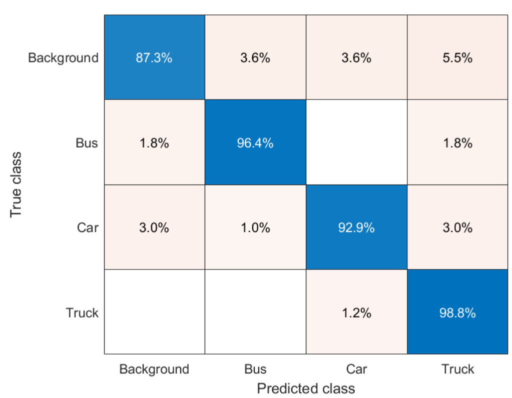


Figure 3.14: *Confusion matrix from trained ResNet101 for vehicle recognition.*

Results for traffic sign recognition (Fig 3.12) show that the model reached 96.1% accuracy with our Canadian traffic sign dataset. The largest values along the main diagonal indicate that the majority of the test sign images were classified correctly. The lowest correct response of 83.3% was obtained for the class *PedestrianCrossover*.

Fig 3.13 illustrates the confusion matrix for traffic light recognition. The results show that the model has reached 96.2% of overall correct classification. As can be seen, the lowest degree of correctly categorized classes belongs to class *NotClear* while classes *Green* and *Red* obtained 98.8% and 99.2% respectively.

The results shown in Figure 3.14 indicate that the vehicle recognizer model achieved 94.8% of overall correct classification. This confusion matrix shows that this model is able to discriminate vehicle objects (i.e. vehicle, bus, and truck) with less than 3% of mislabeling error. The *background* class achieved the least accuracy with 87.3%.

3.5 Conclusion

We conducted a literature review of detection and recognition approaches for four important classes of traffic objects including traffic signs, vehicles, pedestrians and traffic lights. Generally, the availability of suitable and adequate training data is a vital element in the learning process in order to achieve a discriminative model. In this work, we collected over 10,000 object sample images from sequences belonging to the RoadLAB initiative [3]. We also enriched our training data using augmentation and a HEM strategy. We localized the attentional visual area of the driver onto the imaging plane of the forward stereoscopic system, and a framework for the detection and recognition of traffic objects located inside and outside the attentional visual field of drivers was devised. This information helps an ADAS to infer the object seen by the driver when the existing object has fallen inside the driver gaze area. We considered 3, 4, and 19 different classes for vehicles, traffic lights, and traffic signs respectively. The object detection stage was built from a combination of both traditional and deep learning-based models to detect objects at various scales. Finally, in the recognition stage, by means of trained ResNet101 networks, our framework achieved 96.1%, 96.2% and 94.8% of correct classification for traffic signs, traffic lights, and vehicles respectively.

Bibliography

[1] S. Ahmed, M.N. Huda, S. Rajbhandari, C. Saha, M. Elshaw, and S. Kanarachos. “Pedestrian and Cyclist Detection and Intent Estimation for Autonomous Vehicles: A Survey”. In: *Applied Sciences* 9.11 (2019), p. 2335.

[2] T. Almeida, N. Vasconcelos, A. Benicasa, and H. Macedo. “Fuzzy model applied to the recognition of traffic lights signals”. In: *2016 8th Euro American Conference on Telematics and Information Systems (EATIS)*. IEEE. 2016, pp. 1–4.

[3] Steven S Beauchemin, Michael A Bauer, Taha Kowsari, and Ji Cho. “Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality”. In: *IEEE Transactions on Instrumentation and Measurement* 61.2 (2011), pp. 391–401.

[4] K. Behrendt, L. Novak, and R. Botros. “A deep learning approach to traffic lights: Detection, tracking, and classification”. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2017, pp. 1370–1377.

[5] M. Bertozzi, A. Broggi, and S. Castelluccio. “A real-time oriented system for vehicle detection”. In: *Journal of Systems Architecture* 43.1-5 (1997), pp. 317–325.

[6] A. Broggi, M. Bertozzi, A. Fascioli, C Bianco, and A. Piazzi. “Visual perception of obstacles and vehicles for platooning”. In: *IEEE Transactions on Intelligent Transportation Systems* 1.3 (2000), pp. 164–176.

[7] A. Brunetti, D. Buongiorno, G.F. Trotta, and V. Bevilacqua. “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey”. In: *Neurocomputing* 300 (2018), pp. 17–33.

[8] T. Bucher, C. Curio, J. Edelbrunner, C. Igel, D. Kastrup, I. Leefken, G. Lorenz, A. Steinhage, and W. von Seelen. “Image processing and behavior planning for intelligent vehicles”. In: *IEEE Transactions on Industrial electronics* 50.1 (2003), pp. 62–75.

[9] Z. Cai, Y. Li, and M. Gu. “Real-time recognition system of traffic light in urban environment”. In: *2012 IEEE Symposium on Computational Intelligence for Security and Defence Applications*. IEEE. 2012, pp. 1–6.

[10] D. Chen, Y. Lin, and Y. Peng. “Nighttime brake-light detection by Nakagami imaging”. In: *IEEE Transactions on Intelligent Transportation Systems* 13.4 (2012), pp. 1627–1637.

[11] Long Chen, Qingquan Li, Ming Li, and Qingzhou Mao. “Traffic sign detection and recognition for intelligent vehicle”. In: *IEEE Intelligent Vehicles Symposium (IV)*. 2011, pp. 908–913.

[12] Q. Chen, Z. Shi, and Z. Zou. “Robust and real-time traffic light recognition based on hierarchical vision architecture”. In: *2014 7th International Congress on Image and Signal Processing*. IEEE. 2014, pp. 114–119.

[13] Mingxi Cheng, Shahin Nazarian, and Paul Bogdan. "There is hope after all: Quantifying opinion and trustworthiness in neural networks". In: *Frontiers in Artificial Intelligence* 3 (2020), p. 54.

[14] Dan CireAan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. "Multi-column deep neural network for traffic sign classification". In: *Neural networks* 32 (2012), pp. 333–338.

[15] Dan Cireşan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. "A committee of neural networks for traffic sign classification". In: *The 2011 international joint conference on neural networks*. IEEE. 2011, pp. 1918–1921.

[16] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. "R-fcn: Object detection via region-based fully convolutional networks". In: *arXiv preprint arXiv:1605.06409* (2016).

[17] Navneet Dalal and Bill Triggs. "Histograms of oriented gradients for human detection". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 1. 2005, pp. 886–893.

[18] A. De la Escalera, J. Armingol, and M. Mata. "Traffic sign recognition and analysis for intelligent vehicles". In: *Image and vision computing* 21.3 (2003), pp. 247–258.

[19] A. De La Escalera, L. Moreno, M. Salichs, and J.M. Armingol. "Road traffic sign detection and classification". In: *IEEE transactions on industrial electronics* 44.6 (1997), pp. 848–859.

[20] M. Diaz, P. Cerri, G. Pirlo, M.A. Ferrer, and D. Impedovo. "A survey on traffic light detection". In: *International Conference on Image Analysis and Processing*. Springer. 2015, pp. 201–208.

[21] P. Dollár, R. Appel, S. Belongie, and P. Perona. "Fast feature pyramids for object detection". In: *IEEE transactions on pattern analysis and machine intelligence* 36.8 (2014), pp. 1532–1545.

[22] Jan Dürre, Dario Paradzik, and Holger Blume. "A HOG-based real-time and multi-scale pedestrian detector demonstration system on FPGA". In: *Proceedings of the 2018 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. 2018, pp. 163–172.

[23] J.E. Espinosa, S.A. Velastin, and J.W. Branch. "Vehicle detection using alex net and faster R-CNN deep learning models: a comparative study". In: *International Visual Informatics Conference*. Springer. 2017, pp. 3–15.

[24] M. Everingham, A. Eslami, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. "The pascal visual object classes challenge: A retrospective". In: *International journal of computer vision* 111.1 (2015), pp. 98–136.

[25] N. Fairfield and C. Urmson. "Traffic light mapping and detection". In: *2011 IEEE International Conference on Robotics and Automation*. IEEE. 2011, pp. 5421–5426.

[26] U. Franke and I. Kutzbach. "Fast stereo based object detection for stop&go traffic". In: *Proceedings of Conference on Intelligent Vehicles*. IEEE. 1996, pp. 339–344.

[27] C. Fu, W. Liu, A. Ranga, A. Tyagi, and A.C. Berg. "Dssd: Deconvolutional single shot detector". In: *arXiv preprint arXiv:1701.06659* (2017).

[28] Dariu M Gavrila. "Traffic sign recognition revisited". In: *Mustererkennung 1999*. Springer, 1999, pp. 86–93.

[29] R. Girshick. "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448.

[30] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.

[31] C. Goerick, D. Noll, and M. Werner. "Artificial neural networks in real-time car detection and tracking applications". In: *Pattern Recognition Letters* 17.4 (1996), pp. 335–343.

[32] J. Gong, Y. Jiang, G. Xiong, C. Guan, G. Tao, and H. Chen. "The recognition and tracking of traffic lights based on color segmentation and camshift for intelligent vehicles". In: *2010 IEEE Intelligent Vehicles Symposium*. Ieee. 2010, pp. 431–435.

[33] J. Greenhalgh and M. Mirmehdi. "Traffic sign recognition using MSER and random forests". In: *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE. 2012, pp. 1935–1939.

[34] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, and W.V. Seelen. "An image processing system for driver assistance". In: *Image and Vision Computing* 18.5 (2000), pp. 367–376.

[35] M.A. Hannan, A. Hussain, S.A. Samad, and K.A. Ishak. "A unified robust algorithm for detection of human and non-human object in intelligent safety application". In: *International Journal of Computer and Information Engineering* 2.11 (2008).

[36] K. He, G. Gkioxari, P. Dollár, and R. Girshick. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[37] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[38] K. He, X. Zhang, S. Ren, and J. Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 37.9 (2015), pp. 1904–1916.

[39] A. Hechri and A. Mtibaa. "Automatic detection and recognition of road sign for driver assistance system". In: *2012 16th IEEE Mediterranean Electrotechnical Conference*. IEEE. 2012, pp. 888–891.

[40] T. Hoang Ngan Le, Y. Zheng, C. Zhu, K. Luu, and M. Savvides. "Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016, pp. 46–53.

[41] Andrew Hryniowski, Xiao Yu Wang, and Alexander Wong. "Where Does Trust Break Down? A Quantitative Trust Analysis of Deep Neural Networks via Trust Matrix and Conditional Trust Densities". In: *arXiv preprint arXiv:2009.14701* (2020).

[42] S. Hsu, C. Huang, and C. Chuang. "Vehicle detection using simplified fast R-CNN". In: *2018 International Workshop on Advanced Image Technology (IWAIT)*. IEEE. 2018, pp. 1–3.

[43] G.X. Hu, Z. Yang, L. Hu, L. Huang, and J.M. Han. “Small Object Detection with Multiscale Features”. In: *International Journal of Digital Multimedia Broadcasting* 2018 (2018).

[44] Q. Hu, S. Paisitkriangkrai, C. Shen, A. van den Hengel, and F. Porikli. “Fast detection of multiple objects in traffic scenes with a common detection framework”. In: *IEEE Transactions on Intelligent Transportation Systems* 17.4 (2015), pp. 1002–1014.

[45] M.B. Jensen, K. Nasrollahi, and T.B. Moeslund. “Evaluating state-of-the-art object detector on challenging traffic light data”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017, pp. 9–15.

[46] H. Ji, Z. Gao, T. Mei, and Y. Li. “Improved Faster R-CNN With Multiscale Feature Fusion and Homography Augmentation for Vehicle Detection in Remote Sensing Images”. In: *IEEE Geoscience and Remote Sensing Letters* (2019).

[47] V. John, K. Yoneda, Z. Liu, and S. Mita. “Saliency map generation by the convolutional neural network for real-time traffic light detection using template matching”. In: *IEEE transactions on computational imaging* 1.3 (2015), pp. 159–173.

[48] V. John, K. Yoneda, B. Qi, Z. Liu, and S. Mita. “Traffic light recognition in varying illumination using deep learning and saliency map”. In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2014, pp. 2286–2291.

[49] N. Khairdoost, S.A. Monadjemi, and K. Jamshidi. "Front and rear vehicle detection using hypothesis generation and verification". In: *Signal & Image Processing* 4.4 (2013), p. 31.

[50] M. Kobayashi, M. Baba, K. Ohtani, and L. Li. "A method for traffic sign detection and recognition based on genetic algorithm". In: *2015 IEEE/SICE International Symposium on System Integration (SII)*. IEEE. 2015, pp. 455–460.

[51] T. Kowsari, S.S. Beauchemin, M.A. Bauer, D. Laurendeau, and N. Teasdale. "Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems". In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 1245–1250.

[52] A.D. Kumar. "Novel deep learning model for traffic sign detection using capsule networks". In: *arXiv preprint arXiv:1805.04424* (2018).

[53] R. Kumar, A. Kumar, and A. Bhavsar. "Bird region detection in images with multi-scale HOG features and SVM scoring". In: *Proceedings of 2nd International Conference on Computer Vision & Image Processing*. Springer. 2018, pp. 353–364.

[54] W. Kuo and C. Lin. "Two-stage road sign detection and recognition". In: *2007 IEEE international conference on multimedia and expo*. IEEE. 2007, pp. 1427–1430.

[55] Y. Lai, N. Wang, Y. Yang, and L. Lin. "Traffic Signs Recognition and Classification based on Deep Feature Learning." In: *ICPRAM*. 2018, pp. 622–629.

[56] W. Lan, J. Dang, Y. Wang, and S. Wang. “Pedestrian Detection Based on YOLO Network Model”. In: *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE. 2018, pp. 1547–1551.

[57] G. Lee and B.K. Park. “Traffic light recognition using deep neural networks”. In: *2017 IEEE international conference on consumer electronics (ICCE)*. IEEE. 2017, pp. 277–278.

[58] K. Lim, Y. Hong, Y. Choi, and H. Byun. “Real-time traffic sign recognition based on a general purpose GPU and deep-learning”. In: *PLoS one* 12.3 (2017), e0173317.

[59] B. Lin, Y. Chan, L. Fu, P. Hsiao, L. Chuang, S. Huang, and M. Lo. “Integrating appearance and edge features for sedan vehicle detection in the blind-spot area”. In: *IEEE Transactions on Intelligent Transportation Systems* 13.2 (2012), pp. 737–747.

[60] C. Lin and M. Wang. “Road sign recognition with fuzzy adaptive pre-processing models”. In: *Sensors* 12.5 (2012), pp. 6415–6433.

[61] H. Lin. *SVM*. `http://www.work.caltech.edu/~htlin/program/libsvm/`.

[62] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A.C. Berg. “Ssd: Single shot multibox detector”. In: *European conference on computer vision*. Springer. 2016, pp. 21–37.

[63] Y. Liu, Y. Lu, Q. Shi, and J. Ding. “Optical flow based urban road vehicle tracking”. In: *2013 Ninth International Conference on Computational Intelligence and Security*. IEEE. 2013, pp. 391–395.

[64] Markus Mathias, Radu Timofte, Rodrigo Benenson, and Luc Van Gool. “Traffic sign recognition—How far are we from the solution?” In: *The International Joint Conference on Neural Networks (IJCNN)*. 2013, pp. 1–8.

[65] Q. Ming and K. Jo. “Vehicle detection using tail light segmentation”. In: *Proceedings of 2011 6th International Forum on Strategic Technology*. Vol. 2. IEEE. 2011, pp. 729–732.

[66] K. Mu, F. Hui, X. Zhao, and C. Prehofer. “Multiscale edge fusion for vehicle detection based on difference of Gaussian”. In: *Optik* 127.11 (2016), pp. 4794–4798.

[67] M. Najibi, M. Rastegari, and L.S. Davis. “G-cnn: an iterative grid based object detector”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2369–2377.

[68] D. Nandi, S. Saif, P. Prottoy, K. Zubair, and S. Shubho. “Traffic sign detection based on color segmentation of obscure image candidates: a comprehensive study”. In: *International Journal of Modern Education and Computer Science* 10.6 (2018), p. 35.

[69] Image Net. *ImageNet*. `http://www.image-net.org`.

[70] E. Ohn-Bar and M. Trivedi. “Fast and robust object detection using visual subcategories”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2014, pp. 179–184.

[71] M. Omachi and S. Omachi. “Traffic light detection with color and edge information”. In: *2009 2nd IEEE International Conference on Computer Science and Information Technology*. IEEE. 2009, pp. 284–287.

[72] Michael Oren, Constantine Papageorgiou, Pawan Sinha, Edgar Osuna, and Tomaso Poggio. "Pedestrian detection using wavelet templates". In: *Proceedings of IEEE computer society Conference on computer vision and pattern recognition*. IEEE. 1997, pp. 193–199.

[73] A. Pon, O. Adrienko, A. Harakeh, and S.L. Waslander. "A hierarchical deep architecture and mini-batch selection method for joint traffic sign and light detection". In: *2018 15th Conference on Computer and Robot Vision (CRV)*. IEEE. 2018, pp. 102–109.

[74] V. Prisacariu, R. Timofte, K. Zimmermann, I. Reid, and L. Van Gool. "Integrating object detection with 3D tracking towards a better driver assistance system". In: *2010 20th International Conference on Pattern Recognition*. IEEE. 2010, pp. 3344–3347.

[75] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[76] J. Redmon and A. Farhadi. "YOLO9000: better, faster, stronger". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7263–7271.

[77] Joseph Redmon and Ali Farhadi. "Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767* (2018).

[78] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *arXiv preprint arXiv:1506.01497* (2015).

[79] Y. Saadna and A. Behloul. “An overview of traffic sign detection and classification methods”. In: *International Journal of Multimedia Information Retrieval* 6.3 (2017), pp. 193–210.

[80] P. Saxena, N. Gupta, S. Laskar, and P. Borah. “A study on automatic detection and recognition techniques for road signs”. In: *Int. J. Comput. Eng. Res* 5.12 (2015), pp. 24–28.

[81] P. Sermanet and Y. LeCun. “Traffic sign recognition with multi-scale Convolutional Networks.” In: *IJCNN*. 2011, pp. 2809–2813.

[82] Zhiqiang Shen, Zhuang Liu, Jianguo Li, Yu-Gang Jiang, Yurong Chen, and Xiangyang Xue. “Dsod: Learning deeply supervised object detectors from scratch”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 1919–1927.

[83] Z. Shi, Z. Zou, and C. Zhang. “Real-time traffic light detection with adaptive background suppression filter”. In: *IEEE Transactions on Intelligent Transportation Systems* 17.3 (2015), pp. 690–700.

[84] Mohsen Shirpour, Steven S Beauchemin, and Michael A Bauer. “A probabilistic model for visual driver gaze approximation from head pose estimation”. In: *2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS)*. IEEE. 2020, pp. 1–6.

[85] G. Song, K. Lee, and J. Lee. “Vehicle detection by edge-based candidate generation and appearance-based classification”. In: *2008 IEEE Intelligent Vehicles Symposium*. IEEE. 2008, pp. 428–433.

[86] L. Suhao, L. Jinzhao, L. Guoquan, B. Tong, W. Huiqian, and P. Yu. “Vehicle type detection based on deep learning in traffic scene”. In: *Procedia computer science* 131 (2018), pp. 564–572.

[87] Z. Sun, G. Bebis, and R. Miller. “On-road vehicle detection: A review”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.5 (2006), pp. 694–711.

[88] Kenji Takagi, Haruki Kawanaka, Md. Shoaib Bhuiyan, and Koji Oguri. “Estimation of a three-dimensional gaze point and the gaze target from the road images”. In: *Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on*. IEEE. 2011, pp. 526–531.

[89] B. Tian, B.T. Morris, M. Tang, Y. Liu, Y. Yao, C. Gou, D. Shen, and S. Tang. “Hierarchical and Networked Vehicle Surveillance in ITS: A Survey”. In: *IEEE Transactions on Intelligent Transportation Systems* 18.1 (2017), p. 25.

[90] M. Tiwari and R. Singhai. “A review of detection and tracking of object from image and video sequences”. In: *Int. J. Comput. Intell. Res.* 13.5 (2017), pp. 745–765.

[91] J. Torresen, J.W. Bakke, and L. Sekanina. “Efficient recognition of speed limit signs”. In: *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749).* IEEE. 2004, pp. 652–656.

[92] Q. Truong and B. Lee. “Vehicle detection algorithm using hypothesis generation and verification”. In: *International Conference on Intelligent Computing*. Springer. 2009, pp. 534–543.

[93] C. Tzomakas and W. von Seelen. “Vehicle detection in traffic scenes using shadows”. In: *Ir-Ini, Institut fur Nueroinformatik, Ruhr-Universitat.* Citeseer. 1998.

[94] P. Viola, M.J. Jones, and D. Snow. “Detecting pedestrians using patterns of motion and appearance”. In: *International Journal of Computer Vision* 63.2 (2005), pp. 153–161.

[95] Paul Viola and Michael J Jones. “Robust real-time face detection”. In: *International journal of computer vision* 57.2 (2004), pp. 137–154.

[96] S.B. Wali, M.A. Abdullah, M. Hannan, A. Hussain, S. A Samad, P.J. Ker, and M. Mansor. “Vision-based traffic sign detection and recognition systems: current trends and challenges”. In: *Sensors* 19.9 (2019), p. 2093.

[97] X. Wang, T. Han, and S. Yan. “An HOG-LBP human detector with partial occlusion handling”. In: *2009 IEEE 12th international conference on computer vision*. IEEE. 2009, pp. 32–39.

[98] X. Wang, M. Yang, S. Zhu, and Y. Lin. “Regionlets for generic object detection”. In: *Proceedings of the IEEE international conference on computer vision*. 2013, pp. 17–24.

[99] M. Weber, P. Wolf, and J.M. Zöllner. “DeepTLR: A single deep convolutional network for detection and classification of traffic lights”. In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2016, pp. 342–348.

[100] X. Wen, L. Shao, W. Fang, and Y. Xue. “Efficient feature selection and classification for vehicle detection”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 25.3 (2014), pp. 508–517.

[101] Alexander Wong, Andrew Hryniowski, and Xiao Yu Wang. “Insights into Fairness through Trust: Multi-scale Trust Quantification for Financial Deep Learning”. In: *arXiv preprint arXiv:2011.01961* (2020).

[102] Alexander Wong, Xiao Yu Wang, and Andrew Hryniowski. "How Much Can We Really Trust You? Towards Simple, Interpretable Trust Quantification Metrics for Deep Neural Networks". In: *arXiv preprint arXiv:2009.05835* (2020).

[103] M. Xing, M. Chunyang, W. Yan, W. Xiaolong, and C. Xuetao. "Traffic sign detection and recognition using color standardization and Zernike moments". In: *2016 Chinese Control and Decision Conference (CCDC)*. IEEE. 2016, pp. 5195–5198.

[104] T. Xiong and C. Debrunner. "Stochastic car tracking with line-and color-based features". In: *IEEE Transactions on Intelligent Transportation Systems* 5.4 (2004), pp. 324–328.

[105] G. Yan, M. Yu, Y. Yu, and L. Fan. "Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification". In: *Optik* 127.19 (2016), pp. 7941–7951.

[106] S. Yin, P. Ouyang, L. Liu, Y. Guo, and S. Wei. "Fast traffic sign recognition with a rotation invariant binary pattern based feature". In: *Sensors* 15.1 (2015), pp. 2161–2180.

[107] Donggeun Yoo, Sunggyun Park, Joon-Young Lee, Anthony S Paek, and In So Kweon. "Attentionnet: Aggregating weak directions for accurate object detection". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 2659–2667.

[108] S.J. Zabihi, S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. "Detection and recognition of traffic signs inside the attentional visual field of drivers". In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2017, pp. 583–588.

[109] SM Zabihi, Steven S Beauchemin, EAM De Medeiros, and Michael A Bauer. “Frame-rate vehicle detection within the attentional visual area of drivers”. In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 146–150.

[110] H. Zhang, B. Wang, Z. Zheng, and Y. Dai. “A novel detection and recognition system for Chinese traffic signs”. In: *Proceedings of the 32nd Chinese Control Conference*. IEEE. 2013, pp. 8102–8107.

[111] L. Zhang, L. Lin, X. Liang, and K. He. “Is faster r-cnn doing well for pedestrian detection?” In: *European conference on computer vision*. Springer. 2016, pp. 443–457.

[112] Xiaotong Zhao, Wei Li, Yifang Zhang, T Aaron Gulliver, Shuo Chang, and Zhiyong Feng. “A faster RCNN-based pedestrian detection system”. In: *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*. IEEE. 2016, pp. 1–5.

[113] Z. Zhao, P. Zheng, S. Xu, and X. Wu. “Object detection with deep learning: A review”. In: *IEEE transactions on neural networks and learning systems* (2019).

[114] W. Zhiqiang and L. Jun. “A review of object detection based on convolutional neural network”. In: *2017 36th Chinese Control Conference (CCC)*. IEEE. 2017, pp. 11104–11109.

[115] Y. Zhou, Z. Chen, and X. Huang. “A system-on-chip FPGA design for real-time traffic signal recognition system”. In: *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE. 2016, pp. 1778–1781.

[116] Y. Zhou, H. Nejati, T. Do, N. Cheung, and L. Cheah. "Image-based vehicle analysis using deep neural network: A systematic study". In: *2016 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE. 2016, pp. 276–280.

[117] Y. Zhu, C. Zhang, D. Zhou, X. Wang, X. Bai, and W. Liu. "Traffic sign detection and recognition using fully convolutional network guided proposals". In: *Neurocomputing* 214 (2016), pp. 758–766.

[118] Z. Zou, Z. Shi, Y. Guo, and J. Ye. "Object Detection in 20 Years: A Survey". In: *arXiv preprint arXiv:1905.05055* (2019).

Chapter 4

Road Lane Detection and Classification

This Chapter is a reformatted version of the following article:

N. Khairdoost, S.S. Beauchemin, M.A. Bauer, *Road Lane Detection and Classification in Urban and Suburban Areas based on CNNs. in 16th International Conference on Computer Vision Theory and Applications (VISAPP)*, Vienna, Austria, 2021.

Road lane detection systems play a crucial role in the context of Advanced Driver Assistance Systems (ADASs) and autonomous driving. Such systems can lessen road accidents and increase driving safety by alerting the driver in risky traffic situations. Additionally, the detection of ego lanes with their left and right boundaries along with the recognition of their types is of great importance as they provide contextual information. Lane detection is a challenging problem since road conditions and illumination vary while driving. In this contribution, we investigate the use of a CNN-based regression method for detecting lane boundaries. After the lane detection stage, following a projective transformation, the classification stage is performed with a ResNet101 network to verify the detected lanes or a possible road boundary. We applied our framework to real images collected during drives in an urban area with the RoadLAB instrumented vehicle. Our experimental results show that our approach achieved promising results in the detection stage with an accuracy of 94.52% in the lane classification stage.

4.1 Introduction

Nowadays, almost every new vehicle features some type of Advanced Driving Assistance System (ADAS), ranging from adaptive cruise control, blind-spot detection, collision avoidance, traffic sign detection, overtaking assistance, to parking assistance. ADASs generally increase safety and reduce driver workload. Lane detection constitutes one of the fundamental functions found in autonomous driving systems and ADASs. Lane boundaries provide the information required for estimating the lateral position of a vehicle on the road, enabling systems such as lane departure warning, overtaking assistance, intelligent cruise control, and trajectory planning.

Lane detection approaches are categorized into two groups: classical and deep learning methods. The traditional lane detection methods usually employ a number of computer vision and image processing techniques to extract specialized features and to identify the location of lane segments. Subsequently, post-processing techniques remove false detections and join sub-segments to obtain final road lane positions. In general, these traditional approaches suffer from performance issues when they encounter challenging illumination conditions and complex road scenes.

Recently, deep learning-based methods have been employed to provide reliable solutions to the lane detection problem. Methods based on CNNs fall into two categories, namely segmentation-based methods and Generative Adversarial Network based methods (GAN) [26]. Chougule *et al.* [6] proposed a regression-coordinate network based on CNN for lane detection in highway driving scenes in an end-to-end fashion. In this study, we followed their lane detection strategy in environments where there exists a greater variety of lane types as opposed to highways. We classify various types of lanes as they indicate traffic rules relevant for driving. Following the detection stage, we use a two-step algorithm to classify the lane boundaries into eight classes, considering road boundaries (no markings) as one particular type of lane.

The rest of this contribution is organized as follows: In Section 4.2, we review the related literature. Section 4.3 provides a summary of the datasets and the lane model. Results and evaluations are given in Section 4.4. Finally, we summarize our results in Section 4.5.

4.2 Literature Survey

In this section, we survey both traditional and deep learning methods for lane marking recognition and classification.

4.2.1 Traditional Approaches

Most traditional methods extract a combination of visual highly-specialized features using various elements such as color [4], [3], edges [13], ridge features [20], and template matching [5]. These primitive features can also be combined by way of Hough transforms [16], Kalman filters [21], [12], and particle filters [15]. Most of these methods are sensitive to illumination changes and road conditions and thus prone to fail.

4.2.2 Deep Learning-Based Approaches

There are mainly two groups of segmentation methods for lane marker detection: 1) Semantic Segmentation and 2) Instance Segmentation. In the first group, each pixel is classified by a binary label indicating whether it belongs to a lane or not. For instance, in [9], the authors presented a CNN-based framework that utilizes front-view and top-view image regions to detect lanes. Following this, they used a global optimization step to reach a combination of accurate lane lines. Lee *et al.* [14] proposed a Vanishing Point Guided Net (VPGNet) model that simultaneously performs lane detection and road marking recognition under different weather conditions. Their data was captured in a downtown area of Seoul, South Korea.

Conversely, Instance Segmentation approaches differentiate individual instances of each class in an image and identify separate parts of a line as one unit. Pan *et al.* [23] proposed the Spatial CNN (SCNN) to achieve effective information propagation in the spatial domain. This CNN-analogous scheme effectively retains the continuity of long and thin shapes such as road lanes, while its diffusion effects enable it to segment large objects. LaneNet [22] is a branched, instance segmentation architecture that produces a binary lane segmentation mask and pixel embeddings. These are used to cluster lane points. Subsequently, another neural network called H-net with a custom loss function is employed to parameterize lane instances before the lane fitting.

GANs have been used for lane detection. Liu *et al.* [17] presented a style-transfer-based data enhancement approach, which used GANs [8] to create images in low-light conditions that raise the environmental adaptability of the model. Their method does not require additional annotation nor extraneous inference overhead. Ghafoorian *et al.* [7] proposed an Embedding Loss GAN

(EL-GAN) framework for lane boundary segmentation. The discriminator receives the source data, a prediction map, and a ground truth label as inputs and is trained to minimize the difference between the training labels and embeddings of the predictions. In [11], a data augmentation method with GAN was proposed for oversampling minority anomalies in lane detection. The GAN network is employed to address the imbalance problem by synthesizing the anomalous data. It learns the distribution of the falsely detected lane by itself, without domain knowledge.

4.2.3 Approaches for Lane Type Classification

Different types of lane markings exist. Generally, a lane marking is categorized by its color, with dashed or solid, and single or double segments. In [10], a method is presented for road lane detection that discriminates dashed and solid lane markings. Their method outperformed conventional lane detection methods. Several other approaches such as [25], [24], and [1], recognize five lane marking types including Dashed, Dashed-Solid, Double Solid, Solid-Dashed, and Single Solid. In [25], a method that utilizes a two-layer classifier was proposed to classify these lane markings using a customized Region of Interest (ROI) and two derived features, namely; the contour number, and the contour angle. In [24], the authors presented a method to detect lane markers based on a linear parabolic model and geometric constraints. To classify lane markers into the aforementioned five classes, a three-level cascaded classifier consisting of four binary classifiers was developed. In [1], the ROI is divided into two subregions. To identify the lane types, a method based on the Seed Fill algorithm is applied to the location of the lanes. Lo *et al.* [19] proposed two techniques, Feature Size Selection and Degressive Dilation Block to extend an existing semantic segmentation network called EDANet [18] to discriminate the road from four types of lanes, including double solid yellow, single dashed yellow, single solid red, and single solid white.

There has been little previous work done on lane type classification and the majority of studies simply ignore the lane types. As mentioned, in [25], [24], and [1], researchers recognized five lane marking types and in the works [19] and [10], the authors recognized 4 and only 2 lane types, respectively. In contrast, we classify eight different types of lanes. Unlike previous work on classifying lane types, we specifically consider the road boundary as one type of lane when an actual lane marking does not exist. Also, we apply our method in both urban and suburban areas which have been also less studied in the literature where much of the previous work has focused only on highways.

4.3 Proposed Method

In this section, we present our approaches to the problem of lane marking recognition and classification, with their respective datasets extracted from the RoadLAB experiments.

4.3.1 Lane Detection Stage

Regression-Based Lane Detection Model

To identify the ego lane boundaries in the road image, a regression-based network is utilized that outputs two vectors representing the coordinate points of the left and right boundaries from the ego lane. Each coordinate vector consists of 14 coordinates (x, y) on the image plane indicating sampled positions for the ego lane boundary. To construct this model, a pre-trained AlexNet architecture is utilized. First, the last two fully connected layers are removed from the network and then four-level cascaded layers are added to the first six layers of AlexNet to complete the lane detection model. These four-level cascaded layers contain two branches of two back-to-back fully connected layers, a concatenation layer and a regression layer, as shown in Figure 4.1. This branched architecture minimizes misclassifications of the detected lane points [6]. Moreover, this architecture is capable of detecting the road boundary as an assumptive ego lane left/right boundary when there is no actual lane marking.

Our Dataset for Lane Detection

In this section, we introduce our lane detection dataset extracted from the driving sequences, captured with the RoadLAB instrumented vehicle [2], (see Figure 4.2). Our experimental vehicle was used to collect driving sequences from 16 drivers on a pre-determined 28.5km route within the city of London, Ontario, Canada. (see Figure 4.3). Data frames were collected at a rate of 30Hz with a resolution of 320×240. We used 12 driving sequences, as described in Table 4.1, to derive our dataset containing 5782 images along with their corresponding lane annotations. Figure 4.4 illustrates examples from our derived dataset.

An essential element of any deep learning-based system is the availability of large numbers of sample images. Data augmentation is a commonly used strategy to significantly expand an existing dataset by generating unique samples through transformations of images in the dataset. The exploitation of data augmentation strategy reduces overfitting from the network. We employed data augmentation techniques to enrich the dataset, resulting in an improved performance at the lane detection stage.

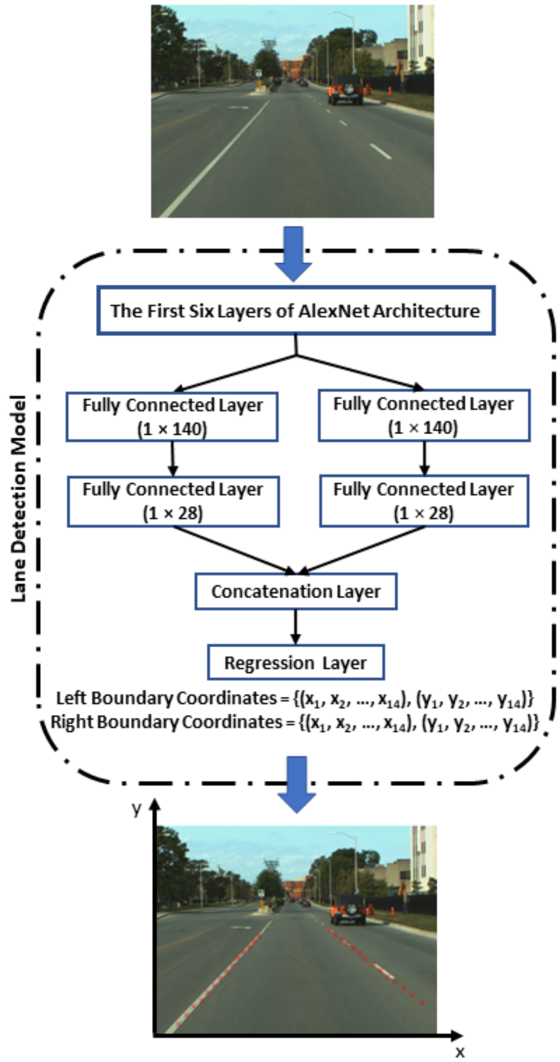


Figure 4.1: *The lane detection model provides two lane vectors, each consisting of 14 coordinates in the image plane that represent the predicted left and right boundaries of the ego lane.*

4.3.2 Lane Type Classification Stage

Lane type information is of great importance in guiding drivers to safely decide either to keep course in the ego lane, to change lane, to overtake, or to turn around. Our goal is to classify the detected ego lane boundaries into eight classes including dashed white, dashed yellow, solid white, solid yellow, double

Figure 4.2: *Forward stereoscopic vision system mounted on rooftop of the Road-LAB experimental vehicle.*

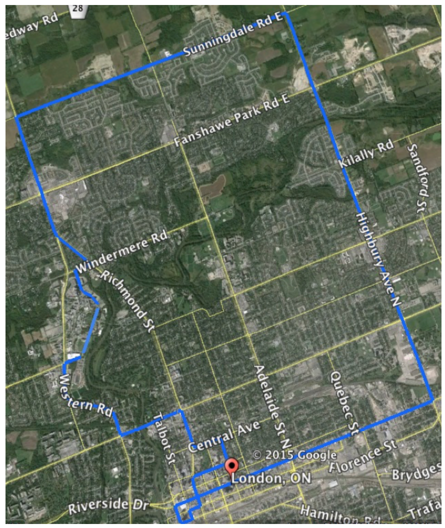


Figure 4.3: *Map of the predetermined course for drivers, located in London, Ontario, Canada. The path includes urban and suburban driving areas and is approximately 28.5 kilometers long.*

Figure 4.4: *Examples of annotated samples of our lane detection dataset.*

Table 4.1: Summary of driving conditions of our data (Each row belongs to one driver.)

Seq. #	Capture Date	Time	Temperature	Weather
2	2012-08-24	15:30	31 °C	Sunny
4	2012-08-31	11:00	24 °C	Sunny
5	2012-09-05	12:05	27 °C	Partially Cloudy
8	2012-09-12	14:45	27 °C	Sunny
9	2012-09-17	13:00	24 °C	Partially Cloudy
10	2012-09-19	09:30	8 °C	Sunny
11	2012-09-19	14:45	12 °C	Sunny
12	2012-09-21	11:45	18 °C	Partially Sunny
13	2012-09-21	14:45	19 °C	Partially Sunny
14	2012-09-24	11:00	7 °C	Sunny
15	2012-09-24	14:00	13 °C	Partially Sunny
16	2012-09-28	10:00	14 °C	Partially Sunny

solid yellow, dashed-solid yellow, solid-dashed yellow, and road boundary. The road boundary type specifies the edge of the road when an actual lane marking does not exist.

ResNet101-Based Lane Type Classification Model

The lane type classification stage receives the output of lane detection (14 coordinates in the image plane for each predicted ego lane boundary) as input. We first identify the ROI for each lane boundary separately. Each ROI fits the detected ego lane boundary as per its corresponding predicted coordinates. Next, we apply a projective transformation to each ROI to obtain an image where the lane marking aligns in the center of the resulting image. Afterwards, we crop the middle rectangular part of the transformed image that contains the lane type information. Finally, we apply our trained ResNet101 network to classify the resulting images obtained for each lane boundary into the aforementioned eight classes. Figure 4.5 illustrates how the lane type classification stage performs the above steps on a sample road image.

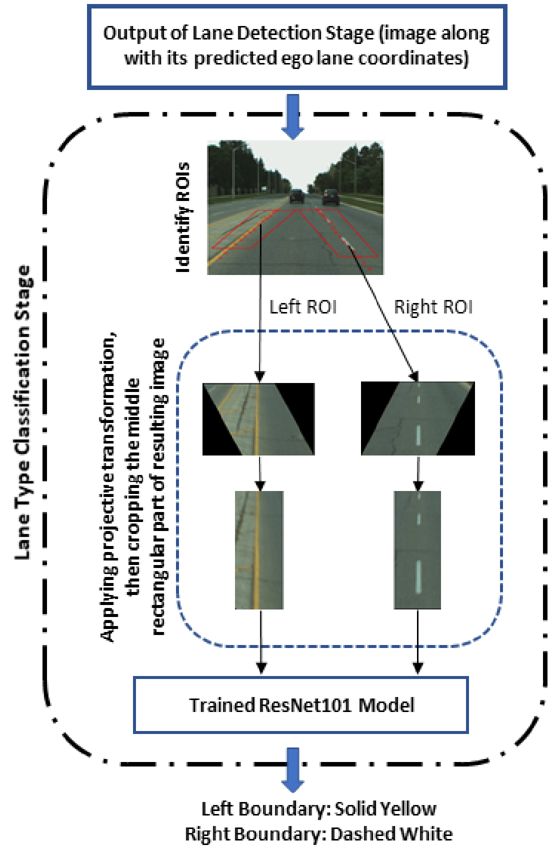


Figure 4.5: *Visualization of the lane type classification stage, from a sample road image to the ego lane boundaries.*

Our Dataset for Lane Boundary Types

In order to train and test our lane type classification model, we collected 10571 sample lane boundary images from the outputs of the lane detection model. These samples are inputs to our ResNet101 model, as they contain the lane type information. Figure 4.6 shows samples of our dataset for the eight lane boundary types.

To further enrich our lane type dataset for training, we employed two different techniques including data augmentation and a boosting method. By means of data augmentation, we expanded our dataset by creating the translated, rotated, sheared, and scaled versions of our original samples. Table 4.2 represents the techniques we have used to augment our data with their descriptions and ranges. To boost the performance of our trained model, we used an advanced learning method called Hard Examples Mining (HEM). HEM refers to the examples that have been misclassified by the current trained version of the model. We trained the ResNet101 model in an iterated procedure, and at each iteration, the model was applied to a number of new samples from the training data. We then added the corrections of misclassified outputs to the training set for the next iteration. Finally, the model is provided with more key samples to increase its robustness.

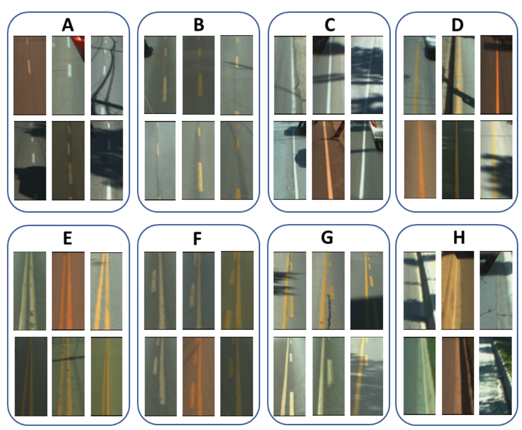


Figure 4.6: *Lane boundary samples of our train-and-test data a) Dashed White, b) Dashed Yellow, c) Solid White, d) Solid Yellow, e) Double Solid Yellow f) Dashed-Solid Yellow, g) Solid-Dashed Yellow, h) Road Boundary*

Table 4.2: DESCRIPTION OF DATA AUGMENTATION

Augmentation Method	**Description**	**Range**
Translate	Each image is translated in the h/v direction by a distance, in pixels	[-20, 20]
Rotate	Each image is rotated by an angle, in degrees	[-25, 25]
Shear	Each image is sheared along the h/v axis by an angle, in degrees	[-25, 25]
Scale	Each image is zoomed in/out in the h/v direction by a factor	[0.5, 1.5]

4.4 Experimental Results

To perform the experiments, we applied the model to the unseen test data extracted from our driving sequences [2]. To evaluate the performance of the lane detection stage, we used a metric suggested by [6]: we compute the mean error between the predicted lane coordinates generated by the lane coordinate model with the corresponding ground truth values as a Euclidean distance (in terms of pixels), for each lane boundary. For each single lane boundary, the Mean Prediction Error (MPE) is computed as follows (see Figure 4.7):

$$\text{MPE} = \frac{1}{14}\sum_{i=1}^{14}\sqrt{(xp_i - xg_i)^2 + (yp_i - yg_i)^2} \tag{4.1}$$

where $(xp_i,\ yp_i)$ and $(xg_i,\ yg_i)$ indicate the predicted lane coordinates and the corresponding ground truth coordinates respectively. Additionally, during network training, we investigated the performance of the following two $L1$ and $L2$ loss functions at the lane detection stage:

$$L1 = \sum_{i=1}^{14}|xp_i - xg_i| + \sum_{i=1}^{14}|yp_i - yg_i| \tag{4.2}$$

$$L2 = \sum_{i=1}^{14}(xp_i - xg_i)^2 + \sum_{i=1}^{14}(yp_i - yg_i)^2 \tag{4.3}$$

where the $L1$ loss computes the absolute differences between the predicted and actual values while the $L2$ loss, also known as the Squared Error Loss, computes the squared differences between the predicted and actual values.

In Table 4.3, we report the performance of the lane detection stage described in Section 4.3.1 for the ego lane left/right boundaries using the aforementioned loss functions. As observed from Table 4.3, the $L1$ loss function is superior to $L2$.

Table 4.3: Description of our lane detection results based on the prediction error

Loss Function	**Ago Lane Boundary**	**MPE**	**Standard Deviation**
$L1$	Left	5.96	4.70
	Right	5.79	4.85
$L2$	Left	7.39	5.55
	Right	7.16	5.42

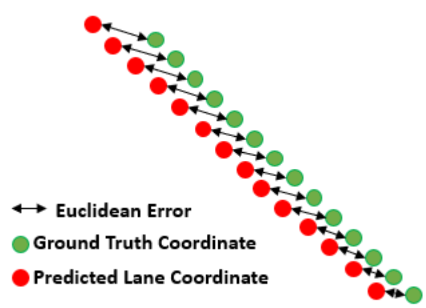


Figure 4.7: *Visualization of the Euclidean error between the predicted lane coordinates and the corresponding ground truth coordinates.*

As described in Section 4.3.2, the lane type classification stage is applied to the output of the lane detection stage to recognize the detected lane boundaries and to provide a classification result. We trained a ResNet101 CNN using our dataset to verify and categorize the localized lane boundaries into eight classes of lane types. To verify the accuracy of the lane type classification stage, we computed the confusion matrix from the ResNet101 model on the test data (See Figure 4.8). The results show that the model reaches 94.52% of overall correct classification. This model is able to discriminate the eight lane types with less than 4.2% of mislabeling error. The lowest degree of correctly categorized classes belongs to class dashed-solid yellow, while class double solid yellow obtained 97.7%. As mentioned, the authors in [24] recognized five lane marking types including dashed, dashed-solid, double solid, solid-dashed, and single solid. They applied their model to three different test data and obtained three corresponding confusion matrices with the overall correct classification of 71.53%, 77.27% and 85.42% which are all less than our overall correct classification. Figure 4.9 displays small portions of the visual outputs from our system for the eight classes of lane boundary types.

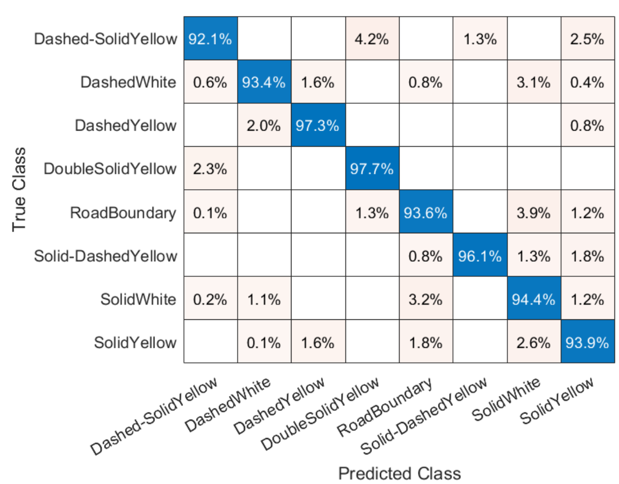


Figure 4.8: *Confusion matrix from ResNet101 for lane type classification.*

Figure 4.9: *Output samples of our experiments on the RoadLAB dataset.*

4.5 Conclusions

In general, in the literature there are little works to classify lane types. In this study, we presented a CNN-based framework to detect and classify lane types in urban and suburban driving environments which have been also less studied in comparison to highways. To perform lane detection and classification stages, we created an image dataset for each from sequences captured in different illumination conditions created by the RoadLAB initiative [2]. We also enriched our training data using data augmentation and a hard example mining strategy. To detect lanes, we used a network which generates lane information in terms of image coordinates in an end-to-end way. In the lane type classification stage, we utilized our trained ResNet101 network to categorize the detected lane boundaries into eight classes including dashed white, dashed yellow, solid white, solid yellow, double solid yellow, dashed-solid yellow, solid-dashed yellow, and road boundary. Finally, our results showed that the ResNet101 model achieved over 94% of correct lane type classifications, which is higher than those of the previous work [24], in addition to the fact that we can recognize three more classes of lane types, especially road boundary type which can be taken into account in urban areas when there is no actual lane marking.

Bibliography

[1] Abduladhem Abdulkareem Ali and Hussein Alaa Hussein. “Real-time lane markings recognition based on seed-fill algorithm”. In: *Proceedings of the International Conference on Information and Communication Technology*. 2019, pp. 190–195.

[2] S.S. Beauchemin, M.A. Bauer, T. Kowsari, and J. Cho. “Portable and Scalable Vision-Based Vehicular Instrumentation for the Analysis of Driver Intentionality”. In: *IEEE Transactions on Instrumentation and Measurement* 61.2 (2012), pp. 391–401.

[3] Hsu-Yung Cheng, Bor-Shenn Jeng, Pei-Ting Tseng, and Kuo-Chin Fan. “Lane detection with moving vehicles in the traffic scenes”. In: *IEEE Transactions on intelligent transportation systems* 7.4 (2006), pp. 571–582.

[4] Kuo-Yu Chiu and Sheng-Fuu Lin. “Lane detection using color-based segmentation”. In: *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.* IEEE. 2005, pp. 706–711.

[5] Hyun-Chul Choi and Se-Young Oh. “Illumination invariant lane color recognition by using road color reference & neural networks”. In: *The 2010 International Joint Conference on Neural Networks (IJCNN).* IEEE. 2010, pp. 1–5.

[6] Shriyash Chougule, Nora Koznek, Asad Ismail, Ganesh Adam, Vikram Narayan, and Matthias Schulze. “Reliable multilane detection and classification by utilizing CNN as a regression network”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

[7] Mohsen Ghafoorian, Cedric Nugteren, Nóra Baka, Olaf Booij, and Michael Hofmann. “El-gan: Embedding loss driven generative adversarial networks for lane detection”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative adversarial nets”. In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[9] Bei He, Rui Ai, Yang Yan, and Xianpeng Lang. “Accurate and robust lane detection based on dual-view convolutional neutral network”. In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2016, pp. 1041–1046.

[10] Toan Minh Hoang, Hyung Gil Hong, Husan Vokhidov, and Kang Ryoung Park. “Road lane detection by discriminating dashed and solid road lanes using a visible light camera sensor”. In: *Sensors* 16.8 (2016), p. 1313.

[11] Hayoung Kim, Jongwon Park, Kyushik Min, and Kunsoo Huh. “Anomaly Monitoring Framework in Lane Detection With a Generative Adversarial Network”. In: *IEEE Transactions on Intelligent Transportation Systems* (2020).

[12] ZuWhan Kim. “Robust lane detection and tracking in challenging scenarios”. In: *IEEE Transactions on Intelligent Transportation Systems* 9.1 (2008), pp. 16–26.

[13] Chanho Lee and Ji-Hyun Moon. “Robust lane detection and tracking for real-time applications”. In: *IEEE Transactions on Intelligent Transportation Systems* 19.12 (2018), pp. 4043–4048.

[14] Seokju Lee, Junsik Kim, Jae Shin Yoon, Seunghak Shin, Oleksandr Bailo, Namil Kim, Tae-Hee Lee, Hyun Seok Hong, Seung-Hoon Han, and In So Kweon. “Vpgnet: Vanishing point guided network for lane and road marking detection and recognition”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 1947–1955.

[15] Andre Linarth and Elli Angelopoulou. “On feature templates for particle filter based lane detection”. In: *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2011, pp. 1721–1726.

[16] Guoliang Liu, Florentin Wörgötter, and Irene Markelić. “Combining statistical hough transform and particle filter for robust lane detection and tracking”. In: *2010 IEEE Intelligent Vehicles Symposium*. IEEE. 2010, pp. 993–997.

[17] Tong Liu, Zhaowei Chen, Yi Yang, Zehao Wu, and Haowei Li. “Lane Detection in Low-light Conditions Using an Efficient Data Enhancement: Light Conditions Style Transfer”. In: *arXiv preprint arXiv:2002.01177* (2020).

[18] Shao-Yuan Lo, Hsueh-Ming Hang, Sheng-Wei Chan, and Jing-Jhih Lin. “Efficient dense modules of asymmetric convolution for real-time se-

mantic segmentation". In: *Proceedings of the ACM Multimedia Asia*. 2019, pp. 1–6.

[19] Shao-Yuan Lo, Hsueh-Ming Hang, Sheng-Wei Chan, and Jing-Jhih Lin. "Multi-Class Lane Semantic Segmentation using Efficient Convolutional Networks". In: *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. 2019, pp. 1–6.

[20] A López, J Serrat, C Canero, F Lumbreras, and T Graf. "Robust lane markings detection and road geometry computation". In: *International Journal of Automotive Technology* 11.3 (2010), pp. 395–407.

[21] Abdelhamid Mammeri, Azzedine Boukerche, and Guangqian Lu. "Lane detection and tracking system based on the MSER algorithm, hough transform and kalman filter". In: *Proceedings of the 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*. 2014, pp. 259–266.

[22] Davy Neven, Bert De Brabandere, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. "Towards end-to-end lane detection: an instance segmentation approach". In: *2018 IEEE intelligent vehicles symposium (IV)*. IEEE. 2018, pp. 286–291.

[23] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. "Spatial as deep: Spatial cnn for traffic scene understanding". In: *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

[24] Mauricio Braga de Paula and Claudio Rosito Jung. "Real-time detection and classification of road lane markings". In: *2013 XXVI Conference on Graphics, Patterns and Images*. IEEE. 2013, pp. 83–90.

[25] Zamani Md Sani, Hadhrami Abd Ghani, Rosli Besar, Azizul Azizan, and Hafiza Abas. "Real-Time Video Processing using Contour Numbers and Angles for Non-urban Road Marker Classification." In: *International Journal of Electrical & Computer Engineering (2088-8708)* 8.4 (2018).

[26] Seungwoo Yoo, Hee Seok Lee, Heesoo Myeong, Sungrack Yun, Hyoungwoo Park, Janghoon Cho, and Duck Hoon Kim. "End-to-End Lane Marker Detection via Row-wise Classification". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, pp. 1006–1007.

Chapter 5

Estimating Average Driver Attention Based on the Visual Field

This Chapter is a reformatted version of the following article:

N. Khairdoost, S.S. Beauchemin, M.A. Bauer, *An Analytical Model for Estimating Average Driver Attention Based on the Visual Field. in 7th International Conference on Signal and Image Processing (ICSIP)*, Suzhou, China, 2022.

The direction of a driver's visual attention plays a crucial role in the context of Advanced Driver Assistance Systems (ADASs) and semi-autonomous driving. The way a driver monitors traffic scene objects partially indicates the level of driver awareness. We propose an analytical method to estimate a driver's average traffic scene attention based on the attentional visual field of the driver in urban and suburban areas. Three metrics are proposed to estimate a driver's average attention. Our model is capable of identifying driver attention with respect to traffic objects including vehicles, traffic lights, traffic signs, and pedestrians within the attentional visual field of the driver at any moment while in the act of driving.

5.1 Introduction

The number of vehicles on the roads increases every day. This fact makes driving safety and road congestion two significant problems. Preventing fatalities and injuries from traffic accidents has become of great importance for governments and vehicle manufacturers around the world. According to the World Health Organization (WHO), the number of people killed in road traffic accidents worldwide was approximately 1.25 million in 2013 and the statistics show that higher-income countries have fewer road fatalities than middle-income countries due to better emergency medical facilities, as well as law enforcement [12]. According to previous studies, driver inattention is one of the main causes of many accidents. Hence, in recent years, real-time analysis of a driver's gaze has attracted the attention of researchers looking to predict driver behavior [9] in order to increase the safety of driving and decrease the number of road accidents.

In this contribution, we propose a new analytical model to estimate a driver's average traffic scene attention. To do this, we utilize YOLOv5 to identify traffic objects in the image plane of the forward stereo system located on the roof of our instrumented vehicle. In addition, our presented model is the first model of its kind that takes advantage of the attentional visual field of the driver to perform its task. This is a significant aspect of a modern ADAS since this allows for the identification of traffic objects seen by the driver.

The rest of this contribution is organized as follows: In Section 5.2, we review the related literature. Section 5.3 explains our proposed method. Section 5.4 describes our instrumented vehicle, the dataset we used, and the results. Finally, we summarize this paper in Section 5.5.

5.2 Literature Survey

In this section, we survey both object detection and driver gaze methods.

5.2.1 Object Detection Methods

Object detection methods can be divided into two major types: traditional and deep learning-based algorithms. Among the traditional object detectors, the approach proposed by Viola and Jones is one that benefits from sliding-windows and AdaBoost classifiers [22]. Another popular framework in this area is Support Vector Machine (SVM) classifier with such features as Histograms of Oriented Gradients (HOG) and Scale Invariant Feature Transforms (SIFT). For example, in [4], authors employed SVM and a multi-scale searching framework with HOG features to detect pedestrians.

Deep learning-based object detection approaches have attracted researchers' attention since they have shown promising results in different applications. We can divide deep learning-based object detection methods into two major categories: Region-based methods and Regression-based methods. The former generates region proposals at the first step and then categorizes them into different object classes. Faster R-CNN [16], R-FCN [3] and SPP-net [5] are some frameworks that follow this strategy. In our laboratory, we have utilized deep neural networks (Faster R-CNN and ResNet) and classical machine learning models (multi-scale HOG-SVM) to detect and recognize traffic objects including traffic signs, vehicles, traffic lights, and pedestrians [21]. However, none of this previous work has provided any analytical approaches related to the traffic objects in the attentional visual field of the driver.

As mentioned, regression-based methods are the second category for object detection based on deep learning, which view the object detection problem as a regression problem and predict locations of objects directly from the whole image. The regression-based methods mainly include YOLOv3 [15], DSOD [18], YOLOv4 [2] as well as YOLOv5 [8]. In this work, we employed YOLOv5 as a traffic object detector along with the attentional visual field of the driver to analyze average driver attention.

5.2.2 Driver Gaze Methods

Driver gaze has been studied in real driving environments and driving simulators for many years. Generally, the driver's gaze is captured by two main instruments including eye glasses/headband and eye trackers. In this section, we provide a short summary of several applications which employed the aforementioned instruments to capture driver gaze.

Eye Glasses/Headband

Some researchers worked on the driver gaze based on eye glasses or headbands. For instance, Jha *et al.* [6] presented an approach using headband based on Gaussian Process Regression (GPR) that predicts the probability of a given point where the driver is looking at. Deep learning-based models have also been used for similar purposes. In [7], a deep learning-based method by means of headband was proposed to predict the driver's visual attention. By gradually upsampling the resolution of the gaze region, the authors increased the accuracy and resolution of the prediction. Palazzi *et al.* [13] introduced the dataset called the DR(eye)VE which was created using eye glasses. They presented a model based on a multi-branch deep network. This model is composed of three branches of convolutional networks for color, motion, and scene semantics and their predictions are integrated to create the final map.

Moreover, using the DR(eye)VE dataset, Lv *et al.* [11] proposed a Reinforced Attention (RA) model that is created directly on top of existing methods as a regulatory mechanism to improve prediction density. Their results showed that the RA model increases the accuracy of gaze prediction on top of existing approaches.

Eye Tracker

Another group of researchers captured the driver gaze information using eye trackers. For instance, a CNN-based model was proposed in [23] for driver gaze estimation in a vehicle environment that combined image information acquired from the front and side cameras into one three-channel image as an input to the model to increase recognition reliability and decrease computational cost. Moreover, in [27], a four-channel gaze estimation model was proposed based on CNN, which was used to estimate the gaze zones of the driver. The authors achieved considerable accuracy in comparison with several other gaze estimation methods. In [24], a novel self-calibrated approach with driver's gaze pattern learning was proposed to automatically obtain the mapping relationship of driver gaze estimation. The new gaze pattern learning algorithm was employed to gradually find typical eye gaze calibration points in a naturalistic driving environment. The authors in [17] proposed a new 3-step deep learning-based method to detect driver head pose class and estimate eye gaze directions. In the first step, the driver's face is detected by a YOLO model. Next, in the second and third steps, CNN-based models were employed to classify a head pose out of seven driver head poses and estimate the eye directions respectively. Rangesh *et al.* [14] presented a method to improve the robustness and generalization of driver gaze estimation on real-world data recorded under extreme conditions. To overcome issues caused by bad lighting, they utilized an IR camera with a suitable equalization/normalization. For the frames that include eyeglasses, the researchers proposed a pre-processing step to remove eyeglasses. In the RoadLAB project, we employed the FaceLAB eye tracker to capture driver gaze information. In our research group, Kowsari *et al.* [10] presented a cross-calibration method to transform the aforementioned driver gaze data from the reference frame of the gaze tracker onto the reference frame of a forward stereoscopic imaging system. Moreover, Shirpour *et al.* [19] employed the RoadLAB gaze data to introduce an approach using a Gaussian Process Regression (GPR) method to estimate the probability of the driver gaze direction. In this work, we also employed the RoadLAB gaze tracker data to analyze average driver attention with respect to traffic objects within the attentional visual field of the driver.

5.3 Proposed Method

In this work, we present a new analytical vision-based model which measures a driver's average attention in their driving environment based on the attentional visual field of the driver and traffic objects. To determine the attentional visual field of the driver, we followed the techniques proposed in our laboratory which are mentioned in Section 5.4. Fig. 5.1 illustrates the attentional visual field of the driver for two sample frames.

In the first step, our model employs a YOLOv5 object detector network to identify traffic objects of interest which are vehicles, traffic lights, traffic signs, and pedestrians in the driving scene images. Afterward, if the object detector identifies that there is at least one object in the image, we establish the attentional visual area of the driver. Next, we can determine whether

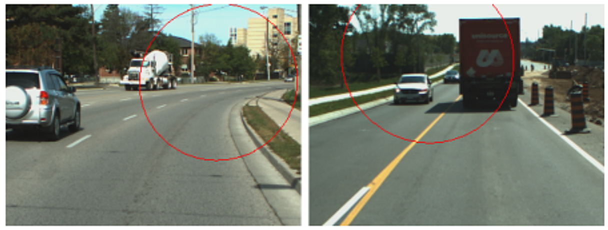


Figure 5.1: *Two samples of attentional visual field of the driver*

the driver is likely to have seen the object or not, namely, when the existing object falls inside the attentional visual field of the driver (see Fig. 5.2). In addition to the objects that are completely located inside the attentional area of the driver, we also need to consider the situations where one or more traffic objects is/are located partially inside the attentional area. In such cases, we consider the object to be partially seen by the driver. Finally, for each object in the frame, we find the percentage of the area of the object that is inside the attentional area. The resulting amount for the Percentage of Inside Area (PIA) for each object can be between 0 and 1. In other words, 0 means the object is completely outside the visual field of the driver, 1 means the object is completely inside the visual field and any other number for PIA means the object is partially located in the visual field; obviously, higher numbers for PIA mean higher overlaps with the visual field of the driver. Fig. 5.3 shows the case of an object is partially located in the attentional area while the other three objects are completely inside the area and our method has obtained 0.28, 1, 1 and 1 for their PIAs respectively.

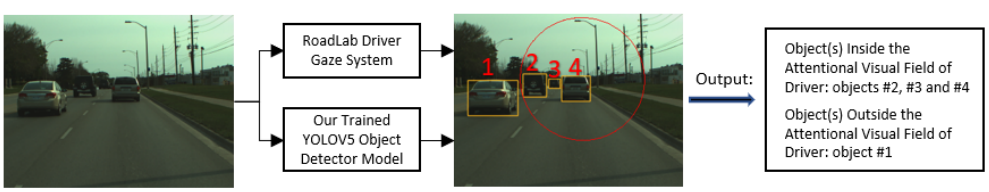


Figure 5.2: *Overview of our model using applying to a sample frame*

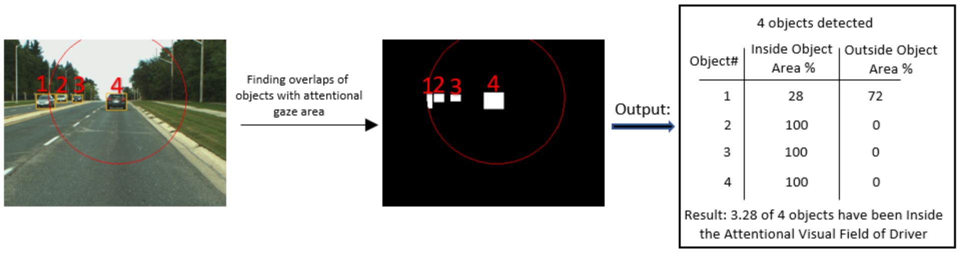


Figure 5.3: *Determining the inside/outside area percentages of the objects based on the attentional field of the driver*

The overall average attention for a driver can be estimated in different ways. We consider three different metrics each making use of the attentional data extracted from an image in the driving sequence of a driver. These metrics can vary between 0 and 1 and are described in the following.

Metric 1 (M1). As the first metric, separately for each of the aforementioned object types, we compute the average PIA of the objects of each type for all frames. Thus we have this metric for each of the individual classes of objects, i.e., since we have four classes of objects M1 will consist of 4 separate measurements. M1 is computed for each object type separately as follows:

$$M1 = \frac{\text{Sum of PIA of Objects of Type i}}{\text{Number of Detected Objects of Type i}} \qquad (5.1)$$
$$\text{i = vehicle, traffic light, traffic sign and pedestrian}$$

Metric 2 (M2). As the second metric, we find the average PIA for all objects ignoring the type of object. In other words, this metric works similar to M1 but views the four traffic object types (vehicles, traffic lights, traffic signs, and pedestrians) as one general traffic object type. M2 is computed as follows:

$$M2 = \frac{\text{Sum of PIA of Traffic Objects}}{\text{Number of Detected Traffic Objects}} \tag{5.2}$$

Metric 3 (M3). This metric, similar to M2, views the four traffic object types as one general traffic object type but unlike M2, determines the average area percentages of the objects which are partially or completely outside the attentional visual area of the driver while driving. This metric simply can be computed as follows:

$$M3 = 1 - M2 \tag{5.3}$$

5.4 Experimental Results

In this section, we provide our vehicle configuration, the data we used for our experiments, and the result for six different drivers.

Our RoadLAB experimental vehicle is equipped with a non-contact gaze tracker. This system consists of a pair of infrared stereo cameras mounted on the dashboard, working at 60Hz. Our instrumented vehicle is equipped with stereo cameras mounted on the vehicle's roof to capture the forward driving environment at a rate of 30Hz. Fig. 5.4 depicts the configuration of the RoadLAB experimental vehicle. Details concerning this configuration were described by [1]. The instrumented vehicle was employed to record data sequences from 16 different drivers on a pre-determined 28.5km course around the city of London, Ontario, Canada. As mentioned, we followed the techniques proposed in our laboratory to establish the attentional visual field of the driver in the image plane of the forward stereo vision system. These techniques have been used in several experiments for various purposes in our laboratory [10], [21], [25], [9], [20], [26].

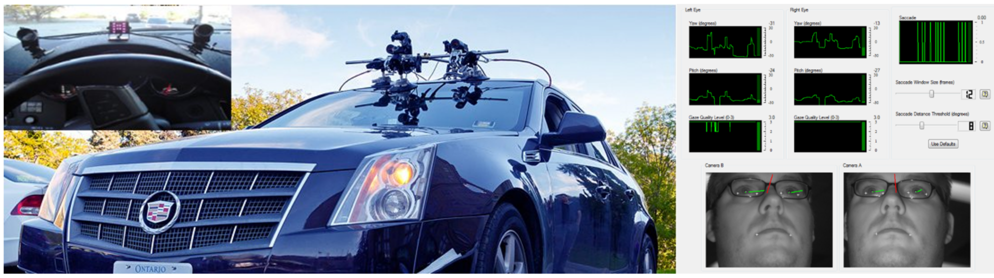


Figure 5.4: *Vehicular instrumentation configuration.* ***(left-top):*** *Infra-red gaze tracker located on the dashboard* ***(left-bottom):*** *Forward stereo vision system mounted on the rooftop* ***(right):*** *The interface of FaceLAB system from Seeing Machines*

To estimate average driver attention based on the attentional visual field of the driver with respect to traffic objects, we employ our method using a YOLOv5 model trained on RoadLAB data and investigate the aforementioned three different metrics for each driver. Table 5.1 provides the details on the sequences that have been gathered by different drivers for our experiments.

Table 5.1: Summary of driving conditions of our data

Driver	Capture Date	Time	Temp.	Weather	Age
3	2012-08-30	12:15	23 °C	Sunny	41
8	2012-09-12	14:45	27 °C	Sunny	21
9	2012-09-17	13:00	24 °C	Partially cloudy	21
12	2012-09-21	11:45	18 °C	Partially sunny	24
13	2012-09-21	14:45	19 °C	Partially sunny	23
15	2012-09-24	14:00	13 °C	Partially sunny	44

The analytical results of our experiments for the drivers have been provided in Table 5.2. In this table, V, TL, TS, and P represent the object types of vehicle, traffic light, traffic sign, and pedestrian, respectively. In general, driver attention while driving can be influenced by various factors such as driving skills, driving habits, distractions, etc. Table 5.2 shows the results for the estimation of average driver attention during driving based on the metrics M1, M2, and M3 which are based on the attentional visual field of the driver. For M1 and M2 which focus on the objects inside the attentional visual field of the driver, higher amounts can indicate higher attentiveness of drivers on average with respect to traffic objects during driving. As mentioned, M1 includes M1-V, M1-TL, M1-TS, and M1-P for four different object types while M2 considers all object types as one object type for processing. As can be seen, the maximum values for M1-V and M1-TL belong to driver 3 while driver 12 and driver 8 gained the maximum values for M1-TS and M1-P respectively. On the other hand, driver 9 was ranked last in terms of two metrics of M1-TL and M1-TS. Similarly, driver 13 placed last for two metrics of M1-V and M1-P. With regarding metric M2, we observe that drivers 3 and 12 achieved the first and second ranks respectively among others. As mentioned M3 similar to M2 considers all object types as one object type but focuses on objects outside of the attentional area of the driver; hence, higher amounts of M3 can indicate higher inattentiveness for drivers. Driver 13 obtained the maximum value for M3. According to Table 5.2, the average for M1-V, M1-TL, M1-TS, M1-P, M2, and M3 are 56.27%, 53.58%, 49.35%, 46.65%, 54.24%, and 45.76% respectively.

5.5 Conclusions

Nowadays, almost every modern vehicle is equipped with some type of ADASs, ranging from collision avoidance system, alcohol ignition interlock devices, anti-lock braking system, to parking assistance system. ADASs generally in-

Figure 5.5: *Output samples of our experiments on the RoadLAB dataset*

crease car and road safety and assist a driver in driving tasks. In this research, we presented an analytical model to estimate average driver attention based on the attentional visual field of the driver using different metrics. For this, we used the RoadLAB dataset obtained from our instrumented vehicle in our experiments. Next, by establishing the attentional field of view of the driver we were able to investigate the average area percentages of the traffic objects including vehicles, traffic lights, traffic signs, and pedestrians, which are inside the driver gaze area while driving. By using our approach we are able to infer the driver's behavior in terms of the driver's attentional visual area. Ultimately, such an augmented approach could enable the driver's gaze information to be integrated into ADAS as a means to determine objects drivers attend to, and those that they do not, and also to be used to predict driver maneuver [9] as well as to detect driver distraction as part of ADASs in the future.

Table 5.2: Analytical results for the attentional visual field of the driver

Driver	**M1-V (%)**	**M1-TL (%)**	**M1-TS (%)**	**M1-P (%)**	**M2 (%)**	**M3 (%)**
3	61.89	67.14	56.20	52.23	61.61	38.39
8	54.92	51.02	54.46	61.27	54.67	45.33
9	56.34	38.94	34.95	39.62	48.95	51.05
12	58.38	64.39	59.53	44.79	58.52	41.48
13	49.47	46.04	38.62	34.18	46.59	53.41
15	56.60	53.94	52.32	47.80	55.07	44.93

Bibliography

[1] Steven S Beauchemin, Michael A Bauer, Taha Kowsari, and Ji Cho. "Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality". In: *IEEE Transactions on Instrumentation and Measurement* 61.2 (2011), pp. 391–401.

[2] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection". In: *arXiv preprint arXiv:2004.10934* (2020).

[3] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. "R-fcn: Object detection via region-based fully convolutional networks". In: *arXiv preprint arXiv:1605.06409* (2016).

[4] Jan Dürre, Dario Paradzik, and Holger Blume. "A HOG-based real-time and multi-scale pedestrian detector demonstration system on FPGA". In: *Proceedings of the 2018 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. 2018, pp. 163–172.

[5] K. He, X. Zhang, S. Ren, and J. Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 37.9 (2015), pp. 1904–1916.

[6] Sumit Jha and Carlos Busso. "Probabilistic estimation of the driver's gaze from head orientation and position". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2017, pp. 1–6.

[7] Sumit Jha and Carlos Busso. "Probabilistic estimation of the gaze region of the driver using dense classification". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2018, pp. 697–702.

[8] Glenn Jocher, Alex Stoken, Jirka Borovec, Liu Changyu, and Adam Hogan. "ultralytics/yolov5: v3. 0". In: *Zenodo* (2020).

[9] N. Khairdoost, M. Shirpour, M.A Bauer, and S. S Beauchemin. "Real-Time Maneuver Prediction Using LSTM". In: *IEEE Transactions on Intelligent Vehicles* (2020).

[10] T. Kowsari, S.S. Beauchemin, M.A. Bauer, D. Laurendeau, and N. Teasdale. "Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems". In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 1245–1250.

[11] Kai Lv, Hao Sheng, Zhang Xiong, Wei Li, and Liang Zheng. "Improving Driver Gaze Prediction with Reinforced Attention". In: *IEEE Transactions on Multimedia* (2020).

[12] World Health Organization et al. *Global status report on road safety 2015*. Tech. rep. World Health Organization, 2015.

[13] A. Palazzi, D. Abati, F. Solera, and R. Cucchiara. "Predicting the Driver's Focus of Attention: the DR (eye) VE Project". In: *IEEE trans-*

actions on pattern analysis and machine intelligence 41.7 (2018), pp. 1720–1733.

[14] Akshay Rangesh, Bowen Zhang, and Mohan M Trivedi. "Driver gaze estimation in the real world: Overcoming the eyeglass challenge". In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2020, pp. 1054–1059.

[15] Joseph Redmon and Ali Farhadi. "Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767* (2018).

[16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *arXiv preprint arXiv:1506.01497* (2015).

[17] Sayyed Mudassar Shah, Zhaoyun Sun, Khalid Zaman, Altaf Hussain, Muhammad Shoaib, and Lili Pei. "A driver gaze estimation method based on deep learning". In: *Sensors* 22.10 (2022), p. 3959.

[18] Zhiqiang Shen, Zhuang Liu, Jianguo Li, Yu-Gang Jiang, Yurong Chen, and Xiangyang Xue. "Dsod: Learning deeply supervised object detectors from scratch". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 1919–1927.

[19] Mohsen Shirpour, Steven S Beauchemin, and Michael A Bauer. "A probabilistic model for visual driver gaze approximation from head pose estimation". In: *2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS)*. IEEE. 2020, pp. 1–6.

[20] Mohsen Shirpour, Steven S Beauchemin, and Michael A Bauer. "What Does Visual Gaze Attend to during Driving?" In: *VEHITS*. 2021, pp. 465–470.

[21] Mohsen Shirpour, Nima Khairdoost, Michael Bauer, and Steven Beauchemin. “Traffic Object Detection and Recognition Based on the Attentional Visual Field of Drivers”. In: *IEEE Transactions on Intelligent Vehicles* (2021), pp. 1–1. DOI: 10.1109/TIV.2021.3133849.

[22] Paul Viola and Michael J Jones. “Robust real-time face detection”. In: *International journal of computer vision* 57.2 (2004), pp. 137–154.

[23] Hyo Sik Yoon, Na Rae Baek, Noi Quang Truong, and Kang Ryoung Park. “Driver gaze detection based on deep residual networks using the combined single image of dual near-infrared cameras”. In: *IEEE Access* 7 (2019), pp. 93448–93461.

[24] Guoliang Yuan, Yafei Wang, Huizhu Yan, and Xianping Fu. “Self-calibrated driver gaze estimation via gaze pattern learning”. In: *Knowledge-Based Systems* 235 (2022), p. 107630.

[25] S.J. Zabihi, S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. “Detection and recognition of traffic signs inside the attentional visual field of drivers”. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2017, pp. 583–588.

[26] S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. “Real-time driving manoeuvre prediction using IO-HMM and driver cephalo-ocular behaviour”. In: *Intelligent Vehicles Symposium (IV), 2017 IEEE*. IEEE. 2017, pp. 875–880.

[27] Yingji Zhang, Xiaohui Yang, and Zhe Ma. “Driver’s Gaze Zone Estimation Method: A Four-channel Convolutional Neural Network Model”. In: *2020 2nd International Conference on Big-data Service and Intelligent Computation*. 2020, pp. 20–24.

Chapter 6

What Has the Driver Gazed at in the Average Percentage of the Driving Time?

In the area of intelligent transportation systems, the role of Advance Driver Assistance Systems (ADASs) is of great importance. In ADAS, many efforts have been done in different areas such as blind-spot detection, collision avoidance systems, traffic sign recognition, lane departure warning systems, etc. Studying how driver gaze information can be leveraged in ADAS is another important area in driving to consider. A driver's gaze during driving can provide an ADAS with insight into the driver's intent or awareness of situations enabling the system to assist the driver or avoid accidents. In this work, we propose an analytical method to measure the percentage of time on average that a driver gazes at different traffic objects in the course of driving including urban and suburban areas. To do this, three metrics are proposed that benefit from the gaze point of the driver with respect to four major types of traffic objects including vehicles, traffic lights, traffic signs, and pedestrians.

6.1 Introduction

Every year, the large number of car collisions leads to both tremendous human and economic costs [39]. According to the global status report on road safety 2018, launched by World Health Organization (WHO) [29], approximately 1.35 million fatalities occur per year in the world because of road traffic accidents, and up to 50 million people are injured. Now, road traffic injury is the leading cause of death among young people and children aged 5-29 years and makes road fatalities the eighth leading cause of death across all age groups. Moreover, drivers are less likely to be involved in an accident in the case of the presence of one or more passengers who can warn them in advance [33]. Obviously, driver error is the main reason for road accidents. To overcome this, efforts are being made by both academic and industrial groups to develop Advanced Driver Assistance Systems (ADASs) in different aspects. These systems attempt to assist the driver's decision-making in the act of driving or even take control of the vehicle by performing automatic actions, improving car, and road safety in general.

In this contribution, we propose a new analytical model based on the Point of Gaze (PoG) of the driver to find the percentage of the time on average in which a driver has gazed at traffic objects in the course of driving. To do this, we employ YOLOv5 to identify traffic objects in the imaging plane of the forward stereo system located on the rooftop of our experimental vehicle. Using our driving sequences, we present the results for percentages of time determining whether a driver's PoG has fallen on a traffic object or not. The resulting insight can be useful in other scenarios involving the analysis of driver gaze behavior and have implications for designing ADASs and for the understanding of driver intent and awareness in the future.

The rest of this contribution is structured as follows: In Section 6.2, we review the related literature. Section 6.3 describes our proposed approach. Section 6.4 explains our experimental vehicle, data, and the results. Finally, we give a summary of this research study in Section 6.5.

6.2 Literature Survey

In general in the literature, many efforts and research have been performed to analyze driver behavior with different views to achieve different goals such as driver distraction detection, driver style identification, driver intent prediction, traffic management, and so on. Here, we give an overview of driver behavior methods and their applications and finally review various research works performed in our research group in this research area.

According to [52], different driver behavior analysis methods based on their applications can be categorized into three classes: vehicle-oriented applications, management-oriented applications, and driver-oriented applications. Vehicle-oriented applications focus mainly on the vehicles to improve the driving task and reduce driver workload by creating advanced systems to assist drivers in different driving situations. Google's first fully autonomous car prototype [8], emergency braking systems [15], [31], lane keeping assistance systems [7], [43], and automatic accident detectors [11], [30], [12] are some examples in this category. Management-oriented applications attempt to optimize vehicle usage mainly including fleet management and traffic modeling. For this, they focus on the management of infrastructure and resources by monitoring the road conditions and the vehicle. These systems identify road conditions based on the driver maneuvers such as accelerations and brakings as well as the data related to three-axes accelerations [34], [6].

Driver-oriented applications consider the driver as the primary factor. Driver attention evaluation, distraction detection, driving style assessment, and driver intent prediction are the main areas of research in this category. Methods for driver attention evaluation analyze the attention of the driver [36], [42], [46], [51] and somnolence of the driver [25], [13]. Regarding distraction detection systems, the degree of driver focus on the road is identified based on driver reactions [18], [27]. Driving style identifiers aim to categorize the driving mode based on a variety of features collected from the vehicle and the driver's actions such as acceleration, steering, speed, braking and GPS [41], [17], [45]. Aggressive style and risky style are the two common styles in this area of research. As for driver intent prediction, these applications aim to anticipate the most probable next maneuver (overtaking, lane change, emergency braking, etc.) of the driver using the methods of automatic prediction of maneuvers [26], [44], [28].

Driver distraction and drowsiness are two main reasons for traffic crashes and the related financial costs throughout the world. Hence, researchers and car manufacturers have been working for more than a decade on analyzing driver behavior to detect his/her inattention while driving. There are four types of driver distraction: visual distraction (caused by driver's eyes off the road), manual distraction (caused by driver's hands off the wheel), auditory distraction (caused by acoustic stimuli or any kind of vocal utterance), and cognitive distraction (caused by driver's mind off the road) [21]. Identifying cognitive distraction can be probably considered as the most difficult distraction type due to the problems related to observing what a driver's brain (as opposed to his/her eyes or hands) is doing [38]. A distracting activity can also involve one or more of the aforementioned distraction types. For instance, the use of a hand-held mobile phone, may involve the four distraction types [9] and increases the risk of an accident significantly [35]. Moreover, regarding the effects related to presence of passengers in the vehicle on driver's performance, there is a debate among researchers. Some of them concluded a reduction in driver's mistakes and violations [33] while some others reported an increase for those [49], [50].

As mentioned, drowsiness detection is an important research area of driver behavior analysis since it is one of the major reasons for road accidents. For instance, according to the National Highway Traffic Safety Administration (NHTSA) [1], approximately 8,000 deaths occur due to drowsy driving annually. Methods have been employed for detecting drowsiness of the driver can be broadly grouped into two categories: methods based on visual features and methods based on non-visual features [20]. Methods based on visual features benefit from computer vision techniques for the detection of drowsiness. Visual feature-based methods attempt to extract facial features such as face, eyes, and mouth. These methods can be mainly divided into four categories including eye state analysis [40], eye blinking analysis [19], mouth and yawning analysis [5], and facial expression analysis [14]. Methods that use non-visual features can be broadly divided into two categories: driver physiological analysis and vehicle parameter analysis. The former usually refers to the brain activity and heart rate of a driver such as electroencephalogram (EEG), heart rate (ECG), and electrooculogram (EOG) [2], [16], [32], whereas methods based on vehicle parameters analysis by analyzing vehicle features such as steering wheel movement, lane keeping, the pressure exerted on the brake, and acceleration pedal movement detect drowsiness of the driver [3], [10].

In the RoadLAB research project, we utilized the FaceLAB eye tracker to record driver gaze data. In our research group, Kowsari *et al.* [24] introduced a cross-calibration technique to transform the aforementioned driver gaze data from the reference frame of the gaze tracker onto the reference frame of a forward imaging system. Moreover, the works which were presented in [22] and [48] employed the RoadLAB gaze data to model driver behavior and predict driver maneuvers using driver cephalo-ocular behavioral and vehicular dynamics information. Also in [37] using the gaze data, we detected and recognized four major types of traffic objects including vehicles, traffic signs, traffic lights and pedestrians inside and outside the visual filed of the driver. Finally, in [23], we studied the driver behavior with respect to the aforementioned traffic objects in terms of the attentional visual field of the driver. In this work, we also attempt to investigate driver behavior in terms of his/her PoG in the course of driving with respect to the aforementioned major classes of traffic objects in percentage of driving time.

6.3 Proposed Method

In this study, we propose a new analytical model which identifies the percentage of the time on average that a driver has gazed at different traffic objects based on PoG of the driver in the course of driving. To determine the PoG of the driver in the image plane of the forward stereo scene system, we followed the techniques proposed in our laboratory. These techniques have been used in different experiments with different purposes in our laboratory [24], [37], [47], [22], [36], [48]. Figure 6.1 illustrates the PoG of the driver for two sample frames. In the first step, our model employs a YOLOv5 object detector network to identify traffic objects of interest which are vehicles, traffic lights, traffic signs, and pedestrians in the driving scene images. Afterward, if the object detector identifies that there is at least one object in the image, we obtain the PoG of the driver. Next, we can determine whether the driver's

PoG has fallen into the object or not. (See Figure 6.2.) To investigate driver behavior in terms of PoG during driving with respect to traffic objects, we present three different metrics each making use of the PoG and traffic objects extracted from an image in the driving sequence of a driver. These metrics can vary between 0 (i.e. 0% of driving time) and 1 (i.e. 100% of driving time) and are explained in the following.

Figure 6.1: *Two samples of PoG of the driver (the red point)*

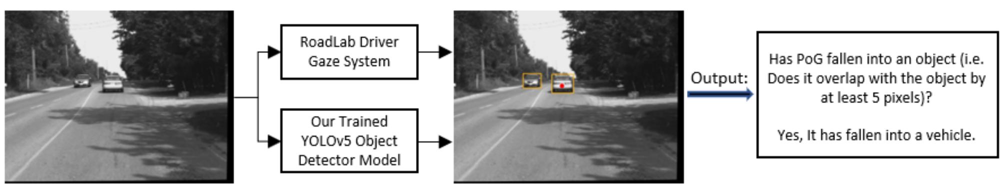


Figure 6.2: *Overview of our model using applying to a sample frame*

Metric 1 (M1) As the first metric, separately for each of the aforementioned object types, we identify the average number of PoGs that have fallen onto the objects of each type for all frames; As a result, since we have four classes of objects M1 will consist of four separate measurements for each of the individual classes of objects. To compute this metric, for each object type separately, we consider the PoG as a circle with a radius of three pixels. Next, for each frame, if the PoG has an overlap with an object bigger than a threshold, the PoG is considered to have fallen onto the object. Otherwise, we conclude that the

PoG is outside the objects. For our experiments, we employed the threshold of five pixels. M1 is computed for each object type separately as follows:

$$M1 = \frac{\text{Number of Frames in which PoG Fell Into Object of Type i}}{\text{Number of Frames}} \qquad (6.1)$$
$$\text{i = vehicle, traffic light, traffic sign and pedestrian}$$

Metric 2 (M2) As the second metric, we obtain the average number of PoGs that have fallen into any traffic objects ignoring the type of object. In other words, this metric works similar to M1 but considers the four traffic object types (vehicles, traffic lights, traffic signs, and pedestrians) as one general traffic object type. M2 is computed as follows:

$$M2 = \frac{\text{Number of Frames in which PoG Fell Into Traffic Object}}{\text{Number of Frames}} \qquad (6.2)$$

Metric 3 (M3) This metric, similar to M2, views the four traffic object types as one general traffic object type but unlike M2, focuses on the average number of PoGs of the driver that have fallen outside all the detected traffic objects while driving. In other words, in this kind of frames we do not know where the driver is gazing at. This metric simply can be computed as follows:

$$M3 = 1 - M2 \qquad (6.3)$$

6.4 Experimental Results

In this section, we provide our vehicle configuration, the data we used for our experiments, and the result for six different drivers.

Our RoadLAB experimental vehicle is equipped with a remote eye-gaze tracker mounted on the dashboard and also stereo cameras placed on the roof of the vehicle to record the frontal driving environment. Details related to this configuration were explained in [4]. Figures 1.2 and 1.3 show the configuration of the RoadLAB vehicle and the pre-determined path of driving respectively.

To investigate PoG behavior of driver with respect to the aforementioned traffic objects during driving, we employ our method using a YOLOv5 model trained on RoadLAB data and measure the aforementioned three different metrics for each driver. Table 5.1 provides the details on the sequences that have been gathered by different drivers for our experiments.

The analytical results of our experiments for the drivers have been provided in Table 6.1. In this table, V, TL, TS, and P stand for the object types of vehicle, traffic light, traffic sign, and pedestrian, respectively. In general, various factors such as driving skills, habits, experience and driver distractions can influence on PoG of the driver during driving. Table 6.1 shows the results for the estimation of the percentage of the time on average in the path of driving based on the metrics M1, M2, and M3 which are based on PoG of the driver. M1 and M2 which focus on the frames in which the PoG has fallen into the object, higher amounts can indicate the driver has spent more percentage of his/her driving time gazing at four types of objects in the path of driving on average. As mentioned, M1 includes M1-V, M1-TL, M1-TS, and M1-P for four different object types while M2 considers all object types as one object type for processing. As can be seen, considering the results related to four measures of M1, the drivers mostly gazed at vehicles in comparison with other traffic objects that is a normal activity of drivers to do in driving task. Regarding M1-V, driver 9 obtained the maximum value of this metric. Controversy, for M1-TL driver 9 has spent almost the minimum amount (quite similar to that of driver 8) as well as the minimum amount for M1-TS. Unlike driver 9, we observe that the maximum values for metrics M1-TL and M1-TS belong to driver 12. Regarding metric M1-P, driver 8 has gazed with more percentage at pedestrians among other drivers while driver 12 performed this as the last rank. Moreover, considering amounts for M1-TL, M1-TS and M1-P (regardless of M1-V) for each driver, it can be seen drivers 8 and 9 have gazed with more percentage at pedestrians in comparison with traffic lights and traffic signs while drivers 3,12 and 13 spent more percentage gazing at traffic lights in comparison with two other object types. In addition, driver 15 has gazed with more percentage at traffic lights and pedestrians (almost equally on average) than traffic signs. With regarding metric M2, we observe that driver 9 has spent more percentage of driving time gazing at traffic objects in comparison with others. As another result for this metric, drivers 8 and 15 have gazed at traffic objects with very similar percentage amount to each other and placed in the second rank among others. As mentioned M3 similar to M2 considers all object types as one object type but focuses on the frames in which the PoG has fallen outside the objects; hence, higher amounts of M3 can indicate higher percentage of the time that the driver has not gazed at the aforementioned object types in the path of driving. Driver 3 obtained the maximum value for M3. According to Table 6.1, the average for M1-V, M1-TL, M1-TS, M1-P, M2, and M3 are 25.13%, 1.02%, 0.36%, 1.18%, 27.42%, and 72.58% respectively. Finally, Figure 6.3 displays a small sample of the visual outputs from the proposed method.

Figure 6.3: *Output samples of our experiments on the RoadLAB dataset*

6.5 Conclusions

Evidence has shown driver error is the main cause of road accidents. In this research, we presented an analytical model to estimate the percentage of time on average in which a driver gazed at different traffic objects using three metrics. For this, we used the naturalistic on-road RoadLAB dataset obtained from our experimental vehicle in our experiments. After obtaining the PoG of the driver, we estimated the percentage of the experimental driving data at which PoG fell into different traffic objects including vehicles, traffic lights, traffic signs, and pedestrians. By using our approach, we can infer the driver's behavior in terms of the driver's PoG in the course of driving. Ultimately, such methods presented in this work can be useful in designing a future ADAS system to understand driver intent in advance as well as to measure driver awareness levels while driving.

Table 6.1: Analytical results for PoG of the driver with respect to the traffic objects

Driver	**M1-V (%)**	**M1-TL (%)**	**M1-TS (%)**	**M1-P (%)**	**M2 (%)**	**M3 (%)**
3	17.12	1.40	0.44	0.84	19.63	80.37
8	25.98	0.27	0.42	2.47	28.90	71.10
9	33.73	0.29	0.23	1.57	35.43	64.57
12	23.44	1.94	0.51	0.31	26.00	74.00
13	23.51	1.32	0.28	0.97	25.60	74.40
15	26.98	0.91	0.30	0.93	28.95	71.05

Bibliography

[1] National Highway Traffic Safety Administration et al. “Asleep at the Wheel: A National Compendium of Efforts to Eliminate Drowsy Driving”. In: *National Highway Traffic Safety Administration: Washington, DC, USA* (2017).

[2] Sadegh Arefnezhad, James Hamet, Arno Eichberger, Matthias Frühwirth, Anja Ischebeck, Ioana Victoria Koglbauer, Maximilian Moser, and Ali Yousefi. “Driver drowsiness estimation using EEG signals with a dynamical encoder–decoder modeling framework”. In: *Scientific reports* 12.1 (2022), pp. 1–18.

[3] Sadegh Arefnezhad, Sajjad Samiee, Arno Eichberger, and Ali Nahvi. “Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection”. In: *Sensors* 19.4 (2019), p. 943.

[4] Steven S Beauchemin, Michael A Bauer, Taha Kowsari, and Ji Cho. “Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality”. In: *IEEE Transactions on Instrumentation and Measurement* 61.2 (2011), pp. 391–401.

[5] GM Bhandari, Archana Durge, Aparna Bidwai, and Urmila Aware. “Yawning analysis for driver drowsiness detection”. In: *Int. J. Res. Eng. Technol.* 3.2 (2014), pp. 502–505.

[6] Ravi Bhoraskar, Nagamanoj Vankadhara, Bhaskaran Raman, and Purushottam Kulkarni. "Wolverine: Traffic and road condition estimation using smartphone sensors". In: *2012 fourth international conference on communication systems and networks (COMSNETS 2012)*. IEEE. 2012, pp. 1–6.

[7] Yougang Bian, Jieyun Ding, Manjiang Hu, Qing Xu, Jianqiang Wang, and Keqiang Li. "An advanced lane-keeping assistance system with switchable assistance modes". In: *IEEE Transactions on Intelligent Transportation Systems* 21.1 (2019), pp. 385–396.

[8] B. Bilger. *Has the self-driving car at last arrived? The New Yorker (2013).* `http://www.newyorker.com/reporting/2013/11/25/131125fa_fact_bilger?currentPage=all`.

[9] Jeanne Breen. *Car telephone use and road safety, an overview prepared for the European Commission.* European Commission, 2009.

[10] Meng Chai et al. "Drowsiness monitoring based on steering wheel status". In: *Transportation research part D: transport and environment* 66 (2019), pp. 95–103.

[11] Nimisha Chaturvedi and Pallika Srivastava. "Automatic vehicle accident detection and messaging system using GSM and GPS modem". In: *Int. Res. J. Eng. Technol.(IRJET)* 5.3 (2018), pp. 252–254.

[12] Jae Gyeong Choi, Chan Woo Kong, Gyeongho Kim, and Sunghoon Lim. "Car crash detection using ensemble deep learning and multimodal data from dashboard cameras". In: *Expert Systems with Applications* 183 (2021), p. 115400.

[13] Md Tanvir Ahammed Dipu, Syeda Sumbul Hossain, Yeasir Arafat, and Fatama Binta Rafiq. "Real-time Driver Drowsiness Detection using Deep Learning". In: *International Journal of Advanced Computer Science and Applications* 12.7 (2021).

[14] Er Manoram Vats1and Er Anil Garg. "Detection and security system for drowsy driver by using artificial neural network technique". In: *International Journal of Applied Science and Advance Technology* 1.1 (2012), pp. 39–43.

[15] G. Griffin, D. Kwiatkowski, and J. Miller. U.S. pat. No. 9248815. Washington, DC: U.S. Patent and Trademark Office. 2016.

[16] Md Mahmudul Hasan, Christopher N Watling, and Grégoire S Larue. "Physiological signal-based drowsiness detection using machine learning: singular and hybrid signal approaches". In: *Journal of safety research* 80 (2022), pp. 215–225.

[17] Jin-Hyuk Hong, Ben Margines, and Anind K Dey. "A smartphone-based sensing platform to model aggressive driving behaviors". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 4047–4056.

[18] Md Uzzol Hossain, Md Ataur Rahman, Md Manowarul Islam, Arnisha Akhter, Md Ashraf Uddin, and Bikash Kumar Paul. "Automatic driver distraction detection using deep convolutional neural networks". In: *Intelligent Systems with Applications* 14 (2022), p. 200075.

[19] Jaeik Jo, Sung Joo Lee, Kang Ryoung Park, Ig-Jae Kim, and Jaihie Kim. "Detecting driver drowsiness using feature-level fusion and

user-specific classification". In: *Expert Systems with Applications* 41.4 (2014), pp. 1139–1152.

[20] Sinan Kaplan, Mehmet Amac Guvensan, Ali Gokhan Yavuz, and Yasin Karalurt. "Driver behavior analysis for safe driving: A survey". In: *IEEE Transactions on Intelligent Transportation Systems* 16.6 (2015), pp. 3017–3032.

[21] Alexey Kashevnik, Roman Shchedrin, Christian Kaiser, and Alexander Stocker. "Driver distraction detection methods: A literature review and framework". In: *IEEE Access* 9 (2021), pp. 60063–60076.

[22] N. Khairdoost, M. Shirpour, M.A Bauer, and S. S Beauchemin. "Real-Time Maneuver Prediction Using LSTM". In: *IEEE Transactions on Intelligent Vehicles* (2020).

[23] Nima Khairdoost, Steven S Beauchemin, and Michael A Bauer. "An Analytical Model for Estimating Average Driver Attention Based on the Visual Field." In: *2022 IEEE 7th International Conference on Signal and Image Processing (ICSIP)*. IEEE. 2022.

[24] T. Kowsari, S.S. Beauchemin, M.A. Bauer, D. Laurendeau, and N. Teasdale. "Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems". In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 1245–1250.

[25] Vijay Kumar, Shivam Sharma, et al. "Driver drowsiness detection using modified deep learning architecture". In: *Evolutionary Intelligence* (2022), pp. 1–10.

[26] Stéphanie Lefèvre, Ashwin Carvalho, Yiqi Gao, H Eric Tseng, and Francesco Borrelli. “Driver models for personalised driving assistance”. In: *Vehicle System Dynamics* 53.12 (2015), pp. 1705–1720.

[27] Tianchi Liu, Yan Yang, Guang-Bin Huang, Yong Kiang Yeo, and Zhiping Lin. “Driver distraction detection using semi-supervised machine learning”. In: *IEEE transactions on intelligent transportation systems* 17.4 (2015), pp. 1108–1120.

[28] Hermes J Mora and Esteban J Pino. “Simplified Prediction Method for Detecting the Emergency Braking Intention Using EEG and a CNN Trained with a 2D Matrices Tensor Arrangement”. In: *International Journal of Human–Computer Interaction* (2022), pp. 1–14.

[29] World Health Organization et al. *Global status report on road safety 2018: Summary*. Tech. rep. World Health Organization, 2018.

[30] Sourav Kumar Panwar, Vivek Solanki, Sachin Gandhi, Sankalp Gupta, and Hitesh Garg. “Vehicle accident detection using IoT and live tracking using geo-coordinates”. In: *Journal of Physics: Conference Series*. Vol. 1706. 1. IOP Publishing. 2020, p. 012152.

[31] D. Parker, K. Cockings, and M. Cund. U.S. pat. NO. 9682689. Washington, DC: U.S. Patent and Trademark Office. 2017.

[32] Anna Persson, Hanna Jonasson, Ingemar Fredriksson, Urban Wiklund, and Christer Ahlström. “Heart rate variability for classification of alert versus sleep deprived drivers in real road driving conditions”. In: *IEEE Transactions on Intelligent Transportation Systems* 22.6 (2020), pp. 3316–3325.

[33] Tova Rosenbloom and Amotz Perlman. “Tendency to commit traffic violations and presence of passengers in the car”. In: *Transportation research part F: traffic psychology and behaviour* 39 (2016), pp. 10–18.

[34] Shahram Sattar, Songnian Li, and Michael Chapman. “Road surface monitoring using smartphone sensors: A review”. In: *Sensors* 18.11 (2018), p. 3845.

[35] Thomas Seacrist, Ethan C Douglas, Chloe Hannan, Rachel Rogers, Aditya Belwadi, and Helen Loeb. “Near crash characteristics among risky drivers using the SHRP2 naturalistic driving study”. In: *Journal of safety research* 73 (2020), pp. 263–269.

[36] Mohsen Shirpour, Steven S Beauchemin, and Michael A Bauer. “What Does Visual Gaze Attend to during Driving?” In: *VEHITS*. 2021, pp. 465–470.

[37] Mohsen Shirpour, Nima Khairdoost, Michael Bauer, and Steven Beauchemin. “Traffic Object Detection and Recognition Based on the Attentional Visual Field of Drivers”. In: *IEEE Transactions on Intelligent Vehicles* (2021), pp. 1–1. DOI: 10.1109/TIV.2021.3133849.

[38] David L Strayer, Jonna Turrill, Joel M Cooper, James R Coleman, Nathan Medeiros-Ward, and Francesco Biondi. “Assessing cognitive distraction in the automobile”. In: *Human factors* 57.8 (2015), pp. 1300–1324.

[39] Gito Sugiyanto and Mina Yumei Santi. “Road traffic accident cost using human capital method (Case study in Purbalingga, Central Java, Indonesia)”. In: *Jurnal Teknologi* 79.2 (2017).

[40] Chao Sun, Jian Hua Li, Yang Song, and Lai Jin. "Real-time driver fatigue detection based on eye state recognition". In: *Applied mechanics and Materials*. Vol. 457. Trans Tech Publ. 2014, pp. 944–952.

[41] Farid Talebloo, Emad A Mohammed, and Behrouz Far. "Deep Learning Approach for Aggressive Driving Behaviour Detection". In: *arXiv preprint arXiv:2111.04794* (2021).

[42] Ashish Tawari, Sayanan Sivaraman, Mohan Manubhai Trivedi, Trevor Shannon, and Mario Tippelhofer. "Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking". In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 115–120.

[43] Qun Wang, Weichao Zhuang, Liangmo Wang, and Fei Ju. *Lane keeping assist for an autonomous vehicle based on deep reinforcement learning*. Tech. rep. SAE Technical Paper, 2020.

[44] Cheng Wei, Fei Hui, and Asad J Khattak. "Driver lane-changing behavior prediction based on deep learning". In: *Journal of advanced transportation* 2021 (2021).

[45] Samuel Würtz and Ulrich Göhner. "Driving Style Analysis Using Recurrent Neural Networks with LSTM Cells". In: *Journal of Advances in Information Technology Vol* 11.1 (2020).

[46] Guoliang Yuan, Yafei Wang, Huizhu Yan, and Xianping Fu. "Self-calibrated driver gaze estimation via gaze pattern learning". In: *Knowledge-Based Systems* 235 (2022), p. 107630.

[47] S.J. Zabihi, S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. "Detection and recognition of traffic signs inside the attentional visual field

of drivers". In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2017, pp. 583–588.

[48] S.M. Zabihi, S.S. Beauchemin, and M.A. Bauer. "Real-time driving manoeuvre prediction using IO-HMM and driver cephalo-ocular behaviour". In: *Intelligent Vehicles Symposium (IV), 2017 IEEE*. IEEE. 2017, pp. 875–880.

[49] Fangda Zhang, Shashank Mehrotra, and Shannon C Roberts. "Driving distracted with friends: Effect of passengers and driver distraction on young drivers' behavior". In: *Accident Analysis & Prevention* 132 (2019), p. 105246.

[50] Lanfang Zhang, Boyu Cui, Minhao Yang, Feng Guo, and Junhua Wang. "Effect of using mobile phones on driver's control behavior based on naturalistic driving data". In: *International journal of environmental research and public health* 16.8 (2019), p. 1464.

[51] Yingji Zhang, Xiaohui Yang, and Zhe Ma. "Driver's Gaze Zone Estimation Method: A Four-channel Convolutional Neural Network Model". In: *2020 2nd International Conference on Big-data Service and Intelligent Computation*. 2020, pp. 20–24.

[52] Kawtar Zinebi, Nissrine Souissi, and Kawtar Tikito. "Driver Behavior Analysis Methods: Applications oriented study". In: *Proceedings of the 3rd International Conference on Big Data, Cloud and Application (BDCA 2018)*. 2018.

Chapter 7

Conclusion and Future Work

7.1 Summary and Conclusion

Evidence has shown that drivers play a crucial role in most driving events, and a significant number of vehicle accidents are due to driver error. Hence, researchers and vehicle manufacturers are making efforts to analyze and model driver behavior with different views in different driving situations as well as to predict the most probable next maneuver and assist the driver in avoiding unsafe maneuvers. In Chapter 2, we developed a deep learning-based model to predict five types of driver maneuvers. For this, our model benefited from driver cephalo-ocular behavioral and vehicular dynamics information to do its task. Our experimental results in this work showed that our LSTM-based model outperformed the traditional IO-HMM-based model. In order to prevent potential accidents, this such a system can offer a possible solution for allowing ADAS to alert the driver at an early stage before making a mistake and performing a dangerous maneuver.

In Chapter 3, we developed a vision-based framework that simultaneously detects and recognizes four important classes of traffic objects including vehicles, pedestrians, traffic signs, and traffic lights inside and outside the attentional visual area of the driver. The object detection stage was constructed by a combination of both traditional and deep learning-based models. Finally, the recognition stage was implemented using ResNet101 models. Nowadays, object detection is widely employed in designing ADASs for not only autonomous driving but also ordinary vehicles. For example, the detection of vehicles can avoid accidents and keep a safe distance from surrounding vehicles. Pedestrian detection is significant in reducing fatalities and injuries. Recognition of traffic signs and lights helps vehicles to comply with traffic rules.

In Chapter 4, we presented a CNN-based model to detect and verify lane types in urban and suburban driving environments. We classified various types of lanes as they provide contextual information and indicate traffic rules relevant to driving. Following the detection stage, we used a two-step method to classify the lane boundaries into eight classes, considering road boundaries as one particular type of lane. These mechanisms can help us in designing ADAS applications such as lane keeping assistance, lane departure warning, overtaking assistance as well as intelligent cruise control.

It is generally accepted that a driver cannot attend to the whole traffic environment because of his/her limited gaze area. Moreover, a driver may miss some critical information because of inappropriate driving habits, driving skills, or distractions that affect the choice of proper driving maneuver. In Chapter 5, we developed an analytical vision-based model to estimate average driver attention based on the attentional visual field of the driver by employing several metrics. For this purpose, we also trained a YOLOv5 object detector model on RoadLAB data to identify traffic objects. By utilizing our approach and considering consecutive small periods of time while driving, it is possible to design an ADAS based on the driver's attentional visual area to infer whether the driver is paying enough attention to the traffic objects or whether he/she has been distracted.

In Chapter 6, we presented an approach to measure an average percentage of the time that a driver has gazed at different traffic objects in the course of driving. To reach this purpose, we benefited from a YOLOv5 object detector trained on RoadLAB data, PoG of the driver as well as our proposed metrics. This approach helps us to understand the driver's behavior in terms of the driver's PoG during driving.

Our contributions to the creation of next generation ADAS are summarized as follows:

1. Developing a novel deep learning-based model to predict driver intent.

2. Developing a model to detect and recognize traffic objects within the attentional visual field of the driver.

3. Collecting and annotating a large dataset for different traffic objects and road lanes.

4. Creating a CNN-based method to detect and classify road lanes in urban and suburban areas.

5. Develop an analytical approach to estimate average driver visual attention based on the visual field of the driver.

6. Introducing an analytical approach to measure the average percentage of the time in which a driver has gazed at different traffic objects based on the driver's point of gaze (PoG).

Collectively, our work addresses a number of related challenges in building models of driver behavior for ADAS. Our maneuver prediction model outperforms, is competitive and more reliable in comparison to previous work. It does this by employing an LSTM to keep long-term temporal dependencies, predicting five maneuver types (which are more than those of previous work), and benefiting from gaze information (which is ignored in many previous works). Our object detection and recognition framework was one of the first to simultaneously detect different major classes of traffic objects and, unlike previous work, we also classify them into their own sub-categories. Our deep learning lane detection and classification model is different from previous work in that we consider urban and suburban roads and eight distinct lane types; most previous studies applied their models to highways and had few types of lanes as well as ignoring the road boundaries. To compute driver attention, in addition to the detection of more traffic object types in comparison to the previous work, our model is the first model of its kind that takes advantage of the attentional visual field of the driver to perform its task. Finally, we consider the driver's PoG behavior for multiple object classes, namely, vehicles, traffic lights, traffic signs, and pedestrians. Our model gives us a better understanding of driver visual behavior about what traffic object (or elsewhere) the driver is gazing at directly while in the act of driving.

7.2 Future Work

Research on driver behavior, intent modeling, and their relationship to ADAS has become of great interest in recent years. Our work has contributed in several ways to methods which can be utilized in ADAS. Given our research, we mention several additional possible research areas that may be undertaken in the future.

1. Objects detected inside the attentional visual field of driver can be employed to analyze driver attention in consecutive small periods of time while driving instead of considering an entire sequence as one time unit. For this, it is possible to define a sliding time period and compute driver attention based on the visual field and investigate in what locations and in what driving situations a driver strengthens his/her attention or is distracted. Moreover, there is also an interest to monitor and analyze the driver's behavior using a dashboard camera observing driver's activities during driving to automatically detect driver distraction. More specifically, these ADAS systems by means of analyzing the face and hands of the driver could detect driver distraction and identify the cause of distractions such as cell phone talking, texting, operating the radio, eating, etc. As a result, for detecting driver distraction a future ADAS which incorporates the two aforementioned methodologies to take advantage of both would be more practical and promising in real driving environments.

2. To make the object detector model more comprehensive, bike and motorcycle objects can be added to the dataset as well. As a result, it can be possible to identify more other objects drivers encounter and attend to while driving.

3. Employing a digital street map along with the vehicle's GPS coordinates can provide an intelligent ADAS with more contextual information. In other words, augmenting the vehicle's GPS coordinates with the street map could enable the ADAS to detect upcoming road artifacts such as

intersections and turns and determine whether a turn maneuver is possible or not. Obviously, this information could then enable a prediction model to anticipate driver maneuvers more effectively and efficiently.

4. By employing the video from the forward imaging system and identifying the side lanes in addition to the ego lane, a future ADAS could determine whether a lane exists on the right side and also on the left side of the vehicle. This contextual information would provide additional information for the driver maneuver prediction system. For instance, when the vehicle is moving in the left-most lane, the only safe maneuvers are going straight and right lane change, unless it is approaching an intersection.

5. The limitations of the employed experimental instruments did not allow us to use them at night and in adverse weather conditions. Such limitations can be eliminated by a proper choice of hardware, requiring further research on driver behavior in these conditions. In other words, to further put ADAS models into real-life use in the future, the models should be applicable and accurate at different times of day and in different weather conditions.

6. Although the experimental instrumentation demonstrates a successful proof of concept, the use of wider angles of viewing for stereo cameras and eye-trackers may be very helpful in enhancing the analysis of gaze across a wider area as well as compensating for head rotations which could enable the system to track driver gaze more comprehensively. The use of multiple cameras could also help in enhancing gaze tracking. Consequently, future ADASs can use this valuable information in different applications such as more accurately assessing the driver's visual atten-

tion.

7. Another direction is collecting a comprehensive and naturalistic dataset to design future ADASs for applications such as driver maneuver prediction and driver attention evaluation. Providing a larger volume of naturalistic driving data should include at least three aspects. First, the data used to build predictive driving models should include data from a range of driving settings and scenarios, including downtown, urban, suburban, and highway (the RoadLAB dataset does not include the highway scenarios). Second, the present studies in this thesis were restricted to only one city (RoadLAB data was collected in London, Ontario, Canada), therefore extending the data collection to other cities would be a further enhancement of the dataset. Third, it would also be interesting to comprehensively investigate the impact of different driving styles. Behavioral modeling approaches could be applied to find driver models considering different behavioral styles such as normal, aggressive, inexperienced, etc. Of course, since such driving data cannot be easily collected for some of those cases, employing realistic driving simulators might need to be used. Collectively, more complex driver behaviors and road structures can be included in the new dataset, making the improved model more suitable and practical for real-world driving situations and scenarios. This ability would be essential for future ADASs to understand and predict driver behavior and accordingly provide the driver with the proper assistance.

Vita

NAME:	Nima Khairdoost
POST-SECONDARY EDUCATION AND DEGREES:	University of Western Ontario London, Canada 2017-2022 PhD in Computer Science
	University of Isfahan Isfahan, Iran 2008-2011 M.Sc. in Computer Eng.
	Ferdowsi University Mashhad, Iran 2004-2008 B.Sc. in Computer Eng.
AWARDS:	WGRS scholarship for 2017-2021, University of Western Ontario
RELATED WORK EXPERIENCE:	Faculty Member in Computer Department Tabaran Institute of Higher Education 2013-2017
	Teaching Assistant University of Western Ontario 2017-2022
	Research Assistant University of Western Ontario 2017-2022

Publications

[1] Zohreh Davarzani, Mohammad-R Akbarzadeh-T, and Nima Khairdoost. "Multiobjective artificial immune algorithm for flexible job shop scheduling problem". In: *International Journal of Hybrid Information Technology* 5.3 (2012), pp. 75–88.

[2] N. Khairdoost, S.A. Monadjemi, and K. Jamshidi. "Front and rear vehicle detection using hypothesis generation and verification". In: *Signal & Image Processing* 4.4 (2013), p. 31.

[3] N. Khairdoost, M. Shirpour, M.A Bauer, and S. S Beauchemin. "Real-Time Maneuver Prediction Using LSTM". In: *IEEE Transactions on Intelligent Vehicles* (2020).

[4] Nima Khairdoost, Steven S Beauchemin, and Michael A Bauer. "An Analytical Model for Estimating Average Driver Attention Based on the Visual Field". In: *7th International Conference on Signal and Image Processing (ICSIP)*. IEEE. 2022.

[5] Nima Khairdoost, Steven S Beauchemin, and Michael A Bauer. "Road Lane Detection and Classification in Urban and Suburban Areas based on CNNs." In: *VISIGRAPP (5: VISAPP)*. 2021, pp. 450–457.

[6] Nima Khairdoost and Nayereh Ghahraman. "Term rewriting for describing constrained policy graph and conflict detection". In: *2010 IEEE International Conference on Progress in Informatics and Computing*. Vol. 1. IEEE. 2010, pp. 645–651.

[7] Nima Khairdoost, S Amirhassan Monadjemi, Zohreh Davarzani, and Kamal Jamshidi. "GA Based PHOG-PCA Feature Weighting for On-RoadVehicle Detection". In: *International Journal of Information and Electronics Engineering* 3.1 (2013), p. 104.

[8] Nima Khairdoost, Mohammad Reza Baghaei Pour, Seyed Ali Mohammadi, and Mohammad Hoseinpour Jajarm. "A ROBUST GA/KNN BASED HYPOTHESIS VERIFICATION SYSTEM FOR VEHICLE DETECTION". In: *International Journal of Artificial Intelligence & Applications* 6.2 (2015), p. 21.

[9] Hoda Khoshnevis and Nima Khairdoost. "Using Simulation Applications for Sustainable Design and Construction". In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. Vol. 33. IAARC Publications. 2016, p. 1.

[10] Hadi Sarvari, Nima Khairdoost, and Abdolvahhab Fetanat. "Harmony search algorithm for simultaneous clustering and feature selection". In: *2010 International Conference of Soft Computing and Pattern Recognition*. IEEE. 2010, pp. 202–207.

[11] Mohsen Shirpour, Nima Khairdoost, Michael Bauer, and Steven Beauchemin. "Traffic Object Detection and Recognition Based on the Attentional Visual Field of Drivers". In: *IEEE Transactions on Intelligent Vehicles* (2021), pp. 1–1. DOI: 10.1109/TIV.2021.3133849.

[12] Alireza Tofighi, Nima Khairdoost, S Amirhassan Monadjemi, and Kamal Jamshidi. "A robust face recognition system in image and video". In: *International Journal of Image, Graphics and Signal Processing* 6.8 (2014), p. 1.