THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Towards Safe Autonomous Driving

From predictive safety evaluation to monitoring of neural networks

Arian Ranjbar

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2021

Towards Safe Autonomous Driving

From predictive safety evaluation to monitoring of neural networks

Arian Ranjbar
ISBN 978-91-7905-604-9

Technical Report No. 5070
ISSN 0346-718X
This thesis has been prepared using LaTeX.

Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000
www.chalmers.se

Printed by Chalmers Reproservice
Gothenburg, Sweden, 2021

Abstract

Autonomous driving is expected to bring several benefits, in particular regarding safety. This thesis aim to contribute towards two questions concerning safety: "What is the potential safety benefit of autonomous driving?" and "How can we ensure safe operation of such vehicles?".

In the first part of the thesis, methods for evaluating the safety benefit are investigated. In particular predictive effectiveness evaluation based on resimulation of accident data, using models to estimate new outcomes in case the safety system had been available. To illustrate the methodology, four examples of gradual increase in model complexity are presented. First, an Autonomous Emergency Braking (AEB) system using a sensor model, decision algorithm, vehicle dynamics model and regression based injury model. This is extended in a Forward Collision Warning (FCW) system which additionally requires a driver model to simulate driver reactions. The third example shows how an active, AEB, and passive, airbag, system can be combined.Finally the fourth example combines several systems to emulate a highly automated vehicle. Apart from predicting the real world performance, this analysis also identifies current safety gaps by studying the residual of the accident set.

Safety benefit estimation using accident data gives an evaluation on the current accident distributions, however, the systems may introduce new accidents if not operated as intended. In the second part of the thesis, safety verification processes with the intent of preventing unsafe operation, are presented. This is particularly challenging for machine learning based components, such as neural networks. In this case, traditional analytical verification approaches are difficult to apply due to the non-linearity and high dimensional parameter spaces. Similarly, statistical safety arguments often require unfeasible amounts of annotated validation data. Instead, monitor functions are investigated as a complement to increase safety during operation. The method presented estimates the similarity of the driving environment, compared to the training data, where decisions inferred from novel data can be considered less reliable. Although not providing a complete safety assurance, the methodology show promising initial results for increasing safety. In addition, it could potentially be used to collect novel data and reduce redundancy in training data.

Keywords: Autonomous driving, safety benefit, effectiveness, predictive evaluation, verification, monitoring, neural networks.

List of Publications

This thesis is based on the following publications:

[A] Olaf Op den Camp, **Arian Ranjbar**, Jeroen Uittenbogaard, Erik Rosen, Stefanie de Hair-Buijssen, "Overview of main accident scenarios in car-to-cyclist accidents for use in AEB-system test protocol". Published in Proceedings of International Cycling Safety Conference, Nov. 2014.

[B] **Arian Ranjbar**, Nils Lubbe, Erik Rosen, Jonas Fredriksson, "Car-to pedestrian forward collision warning revisited: A safety benefit estimation". Submitted for review in journal publication, Nov. 2021.

[C] Rikard Fredriksson, **Arian Ranjbar**, Erik Rosen, "Integrated Bicyclist Protection Systems-Potential of Head Injury Reduction Combining Passive and Active Protection Systems". Published in 24th International Technical Conference on the Enhanced Safety of Vehicles, June 2015.

[D] Nils Lubbe, Hanna Jeppsson, **Arian Ranjbar**, Jonas Fredriksson, Jonas Bärgman, Martin Östling, "Predicted road traffic fatalities in Germany: The potential and limitations of vehicle safety technologies from passive safety to highly automated driving". Published in Proceedings of IRCOBI conference, Sept. 2018.

[E] **Arian Ranjbar**, Chun-Hsiao Yeh, Sascha Hornauer, Stella X. Yu, Ching-Yao Chan, "Scene Novelty Prediction from Unsupervised Discriminative Feature Learning". Published in IEEE 23rd International Conference on Intelligent Transportation Systems, Sept. 2020.

[F] **Arian Ranjbar**, Sascha Hornauer, Jonas Fredriksson, Stella X. Yu, Ching-Yao Chan, "Safety Monitoring of Neural Networks Using Unsupervised Feature Learning and Novelty Estimation". Submitted for review in journal publication, June 2021.

[G] Sascha Hornauer, Baladitya Yellapragada, **Arian Ranjbar**, Stella X. Yu, "Driving Scene Retrieval by Example from Large-Scale Data". Published in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

Acknowledgments

I would like to express my gratitude towards my academic supervisor Prof. Jonas Fredriksson, who has accompanied me on this journey from start to finish, always providing support even in the most turbulent times. I would also like to thank Dr. Erik Rosén and Dr. Nils Lubbe, laying out the foundation of my doctoral studies, in particular safety evaluation research; along with my other industrial supervisors throughout the years. I am also thankful towards my examiner Prof. Jonas Sjöberg, welcoming me to the group, supporting and challenging me through the studies.

This would not have been possible without my employers Autoliv, Zenuity and Zenseact; and my former managers: Dr. Erik Rosén, Benny Nilsson, Camilla Apoy and Dr. Mats Nordlund; thank you. I would also like to thank my coach Dr. Carl Lindberg for your great advice and support, I will remember to celebrate my achievements in the future. In addition, I would like to thank all my colleagues at Autoliv, Zenuity, Zenseact, Chalmers and AI Sweden for the stimulating work environment. Especially, my office-mate and friend Ivo Batkovic, always keeping a great spirit in "skrubben" (our office) even during the toughest times.

I would like to express my deepest gratitude to all my supervisors and colleagues at UC Berkeley, including: Prof. Ching-Yao Chan for welcoming me to his group and Berkeley DeepDrive, both developing my research skills and giving new perspectives; Prof. Stella Yu for your supervision and inspiring talks; and Dr. Sascha Hornauer for your guidance, great collaborations and fruitful discussions.

Finally, I would like to thank all my friends and family who have supported me throughout this time.

Acronyms

ABS:	Anti-lock Braking System
ACC:	Adaptive Cruise
AE:	Autoencoder
AEB:	Autonomous Emergency Braking
AIS:	Abbreviated Injury Scale
CNN:	Convolutional Neural Network
DNN:	Deep Neural Network
ESC:	Electronic Stability Control
FCW:	Forward Collision Warning
GAN:	Generative Adversarial Network
GIDAS:	German In-Depth Accident Study
HAD:	Highly Automated Driving
HIL:	Hardware In the Loop
HMI:	Human Machine Interface
ISA:	Intelligent Speed Adaption
LCA:	Lance Change Assist
LKA:	Lane Keep Assist
MIL:	Model In the Loop
ML:	Machine Learning
NN:	Neural Network
SIL:	Software In the Loop

Contents

Part I

Overview

CHAPTER 1

Introduction

Road traffic accidents are one of the ten largest global health problems according to the World Health Organization (WHO), claiming approximately 1.3 million lives per year, [1]. For young people, aged 15-29, it is the leading cause of death globally. Although the numbers are high, the rate of death per traffic participants are decreasing, as seen in Figure 1.1. This can, to a great extent, be attributed to improvement of infrastructure, building safer roads and the introduction of passive and active safety systems together with legislation.

Passive safety systems have been in development from the early stages of car manufacturing and aim to protect occupants of the vehicle during a crash. Seat belts were predicted to have an effectiveness of over 60% in reducing fatalities during the 1970s, [3]. Since then several advancements such as pre-pretensioner and load limiter have further increased the effectiveness, [4]. The introduction of the driver airbag showed promising results of preventing fatalities by 18%, [5], for frontal crashes. The following development of the passenger airbags, inflatable curtain, knee airbags, windshield airbag and other types of airbags; together with advancements in airbag fabric and inflation techniques have further increased the effectiveness, targeting all types of col-

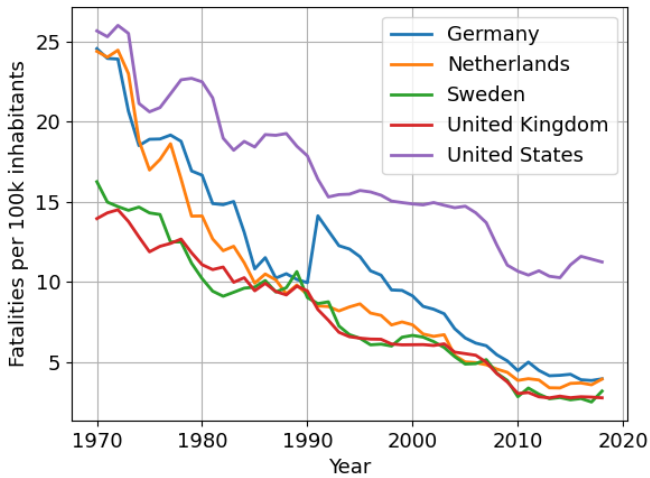

Figure 1.1: Road traffic fatalities over time for some developed countries, obtained from [2].

lisions, [6], [7], [8].

During the last decades there has been an acceleration in development of active systems, further improving safety by preventing accidents from occurring or by mitigating consequences, see e.g. [9] and [10] for comprehensive surveys. One of the first systems introduced under this category is the anti-lock braking system (ABS). ABS prevents brake lock-up by monitoring rotational speed of the wheels. By automatically reducing brake force in case the wheels stops rotating, it allows the vehicle to be steered during hard braking. This may help preventing certain types of accident scenarios, and studies have showed a significant statistical decrease in such accidents with the help of ABS, [11]. Several advancements have been made since the ABS was introduced. Electronic Stability Control (ESC) builds on the concept by implementing sensors to detect skids, allowing for individual braking of each wheel to regain control. In this way ESC reduce single-vehicle crash involvement risk by up to 40%, [12].

More recently several driver assistance systems have been introduced, in addition to the previously discussed vehicle dynamics based control systems.

Adaptive Cruise Control (ACC) helps the driver to maintain distance to vehicles in front, by using a forward looking sensor, [13]. Intelligent Speed Adaption (ISA) extends to adapt to a safe or legally enforced speed, with studies showing a reduction in fatal accidents of 37%-59% by using the system, [14]. Similarly, Lane Keep Assist (LKA) and Lane Change Assist (LCA) uses a forward looking sensor to warn the driver of lane departure or unsafe lane changes; with corresponding fatality reductions of 5%-15% ([10], [15]) and 1%-2% ([10], [16]). Forward Collision Warning (FCW) systems tries to predict potential collisions, warning the driver to take precaution by braking or otherwise interact. Such systems have shown great potential in reducing car-to-car collisions such as front-to-rear end crashes, [17].

An alternative to FCW is to allow the car to automatically brake when potential collisions are detected with high probability or considered unavoidable, using Autonomous Emergency Braking (AEB). Although AEB shows great safety effectiveness potential in several types of accidents such as rear-end ([10], [17]), car-to-pedestrian ([18]) and intersection collisions ([19], [20]); it puts a higher constraint on system verification procedures. A false positive prediction may no longer be ignored by the driver since the vehicle autonomously react to the trigger, potentially leading to hazardous situations. To mitigate such risks several verification procedures have been implemented, such as formal verification of the system, [21], and statistical verification by real world driving or scenario testing, [22], [23].

The next step, to increase safety, is to fully automate the driving. Studies have shown that human errors are the cause of up to 90% of all traffic accidents, [24], [25], [26]; giving such technologies the potential to significantly reduce the number of accidents and road fatalities. Although some studies claim the net result of such technology would be a decrease in traffic safety, in particular during the initial periods of mixed vehicle fleets, [27]; or due to other factors such as increased travel time [28]; others claim there is a need for further evaluation to fully understand the extent of the potential safety benefit, [29], [30]. Due to the limit in data, since no highly automated driving systems are widely available on the market; such investigations can instead be done through predictions on current accident data, [31], in a procedure similar to predicting the effects of single active safety systems.

As for active systems such as the AEB, an autonomous vehicle would require even more comprehensive verification procedures to ensure safe operation. The

safety assurance and verification is one of the key challenges to get autonomous vehicles on the road [30], [32], [33]. The difficulty can be attributed to several sources: infeasibility of complete testing, a fleet of 100 autonomous vehicles would require more than 500 years of driving to demonstrate the failure rate is 20% better than the human driver, with 95% confidence, [34]; controllability challenges, since no driver can be used as backup the software needs to handle all failures and exceptions requiring even higher safety integrity levels, leading to architectural challenges, [35]; non-deterministic and statistical algorithms, and in particular black-box machine learning based systems, [32]; among other things, see e.g. [32] for a detailed survey.

This thesis concern the challenges discussed in the last two paragraphs: evaluating the safety benefit of a highly automated vehicle; and contributing to the safety assurance, with a focus on machine learning based components.

1.1 Problem Formulation

The aim of the work presented in this thesis is to contribute to the answers and methods for answering the following questions:

- What is the safety benefit of a driving function or safety system, and in particular a highly automated vehicle?

- How can we ensure safe driving, i.e. not introduce new accidents, with highly automated vehicles?

The answers to the first question not only provide information on the usefulness of a particular system, assisting regulations and planning research; it may also help directing ongoing development. This can be done by analysing the residuals of the accident scenario coverage, identifying gaps and their corresponding sources. Covering potential gaps combined with ensuring no new accidents are caused by the system, as treated by the second question; both current and potential new accident scenarios are covered. Thereby the second question also helps guiding towards achieving the potential safety benefit and safe autonomous driving.

Delimitations

Due to the difficulty of statistically testing the safety benefit of a highly automated vehicle, as discussed in the introduction, and the lack of available data; the first question is addressed by trying to predict the safety benefit. This is done through the use of accident databases, in particular German accident data; together with approximating models for each subsystem. In particular Paper B and Paper C presents effectiveness estimations of a car-to-pedestrian FCW and AEB, and AEB combined with airbags for car-to-bicyclist accidents; to present typical methodology for such evaluations. While not contributing directly to the first question they provide estimations of potential subsystems of a highly automated vehicle. Paper D directly address the first question, but limits the study to model several safety systems which combined emulate a highly automated car, through a rule based algorithm. Such evaluation only provide a rough estimate to the potential safety benefit, and is only valid for the region represented by the data, although the methodology is generalizable. In addition, due to the use of accident data, no consideration is made towards potential new accidents not yet available. One of the aims of the second question is to minimize the risk of introducing such accidents.

For the second question, this thesis is primarily looking in to ensuring safety of black box machine learning techniques, in particular neural networks, commonly used in many sub-systems of highly automated vehicles. It is further delimited to investigate the potential of using monitoring techniques for this purpose. An elaboration on the choice of methodology can be found in Chapter 3.

Contributions

In relation to the first question, this thesis presents methods for incorporating accident data into the development of safety systems (Paper A) and estimating the potential safety benefit of active safety systems. First for a single active system, which also involves driver interactions (Paper B). Then it is shown how an active and passive system can be combined (Paper C), in order to finally combine several systems to emulate a highly automated vehicle (Paper D). The latter puts an estimate on the potential of reducing road fatalities and also analyses the unresolved scenarios for possible improvements of the system. Chapter 2 of the thesis also provides a formalization of the general

methodology applied in these studies.

Approaching the second question, a method to monitor machine learning based black box methods is developed, based on estimating the novelty of the current driving environment compared to the training data (Paper E & F). Finally, the methodology is also used to retrieve driving scenes from unlabeled data, helping debias and remove redundant data (Paper G).

1.2 Thesis outline

The thesis consists of two parts. Part I serves as a general introduction to Part II and has the following structure. The first chapter contains an introduction and background to safety systems in road vehicles, problem formulation and outline. The second chapter provides a brief introduction to accident databases and safety benefit estimations. This puts the first four appended papers in context and the build up to answering the first question. The third chapter gives an introduction to verification of automated driving functions, and in particular monitoring. This puts the last three papers into the context of contributing to the second question of the problem formulation. Chapter four contains a brief summary of each paper and their findings, with a corresponding discussion in the fourth chapter. Part II presents the main part of the thesis, consisting of the seven papers.

CHAPTER 2

Safety Evaluation

As discussed in the previous chapter, extensive research has shown that passive and active safety systems significantly improved road traffic safety. This chapter will give a brief introduction to different evaluation methods. In particular how accident data can be used for system design (as in Paper A) and predictive assessment of safety systems and highly automated driving (as in Paper B-D). To illustrate the methodology three examples will be given evaluating a single active safety system, a combination of an active and passive system and finally several systems together emulating highly automated driving.

2.1 Methods

Experimental testing

The introduction of passive safety systems brought the need for evaluation methods, both for regulatory purposes and independent rating for consumers. The first evaluations were carried out through experimental testing. Today such experiments involve crash tests performed on dedicated test tracks, using

standardized crash dummies modeling humans. For passive safety systems this allow for measurements of reductions in forces on different body parts; or specially designed quantities such as the Neck Injury Criterion, [36]; through the crash dummies. In Europe experimental testing has been carried out by EuroNCAP on consumer vehicles since 1997, [37].

From 2009 EuroNCAP started to rate ESC, as the first active safety system evaluated through testing. Since then, speed assistance systems and AEB for vulnerable road users are also included in the test protocol, [38]. Such testing is done through a set of exemplary scenarios meant to cover a large proportion of typical accidents, similar to using different load cases in the testing of passive safety. The effectiveness is then measured in the ability of handling each test scenario.

The challenge of experimental testing lies in producing test scenarios representative for typical operation and exposure of traffic situations. For in-crash passive systems this amounts to typical load and acceleration in collisions. However, for active safety the possible configurations and parameter space of pre-crash scenarios is much larger. Typically testing scenarios are derived through the use of statistical methods applied on accident data, [39], further explored in section 2.2.

Simulation

For faster iteration and testing without physical damage, virtual models have been developed for evaluation. For passive safety this include Finite Element (FE) models of hardware, including crash dummies used in the physical experiments. In addition, full human body models have been developed for FE simulations allowing for more accurate modeling of injuries in collisions, [40].

Similarly scenario testing for active safety systems can be done through simulations, requiring environment-, vehicle-, system- and driver models. Environment models are used to simulate a dynamic traffic environment and may be constructed with different levels of detail depending on the system specifications. In case low level sensor models are used a higher constraint is put on the fidelity of the graphical representations. In addition the environment model need to accurately model each traffic participant apart from the host, which may include other drivers and vehicles as well as pedestrians. A variety of simulation environments have been developed with different fidelity for various applications, see e.g. CARLA, [41], and PreScan, [42].

Vehicle models describe the vehicle dynamics of the host car. Depending on the evaluated system a range from bicycle model to advanced approaches including two-track models with non-linear tire-models, see e.g. [43]. System models are used to describe the safety system, which in turn may consist of several subsystems depending on setup; such as sensor models and collision prediction algorithms, see e.g. [43] and [44]. Finally, the driver model is used to emulate driver behaviour in case the system require input, such as for warning based systems. The behaviour may depend on various parameters, for instance distraction and type of Human Machine Interface (HMI), [45], [46].

Statistical Evaluation

Statistical evaluation use accident data to quantify and statistically derive effectiveness of systems. In passive safety, analysis may be carried out by analyzing individual crashes, comparing occupants affected and unaffected by a system; such as belted and unbelted passengers of the same car, [47]; or with and without airbag, [48].

Another statistical technique, suitable for active safety, is evaluating prevalence of cars with and without a system in a specific type of accident. The challenge with this method lies in the statistical ground work, using control groups and corrections to account for different types of driver behaviours etc. In addition, it is most suitable for systems already available in the market where extensive data is obtainable, as for the ESC, [49]. However, as more and more systems get added to the vehicles, the causality in between a specific system and reduction in accident frequency may be harder to derive.

Predictive Assessment

The aim of predictive assessment is to allow for effectiveness evaluations of systems not yet available for the consumer. This can be done through a combination of statistical and simulation evaluation methods, where scenarios from accident databases are resimulated with models for the system under consideration. Such evaluations have been made for systems like AEB, [18], [20]; and other recent technologies where data is too limited to have statistical significance or to establish causality, [29]; making it suitable for the evaluation of highly automated driving.

In the following sections the use of accident data to design and perform predictive assessment will be discussed.

2.2 Accident Statistics

On a fundamental level, consider the set of all *traffic scenarios* $\mathcal{T}$. In general, when developing a safety system, we are interested in a particular subset of traffic scenarios that lead into accidents $\mathcal{A} \subset \mathcal{T}$, for example car-to-pedestrian accidents, or car-to-car collisions in intersections etc. Properties or outcomes of these accidents can be represented by $\xi_m : \mathcal{T} \to I_m$ for some set of values I_m, for example the impact speed $\xi_v : \mathcal{A} \to \mathbb{R}$ or whether a particular participant received a severe injury or not $\xi_{\text{SI}} : \mathcal{T} \to \{0, 1\}$.

In reality, the challenge lies in getting accurate representations of the accident sets $\mathcal{A}$, or samples of the corresponding properties, ξ; as with all statistical studies. The following sections will discuss the use of accident databases and how such data can be used for system development.

Accident databases

During the last decades there has been a significant increase in the use of accident data when designing automotive safety systems. Most European countries collect some form of accident data, see e.g. LAB [50] in France, German national road traffic accident statistics [51], BRON Netherlands national road crash register [52], Swedish Traffic Accident Data Acquisition [53] and STATS19 Road Accident dataset [54] in UK among others. Most of these databases rely on police reports to extract data regarding each accident and/or incident. To increase the information available for each accident as well as increasing the accuracy and depth, some regions have started separate research teams with the purpose of investigating accidents to store in databases. One of the most detailed databases of this kind is the German In-Depth Accident Study (GIDAS), [55].

GIDAS started in 1999 with the goal of collecting at least 2000 accidents per year. The data collection team operates within two areas in Germany, Hanover and Dresden, and operates under a strict sampling plan such that the distribution of the collected accidents accurately represents Germany. This includes the choice of areas which covers both cities, rural areas and different

types of roads. Every time an accident occurs within these regions and the responding police suspects that at least one person is injured, they contact the GIDAS team. The team, consisting of two technicians, a physician and a coordinator; then immediately travel to the accident site to gather information. The data collected are later stored in the database and contains information about the environment, accident events, information about the vehicles and even reconstructions providing impact speeds and geometrical configurations when possible. It also contains detailed information about all participants, particularly all injuries reported through the Abbreviated Injury Scale (AIS); which is a 6 point scale of severity, [56].

In addition, GIDAS carries out further reconstruction on a subset of the accidents including trajectories of the participants (and stationary objects) enabling resimulation off accidents. This information is provided through an extension of the database called the Pre-Crash Matrix (PCM).[57]

System design

Having access to detailed accident databases can greatly assist in developing safety systems. By understanding the context in which the addressed type of accident occur, a more efficient system can be designed. The accident data can be divided into pre-crash, in-crash and post-crash variables. Since the goal of a passive system is to mitigate the effects of a crash, the in-crash and post-crash information is of greatest use. While for active systems, trying to prevent accidents from occurring, the pre-crash information is of most importance.

To illustrate the usefulness of such data an example of an AEB system targeting car-to-bicyclist accidents will be discussed (based on Paper A). One of the first steps in the development chain is to understand the pre-crash scenario of the targeted accidents. Such information determines the required sensor setup, requirements on tracking and prediction collision algorithms etc. Typically there is an infinite amount of pre-crash scenarios with respect to geometrical configuration, road setups and environmental conditions. By defining a set of finite accident classes, the infinite number of setups can be reduced to this finite set of configurations, and their corresponding frequencies can be quantified. A simple way to construct such classes, uniquely assigning one class for every scenario, is illustrated in Figure 2.1. Querying the database with this classification, in this example GIDAS, shows that over 80% of severe injuries and fatalities can be prevented by addressing the scenarios

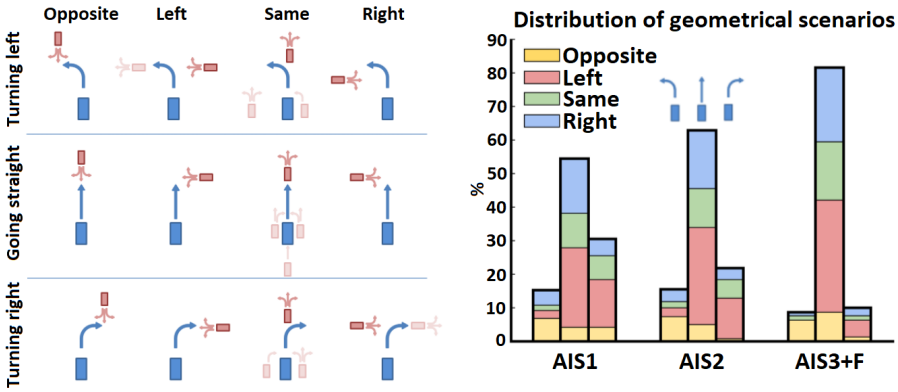

Figure 2.1: Collision classification based on the action of the car and location of the bicyclist pre-collision. The 12 scenario classes are illustrated to the left and the result of the classification can be seen to the right. Each bar correspond to the car turning left, moving straight forward and turning right in that order. Each group of bars corresponds to accidents resulting in a particular AIS injury level.

where the car is going straight, [58]. To expand the contextual understanding, additional information can be added such as weather condition, impact speeds, road type and other environmental variables. Although it is desirable to prevent all accidents, the largest clusters can be addressed first, adding incremental complexity to the system (such as more sensors) to eventually reach full coverage.

2.3 Predictive Safety Benefit Analysis

In the previous section it was shown how accident data can be included in the development of safety systems.This section will present how accident data can be used to predict the real world performance of such a system, by resimulating accidents and evaluating new outcomes.

Introducing a safety system $C : \mathcal{T} \to \mathcal{T}$ we alter the accidents where the system is applied, resulting in new outcomes. For a particular safety system C, the effectiveness in reducing a binary property ξ, is defined as the relative

proportion of ξ with and without the system,

$$E = 1 - \frac{\mathbb{E}_{a \in \mathcal{A}}[\xi(C(a))]}{\mathbb{E}_{a \in \mathcal{A}}[\xi(a)]}, \tag{2.1}$$

assuming $\xi(C(a)) \leq \xi(a)$ for all $a \in \mathcal{T}$. i.e. that the system does not introduce new accidents or make the situation worse with respect to the desired property. Typically ξ is defined as whether a participant of the traffic situation received a (fatal) injury or not.

Again the challenge with this setup is to have an accurate representation of the accident set, while also being able to model the changes in the outcome when applying the safety system. To illustrate the methodology, three examples will be given estimating the safety benefit of one active safety system, an active system together with a passive system and finally several systems together emulating a highly automated vehicle, corresponding to Paper B, C and D.

Safety benefit of an active safety system

In this example the methodology of evaluating the safety benefit of an AEB, C_{AEB}, and FCW, C_{FCW}, system for car-to-pedestrian accidents will be discussed. In particular the effectiveness in reducing severe injuries,ξ_{SI}, of the pedestrians in such collisions.

Again GIDAS will be considered as a sample for the accident set, $A = \{a_1, a_2, ..., a_n\}$. To increase the accuracy of the sample it can further be debiased by looking into German national road traffic accidents. Since the GIDAS investigation team is contacted only when the police suspect that at least one involved participant is injured, there is a slight overrepresentation of accidents with severe injuries. Comparing the severity, weight factors, $\{w_i\}$ can be derived to accurately estimate the effectiveness in equation (2.1) using weighted expected value.

Autonomous Emergency Braking

As discussed in section 2.1, simulation of accident scenario require a set of models for the environment, vehicle dynamics, safety system and; in the case of FCW, a driver model. The environmental model in this case simply consists of geometrical representations in a two dimensional overview of the accident.

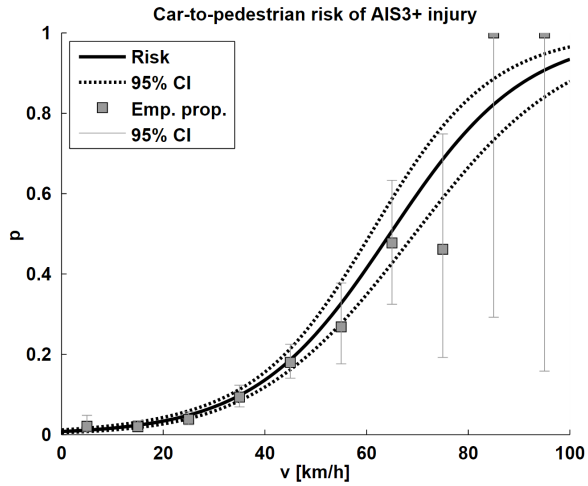

Figure 2.2: Probability of AIS3+ injuries being sustained by the pedestrian with respect to the impact speed of the car.

This allow for a high level model of the sensor to determine whether objects are within field of view of a forward looking sensor, and not obstructed. As vehicle model, a simple bicycle model provide accurate enough results in this particular example. Finally, the AEB system is modeled through a vision system capable of tracking non-occluded pedestrians, visible within at least three consecutive frames (running at 20 Hz). In addition, the decision algorithm consists of a simple geometrical deterministic algorithm predicting collision by extrapolation of travel path. If the car is on a collision path with a pedestrian and reaches a particular time to collision, a decision to brake is made. Using the bicycle model a new trajectory is calculated, using brake profile curves to calculate brake force, to determine new impact speeds, $v^{\text{AEB}} = \hat{\xi}_v$.

The impact speed is of particular interest since it is the main contributing factor to severe injuries in these types of collisions, [59]. By using GIDAS and logistic regression, a model can be derived estimating the probability of receiving a severe injury given the impact speed, $\mathbb{P}(\xi_{\text{SI}} = 1|\xi_v)$, as seen in Figure 2.2. Combining this estimate with equation (2.1), the effectiveness of

the AEB in reducing severe injuries can be estimated through,

$$\hat{E} = 1 - \frac{\sum_{i=1}^{n} w_i \hat{\xi}_{\mathrm{SI}}(v_i^{\mathrm{AEB}})}{\sum_{i=1}^{n} w_i \hat{\xi}_{\mathrm{SI}}(v_i)} \tag{2.2}$$

where v_i is the original impact speed from the GIDAS reconstruction.

Forward Collision Warning

The benefits of using an FCW instead is the possibility of relaxing the trigger requirements. Using more conservative values on the bounding boxes of the traffic participants and greater values for TTC thresholds. Such calculations lead to a greater number of false positives, which is less of a concern when warning the driver rather than emergency braking. In addition the warning system has the possibility of generating brake pressure when the warning is issued, applying full brake force immediately when the driver brakes. A driver model presented in [46] gives different reaction times of a driver depending on attentiveness and HMI. Each scenario is resimulated using different conditions of attentiveness, and the effectiveness is evaluated with the different impact speeds weighted against the probability of the attentiveness,

$$\hat{E} = 1 - \frac{1}{\sum_{i=1}^{n} w_i \hat{\xi}_{\mathrm{SI}}(v_i)} \sum_{i=1}^{n} w_i \Big(\mathbb{P}(\text{attentive}) \hat{\xi}_{\mathrm{SI}}(v_i^{\text{attentive}}) + \mathbb{P}(\text{distracted}) \big(\mathbb{P}(\text{brake}) \hat{\xi}_{\mathrm{SI}}(v_i^{\text{brake}}) + \mathbb{P}(\text{nobrake}) \hat{\xi}_{\mathrm{SI}}(v_i^{\text{nobrake}}) \big) \Big).$$

Safety benefit of an active + passive system

In this example an active and passive safety system is combined, in particular an AEB for car-to-bicyclist accidents combined with a windshield airbag, protecting the bicyclist in case of collision, $C = C_{\mathrm{AEB}} \circ C_{\mathrm{airbag}}$. Since the active system only affect the pre-crash scenario while the passive system affect the in-crash scenario, they can be modelled separately. New impact speeds, v_i, can be estimated as in the AEB simulation in the previous example, but this time with car-to-bicyclist data from GIDAS. This gives us the probability of receiving a severe injury without any passive system, $\hat{xi}_{\mathrm{SI}}(v)$ (see Figure 2.3).

Through empirical derived risk reduction curves, we get the probability that the passive system will prevent a severe injury given the impact speed,

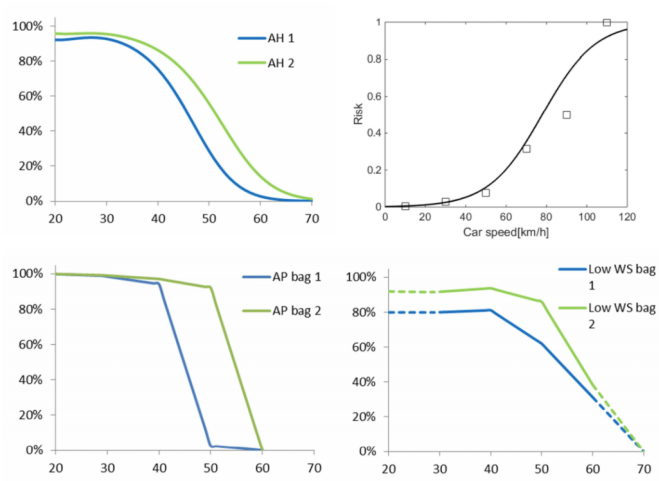

Figure 2.3: Risk reduction of head AIS3+ injury as function of impact speed, for active hood (top left) and windshield airbag in A-pillar (below left) and lower windshield (below right). Probability of AIS3+ head injury given impact speed without passive system (top right).

$p_{\text{passive}}(v)$ (see Figure 2.3), [60]. This finally gives the combined probability of receiving a severe injury for the bicyclist of $\hat{\xi}_{\text{SI}}^{\text{total}} = (1 - p_{\text{passive}}(v))\hat{\xi}_{\text{SI}}(v)$, and the effectiveness can be estimated through equation (2.2).

Safety benefit of highly automated driving

In this example the avoidability of accidents using a highly automated vehicle will be considered. In general, highly automated driving can be seen as a combination of several active and passive systems, $C = C_1 \circ ... \circ C_k$. Resimulation of accident scenarios drastically increase in complexity with the number of subsystems involved. In addition, combined effects such as in the previous example are much harder to derive when several systems are interacting. When combining a single active and passive system, only one system is engaged at a time, the active system during the pre-crash and the passive system during in-crash, $C(a) = C_{\text{AEB}} \circ C_{\text{airbag}}(a) = C_{\text{airbag}}(C_{\text{AEB}}(a))$. For this case, several systems are engaged at the same time.

Rather than relying on resimulation using dynamic models, each system can instead be represented by a set of simple deterministic logical rules. Each

Frontal airbags	klassece == 1 vdi2 == 1 airbf == 0 sreihe == 1 & (squer == 1 | squer == 9) ais98reg1 == [1:6] | ais98reg2 == [1:6] (vdi6 <= 3 & vdi6 >= 0)	M1 vehicle & Frontal impact & airbag not present & first-row occupant & injury to head present & non-catastrophic vehicle damage
AEB cyclist	klassece == 1 artteil == 5 v0 <= 40 & v0 >= 5 v0 <= 30 sichtbv ~= [1, 4] strob ~= [6,7] strzb ~= 4 schleu == 2 nied == 2 & (wolk ~= 7 & nebelm ~= [3,4])	M1 vehicle & opponent is a cyclist 5 km/h <= own driving speed <= 40 km/h & cyclist speed <= 30 km/h no visual obstruction & no ice and snow on road & no poor road condition & no unstable vehicle condition & fine weather
LKA (Lane Keep Assist)	klassece == 1 spverla = [1, 3, 4, 8] v0 >= 60 & v0 < 500 mark == [1,3,4,5,6,7,8,10,11,12] & strob ~= [6,7] strzb ~= 4 schleu == 2 nied == 2 & (wolk ~= 7 & nebelm ~= [3,4]) hursau ~= [12, 13]	M1 vehicle & unintentionally leaving lane before crash & own driving speed >= 60 km/h & markings present & no ice and snow on road & no poor road condition & no unstable vehicle condition & fine weather accident not caused by speeding

Figure 2.4: Example of rulesets for three different systems. The first column contains the name of the system, the second the actual GIDAS queries and the third the corresponding rules under which the system could avoid the accident.

ruleset correspond to conditions under which a particular system theoretically could avoid an accident, $\hat{\xi}^{C_i}_{\text{collision}}$. Applying each ruleset to the accident set we get an underapproximation for the effectiveness in avoiding accidents with the system.

Discussion

In section 2.1 different methods of evaluation are introduced, in particular experimental, simulation based, statistical and predictive. For highly automated driving, experimental evaluation is unfeasible for a large scenario coverage. On the other hand, there is insufficient data available to perform retrospective statistical analysis; due to the limited number of highly automated vehicles in regular traffic. Consequently predictive statistical analysis is performed to evaluate highly automated driving, however, such analysis have several limitations. Errors may arise from uncertainties in accident data, due to the reconstruction or collection procedure; uncertainties in environmental,

vehicle, system or driver models and errors from outcome estimations.

Apart from data uncertainties or model errors, predictive analysis focuses on the current set of accident distributions, provided by a snapshot of the current or historical situation. Any shift in transportation modalities, or other trends skewing the distribution, cannot be accounted for. This include the possibility of the system introducing new types of accidents, currently not present in the accident data. This is particularly true for highly automated driving where many of the subsystems are still not available in the market. The next chapter will discuss how to verify new systems, to ensure that they will operate as intended and not cause new accidents; in order to achieve the potential safety benefit.

CHAPTER 3

Safety Assurance

Chapter 2 gave an introduction to safety benefit estimation and a way of estimating the potential of reducing road fatalities with highly automated vehicles. This however assumes that the system operates safely, i.e. be able to handle complex traffic scenarios and not introduce any new accidents. In this chapter a brief introduction to verification of automated driving functions will be given. First presenting some of the most common techniques, and then their applicability to black box machine learning based driving functions such as neural networks.

3.1 Verification

One of the greatest challenges in the development of fully automated vehicles is to ensure safe operation. This proves to be difficult since it needs to hold for every potential scenario, while being subject to noisy sensor measurements, uncertainty in intentions of other traffic participants and false ego-state perception etc. To simplify the process the vehicle is often divided into different subsystems, each verified individually to comply with the intended functionality. A common division is between perception, using sensors to build a

model of the surroundings, see e.g. [61] and [62]; planning, constructing both long term and short term trajectories for the vehicle to follow, see e.g. [63]; and control, executing the commands necessary for the vehicle to follow the plan, see e.g. [64]. Other examples include division between driving modes or operational design domains, see e.g. [65]. In the following subsections different evaluation methods based on analytical and statistical approaches will be discussed.

Analytical Methods

Analytical methods are based on mathematically proving safety constraints. Such methods demands accurate mathematical models, describing the corresponding systems. To be computationally viable such models are often required to be simple, i.e. linear and few state variables, limiting the applicability to subsystems where such limitations are viable. For example, formal methods define the evaluated system with the help of formal language based on logical rules, in order to logically conclude no failure states are possible, [66], [67], [68].

Statistical Methods and Testing

Data driven, learning based or other high complexity systems i.e. highly nonlinear and with state spaces order of magnitudes higher than the previously discussed subsystems, are challenging to simplify to the level where analytical methods are applicable. The same is true in systems with hidden information and random processes. Instead testing and verification can be done using statistical approaches, based on stochastic modelling and often extensive data collections and annotations.

Example of statistical methods include real world driving, where a certain mileage have to be covered to ensure the failure rate is below the required level with high enough confidence. This in particular is regulated in the ISO 26262 standard [69], which puts a very low limit on the failure rate, requiring a very high mileage to satisfy the requirements, [34]. In addition different methods have been developed to get a statistical representative sampling of traffic scenarios, [23], [70]. Other techniques aim to reduce the amount of test data required by using statistical tools such as extreme value theory, [71].

As a complementary some techniques involve adding additional test cases

artificially. Either through directed testing on a test track [70], by augmenting data [72] or through simulation. Directed testing allows to test the whole vehicle and real system under worst case scenarios, such as extreme environmental conditions or otherwise dangerous situations; or rare case scenarios not commonly encountered in real traffic. However, these methods again suffer from the difficulty in creating scenarios representative for real world driving, covering a meaningful amount of situations, [73]. Augmentation and simulation on the other hand allows for faster testing on larger amounts of scenarios, either running with parts of the software (SIL) or model (MIL) in the loop, as in [70], [74], [75]; or parts of hardware in the loop (HIL) as in [76]. While being faster, these types of evaluation still suffer from the same challenge that all possible configurations of traffic scenarios can never be captured. In addition the simulation models and environments needs to be validated.

Monitoring

To increase safety it is sometimes possible to add online methods of verification, running in parallel to the driving functions, a so called monitoring systems. In reachability analysis the ego states of the vehicle is monitored together with the environment and surrounding traffic participants, continuously propagated forward in time in order to ensure no set of unsafe states is reachable, see e.g. [77] and [78]; or rejecting trajectories during high levels of input noise and uncertainty, [79]. The next section will discuss how online techniques are applicable in machine learning, and particular neural network based methods, rejecting predictions inferred from data of unfavorable conditions.

3.2 Neural networks

Verification of neural networks pose several challenges due to their nonlinearity and high dimensional parameter space, often described as being black-box models. Attempts of using formal verification on neural networks have been made, but they are typically done for simple models with a few parameters, few layers and in other ways limited architectures, [80]. Instead safety arguments rely on extensive statistical testing, using test data sets to show performance fulfill a certain statistical limit. As previously discussed such

testing is hard to perform up to the level of safety required for an automated vehicle, due to the infinite amount of possible driving scenarios. Studies have also shown that neural networks may give false results with high confidence even in environments or domains they were not trained or designed for, [81].

Several attempts have been made on architectural changes in order to improve the confidence estimations of the networks using the available data, such as Bayesian neural networks [82] and ensemble networks [83]; or changes to the training procedure such as the us of re-sampled training datasets, [84]. However, these methods still operate within the training domain. Statistical tools have also been applied to detect adversarial changes to input data, which may lead to false results. Successful attempts such as using influence functions, [85], are able to trace predictions back to the training data, detecting errors and weak predictions.

Instead of doing architectural changes monitoring techniques aim to reject predictions considered to be erroneous or of low confidence. Attempts on monitoring neural networks were made already in the 90s, using input reconstruction reliability estimation, [86]. As the name suggest, features from a driving network are used together with a decoder reconstructing the input data, using the reconstruction error as a measure of confidence. Due to the computational limits at the time, this is done for low resolution images with a network using few parameters. Later on similar methodologies have been reinvestigated and extended using autoencoders. Autoencoders consist of an encoder and a decoder, where the encoder compresses the input data and the decoder tries to reconstruct the original input from the compression. Again the reconstruction error can be used as an indication on the novelty of the input data, compared to the training data. This provides a monitoring function independent of the underlying driving task, where predictions on novel environments are assumed to be of lower quality, [87]. In the following section different methods for developing such a monitoring function will be investigated.

Novelty Estimation

Extensive research have been made in the field of unsupervised anomaly detection, with the potential of being applicable to novelty estimation. In the following sections three widely applied techniques will be presented and discussed.

Autoencoders

As previously mentioned, autoencoders consist of an encoder compressing input data and a decoder with the task of reconstructing the original input data. Although a simple autoencoder with only three fully connected layers showed success in the automotive setting, [87], it was applied to a very simple constructed environment using low resolution input.

Several modifications have been made to autoencoders in order to improve performance for anomaly detection on image data. For higher resolution images, convolutional layers have been implemented as in [88] and [89]. Others have suggested using density estimation in the latent space, using a Gaussian mixture model [90], or autoregressive model [91], instead of using reconstruction error as novelty measurement.

GANs

Another technique commonly used in anomaly detection is Generative Adversarial Networks (GANs). Instead of an encoder and a decoder, a GAN consists of a generator and discriminator as antagonists to generate input candidates with the former and judge them with the latter, [92]. Either the generator or discriminator can then be used for anomaly detection depending on the training procedure and underlying architectural design. The generator by its ability to represent a particular test image, by for example finding the optimal feature representation in the corresponding latent space as in [93].

Metric Learning

Metric learning distinguishes from the previous methods in that it learns a metric in between data points, rather than generating or reconstructing data. In [94] a neural network is learned, mapping the training data into a feature space with the goal of enclosing the feature within a minimal hyper-sphere, considering test input maped outside of the sphere as anomalies.

Lately advances within metric learning have led to significant improvements on tasks such as image classification, [95]. In [96] a CNN-backbone is trained using a non-parametric softmax for unsupervised image classification, reaching close to supervised performance.

3.3 Discussion

In this chapter a brief introduction to commonly used techniques of safety verification have been presented, broadly categorised under analytical and statistical methods. In particular the strengths and weaknesses and how online monitoring as a complement may help covering up for some of their limitations.

For neural networks novelty estimation of input data seem to show promising results as an unsupervised confidence measurement. Although most techniques so far have been using autoencoders or GANs, metric learning has the added benefit of intrinsically providing a metric between training and test instances. This metric could in theory improve novelty measurements, as seen in Paper E and F; while also providing deeper insights into the training data allowing for data retrieval and filtering, as seen in Paper G.

In the end a combination of all methods are most likely needed, to ensure safe operation of a neural network, and especially to verify a highly automated vehicle.

CHAPTER 4

Summary of included papers

This chapter provides a summary of the included papers.

4.1 Paper A

Olaf Op den Camp, **Arian Ranjbar**, Jeroen Uittenbogaard, Erik Rosen, Stefanie de Hair-Buijssen
Overview of main accident scenarios in car-to-cyclist accidents for use in AEB-system test protocol
Published in Proceedings of International Cycling Safety Conference, Nov. 2014. .

At the time of publishing this paper the general trend of road traffic accidents resulting in fatalities was decreasing, however not for bicyclists. The automotive industry thus made significant efforts to mitigate such accidents by introducing active safety systems, in particular AEB; previously implemented for other vulnerable road users such as pedestrians.

This paper introduce a way of classifying car-to-bicyclist accidents in order to quantify different types of configurations. The classification procedure is

applied to several accident databases around Europe, presenting statistics in order to help the development of AEB systems. In particular the aggregated data show the most typical scenarios involve crossing or longitudinal configurations, covering 63% of all accidents resulting in severe injuries and 78% of all accidents resulting in fatalities.

The thesis author contributed with the problem formulation; statistics methodology, i.e. classification procedure; implementation, in particular for the German accident data but also compiling statistics from all databases; analysis and parts of the writing.

4.2 Paper B

Arian Ranjbar, Nils Lubbe, Erik Rosen, Jonas Fredriksson
Car-to pedestrian forward collision warning revisited: A safety benefit estimation
November 2021. *Submitted for review in journal publication* .

This paper presents a predictive safety benefit analysis on a Forward Collision Warning (FCW) system for car-to-pedestrian accidents, in terms of mitigating severe injuries. In addition, it is compared against an Autonomous Emergency Braking (AEB) system addressing the same type of accidents. The evaluation is made through resimulations of accidents from the German In-Depth Accident Study (GIDAS) database. As discussed in section 2.3, the modelling is done using a simple environmental model, bicycle model for vehicle dynamics, collision prediction algorithm and driver behavior model (handling the reaction time and response of the driver). Utilizing the new impact speeds between the car and pedestrian, the probability of severe injuries are estimated through a logistic regression model. Comparing the injury estimates from impact speeds with and without the system provides the effectiveness in terms of mitigating severe injuries.

The study shows that FCW may serve as a great alternative or complement to AEB, providing efficiency of 34%-54% compared to 43% for the AEB. However, as the results indicate, FCW is heavily dependent on designing an effective HMI to be competitive.

The thesis author contributed with problem formulation, implementation, analysis and writing the paper.

4.3 Paper C

Rikard Fredriksson, **Arian Ranjbar**, Erik Rosen
Integrated Bicyclist Protection Systems-Potential of Head Injury Reduction Combining Passive and Active Protection Systems
Published in 24th International Technical Conference on the Enhanced Safety of Vehicles,
2015. .

This paper expands on the idea of predictive safety benefit evaluation of an active safety system, by combination with a passive system. In particular an AEB for car-to-bicyclist accidents together with a windshield airbag or deployable hood system, is investigated. As discussed in section 2.3, a similar model as in the previous paper is used for the evaluation, with the addition of risk reduction curves modeling the passive system.

The evaluation show that the AEB have an effectiveness of 26%-48% independently, depending on system configuration. The passive systems an effectiveness in reducing severe head injuries of 21%-38% independently, again depending on system configuration. Combining the systems has an effectiveness of 38%-62%, showing there is a potential safety benefit in combining such systems.

The thesis author contributed with problem formulation, methodology, analysis and parts of the writing. In particular the development of the injury risk model, both for the active system and the combination with the pre-developed passive injury risk reduction functions; implementation of the active safety system simulations and effectiveness evaluation, with and without the passive system.

4.4 Paper D

Nils Lubbe, Hanna Jeppsson, **Arian Ranjbar**, Jonas Fredriksson, Jonas Bärgman, Martin Östling
Predicted road traffic fatalities in Germany: The potential and limitations of vehicle safety technologies from passive safety to highly automated driving
Published in Proceedings of IRCOBI conference,
Sept. 2018. .

The aim of this paper is to evaluate the safety benefit of a highly automated driving in terms of preventing fatalities, through predictive assessment. As with the previous studies, the data used is from GIDAS. The highly automated vehicle is emulated through all potential safety subsystems which may be included. In particular a rule based approach is implemented, modelling the capabilities of each subsystem according to their specification. The specifications used are a combination of EU regulations and previous literature; verified by in-depth studies on randomly selected accidents in GIDAS, effectiveness comparisons against previous research and sensitivity analysis. Two rule sets are derived corresponding to an optimistic and conservative view of the systems potential. In addition the effectiveness analysis is done by adding more advanced systems in five steps, where the last step correspond to autonomous driving.

The results show a potential of reducing road fatalities by 45%-63%. The remaining accidents can mainly be explained by the study focusing on passenger cars, defined as m1-vehicles. Most remaining fatalities were caused in accidents not involving passenger cars.

The thesis author contributed with problem formulation, derivation of the rule sets with complementary literature review and case study, implementation of the rule set methodology and database management, statistical analysis including sensitivity and a significant part of the writing.

4.5 Paper E

Arian Ranjbar, Chun-Hsiao Yeh, Sascha Hornauer, Stella X. Yu, Ching-Yao Chan
Scene Novelty Prediction from Unsupervised Discriminative Feature Learning
Published in IEEE 23rd International Conference on Intelligent Transportation Systems,
Sept. 2020. .

The previous paper presented methods for evaluating new systems using accident data. This paper instead aim to prevent new types of accident caused by such systems, in particular machine learning based subsystems. By estimating the novelty of incoming sensor data in relation to training data, a measure of confidence can be introduced. The approach is build up on unsupervised

feature learning, mapping training instances onto a feature space most discriminative among them. In this feature space, the training set is modeled through a Gaussian distribution. The confidence can then be set in relation to the probability of new input data being sampled from the same distribution; or equivalently the Mahalanobis distance between the input features and the Gaussian distribution.

Previous such techniques often rely on other unsupervised methods such as autoencoders. The presented approach outperforms state of the art, both on anomaly detection on typical image benchmark datasets; but in particular for autonomous driving based datasets such as BDD100k and KITTI. In addition an experiment is presented predicting the performance degradation in a image segmentation network, to illustrate a typical use case for autonomous driving.

The thesis author contributed with problem formulation, implementation, analysis and writing the paper.

4.6 Paper F

Arian Ranjbar, Sascha Hornauer, Jonas Fredriksson, Stella X. Yu, Ching-Yao Chan
Safety Monitoring of Neural Networks Using Unsupervised Feature Learning and Novelty Estimation
June 2021. *Submitted for review in journal publication*
October 2021. *Major revision submitted for potential journal publication.*

This paper expands upon the concepts from the previous paper, providing a theoretical foundation for the novelty estimation of test data. To improve the performance, the unsupervised feature learning is instead implemented on a spherical feature space, where von Mises-Fisher distributions are used to model the training dataset. In addition the training methodology is expanded upon, allowing for additional information to be incorporated into a training instance, such as several consecutive frames and driving actions. Driving actions in particular can be used to improve the training procedure, where action prediction may be used as a proxy task for model evaluation. Finally, a method for evaluating what segments of a test instance contribute to the novelty is presented, through the use of unsupervised segmentation.

The experiments presented show state of the art performance, both on gen-

eral image benchmarking datasets for anomaly detection, such as CIFAR100; and autonomous driving datasets, such as BDD100k. For the driving datasets more challenging experiments are performed by omitting smaller objects in the training data.

The thesis author contributed with problem formulation, implementation, analysis and writing the paper.

4.7 Paper G

Sascha Hornauer, Baladitya Yellapragada, **Arian Ranjbar**, Stella X. Yu
Driving Scene Retrieval by Example from Large-Scale Data
Published in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,
2019.
.

This paper investigates the possibility of retrieving driving scenes from unlabeled driving datasets, by querying examples. As in the previous paper, data can be mapped onto a spherical feature space using discriminative feature learning; providing a metric of similarity in between instances. Retrieving driving scenes with desired properties can be done without labels, by mapping an example query into the feature space and find nearest neighbours. The same technique can also be used to filter or remove bias and removing redundant data.

The thesis author contributed with the problem formulation, in particular the use of unsupervised methods as in the previous papers for the use in data retrieval or filtering purposes; implementation of the unsupervised training using unsupervised feature learning; analysis, for both parts but in particular for the section without action inclusion; and parts of the writing, in particular parts of the related works and method sections.

CHAPTER 5

Conclusion and Discussion

This thesis presents work towards the safety evaluation and safety assurance of highly automated driving. Paper A-C contribute with safety benefit estimates for subsystems of a highly automated vehicle while providing methodology development. Paper D tries to directly provide an estimate of the effectiveness of HAD in reducing fatalities, which corresponds to the first research question presented in section 1.1. Although providing initial estimates the accuracy may be improved in the future by implementing more advanced models and simulation environments.

Paper E and F contribute towards the second question in the problem formulation. While not ensuring safe driving for a complete system, they investigate the possibility of increasing the safety when using machine learning. Paper G uses the same methodology for data retrieval, filtering and collection; which may help increase performance and thereby also contribute to the safety of the system.

5.1 Safety Evaluation

In this section a brief discussion regarding the first four papers is presented. Paper A, as discussed in section 2.2, provides classification and frequency of car-to-bicyclist accident scenarios, showing potential coverage depending on system setup. The results later on served as input for the development of the EuroNCAP test protocol for car-to-bicyclist AEB, [97]. Paper C puts the potential in relation to the predicted real world performance, showing effectiveness in reducing severe head injuries with AEB and AEB together with a passive system. Since then effectiveness for car-to-bicyclist AEB in particular have been further evaluated using various methods, see e.g. [98] and [99]. In [99], data reconstructed from dash cameras is used with a similar assessment method, showing consistent results with the results presented in this thesis. In addition, they concluded that the effectiveness is highly sensitive to system configuration, in particular field of view of the sensor, due to the high amount of crossing accidents; consistent with both Paper A and C.

Papers A-C also helps putting the methodology of Paper D in relation to more advanced modelling techniques for e.g. AEB, which is one of the subsystems. Not only confirming the accuracy of the individual performance predictions, by for example combining the AEB performance from Paper B and C; but also illustrating the limitations since no combinatorial effects are taken into account. Paper D directly aim to contribute towards answering the first question stated in section 1.1, providing initial estimates of the potential safety benefit. In addition to the limitations discussed above, Paper D include uncertainties in the system modeling. In particular the ruleset is only defined to define an accident as avoidable or unavoidable, i.e. no consideration is taken to whether the active systems are able to lower impact speeds or in other ways mitigate injuries. As previously discussed, new accidents scenarios introduced by the systems are not covered either; since they are not represented in the accident data.

Recently HAD have been further evaluated using more advanced models, either for the systems or environment, to overcome some of the limitations and increase accuracy. Notably [100] evaluate an autonomous driving system with software in the loop in a high fidelity simulation environment, using reconstructed severe injuries from police reports. This study showed an effectiveness, even higher than the results of Paper D, between 82%-100% depending on whether the crash initiator or responder were host for the system.

5.2 Safety Assurance

The last two papers present a method of increasing safety, by using a monitoring framework for machine learning based subsystems in an automated vehicle. Although not providing a complete answer to the second research question of section 1.1, it shows promising performance on benchmarking datasets and in finding novel scenes from driving data; contributing to the safety assurance of the full system.

One of the main limitations in the current framework is the ability of explaining what contributes to the novelty. Although Paper F include initial tries of solving the problem, the difficulty lies in benchmarking the methods. Attempts have been made to construct such datasets by using ground truth segmentation labels, [101]. However, it is targeted towards supervised learning settings. On the other hand pure segmentation performance, using metrics such as Intersection-Over-Union, [102]; may not necessarily be optimal for understanding the context of a novelty.

The last paper show the applicability of using the monitor framework for data retrieval and filtering. In the same way the monitoring framework could potentially be used for data collection. Since most data collected is redundant, novel scenes could automatically be flagged for labeling. This is similar to the development of active learning, which finds optimal data for training a particular network, [103].

5.3 Future Work

As the limitation in current methods indicate, there is a need for better evaluation and verification methods. Here are some topics inspired from this work:

Models for safety benefit evaluation

As previously discussed, the safety benefit estimation of Paper D provide quite limited models. And although [100] show great advances in the modelling, all the data and models used are company internal. As open source models like CARLA gets widely available, more replicable alternatives could be investigated.

Dedicated benchmarks for novelty estimation in autonomous driving

The current method for benchmarking novelty estimation in autonomous driving, rely on artificially creating datasets by omitting certain labeled objects. The labeling of the original datasets are often done with other tasks in mind, e.g. only the largest object is annotated, or the other way around; a small driving related object is annotated where a larger important anomaly is unlabeled. Developing new datasets would also allow for ground truth labeling with respect to testing novelty estimation.

Advanced novelty detection

Paper F presents an extension of the methodology to work on segments of images, to find novel objects or segments particularly contributing to the novelty. At the current stage, this needs further development to be used in addition to monitoring. In particular, by developing specific datasets for the task,better quantitative studies could be performed, speeding up the development towards safe autonomous driving.

References

[1] World Health Organization, "Global status report on road safety 2018: Summary," World Health Organization, Tech. Rep., 2018.

[2] OECD, *Oecd statistics*, stats.oecd.org, Accessed: Oct 2021, 2021.

[3] L. S. Robertson, "Estimates of motor vehicle seat belt effectiveness and use: Implications for occupant crash protection.," *American Journal of Public Health*, vol. 66, no. 9, pp. 859–864, 1976.

[4] Y. Håland, "The evolution of the three point seat belt from yesterday to tomorrow," in *IRCOBI Conference*, 2006.

[5] L. Evans, "Airbag effectiveness in preventing fatalities predicted according to type of crash, driver age, and blood alcohol concentration," *Accident Analysis & Prevention*, vol. 23, no. 6, pp. 531–541, 1991.

[6] O. Bostrom, H. C. Gabler, K. Digges, B. Fildes, and C. Sunnevang, "Injury reduction opportunities of far side impact countermeasures," in *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, Association for the Advancement of Automotive Medicine, vol. 52, 2008, p. 289.

[7] N. Yoganandan, F. A. Pintar, J. Zhang, and T. A. Gennarelli, "Lateral impact injuries with side airbag deployments—a descriptive study," *Accident Analysis & Prevention*, vol. 39, no. 1, pp. 22–27, 2007.

[8] R. Fredriksson, Y. Håland, and J. Yang, "Evaluation of a new pedestrian head injury protection system with a sensor in the bumper and lifting of the bonnet's rear part," SAE Technical Paper, Tech. Rep., 2001.

[9] M. Bayly, B. Fildes, M. Regan, and K. Young, "Review of crash effectiveness of intelligent transport systems," *Emergency*, vol. 3, p. 14, 2007.

[10] T. Hummel, M. Kühn, J. Bende, and A. Lang, "Advanced driver assistance systems," *German Insurance Association Insurers Accident Research. Available on www. udv. de, accessed at*, vol. 6, no. 01, p. 2015, 2011.

[11] D. Burton, A. Delaney, S. Newstead, D. Logan, and B. Fildes, "Evaluation of anti-lock braking systems effectiveness," *Accident Analysis and Prevention*, vol. 29, no. 6, pp. 745–757, 1997.

[12] C. M. Farmer, "Effect of electronic stability control on automobile crash risk," *Traffic injury prevention*, vol. 5, no. 4, pp. 317–325, 2004.

[13] L. Xiao and F. Gao, "A comprehensive review of the development of adaptive cruise control systems," *Vehicle system dynamics*, vol. 48, no. 10, pp. 1167–1192, 2010.

[14] O. M. Carsten and F. Tate, "Intelligent speed adaptation: Accident savings and cost–benefit analysis," *Accident Analysis & Prevention*, vol. 37, no. 3, pp. 407–416, 2005.

[15] S. Sternlund, J. Strandroth, M. Rizzi, A. Lie, and C. Tingvall, "The effectiveness of lane departure warning systems—a reduction in real-world passenger car injury crashes," *Traffic injury prevention*, vol. 18, no. 2, pp. 225–229, 2017.

[16] R. Anderson, T. Hutchinson, B. Linke, and G. Ponte, "Analysis of crash data to estimate the benefits of emerging vehicle technology," *Centre for Automotive safety Research, The University of Adelaide*, 2010.

[17] J. B. Cicchino, "Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates," *Accident Analysis & Prevention*, vol. 99, pp. 142–152, 2017.

[18] E. Rosen, "Autonomous emergency braking for vulnerable road users," in *Proceedings of IRCOBI conference*, 2013, pp. 618–627.

[19] J. M. Scanlon, R. Sherony, and H. C. Gabler, "Injury mitigation estimates for an intersection driver assistance system in straight crossing path crashes in the united states," *Traffic injury prevention*, vol. 18, no. sup1, S9–S17, 2017.

[20] U. Sander, "Opportunities and limitations for intersection collision intervention—a study of real world 'left turn across path'accidents," *Accident Analysis & Prevention*, vol. 99, pp. 342–355, 2017.

[21] J. Nilsson, J. Fredriksson, and A. C. Ödblom, "Verification of collision avoidance systems using reachability analysis," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 10 676–10 681, 2014.

[22] E. Coelingh, L. Jakobsson, H. Lind, and M. Lindman, "Collision warning with auto brake: A real-life safety perspective," *Innovations for Safety: Opportunities and Challenges*, 2007.

[23] M. Distner, M. Bengtsson, T. Broberg, and L. Jakobsson, "City safety—a system addressing rear-end collisions at low speeds," in *Proc. 21st International Technical Conference on the Enhanced Safety of Vehicles*, 2009.

[24] J. R. Treat, N. Tumbas, S. McDonald, *et al.*, "Tri-level study of the causes of traffic accidents: Final report. executive summary.," Indiana University, Bloomington, Institute for Research in Public Safety, Tech. Rep., 1979.

[25] V. L. Neale, T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman, "An overview of the 100-car naturalistic study and findings," *National Highway Traffic Safety Administration, Paper*, vol. 5, p. 0400, 2005.

[26] T. A. Dingus, S. G. Klauer, V. L. Neale, *et al.*, "The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment," United States. Department of Transportation. National Highway Traffic Safety ..., Tech. Rep., 2006.

[27] M. Sivak and B. Schoettle, "Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles," University of Michigan, Ann Arbor, Transportation Research Institute, Tech. Rep., 2015.

[28] N. Kalra and D. G. Groves, *The enemy of good: Estimating the cost of waiting for nearly perfect automated vehicles*. Rand Corporation, 2017.

[29] T. Winkle, "Development and approval of automated vehicles: Considerations of technical, legal, and economic risks," in *Autonomous Driving*, Springer, 2016, pp. 589–618.

[30] T. Litman, *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, Canada, 2017.

[31] T. Winkle, "Safety benefits of automated vehicles: Extended findings from accident research for development, validation and testing," in *Autonomous driving*, Springer, 2016, pp. 335–364.

[32] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, 2016.

[33] M. Martinez-Diaz and F. Soriguera, "Autonomous vehicles: Theoretical and practical challenges," *Transportation Research Procedia*, vol. 33, pp. 275–282, 2018.

[34] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016.

[35] I. ISO, "26262: Road vehicles-functional safety," *International Standard ISO/FDIS*, vol. 26262, 2011.

[36] O. Boström, M. Y. Svensson, B. Aldman, *et al.*, "A new neck injury criterion candidate-based on injury findings in the cervical spinal ganglia after experimental neck extension trauma," in *Proceedings of The 1996 International Ircobi Conference On The Biomechanics Of Impact, September 11-13, Dublin, Ireland*, 1996, pp. 123–136.

[37] M. Van Ratingen, A. Williams, A. Lie, *et al.*, "The european new car assessment programme: A historical review," *Chinese journal of traumatology*, vol. 19, no. 2, pp. 63–69, 2016.

[38] EuroNCAP, "Euro ncap rating review 2018," *Report from the Ratings Group*, 2018.

[39] C. Grover, M. Avery, and I. Knight, "The development of a consumer test procedure for pedestrian sensitive aeb," in *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration*, 2015.

[40] R. Watanabe, T. Katsuhara, H. Miyazaki, Y. Kitagawa, and T. Yasuki, "Research of the relationship of pedestrian injury to collision speed, car-type, impact location and pedestrian sizes using human fe model (thums version 4)," *Stapp car crash journal*, vol. 56, p. 269, 2012.

[41] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*, PMLR, 2017, pp. 1–16.

[42] O. J. Gietelink, D. Verburg, K. Labibes, and A. Oostendorp, "Pre-crash system validation with prescan and vehil," in *IEEE Intelligent Vehicles Symposium, 2004*, IEEE, 2004, pp. 913–918.

[43] U. Sander, *Predicting Safety Benefits of Automated Emergency Braking at Intersections-Virtual simulations based on real-world accident data.* Chalmers University of Technology, 2018.

[44] J. Nilsson, *Computational verification methods for automotive safety systems.* Chalmers Tekniska Hogskola (Sweden), 2014.

[45] N. Lubbe and E. Rosén, "Pedestrian crossing situations: Quantification of comfort boundaries to guide intervention timing," *Accident Analysis & Prevention*, vol. 71, pp. 261–266, 2014.

[46] N. Lubbe, "Brake reactions of distracted drivers to pedestrian forward collision warning systems," *Journal of safety research*, vol. 61, pp. 23–32, 2017.

[47] P. Cummings, J. D. Wells, and F. P. Rivara, "Estimating seat belt effectiveness using matched-pair cohort methods," *Accident Analysis & Prevention*, vol. 35, no. 1, pp. 143–149, 2003.

[48] C. S. Crandall, L. M. Olson, and D. P. Sklar, "Mortality reduction with air bag and seat belt use in head-on passenger car collisions," *American journal of epidemiology*, vol. 153, no. 3, pp. 219–224, 2001.

[49] J. N. Dang, "Preliminary results analyzing the effectiveness of electronic stability control (esc) systems," US Department of Transportation, National Highway Traffic Safety Administration, Tech. Rep., 2004.

[50] F. Leopold, P. Lesire, and C. Chauvel, "Voiesur: French research project on global road safety, focus on child safety specificities," in *Protection of Children in Cars - 10th International Conference, Munich, Germany*, 2012.

[51] *German national road traffic accident statistics provided by destatis*, `https://www.destatis.de/EN/Themes/Society-Environment/Traffic-Accidents/_node.html`, Accessed: 2014.

[52] *Bron: Netherlands national road crash register*, `www.swov.nl`, Accessed: 2014.

[53] *Swedish traffic accident data acquisition*, `https://www.transportstyrelsen.se/en/road/STRADA/`, Accessed: 2014.

[54] *Stats19, access tp great britain's official road traffic casualty database*, `http://www.adls.ac.uk/department-for-transport/stats19-road-accident-dataset/`, Accessed: 2014.

[55] D. Otte, C. Krettek, H. Brunner, and H. Zwipp, "Scientific approach and methodology of a new in-depth investigation study in germany called gidas," in *Proceedings: International Technical Conference on the Enhanced Safety of Vehicles*, National Highway Traffic Safety Administration, vol. 2003, 2003, 10–p.

[56] T. A. Gennarelli, E. Wodzin, *et al.*, "The abbreviated injury scale 2005," *Update*, vol. 20082008, 2008.

[57] A. Schubert, C. Erbsmehl, and L. Hannawald, "Standardized pre-crash-scenarios in digital format on the basis of the vufo simulation," 2013.

[58] A. Ranjbar, "Active safety for car-to-bicyclist accidents," M.S. thesis, 2014.

[59] E. Rosén, J.-E. Källhammer, D. Eriksson, M. Nentwich, R. Fredriksson, and K. Smith, "Pedestrian injury mitigation by autonomous braking," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1949–1957, 2010.

[60] R. Fredriksson and E. Rosén, "Head injury reduction potential of integrated pedestrian protection systems based on accident and experimental data–benefit of combining passive and active systems," in *IRCOBI (International Research Council On the Biomechanics of Impact) Conference. Berlin, Germany*, 2014, pp. 603–613.

[61] M. P. Muresan, I. Giosan, and S. Nedevschi, "Stabilization and validation of 3d object position using multimodal sensor fusion and semantic segmentation," *Sensors*, vol. 20, no. 4, p. 1110, 2020.

[62] T. Sattler, W. Maddern, C. Toft, *et al.*, "Benchmarking 6dof outdoor visual localization in changing conditions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8601–8610.

[63] H. Kim, J. Cho, D. Kim, and K. Huh, "Intervention minimized semi-autonomous control using decoupled model predictive control," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2017, pp. 618–623.

[64] I. Batkovic, M. Zanon, M. Ali, and P. Falcone, "Real-time constrained trajectory planning and vehicle control for proactive autonomous driving with road users," in *2019 18th European Control Conference (ECC)*, 2019, pp. 256–262.

[65] M. Gyllenhammar, R. Johansson, F. Warg, *et al.*, "Towards an operational design domain that supports the safety argumentation of an automated driving system," in *10th European Congress on Embedded Real Time Systems (ERTS 2020)*, 2020.

[66] S. M. Loos, D. Witmer, P. Steenkiste, and A. Platzer, "Efficiency analysis of formally verified adaptive cruise controllers," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, IEEE, 2013, pp. 1565–1570.

[67] M. Luckcuck, M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher, "Formal specification and verification of autonomous robotic systems: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 1–41, 2019.

[68] J. Woodcock, P. G. Larsen, J. Bicarregui, and J. Fitzgerald, "Formal methods: Practice and experience," *ACM computing surveys (CSUR)*, vol. 41, no. 4, pp. 1–36, 2009.

[69] *Iso 26262-8: Road vehicles - functional safety*, International Organization for Standardization. Geneva, Switzerland., 2017.

[70] E. Coelingh, H. Lind, W. Birk, M. Distner, and D. Wetterberg, "Collision warning with auto brake," in *FISITA 2006 World Automotive Congress: 22/10/2006-27/10/2006*, JSAE, 2006.

[71] D. Åsljung, J. Nilsson, and J. Fredriksson, "Using extreme value theory for vehicle level safety validation and implications for autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 4, pp. 288–297, 2017.

[72] J. Nilsson, P. Andersson, I. Y.-H. Gu, and J. Fredriksson, "Pedestrian detection using augmented training data," in *2014 22nd International Conference on Pattern Recognition*, IEEE, 2014, pp. 4548–4553.

[73] D. Åsljung, *On Safety Validation of Automated Driving Systems using Extreme Value Theory*. Chalmers Tekniska Hogskola (Sweden), 2017.

[74] J. Hillenbrand and K. Kroschel, "A study on the performance of uncooperative collision mitigation systems at intersection-like traffic situations," in *2006 IEEE Conference on Cybernetics and Intelligent Systems*, IEEE, 2006, pp. 1–6.

[75] D. Gruyer, S. Choi, C. Boussard, and B. d'Andréa-Novel, "From virtual to reality, how to prototype, test and evaluate new adas: Application to automatic car parking," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, IEEE, 2014, pp. 261–267.

[76] O. Gietelink, J. Ploeg, B. De Schutter, and M. Verhaegen, "Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations," *Vehicle System Dynamics*, vol. 44, no. 7, pp. 569–590, 2006.

[77] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, 2014.

[78] I. Batkovic, M. Ali, P. Falcone, and M. Zanon, "Safe trajectory tracking in uncertain environments," *arXiv preprint arXiv:2001.11602*, 2020.

[79] S. Kojchev, E. Klintberg, and J. Fredriksson, "A safety monitoring concept for fully automated driving," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2020, pp. 1–7.

[80] A. Boopathy, T.-W. Weng, P.-Y. Chen, S. Liu, and L. Daniel, "Cnn-cert: An efficient framework for certifying robustness of convolutional neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3240–3247.

[81] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," *arXiv preprint arXiv:1611.01236*, 2016.

[82] D. J. MacKay, "Probable networks and plausible predictions-a review of practical bayesian methods for supervised neural networks," *Network: computation in neural systems*, vol. 6, no. 3, p. 469, 1995.

[83] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *arXiv preprint arXiv:1612.01474*, 2016.

[84] S. Reed, H. Lee, D. Anguelov, C. Szegedy, D. Erhan, and A. Rabinovich, "Training deep neural networks on noisy labels with bootstrapping," *arXiv preprint arXiv:1412.6596*, 2014.

[85] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *International Conference on Machine Learning*, PMLR, 2017, pp. 1885–1894.

[86] D. A. Pomerleau, "Input reconstruction reliability estimation," in *Neural Network Perception for Mobile Robot Guidance*, Springer, 1993, pp. 133–150.

[87] C. Richter and N. Roy, "Safe visual navigation via deep learning and novelty detection," 2017.

[88] A. Makhzani and B. Frey, "Winner-take-all autoencoders," *arXiv preprint arXiv:1409.2752*, 2014.

[89] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes," *Computer Vision and Image Understanding*, vol. 172, pp. 88–97, 2018.

[90] B. Zong, Q. Song, M. R. Min, *et al.*, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[91] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, "Latent space autoregression for novelty detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 481–490.

[92] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[93] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International conference on information processing in medical imaging*, Springer, 2017, pp. 146–157.

[94] L. Ruff, R. Vandermeulen, N. Goernitz, *et al.*, "Deep one-class classification," in *International conference on machine learning*, PMLR, 2018, pp. 4393–4402.

[95] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *2012 IEEE conference on computer vision and pattern recognition*, IEEE, 2012, pp. 2288–2295.

[96] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3733–3742.

[97] O. O. Den Camp, S. van Montfort, J. Uittenbogaard, and J. Welten, "Cyclist target and test setup for evaluation of cyclist-autonomous emergency braking," *International Journal of Automotive Technology*, vol. 18, no. 6, pp. 1085–1097, 2017.

[98] J. Lenard, R. Welsh, and R. Danton, "Time-to-collision analysis of pedestrian and pedal-cycle accidents for the development of autonomous emergency braking systems," *Accident Analysis & Prevention*, vol. 115, pp. 128–136, 2018.

[99] Y. Zhao, D. Ito, and K. Mizuno, "Aeb effectiveness evaluation based on car-to-cyclist accident reconstructions using video of drive recorder," *Traffic injury prevention*, vol. 20, no. 1, pp. 100–106, 2019.

[100] J. M. Scanlon, K. D. Kusano, T. Daniel, C. Alderson, A. Ogle, and T. Victor, *Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain*, 2021.

[101] D. Hendrycks, S. Basart, M. Mazeika, M. Mostajabi, J. Steinhardt, and D. Song, "Scaling out-of-distribution detection for real-world settings," *arXiv preprint arXiv:1911.11132*, 2019.

[102] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.

[103] B. Settles, "Active learning literature survey," 2009.