Defining Object Detection Requirements for Safe Automated Driving

Am Fachbereich Maschinenbau an der
Technischen Universität Darmstadt
zur Erlangung des Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte

Dissertation

vorgelegt von

Ken Thaddäus Mori M. Sc.

aus Bensheim

Erstgutachter:in:	Prof. Dr. -Ing. Steven Peters
Zweitgutachter:in:	Prof. Dr. Arjan Kuijper

Tag der Einreichung:	04.12.2023
Tag der mündlichen Prüfung:	30.01.2024

Darmstadt 2024

D 17

Ken Thaddäus Mori M. Sc.: Defining Object Detection Requirements for Safe Automated Driving
Darmstadt, Technische Universität Darmstadt
Tag der mündlichen Prüfung: 30.01.2024

Dieses Dokument wird bereitgestellt von TUprints – Publikationsservice der TU Darmstadt.
https://tuprints.ulb.tu-darmstadt.de/
Jahr der Veröffentlichung der Dissertation auf TUprints: 2024

Bitte verweisen Sie auf:
URN: urn:nbn:de:tuda-tuprints-266228
URI: https://tuprints.ulb.tu-darmstadt.de/id/eprint/26622

Preface

This dissertation was completed during my time at the Institute of Automotive Engineering (FZD) of the Technical University Darmstadt while working on the VIVID project. The project was funded by the German Federal Ministry of Education and Research, based on a decision of the Deutsche Bundestag with the grant number 16ME0173. While this dissertation was written by me, it was made possible by many others to whom I owe my gratitude.

Firstly, I extend my thanks to Prof. Dr. rer. nat. Hermann Winner for offering me the opportunity to begin working at FZD despite my lack of background in the field of automotive engineering. Special thanks go to my doctoral supervisor Prof. Dr.-Ing. Steven Peters. He supported all doctoral candidates, including myself, while navigating FZD through a difficult time of change. Additionally, I would like to thank Prof. Dr. Arjan Kuijper for taking over the role of the second examiner.

I would also like to thank all my friends and colleagues at FZD. All of you have been nothing but supportive, and the work culture at FZD is amazing. In particular, I am grateful to Timm, who offered me a Master's thesis opportunity. That got me started both in the field of computer vision and at FZD with which I had no prior contact. A special mention also goes to Kai, with whom I suffered through endless discussions on the relevance topic. I owe a big thank you for the good results which found their way into publications and both of our dissertations. Another special mention goes to Lukas who worked on the VIVID project with me. I have to thank you for the great collaboration and for keeping a lot of tasks off my back.

Of course, my gratitude also extends to all of my friends outside of FZD. Special thanks also go to my girlfriend Nicki for putting with me. Finally, but most importantly, I would like to thank my family, who are always there for me. You believed in me from the start and now I get to prove you right. Also, big thanks to my mother for actually proofreading the entire document.

Darmstadt, November 2023.

Table of Contents

List of Symbols and Indices

Latin formula symbols

Symbol	Unit	Description
a	$\frac{\mathrm{m}}{\mathrm{s}^2}$	Acceleration
d	m	Distance
e	-	Unit vector
r	m	Direction vector or projection onto reference
s	m	Size
t	s	Time
t_r	s	Reaction time
v	$\frac{\mathrm{m}}{\mathrm{s}}$	Velocity

Greek formula symbols

Symbol	Unit	Description
α	$^\circ$	Angle between ego heading and direction of other vehicle
β	$^\circ$	Object orientation relative to ego heading
Δ	-	Difference
τ	s	Time difference
$\dot{\Theta}$	$\frac{^\circ}{\mathrm{s}}$	Angular velocity

Indices

Symbol	Description
$\parallel$	Parallel
$\perp$	Perpendicular
0	Initial
1	Ego
2	Object of interest (OOI)
3	Static object
a	Accelerating
b	Braking
c	Change lateral acceleration
d	Desired
e	End
g	Guaranteed acceleration
GT	Ground truth
h	Halting
l	Lateral
lim	Speed limit
low	Lower error bound
M	Matching
max	Maximum
min	Minimum
PRED	Predicted
r	Radial
s	Starts moving laterally
t	Tangential

Accents and Operators

Symbol	Description
$\vec{x}$	Vector
t'	Substitute with other time reference

List of Abbreviations

AD	automated driving
ADE	average distance error
AP	average precision
BEV	birds-eye view
COCO	Common Objects in Context
DNN	deep neural network
ECDF	empirical cumulative distribution function
FN	false negative
FP	false positive
GT	ground truth
IoU	intersection over union
iTTC	inverse time to collision
mAP	mean average precision
ML	machine learning
MOT	multi object tracking
MOTA	multi object tracking accuracy
NN	neural network
OOI	object of interest
PKL	Planning Kullback-Leibler Divergence
RQ	research question
RSS	Responsibility-Sensitive Safety
sAMOTA	scaled average multi object tracking accuracy
SOTIF	safety of the intended functionality
SUT	system under test
TP	true positive
TTC	time to collision
VRU	vulnerable road user

List of Figures

Kurzzusammenfassung

Automatisiertes Fahren erfordert eine zuverlässige Umgebungswahrnehmung zur Gewährleistung der Sicherheit. Eine häufige Wahrnehmungsaufgabe ist die 3D-Objekterkennung, die darauf abzielt, den Ort und die Eigenschaften dynamischer Objekte wahrzunehmen. Die Bewertung der Objekterkennung stützt sich typischerweise auf Datensätze. Obwohl diese Datensätze Bewertungsmetriken bereitstellen, gelingt es ihnen nicht, die Erkennungsergebnisse mit Sicherheitsaspekten wie Unfällen zu verknüpfen. Daher fehlen klare Anforderungen an die Objekterkennung, welche die Sicherheit der Fahraufgabe berücksichtigen.

Dementsprechend ist das Ziel dieser Arbeit, Anforderungen für die 3D-Objekterkennung zu identifizieren, welche die Sicherheit berücksichtigen. Darüber hinaus ist es wünschenswert, interpretierbare Anforderungen zu erhalten. Daher wird die übergeordnete Forschungsfrage bezüglich der Anforderungen für sichere Objekterkennung dekomponiert. Anforderungen für die drei interpretierbaren Aspekte der Klassifizierung, Relevanz und Attribute von Objekten werden separat behandelt. Schließlich besteht das letzte Ziel darin, eine Methode zur Bewertung und Validierung der verschiedenen Anforderungen für jeden dieser Aspekte bereitzustellen.

Die Gesamtmethodik dieser Arbeit identifiziert zunächst gemeinsame Prinzipien, die das übergeordnete Ziel der Anforderungen für Sicherheit spezifizieren. Die vier Prinzipien sind Interpretierbarkeit, rechtliche Anforderungen, Sicherheitsanforderungen und die menschliche Baseline. Durch die Anwendung dieser Prinzipien ist es möglich, unterschiedliche Methoden für die Aspekte Klassifizierung, Relevanz und Attribute zu entwickeln. In dieser Arbeit werden die erforderlichen Objektkategorien, Kriterien für die Relevanz und Anforderungen für Attribute erfolgreich identifiziert. Darüber hinaus wird eine neuartige Validierungsmethode vorgestellt, die auf einer Prädiktion von Trajektorien unter Verwendung eines tiefen neuronalen Netzwerks basiert. Die Anwendung dieser Validierungsmethode auf die in dieser Arbeit vorgeschlagenen Anforderungen ist erfolgreich und unterstützt somit die Ergebnisse. Verbleibende Einschränkungen der vorgeschlagenen Methodik bezüglich verfügbarer Daten und Algorithmen werden identifiziert und diskutiert. Darüber hinaus werden die Auswirkungen der neuen Anforderungen auf Datensätze, Algorithmen und Sensoraufbauten für die 3D-Objekterkennung berücksichtigt.

Die Gesamtmethodik präsentiert einen zweiteiligen Ansatz zur Anforderungsdefinition. Zunächst werden interpretierbare Anforderungen auf der Grundlage einer Sicherheitsargumentation entwickelt. Diese werden dann durch die unabhängige Validierungsmethode untermauert. Die Ergebnisse liefern somit die erforderlichen Anforderungen, um Objekterkennungen im Kontext der Sicherheit für die Aufgabe des automatisierten Fahrens zu testen und zu validieren. Der Autor hofft, dass die Anforderungen die explizite Bewertung und Verbesserung der Sicherheit für die zukünftige Objekterkennung fördern.

Abstract

Automated driving (AD) requires reliable environment perception to ensure safety. One common perception task is 3D object detection, which aims at perceiving location and attributes of dynamic objects. The evaluation of object detection typically relies on datasets. While these datasets provide evaluation metrics, they fail to connect detection results to safety outcomes such as accidents. Therefore, there is a lack of clear requirements for object detection which consider the safety of the driving task.

Accordingly, the objective of this work is to identify requirements for 3D object detection which consider safety. Furthermore, it is desirable to obtain requirements which are interpretable. Therefore, the overall research question regarding requirements for safe object detection is decomposed. Requirements for the three interpretable aspects of classification, relevance and attributes of objects are treated separately. Finally, the last objective is to provide a method to evaluate and validate the different requirements for each of these aspects.

The methodology of this work first identifies common principles which further specify the overall objective of requirements for safety. The four principles are interpretability, legal requirements, safety requirements and the human baseline. Applying these principles allows developing different methods for the aspects classification, relevance and attributes, respectively. In this work, the required object categories, criteria for relevance and attribute requirements are successfully identified. In addition, a novel validation method based on a motion prediction leveraging a deep neural network is presented. Applying this validation method to the requirements proposed in this work is successful, thus supporting the results. Remaining limitations of the proposed methodology including the available data and algorithms are identified and discussed. Furthermore, the implications of the novel requirements on datasets, algorithms and sensor setups for 3D object detection are considered.

The overall methodology presents a two-pronged approach to requirement definition. Firstly, simple and interpretable requirements are developed based on a safety argumentation. These requirements are then additionally substantiated by the validation method, which relies upon a deep neural network. The results thus provide the requirements, which are required to test and validate object detectors in the context of safety for the task of AD. The author hopes that the requirements encourage the explicit evaluation and improvement of safety for future object detection.

1 Introduction

This chapter serves as an introduction by first providing the motivation of this work. Subsequently, the objectives and the scope are specified. Next, the research questions (RQs) pursued in this work are introduced. Finally, the structure of the remaining document is presented.

1.1 Motivation

The topic of AD has recently been receiving considerable interest.[1] Potential benefits of this technology include improvements with regards to safety, accessibility and efficiency.[2] As demanded by an Ethics Commission appointed by the German Federal Minister of Transport and Digital Infrastructure, introducing such a technology should be accompanied by an improvement in safety.[3] However, validating the safety of an AD system remains challenging, especially with regards to the validation effort.[4]

One task required to ensure the safety of an AD system is a reliable environment perception. In typical modular AD system architectures, the perception component is separated from the path planning.[5] Testing such a modular AD system with respect to its internal structure improves the understanding of performance bounds.[6] Furthermore, testing a perception component separately offers potential reductions of the overall testing effort for AD systems.[5] Therefore, it is beneficial to test the perception component separately.

Perception components typically incorporate deep neural networks (DNNs), which have shown large success for different perception tasks.[7] These machine learning (ML) methods use datasets to provide samples for training. Furthermore, data is also used to verify compliance with requirements.[8] Therefore, different public perception datasets[9,10,11,12] have emerged as common

1 Kang, Y. et al.: Test Your Self-Driving Algorithm (2019), p. 171-172.

2 Bagloee, S. A. et al.: Autonomous vehicles (2016), p. 284.

3 Ethics Comission: Automated and Connected Driving: Report (2017), p. 10.

4 Junietz, P. et al.: Criticality Metric for Safety Validation (2018), p. 493-495.

5 Amersbach, C. T.: Functional Decomposition Approach (2020), p. 41-44.

6 Thorn, E. et al.: A Framework for Automated Driving System Testable Cases and Scenarios (2018), p. 64-65.

7 Feng, D. et al.: Deep Multi-Modal Object Detection and Semantic Segmentation (2020), p. 3.

8 Ashmore, R. et al.: Assuring the Machine Learning Lifecycle (2022), p. 3-4.

9 Deng, J. et al.: ImageNet: A large-scale hierarchical image database (2009).

10 Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010).

11 Lin, T.-Y. et al.: Microsoft COCO: Common Objects in Context (2014)

12 Geiger, A. et al.: Are we ready for Autonomous Driving? (2012).

benchmarks. These benchmarks provide the required data as well as the opportunity to evaluate and compare different perception algorithms under equal conditions. Datasets are important for research on perception, because they define the task objectives and the requirements.[13]

Since ML methods may fail due to underspecification of the task,[14] it is important to sufficiently specify perception requirements with regards to safety. Common safety evaluations of AD typically rely on failure rates such as accident rates.[15,16,17] Safety is hereby exclusively quantified in terms of observable behavioral outcomes of the entire AD system. However, perception components are not directly connected to behavioral outputs of the system which can be evaluated in this manner. This means that common requirements for AD safety are not applicable to perception components. Conversely, the field of perception relies on distinct evaluation procedures which insufficiently consider the context of driving. For instance, the current dataset perception metrics insufficiently consider legal and safety requirements for the driving task.[18] In addition, common perception metrics lack interpretability.[19,20] To better incorporate safety, proposals have been made to evaluate perception functions together with the downstream planner.[21] However, this approach fails to leverage the aforementioned benefits of separate perception testing and is not interpretable. Furthermore, all available evaluation methods fail to specify requirements including thresholds which define sufficient performance. While the human baseline may provide a reference for such a threshold, human perception performance is currently unclear[22].

Overall, safety requirements are necessary and directly evaluating safety for perception is likely beneficial. Nevertheless, there is a lack of perception requirements which consider the safety in driving context.

1.2 Objectives and Scope

Having provided the overall motivation, the objective and scope of this work are first specified on a high level in this section. Further specification of the scope in the form of RQs is provided in the subsequent section 1.3.

[13] Lin, T.-Y. et al.: Microsoft COCO: Common Objects in Context (2014), p. 742.
[14] D'Amour, A. et al.: Underspecification Presents Challenges (2022), p. 30-31.
[15] Junietz, P. et al.: Macroscopic Safety Requirements (2019), p. 3-4.
[16] Kalra, N.; Paddock, S. M.: Driving to Safety (2016), p. 191.
[17] Liu, P. et al.: How Safe Is Safe Enough for Self-Driving Vehicles? (2019), p. 317.
[18] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 1-2.
[19] Oksuz, K. et al.: One Metric to Measure them All (2020), p. 9446.
[20] Luiten, J. et al.: HOTA: A Higher Order Metric for Evaluating Multi-object Tracking (2021), p. 566.
[21] Philion, J.; Fidler, S.: Lift, Splat, Shoot (2020).
[22] Qi, C. R. et al.: Offboard 3D Object Detection (2021), p. 6135.

Based on the motivation, the objective of this work is to determine perception requirements with respect to the safety of the driving task. This work targets perception for the whole of the AD task. This means that the results are intended for use in vehicles with higher levels of AD at level 4 and above according to SAE[23]. Therefore, no limitations regarding the operational design domain are assumed. A requirement is considered to be specified if both a metric and a threshold for acceptable performance are defined. Furthermore, it is desirable to elicit requirements which are interpretable[24,25] and generalize to different scenarios and downstream planners. Finally, the last objective is to provide a method which can validate the perception requirements.

In order to reduce the complexity of this work, some limitations on the scope are introduced. Since this work considers perception, safety is mainly considered in terms of the safety of the intended functionality (SOTIF)[26] rather than in terms of functional safety[27]. Other limitations pertain to the high-level architecture of the AD system as well as to the perception task. A modular sense-plan-act architecture is assumed, since independently testable modules offer the potential of reducing the testing effort. The sense component includes both the sensor as well as the perception pipeline.[28] In order to leverage this potential and improve generalization, no specific downstream planner implementation is assumed. This assumption also acknowledges the fact that the final implementation of the downstream task may not be available. The task of environment perception may include different tasks such as semantic segmentation or object detection.[29] This work focuses on the task of 3D object detection in the context of collision avoidance. One reason is the direct connection of collisions with dynamic objects to the objective of safety for the driving task. Another reason is the fact that an object list is a common interface between a perception and a planner.[30]

1.3 Research Questions

After describing the objective as well as the general scope, it is possible to formulate the RQs of this work. The RQs further specify the overall scope and provide the structure for the remainder of this document. An overview of the structure is provided in the subsequent section after introducing the RQs in this section.

---

23 On-Road Automated Driving (ORAD) committee: Taxonomy and Definitions (30.04.2021), p. 30-34.

24 Oksuz, K. et al.: One Metric to Measure them All (2020), p. 9446.

25 Ivanovic, B.; Pavone, M.: Injecting Planning-Awareness (2022), p. 822-823.

26 ISO/TC 22/SC 32: ISO/PAS 21448 SOTIF (06/2022), p. vi.

27 ISO/TC 22/SC 32: ISO 26262-1 Functional Safety (12/2018).

28 Amersbach, C. T.: Functional Decomposition Approach (2020), p. 41-48.

29 Feng, D. et al.: Deep Multi-Modal Object Detection and Semantic Segmentation (2020), p. 2.

30 Hoss, M. et al.: A Review of Testing Object-Based Environment Perception (2022), p. 229.

First, the overall RQ encompassing the entirety of this work is presented. Considering the objective and the scope of this work, the overall RQ regarding safety is:

RQ: *Is it possible to specify requirements for 3D object detection which consider the safety of the driving task without fully specifying a downstream planning module?*

This overall RQ contains different aspects pertaining to the task of object detection. To understand detection performance, interpretable components are desirable.[31] Therefore, the overall RQ is decomposed into different RQs numbered from one to four. Each RQ is derived from the overall RQ to further specify it and independently treats one distinct aspect. A visual overview of the structure including the different RQs is provided in Fig. 1-1 in the next section. The decomposition of the overall RQ into the four questions is presented in the following.

Object detection aims to distinguish a selected set of classes from the background.[32] Therefore, the first question when considering object detection is which classes are to be identified. Accordingly, the following RQ regarding classification is raised:

RQ1: *Is it possible to systematically identify categories which must be distinguished to safely perform the driving task?*

The next question after specifying the object categories is which instances of these classes require detection. This is the question of object relevance, equivalent to defining which objects are included in the ground truth (GT) for a detection dataset. While most datasets use arbitrary heuristics for object relevance,[33,34] a more structured approach is desirable. At the same time, reliable information on the ego intention or the road environment may not be available to the object detection component. Therefore, the corresponding RQ regarding relevance is:

RQ2: *Is it possible to systematically identify objects relevant for detection based only on information contained in the object list while considering the safety in driving context?*

The aspects of classification and relevance only consider which objects must be identified. However, the detection quality of the relevant objects remains unspecified. Therefore, the next questions are what physical quantities of these objects require detection and what the required detection quality is. As stated in the objectives, a requirement is specified if metrics including a threshold are identified while considering the safety of the driving task. Accordingly, the corresponding RQ regarding attributes is:

RQ3: *Is it possible to define requirements for the detected object attributes considering the safety of the driving task?*

31 Oksuz, K. et al.: One Metric to Measure them All (2020), p. 9446.

32 Wu, X. et al.: Recent Advances in Deep Learning for Object Detection (2019), p. 43.

33 Caesar, H. et al.: nuScenes (2020), p. 11621.

34 Sun, P. et al.: Scalability in Perception (2020), p. 2447.

While various definitions of detection metrics are available in literature, the validation of metrics and requirements remains an open question. However, the evaluation and safety validation of any detection requirements presented in this work is necessary. Accordingly, the final RQ regarding validation is:

RQ4: *How can detection requirements be validated with respect to the safety of the driving task?*

1.4 Structure of the Document

With this, the RQs considered in this work are specified. The RQs provide the framework which structures later chapters of this document. An illustration of the relation of the different RQs and the corresponding document structure is provided in Fig. 1-1. In the visualization, each RQ is abbreviated with a single descriptive term which is also used for later section and chapter titles. The next chapter 2 discusses basics which relate to all of the RQs and in particular the overall RQ. This is followed by related work which is associated more specifically with each of the four RQs raised in this work. The overall methodology of this work to answer all RQs including the overall RQ is presented in chapter 4. This chapter also provides further information regarding the structure of the subsequent chapters. Finally, chapter 9 provides a discussion in which the answers to all RQs are provided.

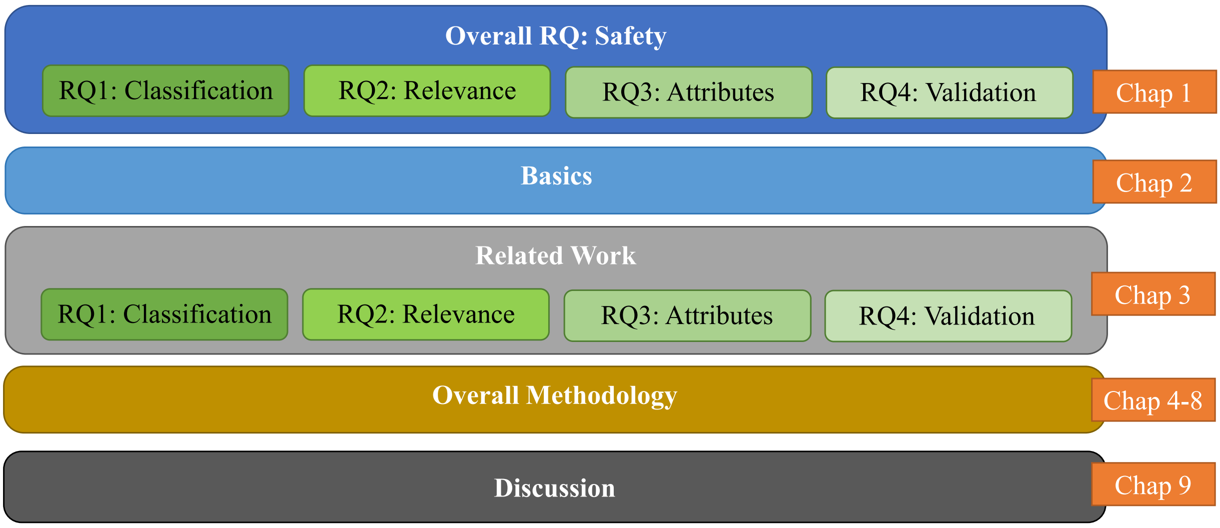

Figure 1-1: Overview of the structure of the RQs and the chapters in this work.

2 Basics

This chapter introduces the background and the basic concepts which are relevant to all RQs, in particular the overall RQ regarding safety. All concepts presented here are already established within literature and serve as basis for the following chapters. Previous work which relates more specifically to the individual RQs 1-4 of this work is presented in chapter 3. First, some general considerations on AD functions are introduced. This is followed by different strategies for testing AD functions. Finally, typical evaluation procedures specifically for perception are discussed.

2.1 Automated Driving Functions

This section presents general considerations regarding AD functions. Firstly, high-level architectures which have previously been applied to the driving task are covered. Next, the role of ML as well as resulting implications are discussed.

2.1.1 Architecture

For the task of AD, two categories of system architectures are typically distinguished. The first consists of end-to-end systems, which directly map sensory input to outputs or trajectories. The second category are modular designs consisting of modules with distinct interfaces.[35] The typical high-level modular architecture is sense-plan-act, which is extended by multiple more complex architectures.[36] As previously indicated, the scope of this work is limited to modular sense-plan-act architectures.

An advantage of modular structures is the possibility of evaluating the internal structure of the system. This means that the understanding is improved during testing when compared to black box systems.[37] However, this necessitates deriving particular criteria for each module from the general safety goals.[38] A modular architecture is also considered to be more robust in addition to the improved interpretability.[39] However, this is also influenced by the choice of the intermediate representation between perception and planning. Finding such a suitable intermediate representation is still an open problem.[40] Therefore, various representations are

[35] Le Mero, L. et al.: Survey on Imitation Learning Techniques (2022), p. 14128.

[36] Amersbach, C. T.: Functional Decomposition Approach (2020), p. 41-44.

[37] Thorn, E. et al.: A Framework for Automated Driving System Testable Cases and Scenarios (2018), p. 64-65.

[38] Klamann, B. et al.: Defining Pass-/Fail-Criteria for Particular Tests (2019), p. 169-174.

[39] Bansal, M. et al.: ChauffeurNet (2018), p. 5.

[40] Hu, Y. et al.: Scenario-Transferable Semantic Graph Reasoning (2022), p. 23212.

found in literature.[41] Only the task of 3D object detection which outputs a 3D object list as intermediate representation is considered in this work.

End-to-end architectures are credited with higher computation efficiency compared with modular designs.[42a] On the other hand, end-to-end architectures typically leverage highly abstract implicitly learned internal representations.[43] However, the work “Lift, Splat, Shoot” shows that end-to-end learning is also applicable to interpretable representations.[44] Nevertheless, these approaches may lack transferability,[45] especially when combined with the common practice of training agents in simulations[46]. These factors constitute challenges when applying end-to-end to safety critical tasks such as AD.[42b]

2.1.2 Machine Learning

The state of the art regarding different tasks such as perception and prediction is achieved through the application of DNNs.[47] Applying a neural network (NN) allows approximating an arbitrary unknown function.[48] NNs and more specifically DNNs rely on data to specify samples for the desired output. These samples are used to train ML components and verify that they meet requirements.[49] To apply these methods, sufficient data including labels for the desired output are required. An advantage of these methods is that they are applicable to tasks such as perception, where it is difficult to specify how to obtain this desired output.

While DNNs have been applied for different tasks of AD, application to perception is especially common. Benchmark performance for general purpose visual perception such as Common Objects in Context (COCO) and Pascal Visual Object Classes is currently reached by DNNs.[50] Here, DNNs outperform traditional methods by a large margin.[51] NNs are successfully applied to a growing number of perception tasks[52] including benchmarks in the context of AD[47].

[41] Ilievski, M. et al.: Design Space of Behaviour Planning for Autonomous Driving (2019), p. 2-3.

[42] Le Mero, L. et al.: Survey on Imitation Learning Techniques (2022), a: p. 14128, b: p. 14145.

[43] Hu, Y. et al.: Scenario-Transferable Semantic Graph Reasoning (2022), p. 23214.

[44] Philion, J.; Fidler, S.: Lift, Splat, Shoot (2020), p. 5-7.

[45] Srikanth, S. et al.: INFER: INtermediate representations for FuturE pRediction (2019), p. 943.

[46] Moghadam, M.; Elkaim, G. H.: A Hierarchical Architecture (2019), p. 2.

[47] Huang, Y.; Chen, Y.: Autonomous Driving with Deep Learning (2020), p. 221-222.

[48] Mehrabi, M. et al.: Bounds on the Approximation Power (2018), p. 1.

[49] Ashmore, R. et al.: Assuring the Machine Learning Lifecycle (2022), p. 3-4.

[50] Feng, D. et al.: Deep Multi-Modal Object Detection and Semantic Segmentation (2020), p. 1241-1343.

[51] Wu, C. et al.: A hierarchical loss (2019), p. 2.

[52] Houben, S. et al.: Inspect, Understand, Overcome (2022), p. 5.

Another task in AD is the prediction of trajectories of other traffic participants. While physics-based methods without ML are available, they are generally limited to simple cases. Therefore, recent methods incorporate ML and in particular deep learning to handle complex situations including interactions to achieve state of the art performance.[53,54]

For the task of planning, DNNs are applied less frequently. Contrary to perception, attempts have been made to formalize the planning task such as with the Responsibility-Sensitive Safety (RSS) model.[55] Nevertheless, deep learning is also increasingly applied to the task of planning. Here, the focus is primarily on imitation learning where human expert demonstrations are learned from large scale datasets.[42a]

However, despite their performance, DNNs display model-immanent insufficiencies regarding safety.[56a] One of the concerns is their lack of interpretability.[57] Another major concern is the fact that DNNs exhibit a lack of robustness towards perturbations in the input.[56b,57b,58] This issue has been conceptualized as underspecification due to the complexity of the task which is only specified by samples.[59,60] These shortcomings of DNNs emphasize the importance of thorough safety evaluation driven by clear perception requirements.

2.2 Testing Automated Driving Functions

This section covers basics regarding the testing of complete AD functions. Testing approaches specifically for perception components are discussed separately in section 2.3. Firstly, some general considerations regarding safety are introduced. This is followed by different testing approaches for AD.

2.2.1 Safety

This section considers the topic of safety. Applicable regulations as well as definitions relevant to the topic are discussed.

The ISO 26262 considers the topic of functional safety of electrical and electronic systems within road vehicles. It covers systematic as well as random hardware failures and their consequences. A

[53] Huang, Y. et al.: A Survey on Trajectory-Prediction Methods for Autonomous Driving (2022), p. 652-663.

[54] nuScenes: nuScenes prediction task: Leaderboard (2020).

[55] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017), p. 1.

[56] Houben, S. et al.: Inspect, Understand, Overcome (2022), a: p. 5., b: p. 6.

[57] Willers, O. et al.: Safety Concerns and Mitigation Approaches (2020), a: p. 339, b: p. 341.

[58] Gopinath, D. et al.: Property Inference for Deep Neural Networks (2019), p. 797.

[59] Spanfelner, B. et al.: Challenges in applying the ISO 26262 for driver assistance systems (2012), p. 10-11.

[60] D'Amour, A. et al.: Underspecification Presents Challenges (2022), p. 30-31.

failure is hereby defined as termination of the intended behavior.[61a] According to ISO 26262, safety is defined as the absence of unreasonable risk. Unreasonable is defined as being unacceptable according to societal moral concepts.[61b] Here, risk is a combination of the likelihood of occurrence and the severity of a harm. Harm refers to injury or the damage to the health of people.[61c] If a system possesses additional means to perform a required function beyond what is sufficient, the system is redundant.[61c]

Another complementary standard is the ISO/PAS 21448 which considers the SOTIF. The SOTIF considers systems free from faults as per ISO 26262. However, further hazards may originate from the intended functionality or performance limitations. Reasons may include the lack of situation comprehension or insufficient robustness.[62a] The ISO/PAS 21448 is applied where situational awareness is critical to safety, including complex processing algorithms for sensor data. It is applicable to automation levels of 1-5 according to SAE, thus including fully automated driving.[62b,63] However, a functional description including the intended functionality, the system design and performance targets are required.[62c]

2.2.2 Safety Requirements

This section covers existing approaches in literature to specify the safety requirements for the driving task.

The most common approach is to consider the risk of accidents. Some works focus on fatality rates[64,65] while other works consider accidents of different severity[66,67a]. Typically, these works consider the human performance for the respective accident type as a baseline.[68,65] It has however been argued that contribution of the vehicles should also be assessed to account for rule violations or dangerous behavior by other traffic participants.[69] Furthermore, risk acceptance also has a subjective element based on factors such as controllability, personal benefit or whether the risk is natural or synthetic.[64] In addition, a survey shows that the acceptable risk may even be surpassed by prevalent technology such as human driven vehicles.[67b]

[61] ISO/TC 22/SC 32: ISO 26262-1 Functional Safety (12/2018), a: p. vi-14, b: p. 21-26, c: p. 15-21.

[62] ISO/TC 22/SC 32: ISO/PAS 21448 SOTIF (06/2022), a: p. vi, b: p. 1, c: p.21-22.

[63] On-Road Automated Driving (ORAD) committee: Taxonomy and Definitions (30.04.2021).

[64] Junietz, P. et al.: Macroscopic Safety Requirements (2019), p. 3-4.

[65] Kalra, N.; Paddock, S. M.: Driving to Safety (2016), p. 191.

[66] Junietz, P. et al.: Evaluation of Different Approaches to Address Safety Validation (2018), p. 493.

[67] Liu, P. et al.: How Safe Is Safe Enough for Self-Driving Vehicles? (2019), a: p. 317, b: p. 319-322.

[68] PEGASUS Project: PEGASUS Method: An Overview (2019), p. 5.

[69] Victor, T. et al.: Safety Performance of the Waymo System (2023), p. 10.

Other works define conditions beside accidents which are considered to be unsafe. Off-road driving has been penalized in the loss function of planners[70] and in planning benchmarks[71]. The planning benchmark nuPlan also penalizes traffic rule violations and encourages human similarity.[71]

Overall, the focus of safety requirements generally lies on the behavioral level. These behaviors are easily observed and evaluated as output of the vehicle. However, behavioral requirements are not directly applicable on intermediate interfaces of modular architectures as are prevalent in AD.

2.2.3 Testing Approaches

Verification and validation are intended to provide an argumentation that the risk is below an acceptable level.[72] Even for clearly specified requirements such as fatality rates, this proof is challenging. This section gives an overview over the approaches most prevalent in literature.

The first is the distance based approach, also referred to as real world testing. Here, the system is deployed and the safety is observed using the frequency of accidents.[66] The macroscopic safety requirements as accident rates can be directly applied as reference.[73] However, the statistical proof of decreased fatality rates requires distances in the order of billions of kilometers and is therefore considered infeasible.[74a,75] An alternative is extreme value theory which extrapolates to rare accident events based on more frequent critical events. However, obtaining generally applicable surrogate metrics to define critical events remains a challenge.[74a]

Another method is the function based testing, where different system functions and their requirements are specified and then tested. However, a general description of the AD functionality is considered infeasible.[76]

One approach to overcome the difficulty of functional description is to decompose the behavior into scenarios.[76] The testing of the scenario can be performed in different environments including simulation.[73b] The reaction of the system to a repeatable predefined scenario is then assessed. However, the parameter space for the task of AD is large and no link to distance is available.[74b] Currently, the state of the art is to use random real world data, where coverage is difficult to achieve.[77] In addition, an argumentation for interpolation between two scenarios and for

70 Bansal, M. et al.: ChauffeurNet (2018), p. 8-12.

71 Caesar, H. et al.: nuPlan: A closed-loop ML-based planning benchmark (2021), p. 4.

72 ISO/TC 22/SC 32: ISO 26262-1 Functional Safety (12/2018), p. 10.

73 Amersbach, C. T.: Functional Decomposition Approach (2020), a: p. 9-10, b: p. 14-20., c: p. 45, d: p. 97-98.

74 Junietz, P. et al.: Evaluation of Different Approaches to Address Safety Validation (2018), a: p. 493, b: p. 494.

75 Kalra, N.; Paddock, S. M.: Driving to Safety (2016), p. 191.

76 Riedmaier, S. et al.: Survey on Scenario-Based Safety Assessment (2020), p. 87458-87472.

77 Schwalbe, G. et al.: Structuring the Safety Argumentation for Perception (2020), p. 392.

extrapolation to novel scenarios is generally required. Previous work coauthored by the author has shown that for DNN based components, such an interpolation may not be applicable.[78]

Another approach is silent testing, where simulated driving functions are compared to real driver trajectories in a real vehicle.[79,80] Similar approaches have also been applied by automotive companies.[81] However, one disadvantage of this approach is the lack of consideration of interactive behavior.[76]

To conclude, the testing of AD functions remains challenging even for well specified requirements. While different approaches are currently considered, no generally accepted procedure for safety validation is available.

2.2.4 Modular Decomposition

In this section, the approach of modular decomposition is introduced. This methodology is applicable for system design and testing.

Modular decomposition is the idea of decomposing the system under test (SUT) into functional layers. Each of these functional layers provides standardized interfaces which are evaluated separately at testing time.[73c] Previous work has shown that particular tests for each module as well as less complex subsystems promise a reduction potential of test cases by an order of one or two magnitudes.[73d] It should be noted that typical AD architectures consist of sense-plan-act modules. Therefore, the scope of this work was also limited to such architectures. For these architectures, the modular decomposition approach for testing is directly applicable.

2.3 Common Perception Evaluation

The approaches for testing the safety of AD functions generally rely on requirements specified on behavioral level. However, testing a perception module separately requires particular tests on the perception interface. Since safety requirements on a behavioral level are inapplicable, other approaches have become the de-facto standard in perception testing. This section first outlines different perception tasks, followed by perception datasets as well as common evaluation metrics.

---

78 Mori, K. T. et al.: The Inadequacy of Discrete Scenarios (2022), p. 118240-118241.

79 Wachenfeld, W.; Winner, H.: Virtual Assessment of Automation in Field Operation (2015), p. 2-4.

80 Wang, C.; Winner, H.: Overcoming Challenges of Validation (2019), p. 2640-2641.

81 Tesla: Upgrading Autopilot: Seeing the World in Radar (2016).

2.3.1 Perception Tasks

In this section, different perception tasks including alternatives to object detection are introduced. Each task provides distinct objectives and representations.

One task typically considered for the image domain is the classification task. The objective is to determine the presence or absence of an object[82], commonly including a confidence score for compatibility with evaluation metrics[83]. The 2D object detection task aims to predict both the class and location of objects with a bounding box in image space.[82] Similarly, the task can be formulated to detect 3D objects in a scene. In this case, objects are typically represented as 3D cuboid with class, location, size, rotation and potentially velocity.[84] While object detection focuses on object classes, semantic segmentation aims at pixel wise annotation for stuff classes.[85] Other variants include the instance segmentation with a pixel-wise mask per instance or panoptic segmentation which provides both a class and instance ID per pixel.[86] This task has been extended to 3D by performing semantic and panoptic segmentation on 3D sensor data.[87a] Another alternative is the online prediction of a semantic map in a birds-eye view (BEV) perspective.[88,89] In any of these variants, segmentation emphasizes global context and scene understanding.[85] The previous tasks are typically evaluated per frame without considering temporal dynamics. Therefore, the alternative task of multi object tracking associates detections in a sequence either in 2D image space or in 3D space.[90,91,87b] Overall, the object list is currently the common interface between perception and planning. Tracked objects are more common than untracked object lists.[92] Considering this fact as well as the scope of this work, the following sections mainly focus on 3D object detection.

2.3.2 3D Object Detection Datasets

For ML components including DNN, data is used for training as well as evaluation.[93] Therefore, publicly available datasets are important since they often determine the objectives for research.[94]

82 Wu, C. et al.: A hierarchical loss (2019), p. 1-2.
83 Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010), p. 305.
84 Mao, J. et al.: 3D Object Detection for Autonomous Driving (2023), p. 1910.
85 Caesar, H. et al.: COCO-Stuff (2018), p. 1209.
86 Kirillov, A. et al.: Panoptic Segmentation (2019), p. 9396-9397.
87 Fong, W. K. et al.: Panoptic nuScenes (2022), a: p. 3795, b: p. 3798.
88 Pan, B. et al.: Cross-view Semantic Segmentation (2020), p. 4876.
89 Hendy, N. et al.: FISHING Net (2020), p. 1-2.
90 Weng, X. et al.: 3D Multi-Object Tracking (2020), p. 10360.
91 Caesar, H. et al.: nuScenes (2020), p. 11623.
92 Hoss, M. et al.: A Review of Testing Object-Based Environment Perception (2022), p. 229.
93 Ashmore, R. et al.: Assuring the Machine Learning Lifecycle (2022), p. 3-4.
94 Lin, T.-Y. et al.: Microsoft COCO: Common Objects in Context (2014), p. 742.

This section provides a brief overview of 3D object detection datasets while the detailed implications are discussed in later sections. Popular datasets such as KITTI[95], nuScenes[96] and Waymo Open[97] are discussed with respect to their general similarities.

One aspect that these datasets share is the data collection procedure. Early image based datasets such as Pascal VOC or ImageNet typically used images collected over the internet.[98,99] This approach is no longer applicable for the automotive domain. For multimodal data as is typically acquired in the automotive domain, additional effort for sensor integration and calibration is required.[96a] Therefore, such datasets rely on test drives using vehicles equipped with designated sensor setups. The sensor setups are typically described without further argumentation for the selected placement. Many datasets utilize at least one 360° lidar mounted on the vehicle roof which is supplemented with additional cameras. Nevertheless, the setups differ with regards to sensor modalities, number of sensors, field of view and resolution.[95a, 96a, 97b, 100,101,102,103,104,105,106] Surveying the available information from safety reports of companies holding permits for autonomous vehicle testing in California [107] shows similar trends [108,109,110]. Overall, sensor setups show variations which are difficult to evaluate due to a lack of established evaluation procedures. Existing evaluation protocols are only applicable for given GT and sensor data which presupposes a fixed sensor setup.

Once the real world data is collected, the tasks and objectives are defined. Firstly, representative or interesting scenes are selected to limit the labeling effort.[95a, 96a,111] For the selected scenes, datasets typically provide labels for multiple different perception tasks. These different tasks such as 3D object detection and tracking typically coexist and are solved and evaluated separately.[96b, 97a] In any case, the dataset arbitrarily defines the classes considered relevant for the respective task. For the object detection task, only the selected classes are considered objects, while all other objects are

[95] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), a: p. 3355-3357.
[96] Caesar, H. et al.: nuScenes (2020), a: p. 11620-11621, b: p. 1622-1623.
[97] Sun, P. et al.: Scalability in Perception (2020), a: p. 2447-2448, b: p. 2444-2445.
[98] Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010), p. 305.
[99] Deng, J. et al.: ImageNet: A large-scale hierarchical image database (2009), p. 251.
[100] Geyer, J. et al.: A2D2: Audi Autonomous Driving Dataset (2020), p. 3-4.
[101] Patil, A. et al.: The H3D Dataset (2019), p. 9254-9255.
[102] Chang, M. et al.: Argoverse (2019), p. 8742-8743.
[103] Pham, Q.-H. et al.: A 3D Dataset (2020), p. 2269.
[104] Sun, P. et al.: Scalability in Perception (2020), p. 2444-2445.
[105] Mao, J. et al.: One Million Scenes for Autonomous Driving: ONCE Dataset (2021), p. 4.
[106] ApolloAuto: apollo (2021).
[107] State of California Department of Motor Vehicles: Autonomous Vehicle Testing Permit Holders (2022).
[108] General Motors: Self Driving Safety Report (2018), p. 7.
[109] nuro: Delivering Safety (2021), p. 9-10.
[110] Waymo LLC: Waymo Safety Report (2020), p. 14.
[111] Cordts, M. et al.: The Cityscapes Dataset (2016), p. 3214.

considered background.[112] Therefore, current datasets exhibit ambiguous class definitions[113,114] as well as inconsistent or conflicting taxonomies between different datasets.[115,116] In addition, not all objects belonging to these classes are considered relevant. To facilitate labeling and perception, heuristics are applied to determine which objects of the selected classes are included in the GT. For instance, nuScenes demands the presence of sensor detection points for object annotation[117] while the Waymo Open Dataset limits the annotation range[118a]. Overall, datasets exhibit various arbitrary design choices in their specification of the perception task.

Once the tasks and objectives are clarified, the GT labels are created. For image data, datasets such as ImageNet and COCO relied on labeling with crowdsourcing and additional measures to ensure label quality.[119,120] Datasets for 3D tasks in the automotive domain such as KITTI or nuScenes generally rely on expert annotators.[117, 121] The Waymo Open dataset additionally relies on labeling tools to support the human annotators.[118b] It has however been noted that the human performance for the 3D detection task is unclear.[122]

Overall, existing object detectors heavily rely on public datasets for training and evaluation. However, these datasets include many arbitrary design choices during their creation.

2.3.3 Common Object Detection Metrics

When a dataset is published, it typically includes evaluation metrics which establish a benchmark ranking. This section provides a brief overview over the distinct metrics of the object detection and tracking task. The overview begins with general detection and tracking metrics before focusing on the context of driving. Finally, the aspect of redundancy in perception is considered.

Evaluating tracked or untracked object lists requires matching the perceived objects with the ground truth objects. The matching defines positive and negative samples according to a distance metric using arbitrary thresholds.[123] The most common metric for object detection is the average precision (AP) metric[124] which originated from 2D object detection. AP uses the bounding box

[112] Wu, C. et al.: A hierarchical loss (2019), p. 5.

[113] Zendel, O. et al.: WildDash - Creating Hazard-Aware Benchmarks (2018), p. 411-412.

[114] Huang, X. et al.: The ApolloScape Open Dataset (2020), p. 5.

[115] Lambert, J. et al.: MSeg: A Composite Dataset (2020), p. 2877.

[116] Bevandic, P. et al.: Multi-domain semantic segmentation with overlapping labels (2022), p. 2422.

[117] Caesar, H. et al.: nuScenes (2020), p. 11620-11621.

[118] Sun, P. et al.: Scalability in Perception (2020), a: p. 2447, b: p. 2446.

[119] Deng, J. et al.: ImageNet: A large-scale hierarchical image database (2009), p. 251-252.

[120] Lin, T.-Y. et al.: Microsoft COCO: Common Objects in Context (2014), p. 745-748.

[121] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3358.

[122] Qi, C. R. et al.: Offboard 3D Object Detection (2021), p. 6135.

[123] Hoss, M. et al.: A Review of Testing Object-Based Environment Perception (2022), p. 229.

[124] Oksuz, K. et al.: One Metric to Measure them All (2020), p. 4.

overlap calculated as intersection over union (IoU) for matching. For these samples, the average over an interpolated precision/recall curve is calculated.[125] The AP is most often applied per class and then averaged across all classes to yield the mean average precision (mAP).[126]

For the tracking task, the evaluation additionally includes the unique identification over time. Here, the most common metric is the multi object tracking accuracy (MOTA) introduced by the CLEAR multi object tracking (MOT) metrics.[127] This metric sums false negatives (FNs), false positives (FPs) and ID mismatches and normalizes them with the number of GT objects.[128] Proposed modifications include the average multi object tracking accuracy which integrates the MOTA across multiple recall values while the scaled multi object tracking accuracy scales the MOTA with the recall. Combining the two approaches yields the scaled average multi object tracking accuracy (sAMOTA).[129] Another alternative is proposed by the higher order tracking accuracy metric which re-weights different interpretable error components.[130]

In addition to matching based metrics, it is also possible to consider true positive metrics which evaluate the accuracy of the localization for matched objects. Localization accuracy can be indirectly included in AP by averaging over multiple IoU thresholds as introduced by the COCO dataset.[131] However, the localization can also be explicitly considered by evaluating metrics such as multi object tracking precision, which assesses the precision of the localization.[128.] Overall, 2D detection metrics consider average performance mostly in terms of average location offsets.

When surveying detection metrics in driving context, the general tendency is to adopt metrics from 2D with little modification. The KITTI dataset follows the Pascal AP[132] which is also adopted by the A*3D dataset[133] Other datasets have proposed different modifications. nuScenes proposes to match objects by center distance and to omit recall and precision values below 10% to minimize the impact of noise.[134] Waymo Open weighs the true positives (TPs) according to the heading score to emphasize the importance of heading. Additionally, the matching relies upon global optimization instead of greedy matching.[135] The ONCE dataset modifies the matching procedure from KITTI by using class specific IoU thresholds and including the orientation in the matching.[136] Detection and tracking typically exist as independent tasks on driving datasets. Argoverse directly

[125] Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010), p. 313-314.

[126] Liu, Y. et al.: 1st Place Solutions for OpenImage2019 (2020), p. 271.

[127] Luo, W. et al.: Multiple Object Tracking: A Literature Review (2021), p. 11.

[128] Bernardin, K.; Stiefelhagen, R.: Evaluating Multiple Object Tracking Performance (2008).

[129] Weng, X. et al.: 3D Multi-Object Tracking (2020), p. 10362-10363.

[130] Luiten, J. et al.: HOTA: A Higher Order Metric for Evaluating Multi-object Tracking (2021), p. 548-556.

[131] COCO Consortium: COCO Common Objects in Context: Detection Evaluation (2015).

[132] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3358.

[133] Pham, Q.-H. et al.: A 3D Dataset (2020), p. 2271.

[134] Caesar, H. et al.: nuScenes (2020), p. 11622.

[135] Sun, P. et al.: Scalability in Perception (2020), p. 2447.

[136] Mao, J. et al.: One Million Scenes for Autonomous Driving: ONCE Dataset (2021), p. 21.

adopts the CLEAR metrics such as MOTA[137] for tracking while nuScenes additionally relies on its adaptation sAMOTA[138]. Some datasets further adopt true positive metrics for orientation[139a] or various attributes[138b]. In addition to the arbitrary design choices in the dataset creation, evaluation metrics may introduce further arbitrary design choices when specifying the perception task. For instance, KITTI only considers objects above a certain height when projected to the image plane for evaluation.[139b] Alternatively, the nuScenes[138] dataset has distance thresholds specific to each class.[140] Overall, detection performance is fundamentally conceptualized in terms of average location offsets. Accordingly, there is a lack of clear requirements for object detection.

Another aspect besides the object detection performance under normal operating conditions is the response to hardware failures. If alternative means to fulfill detection requirements are present even for a sensor hardware failure, a system is considered to be redundant.[141] However, in the absence of clear detection requirements, the concept of redundancy remains underspecified. Previous works therefore evaluate the robustness of detection performance under artificial sensor failures such as sensor dropout or misalignment.[142a,143,144,145] Existing lidar and lidar-camera detectors exhibit low performance in these test cases, particularly for simulated lidar failures.[142b] However, correlation of failures which may occur due to common mode and common cause failures[146] is not considered. Therefore, redundancy of different detectors, sensors and modalities remains insufficiently understood.

Overall, typical perception metrics evaluate the average performance without specifying requirements including thresholds. Safety aspects relating to the driving task or to potential sensor failures are insufficiently considered.[147,148] Furthermore, many common metrics provide aggregated scores which lack interpretability.

---

[137] Chang, M. et al.: Argoverse (2019), p. 8745.

[138] Caesar, H. et al.: nuScenes (2020), a: p. 11623., b: p. 11622.

[139] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), a: p. 3358, b: p. 3360.

[140] nuScenes: nuScenes prediction task: Leaderboard (2020).

[141] ISO/TC 22/SC 32: ISO 26262-1 Functional Safety (12/2018), p. 20.

[142] Yu, K. et al.: Benchmarking the Robustness (2023), a: p. 3191-3193, b: p. 3193-3195.

[143] Philion, J. et al.: Learning to Evaluate Perception (2020), p. 203-205.

[144] Bai, X. et al.: TransFusion (2022), p. 1086-1087.

[145] Mohta, A. et al.: Investigating the Effect of Sensor Modalities (2021), p. 1-4.

[146] Stapelberg, R. F.: Handbook of Reliability (2009), p. 621-623.

[147] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 1-2.

[148] Willers, O. et al.: Safety Concerns and Mitigation Approaches (2020), p. 342.

3 Related Work

After introducing basics mainly relating to the overall RQ of safety in the previous chapter, this chapter introduces literature which relates more closely to the specific RQs 1-4. The content is distinct from the previous chapter in that these works cannot yet be considered established and are not part of standard evaluation procedures. As already indicated in section 1.3, each of the RQs 1-4 relates to one of the aspects of classification, relevance, attributes and validation. Each section in this chapter discusses one of these aspects for the context of perception with focus on object detection in AD. Finally, a brief summary of the state of the art is provided.

3.1 Classification

As shown in previous sections, object detection aims to distinguish the predefined set of classes from the background[149] and provide a likelihood for the predicted classes[150,151a]. As indicated previously, common metrics explicitly consider the class for matching during the evaluation.[151b] Therefore, the correct category is a prerequisite for correct matching.[149] Thus, the predefined categories and their correct prediction are effectively treated as equally important as detecting the existence of an object. Due to their large impact on the evaluation procedure, different classification schemes and metrics are elaborated in this section. This content of this section is taken from a prior publication[152] coauthored by the author with little modification.

3.1.1 Dataset Categories

This section jointly considers different classification-relevant tasks and datasets with focus on the automotive domain.

Large-scale datasets which are popular for the task of classification[153,154] may include as many as thousands of image categories. For other perception tasks, the number of classes is typically smaller by orders of magnitude. For instance, the 2D object detection task only reaches a maximum

[149] Wu, X. et al.: Recent Advances in Deep Learning for Object Detection (2019), p. 5.

[150] Zhou, X. et al.: Probabilistic two-stage detection (2021), p. 3.

[151] Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010), a: p. 306, b: p. 313.

[152] Mori, K. T. et al.: Systematic Classification Requirements (2023) © 2023 IEEE.

[153] Deng, J. et al.: ImageNet: A large-scale hierarchical image database (2009), p. 248-249.

[154] Kuznetsova, A. et al.: The Open Images Dataset V4 (2020), p. 1958-1960.

of 600 classes provided by the Open Images Dataset[155]. Other datasets for the same task[156,157] number fewer categories.

Specifically within the context of driving, the nuScenes dataset with 23 classes[158] and Argoverse 2 with 30 classes[159] provide the highest number of classes for object detection. This applies both to 2D[160,161] and to 3D detection within driving datasets[158,162,163,164,165a,166]. Semantic segmentation annotations typically provide a higher number of categories, ranging between 24-66.[161b,167,168,169] In some cases, additional tasks are annotated and separately evaluated. Examples include recognition of semantic maps[158a,165b], driveable area[170,171] or lane markings[161b,168]. The discrepancy between object detection and semantic segmentation categories results from the stuff categories which do not possess object properties.[172,173] Overall, the class definitions show ambiguities and inconsistencies.[168,174] Therefore, no generally accepted class structure is available.

3.1.2 Existing Class Structures

Dataset classes are typically presented as flat taxonomies without further structure. Similar flat taxonomies are proposed by works unifying classification taxonomies across domains for compatibility with standard training procedures.[175,176,177]

155 Kuznetsova, A. et al.: The Open Images Dataset V4 (2020), p. 1958-1960.

156 Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010), p. 305.

157 Lin, T.-Y. et al.: Microsoft COCO: Common Objects in Context (2014), p. 746.

158 Caesar, H. et al.: nuScenes (2020), p. 11621.

159 Wilson, B. et al.: Argoverse 2 (2021), p. 4.

160 Geyer, J. et al.: A2D2: Audi Autonomous Driving Dataset (2020), p. 6-7.

161 Yu, F. et al.: BDD100K: A Diverse Driving Dataset (2020), a: p. 2635, b: p.2636.

162 Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3357.

163 Sun, P. et al.: Scalability in Perception (2020), p. 2447.

164 Chang, M. et al.: Argoverse (2019), p. 8744.

165 Houston, J. et al.: One Thousand and One Hours (2020), a: p. 5, b: p. 6.

166 Mao, J. et al.: One Million Scenes for Autonomous Driving: ONCE Dataset (2021), p. 5.

167 Cordts, M. et al.: The Cityscapes Dataset (2016), p. 3214-3215.

168 Huang, X. et al.: The ApolloScape Open Dataset (2020), p. 2706-2708.

169 Neuhold, G. et al.: The Mapillary Vistas Dataset (2017), p. 5002-5003.

170 Pan, B. et al.: Cross-view Semantic Segmentation (2020), p. 4867.

171 Hendy, N. et al.: FISHING Net (2020), p. 1-2.

172 Caesar, H. et al.: COCO-Stuff (2018), p. 1210.

173 Kirillov, A. et al.: Panoptic Segmentation (2019), p. 9396-9397.

174 Zendel, O. et al.: WildDash - Creating Hazard-Aware Benchmarks (2018), p. 411-412.

175 Zendel, O. et al.: Unifying Panoptic Segmentation for Autonomous Driving (2022), p. 21353.

176 Lambert, J. et al.: MSeg: A Composite Dataset (2020), p. 2878.

177 Bevandic, P. et al.: Multi-domain semantic segmentation with overlapping labels (2022), p. 2424-2425.

However, some approaches to structure dataset categories have been proposed within literature. A semantically expressive formal specification of concepts including entities and their relations is given by a so called ontology.[178] Previous works on datasets[179,180,181a] or unified taxonomies across multiple datasets[182] implicitly define such an ontology. However, they generally do not address the concept of ontologies explicitly and fail to provide an argumentation for the selected semantic hierarchy.

Some works also consider ontologies in a more explicit manner. A prominent example is the ImageNet dataset for image classification. This dataset creates a semantic structure based on an ontology into which the images are integrated.[183] The ontology is provided by the external source WordNet which structures concepts including synonyms and semantic relations as a lexical dataset.[184] YOLO9000 also relies on ImageNet to scale up the object detector to a higher number of classes.[185] In the context of driving, ontologies have been applied to create and structure scenarios in scenario based testing approaches for AD.[178,186] However, the focus in these works lies on behavioral aspects rather than the classification of the perception component.

3.1.3 Classification Metrics

As indicated in the previous section, datasets typically do not consider the structure of classes. Accordingly, a similar trend of neglecting class structure is observed in typical evaluation metrics.[187] One exception is the OpenImages dataset, which explicitly considers the hierarchy for its AP calculation by aggregating leaf nodes for higher level categories.[181b] The ImageNet dataset originally proposed a hierarchical error based on the height of the lowest shared node. However, Top-5 accuracy was preferred for practical use since different metrics show high agreement.[188]

A different type of evaluation is given by loss functions which are used at training time to optimize the neural networks. Cross-entropy loss is the standard loss typically used for classification tasks. Similar to prevalent classification metrics, this loss neglects the structure between classes.[187] However, some previous works make proposals to incorporate class structure into loss functions.

[178] Klueck, F. et al.: Using Ontologies for Test Suites Generation (2018), p. 120.
[179] Cordts, M. et al.: The Cityscapes Dataset (2016), p. 3214-3215.
[180] Lin, T.-Y. et al.: Microsoft COCO: Common Objects in Context (2014), p. 746.
[181] Kuznetsova, A. et al.: The Open Images Dataset V4 (2020), a: p. 1959-1960, b: p. 1974.
[182] Meletis, P.; Dubbelman, G.: Training on Multiple Heterogeneous Datasets (2018), p. 1046-1047.
[183] Deng, J. et al.: ImageNet: A large-scale hierarchical image database (2009), p. 248-249.
[184] Miller, G. A.: WordNet: a lexical database for English (1995), p. 39.
[185] Redmon, J.; Farhadi, A.: YOLO9000: Better, Faster, Stronger (2017), p. 6522-6524.
[186] Bagschik, G. et al.: Ontology based Scene Creation (2018), p. 1814-1819.
[187] Wu, C. et al.: A hierarchical loss (2019), p. 1.
[188] Russakovsky, O. et al.: ImageNet Large Scale Visual Recognition Challenge (2015), p. 225.

One idea is to incorporate the similarity of classes.[189] Other proposals consider the consequences of misclassifications by considering semantic structure[190a] or the impact on collision safety[191].

Overall, these metrics assume a hierarchical class structure for evaluation. Class similarity is conceptualized with respect to this hierarchy. Variants include using a distance in the structure[190b] or the depth of the lowest shared node[192]. It has been noted that loss functions which consider the class structure poss difficulties for the optimization.[190c] In addition, existing approaches tend to focus on semantic similarity. However, aspects such as regulations and considerations regarding safety in the context of traffic are neglected.

3.2 Relevance

After considering the target classes, it is necessary to determine which objects of these classes are relevant for the perception task. Considering object relevance for the perception safety evaluation is necessary.[193] As shown in section 2.3.3, relevance is only implicitly considered by datasets by inclusion in the GT using heuristics. This section provides a detailed overview of more explicit conceptualizations of relevance. Approaches are categorized into those relying on heuristics, those relying on formal specifications of the driving task and approaches incorporating a specific downstream task implementation. In addition, the two related topics of criticality metrics and relevance in vision are discussed. The author coauthored two prior publications on relevance[194,195] which include a literature overview. The content of this section is therefore taken from these prior publications with minor modifications.

3.2.1 Heuristics

Heuristics are mainly used in cases where relevance is only implicitly considered. Typically, simple arbitrary criteria are used to exclude certain objects.

One example already mentioned in previous sections are datasets for perception testing. Some datasets restrict the annotated objects to objects within specific distance[196] or to objects visible in lidar or radar[197]. In either case, the heuristic determines if an object is included into the GT

[189] Kobs, K. et al.: SimLoss: Class Similarities in Cross Entropy (2020), p. 431-439.

[190] Wu, C. et al.: A hierarchical loss (2019), a: p. 3-4, b: p. 1-4, c: p. 5

[191] Liu, X.-F. et al.: Reinforced Wasserstein Training (2020), p.12563-12564.

[192] Russakovsky, O. et al.: ImageNet Large Scale Visual Recognition Challenge (2015), p. 225.

[193] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 3-4.

[194] Mori, K. et al.: Conservative Estimation of Perception Relevance (2023) © 2023 IEEE.

[195] Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023).

[196] Sun, P. et al.: Scalability in Perception (2020), p. 2447.

[197] Caesar, H. et al.: nuScenes (2020), p. 11621.

object list. In other cases, heuristics are applied at evaluation time by the perception metrics.[198] Dataset metrics use simple criteria such as the object distance[199]. The KITTI dataset uses criteria based on visibility in camera images by considering the object height in the camera image as well as occlusion.[200]

A different application of simple heuristics is provided by the downstream path planning task. While the processing within the network remains a blackbox, it is possible to analyze the network inputs. Neural planners typically limit the objects input to the network to reduce computational effort. Any object not contained in the input is thus implicitly declared irrelevant. The filtering criteria applied depend on the type of input ingested by the planner. Convolutional networks which receive a BEV grid as input are restricted to a rectangular region with predefined dimensions.[201,202,203] Similar ideas are found in context of simulation to test AD functions.[204] Other planners which receive object lists directly restrict the number of objects.[205,206] Similar considerations are applied for the motion prediction task where the number of objects is limited.[207,208,209]

3.2.2 Formal Specification

Formal approaches specify the potential behaviors and behavioral requirements to explicitly derive relevant objects. Two common approaches to incorporate behavioral safety into relevance are reachability and formal planners.[198] The concept of reachability analysis considers physical limitations such as kinematic constraints to define potentially unsafe states regarding an object.[210] Objects are then considered relevant depending on whether an unsafe state is possible or not within a specified time horizon.[211] Another application of a similar concept distinguishes potential and imminent collision objects for purposes of perception evaluation.[212] However, it is not clear how to define a suitable time horizon. Formal planners leverage explicit specification of planning behavior. One work utilizes the RSS model, a formal planner from prior work, for this

[198] Hoss, M. et al.: A Review of Testing Object-Based Environment Perception (2022), p. 233.
[199] nuScenes: nuScenes Detection Task: Leaderboard (2020).
[200] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3360.
[201] Bansal, M. et al.: ChauffeurNet (2018), p. 4, 12.
[202] Sadat, A. et al.: Perceive, Predict, and Plan (2020), p. 6-9.
[203] Philion, J.; Fidler, S.: Lift, Splat, Shoot (2020), p. 198-199.
[204] Hallerbach, S. et al.: Simulation-Based Identification of Critical Scenarios (2018), p. 97-98.
[205] Xu, Z. et al.: Zero-shot Deep Reinforcement Learning Driving (2018), p. 2867.
[206] Cho, K. et al.: Deep Predictive Autonomous Driving (2019), p. 2077-2078.
[207] Ettinger, S. M. et al.: The Waymo Open Motion Dataset (2021), p. 9712.
[208] Houston, J. et al.: One Thousand and One Hours (2020), p. 7-8.
[209] Vázquez, J. L. et al.: Deep Interactive Motion Prediction and Planning (2022), p. 8.
[210] Althoff, M.: Reachability Analysis and its Application (2010), p. 123-124.
[211] Topan, S. et al.: Interaction-Dynamics-Aware Perception Zones (2022), p. 1201-1206.
[212] Bansal, A. et al.: Risk Ranked Recall (2021), p. 2-3.

purpose.[213] Another option is to newly derive behavioral requirements for the specific driving context considered.[214,215] However, the current approaches have limitations to their applicability. The first is that commonly, stopping is considered to be valid behavior[214,216,217] which may not always be the case. Other works require information regarding the ego intention or the static road environment.[214,215] However, the planned ego trajectory may not be known when evaluating a perception function. In addition, information regarding roads obtained from maps or perception functions may either be unavailable or unreliable.

3.2.3 Downstream Task Implementation

In order to avoid manual specification of relevance, Planning Kullback-Leibler Divergence (PKL)[218] and following work[219,220] propose to leverage neural planners. These works directly use a specific implementation of the downstream planning task to evaluate the upstream detection. Therefore, the effects of perturbations in the detection results on the downstream task can directly be observed.[218a,219a,220a] Relevance is accordingly conceptualized as the magnitude of the observed effect on the downstream task implementation.[??a]

In this case, validity is only ensured for the specific implementation used for the planner.[220b] Additionally, ensuring the validity of the planner itself remains difficult.[221] Furthermore, results are not interpretable due to the black box nature of neural planners. It should be noted that formal relevance criteria and downstream implementation based criteria generally reach different conclusions. Currently, these approaches lack a unified approach with the ability to reconcile these different paradigms. Therefore, some additional concepts which relate to the concept of relevance are discussed in the following subsections.

3.2.4 Criticality Metrics

One domain which is conceptually situated between heuristics and formal planning specification is the field of criticality metrics. These metrics provide surrogate measures for driving safety when attempting validation.[222] Overviews across different criticality metrics and their respective

[213] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 3-4.

[214] Schönemann, V. et al.: Scenario-Based Functional Safety (2019), p. 345-349.

[215] Philipp, R. et al.: Systematization of Relevant Road Users (2022) p. 2-6.

[216] Topan, S. et al.: Interaction-Dynamics-Aware Perception Zones (2022), p. 1201-1206.

[217] Bansal, A. et al.: Risk Ranked Recall (2021), p. 2-3.

[218] Philion, J. et al.: Learning to Evaluate Perception (2020), a: p. 14054-14058.

[219] Henze, F. et al.: Admissible Uncertainty Bounds for Planning Algorithms (2021), a: p. 3129-3130.

[220] Philipp, R. et al.: Accuracy Requirements for Environmental Perception (2021), a: p. 131-133, b: p. 141-142.

[221] Mao, J. et al.: 3D Object Detection for Autonomous Driving (2023), p. 1914.

[222] Junietz, P. et al.: Criticality Metric for Safety Validation (2018), p. 60-65.

attributes are already available in literature[223,224] and are thus not reiterated here. However, it can be noted that various criticality metrics are available. Many of these metrics are only applicable to specific driving scenarios since they depend on driving context. Overall, there are no generally accepted metrics or thresholds to distinguish critical objects.

3.2.5 Relevance in Vision

Within the domain of computer vision in images, explicit concepts of relevance are available. Typically, relevance is either considered as saliency or as eye fixation.

Saliency attempts to distinguish important objects or regions of an image. Often, this task is specified as a segmentation task.[225a] Difficulties occur when attempting to annotate the ground truth information. Here, the subjective element of the task leads to variance between different human annotators. To mitigate these issues, it is possible to use multiple human annotations. The results can then be unified using voting schemes or images with substantial disagreement may be discarded.[226,227] When evaluating driving scenes, discarding scenes with disagreement is not applicable.

Fixation or gaze prediction is a different task where eye fixations are regarded to be a useful proxy for visual attention. Therefore, fixations provide information regarding the information content of image regions.[225b,228a] However, studies show that viewing conditions such as the difference in passive viewing and actively driving can impact the resulting fixations.[228b] Additional complications arise from the fact that gaze is also influenced by the intention of the driver.[229] Furthermore, humans are capable of interpreting their peripheral vision even without fixations. Exclusive focus on eye fixations may therefore wrongfully neglect roadside objects.[230] Fixation is also limited to a human field of view. Occluded objects, objects outside the field of view and 3D space are insufficiently considered.

[223] Mahmud, S. S. et al.: Application of proximal surrogate indicators (2017).

[224] Westhofen, L. et al.: Criticality Metrics for Automated Driving (2023).

[225] Ullah, I. et al.: A brief survey of visual saliency detection (2020), a: p. 34606, b: p. 34608.

[226] Zhang, Y. et al.: Key Issues of Salient Object Datasets (2020), p. 118-121.

[227] Li, G.; Yu, Y.: Visual Saliency Based on Multiscale Deep Features (2015), p. 5.

[228] Chapman, P.; Underwood, G.: Visual Search of Driving Situations (1998), a: p. 951, b: p. 952.

[229] Makrigiorgos, A. et al.: Human Visual Attention Prediction (2019), p. 3-6.

[230] Palazzi, A. et al.: Predicting the Driver's Focus of Attention (2018), p. 1724.

3.3 Attributes

The previous sections focused on classification and relevance which relate to the existence of objects. For this section, it is assumed that the relevant classes and objects have already been identified. The requirements for different attributes of the relevant objects are considered in the following. Only metrics and requirements which attempt to incorporate safety aspects into detection are considered for this section. This section is mostly taken from a prior publication coauthored by the author.[231] As for the relevance topic, heuristics and downstream task implementations are applicable.

3.3.1 Heuristics

One possibility when attempting to incorporate safety into perception evaluation is to rely upon manually designed heuristics. Temporal aspects relating to safety include evaluation of the perception time[232] or additionally evaluating the time between two detections belonging to the same object[233]. Proximity to collision can also be incorporated by relying on a criticality metric such as time to collision (TTC). Such approaches have been used to re-weight common perception metrics[234] or to visually compare results for different subsets of objects[235]. Another aspect is the association of GT and detected objects which requires defining an association threshold. Proposals from literature include utilizing egocentric distance[236] or longitudinal and lateral distance relative to the ego vehicle[237].

Most metrics essentially maintain the same evaluation procedure as standard AP metrics. The focus lies upon re-weighting previous metric scores. However, requirements which define clear thresholds are either absent or arbitrary.

3.3.2 Downstream Task Implementation

Similar as with the topic of object relevance, it has been proposed to directly consider a specific implementation of a downstream planner. The PKL metric is applicable to different types of detection errors, where the severity of an error is judged by the impact on a planner.[238] Since its

[231] Mori, K. T.; Peters, S.: SHARD (2023).

[232] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 5.

[233] Caesar, H. et al.: nuScenes (2020). p. 11623.

[234] Wolf, M. et al.: Safety-Aware Metric for People Detection (2021), b: p. 2760.

[235] Lyssenko, M. et al.: Towards Safety-Aware Pedestrian Detection (2022), p. 297-299.

[236] Bansal, A. et al.: Verifiable Obstacle Detection (2022), p. 64-69.

[237] Deng, B. et al.: Revisiting 3D Object Detection From an Egocentric Perspective (2021), p.4-5.

[238] Philion, J. et al.: Learning to Evaluate Perception (2020), p. 14053-14055.

introduction, it has also been included into the popular nuScenes detection benchmark.[239] Other works apply similar ideas of perturbing inputs to a planner. In this case, the sensitivity regarding these perturbations is analyzed.[240,241] A further variant is provided by deriving perturbations which still produce acceptable planning behavior. However, this approach is limited to a single scenario.[242a]

The downsides of this approach remain identical to the discussion regarding relevance. Firstly, the planner and its specific implementation must be available, which may not always be the case.[243] The objectives of the planning task are ambiguous, further increasing the difficulty of the task.[244] In addition, the validity of the approach is restricted to the specific planner which is applied.[242b] Furthermore, current approaches fail to provide generally applicable thresholds for perception.

3.4 Validation

In this section, possibilities of verifying and validating perception metrics are discussed to the degree in which they are present in literature. First, the lack of generally accepted validation methods is discussed. This is followed by a discussion of available and related validation approaches referring to either human performance or the downstream driving task.

3.4.1 Lack of Validation

Typical perception datasets do not consider the validity of the metrics applied.[245,246,247] Other datasets provide a discussion of the attributes of the metrics.[248,249] However, only the ability of the metrics to produce a ranking is validated[249], while the safety of the driving task is not.

[239] nuScenes: nuScenes Detection Task: Leaderboard (2020).

[240] Zhao, H. et al.: Suraksha: A Quantitative AV Safety Evaluation Framework (2021), p. 35-38.

[241] Henze, F. et al.: Admissible Uncertainty Bounds for Planning Algorithms (2021), p. 3129-3130.

[242] Philipp, R. et al.: Accuracy Requirements for Environmental Perception (2021), a: p. 131-133, b: p. 141-142.

[243] Wolf, M. et al.: Safety-Aware Metric for People Detection (2021), p. 2760.

[244] Guo, Y. et al.: CS-R-FCN: Cross-supervised Learning for Large-Scale Object Detection (2020), p. 2.

[245] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012).

[246] Sun, P. et al.: Scalability in Perception (2020).

[247] Wilson, B. et al.: Argoverse 2 (2021).

[248] Mao, J. et al.: One Million Scenes for Autonomous Driving: ONCE Dataset (2021), p. 21.

[249] Caesar, H. et al.: nuScenes (2020), p. 11624.

Other works rely on plausibilization of results without attempting validation. A common approach is to visualize results for exemplary scenarios,[250,251,252] sometimes explicitly designed simple scenarios for verification[253]. Further arguments consider spatial proximity and object velocity,[254a,255] thereby appealing to the intuition of the reader. Other ideas include the comparison with existing perception metrics.[254a]

Overall, it is observed that perception metrics generally lack validation. Furthermore, there are no generally accepted methods to validate perception metrics. The following sections describe related ideas which may assist future validation.

3.4.2 Human Performance

In driving context, the current baseline for driving performance is provided by human drivers. Therefore, the human driving performance has been referenced as the requirement for AD functions in the PEGASUS project which considers the safety assurance process.[256] Similar approaches are found when considering requirements for accident rates to assess the safety of the AD system.[257,258]

In the case of current datasets, the presence of human labels defining a perfect performance means that the human baseline is also implicitly present. However, the human performance for 3D recognition is currently unclear[259] since it is unknown how well different humans agree. In addition, dataset annotation occurs in an offline setting removed from the driving task or time constraints and includes lidar sensor data. Therefore, it is doubtful if this setting is suitable to quantify the human perception performance.

Human evaluation has previously also been used to compare different perception metrics. For this purpose, human subjects were asked to indicate their preference for a detection results in cases where to different metrics disagree.[254b,260] While this allows showing a preference of human subjects for either metric, it is unclear what level of agreement is required to consider one metric more valid than the other. In addition, this evaluation only allows for comparison and does not necessarily indicate that either of the metrics is valid.

[250] Topan, S. et al.: Interaction-Dynamics-Aware Perception Zones (2022), p. 1208-1209.

[251] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 6-7.

[252] Philipp, R. et al.: Systematization of Relevant Road Users (2022), p. 6-7.

[253] Ivanovic, B.; Pavone, M.: Injecting Planning-Awareness (2022), p. 824-825.

[254] Philion, J. et al.: Learning to Evaluate Perception (2020), a: p. 14055-14057, b: p. 14057-14058.

[255] Deng, B. et al.: Revisiting 3D Object Detection From an Egocentric Perspective (2021), p. 1-4.

[256] PEGASUS Project: PEGASUS Method: An Overview (2019), p. 5.

[257] Junietz, P. et al.: Macroscopic Safety Requirements (2019), p. 3-4.

[258] Liu, P. et al.: How Safe Is Safe Enough for Self-Driving Vehicles? (2019), p. 317.

[259] Qi, C. R. et al.: Offboard 3D Object Detection (2021), p. 6135.

[260] Luiten, J. et al.: HOTA: A Higher Order Metric for Evaluating Multi-object Tracking (2021), p. 573-575.

3.4.3 Downstream Planner

As shown in previous sections, it is possible to use an implementation of a downstream planner for perception evaluation purposes. Some works define permissible perception error bounds by considering deviations in planned trajectories upon error injection.[261,262,263] In these cases the downstream planner is assumed to be valid. Other works directly consider the correlation with safety goals such as likelihood of accidents in a simulation.[264,265] While this approach directly uses safety goals to assess the planner, validity of the simulation is merely assumed.

However, the possibility of errors in the planner poses a problem in validating metrics relying on a downstream planner.[266] In addition, the utilization of a single specific planner limits the generality and the practical applicability of the approach.[267]

3.5 Summary

So far, the state of the art regarding different aspects of this work has been discussed separately. The objective in this section is to provide a concise summary over all aspects.

As shown in previous sections, the state of the art for perception evaluation is mostly defined by benchmarks provided by datasets. These datasets are typically dominated by arbitrary design choices or heuristics. This includes the selection of categories, the instances included in the ground truth as well as the metrics for object detection. Most notably, these datasets show a lack of consideration with regards to legal and safety requirements of the driving task. Furthermore, metrics typically evaluate performance without providing explicit requirements including thresholds. These issues with arbitrary heuristics and lack of requirements also extend to other metrics which have been proposed to better incorporate safety aspects.

However, some previous works also use more well-founded approaches to incorporate safety aspects of the driving task. These approaches can be broadly categorized in two classes. The first class of approaches relies on an argumentation while the second class leverages an implementation of a downstream task. Approaches belonging to the former category are based on an argumentation or formal specification. However, such approaches are currently only available for the topic object relevance. While different relevance approaches are available, they typically require information which may not be reliable or available in practice. Additionally, simple assumptions such as

[261] Henze, F. et al.: Admissible Uncertainty Bounds for Planning Algorithms (2021), p. 3129-3130.

[262] Philipp, R. et al.: Accuracy Requirements for Environmental Perception (2021), p. 131-133.

[263] Zhao, H. et al.: Suraksha: A Quantitative AV Safety Evaluation Framework (2021), p. 35-38.

[264] Jha, S. et al.: Watch Out for the Safety-Threatening Actors (2022), p. 6-9.

[265] Piazzoni, A. et al.: Modeling Perception Errors (2020), p. 3497-3499.

[266] Mao, J. et al.: One Million Scenes for Autonomous Driving: ONCE Dataset (2021), p. 1914.

[267] Ivanovic, B.; Pavone, M.: Injecting Planning-Awareness (2022), p. 823.

stopping always being considered valid behavior neglect the context dependency of behavioral requirements. Generally, argumentation based approaches require additional validation regarding the safety and context of the driving task. However, no generally accepted validation methods for perception metrics are available.

The alternative class of approaches proposes the application of a downstream neural planner implementation. In this case, perception requirements are indirectly defined by considering the downstream impact. These approaches have previously been applied to object relevance as well as to define detection metrics. One advantage is that planning based approaches directly incorporate the driving task. However, these approaches suffer from a lack of interpretability and require availability of the downstream planner. Additionally, they too suffer from a lack of validity and available validation methods. Firstly, the results are specific to a single implementation of a planner and therefore lack generality. Furthermore, neural planners themselves lack validity.

Overall, the consideration of perception requirements with respect to safety in driving context is lacking. Firstly, no structured approaches to define the classification requirements are identified. For relevance and object detection metrics, some approaches exist. However, they generally fail to provide clear requirements including thresholds. Furthermore, the two complimentary approaches of providing an argumentation and leveraging a downstream planner coexist. Due to differences in methods and conclusions, no reconciliation of the two approaches is available thus far. In addition, it remains an open question how to validate such perception requirements.

4 Methodology

Having established the deficiencies of existing approaches, this chapter provides a broad overview of the methodology of this work. Firstly, the overall methodology applied in this work is presented. This is followed by a description of the common principles which are introduced as part of the overall methodology.

4.1 Overall Methodology

This section outlines the overall methodology of this work. A visual overview of the overall structure of the methodology and this document is provided in Fig. 4-1.

The upper part of the image contains the RQs which were already introduced in section 1.3. There, the overall RQ considering safety requirements for detection was decomposed into four different RQs. Each RQ pertains to one of the aspects of classification, relevance, attributes or validation. These distinct aspects continue to serve as general structure for the remainder of this work.

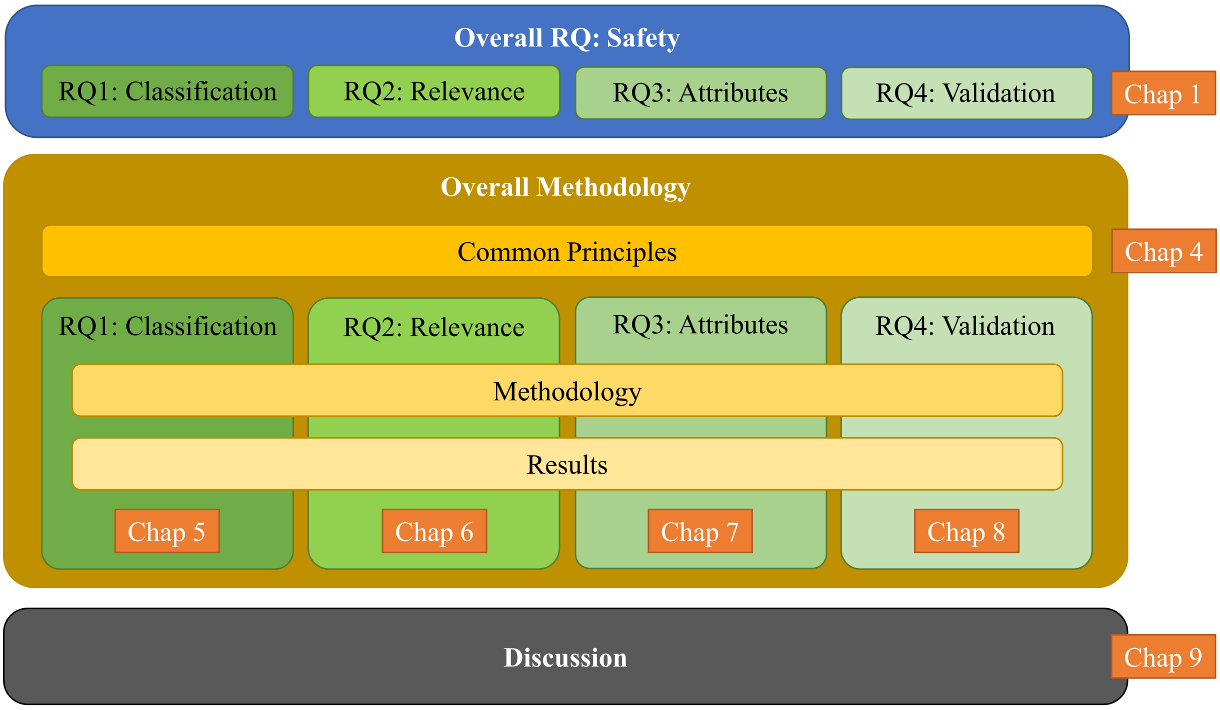

Figure 4-1: Overview of the structure of this work.

As shown in later sections, each of the aspects requires a distinct method. However, the presence of the overall RQ makes it possible to derive a set of common principles. These common principles apply to all RQs and are introduced in the following section. This is followed by four chapters which each separately treat one of the RQs 1-4. Within each of these chapters, the corresponding objectives, assumptions and the method are specified. Subsequently, results from applying the method are provided. While these results each present a proposal on how to answer the respective RQs, providing a final answer to the RQs is left for the discussion. This discussion is presented in the later chapter 9 to jointly consider all results including the validation. Finally, the overall RQ is discussed after discussing the other four RQs.

4.2 Common Principles

As introduced in the previous section, the different methods for each RQ all adhere to common principles resulting from the overall RQ. These common principles are therefore derived and presented in the following. The principles provide the basis which guides the methods developed in subsequent sections. Interpretability, legal and safety requirements as well as the human baseline are considered.

4.2.1 Interpretability

One requirement for both the methodology as well as the resulting criteria is interpretability. Interpretability is desirable since it offers insights into the SUT as well as its performance bounds.[268,269] Interpretable error types have been previously suggested[270] and applied for evaluation[271] or plausibilization[272] in literature.

Interpretability is considered in this work regarding two aspects. Firstly, the perception component is evaluated on the interface separately from the other components of the modular AD system. This defined system structure improves system understanding with respect to performance limitations.[273] In addition, intermediate representations have been shown to positively impact driving performance and transfer.[274] Evaluating the interface in a modular manner means that no specific downstream planner implementation is assumed to be available.

[268] Oksuz, K. et al.: One Metric to Measure them All (2020), p. 9446.

[269] Ivanovic, B.; Pavone, M.: Injecting Planning-Awareness (2022), p. 822-823.

[270] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017), p. 35-36.

[271] Caesar, H. et al.: nuScenes (2020), p. 11622-11623.

[272] Philion, J.; Fidler, S.: Lift, Splat, Shoot (2020), p.14056-14057.

[273] Thorn, E. et al.: A Framework for Automated Driving System Testable Cases and Scenarios (2018), p. 64-65.

[274] Zhou, X. et al.: Probabilistic two-stage detection (2021), p. 1-2.

The other aspect of interpretability concerns the structure of the requirements. In this work, interpretability is already reflected in the decomposition of the overall RQ into interpretable components. Similarly, interpretability also applies to the RQ regarding attributes. Later sections will show that it is possible to derive interpretable requirements for different attributes. This is one important reason why this work rejects neural planner based metrics. Besides their lack of validity, neural planners combine different error components in a black box manner. The alternative approach of providing interpretable requirements is likely to improve acceptance by the public as well as regulatory institutions.

4.2.2 Legal Requirements

One source of requirements is provided by applicable regulations or legal texts. Legal requirements apply to any AD system and must be satisfied to safely fulfill the driving task. Currently, the regulation for the type-approval of AD systems by the European Commission demands that the system complies with traffic rules in the country of operation.[275] The German regulation for the type-approval of vehicles with AD function also demands the compliance with existing legal requirements.[276] Similar concepts are also reflected in further proposals for regulations specifically for artificial intelligence[277] and for the safety approval of AD[278]. While these proposals are are not yet in effect, they too demand compliance with existing legal regulations.[277a,278a]

For the purpose of this work, it is therefore assumed that legal texts such as current road regulations are present and applicable to AD. It is further assumed that these regulations are formulated in natural language as it is presently the case. More specifically, this work relies upon the German road regulations StVO[279] as exemplary legal text. This regulation is used whenever applying an abstract method to a concrete example.

The benefit of legal requirements is the fact that they provide definitive requirements which any AD system must satisfy. However, these requirements are typically not sufficiently specific for the purpose of this work. Existing road regulations mainly specify behavioral requirements such as yielding aimed at human drivers. Therefore, they lack specificity with regards to uncertainties and perception. In conclusion, legal requirements in themselves are insufficient to specify perception requirements as intended by this work. Accordingly, additional interpretation and specification are required for their application.

[275] European Commission: Implementing Regulation (EU) 2022/1426 (05.07.2022), p. 11

[276] Bundesministerium der Justiz: AFGBV (24.06.2022), p. 13

[277] European Commission: Artificial Intelligence Act (2021), a: p. 3.

[278] European Commission: Annexes to Regulation (EU) 2019/2144 (2022) a: p. 6

[279] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013).

4.2.3 Safety Requirements

When further specifying and interpreting legal requirements, additional guiding principles are required. Both the regulations for type approval of AD systems on European and on German national level demand compliance with safety standards regarding functional safety and the SOTIF.[280,281] Similarly, proposals for future regulations for road vehicles[282], artificial intelligence[283] and AD systems[284] also demand conformity with safety regulations. Therefore, safety regulations and particularly the SOTIF[285] with its direct relation to perception are considered throughout the process of eliciting object detection requirements.

As with legal requirements, safety requirements also focus on observable outcomes in terms of behavior. This means that applying safety to the detection output is not straightforward. Rather, consideration of the downstream planning task is required. Therefore, safety is considered by connecting detection errors to safety outcomes of the driving task.

For instance, overestimating the distance to an object may cause an accident, while this may not be the case for distance underestimation. Thus, different types of detection errors may lead to different safety outcomes.[286] The risk of negative safety outcomes is considered in terms of severity and probability of harm in accordance with the ISO 26262[287]. Additionally, uncertainty is present regarding various aspects. Examples include future trajectories of other objects, the implementation of the downstream planner and the thresholds of the perception requirements. In any case, these uncertainties are addressed by using conservative estimates. This means that worst-case assumptions with respect to safety outcomes are used in presence of uncertainty.

The object list represents the interface between the object detection and the planning modules. However, any requirements imposed upon this interface reflects a compromise between these two modules. While higher accuracy will benefit planning, this makes the detection more difficult. Conversely, lower accuracy may simplify the detection task while simultaneously increasing the difficulty of the planning task. Since safety principles only consider behavioral outcomes, they are insufficient to derive this compromise.

[280] European Commission: Implementing Regulation (EU) 2022/1426 (05.07.2022), p. 14,36.

[281] Bundesministerium der Justiz: AFGBV (24.06.2022) p. 17.

[282] ISO/TC22/SC32/WG14: ISO PAS 8800 Road Vehicles (2021).

[283] European Commission: Artificial Intelligence Act (2021)

[284] European Commission: Annexes to Regulation (EU) 2019/2144 (2022)

[285] ISO/TC 22/SC 32: ISO/PAS 21448 SOTIF (06/2022).

[286] Bansal, A. et al.: Verifiable Obstacle Detection (2022), p. 63-64.

[287] ISO/TC 22/SC 32: ISO 26262-1 Functional Safety (12/2018), p. 15-24.

4.2.4 Human Baseline

In order to quantitatively specify requirements as compromise for the detection and the planning, an additional guiding principle is required. Existing road regulations presuppose human drivers. While perception requirements are not explicitly specified, the human baseline is implicitly considered acceptable. Therefore, the human detection performance is applied as final principle.

As shown in section 3.4.2, the human performance has previously been applied to acceptable accident rates, dataset labels and for plausibilization of metrics. Choosing human baselines for both perception and planning separately is likely not the only possible solution. However, it represents one interpretable and quantifiable option. Therefore, this work attempts to supplement legal and safety requirements with explicit quantification of human perception performance. The same performance is then used as requirement for the perception component to define thresholds specifying sufficient performance.

5 Classification

Having outlined the overall method, this chapter as well as the following chapters each presents one research question. This chapter considers RQ 1.3 pertaining to the required categories:

Is it possible to systematically identify categories which must be identified to safely perform the driving task?

The content of this chapter was previously published by the author[288] and is adapted here with minimal modification. First, the objectives regarding the classification structure and the corresponding method are discussed. This is followed by a detailed description of each of the three steps of the proposed method and its application. Finally, a conclusion of the chapter is presented. A full discussion of the results is postponed until chapter 9 to include validation results.

5.1 Objectives

This section discusses the objectives for eliciting classification requirements for detection. The method should satisfy the common principles of interpretability, legal requirements, safety and the human baseline.

In order to achieve interpretability, the categories are structured. When considering the task of object detection, the first question is which categories to consider as objects. Within this work, WordNet[289] is leveraged to define objects. If the word "object" is among the inherited hypernyms in WordNet, a category is considered to be an object. To ensure a structured approach, the concept of an ontology is applied. An ontology demands an acyclic graph[290] which is also the approach followed by ImageNet[291]. Therefore, the objective is formulated as constructing a hierarchical tree structure containing all required categories as nodes.

[288] Mori, K. T. et al.: Systematic Classification Requirements (2023) © 2023 IEEE.

[289] Miller, G. A.: WordNet: a lexical database for English (1995).

[290] Klueck, F. et al.: Using Ontologies for Test Suites Generation (2018), p. 120.

[291] Deng, J. et al.: ImageNet: A large-scale hierarchical image database (2009), p. 248-249.

5.2 Method

Having clarified the objective, the next sections successively expand a tree structure in three steps. Each step corresponds to one of legal requirements, safety requirements and the human baseline for perception as laid down in the common principles. Additional categories or nodes are added to the structure by each step. The terms “categories” and “nodes” are used interchangeably in the following.

The legal categories are added according to the behavioral requirements stated in regulations. If one object has different requirements than another, the two categories must be distinguished. An example is distinguishing traffic participants where yielding may be required from wild animals. The safety requirements are necessary to distinguish between different types of collisions. As example, a collision between a moose or a bird may entail different safety outcomes despite no regulatory distinction being present. Finally, the aforementioned categories may not distinguish perceptually dissimilar objects occupying the same category. To improve interpretability while considering human perception, additional categories are added.

5.2.1 Legal Structure

Since this section requires precise terminology to derive category names, an English translation[292] is used in addition to the German StVO[293]. It should be noted that this is an exemplary regulation while the method is applicable to any human readable regulatory text. For this exemplary application of the method, the ego vehicle is assumed to have no special rights. In addition it is assumed to weigh less than 2.8 tonnes, since additional restrictions apply for heavier vehicles.[293a]

The method for the legal structure consists of three steps. First, all objects and their respective requirements are extracted from the legal text. Next, the object categories are derived. Finally, these categories are structured hierarchically.

Extracting Requirements

Previous work in literature has already applied the general concept of deriving an ontology from traffic laws.[294] The behavior-semantic scenery description is another related method which considers behavioral requirements relating to the static scenario.[295] However, both works only consider a limited scope mostly emphasizing the static scenery. This work differs from these approaches in the objective of a hierarchical structure including all objects which require classification.

[292] Böttcher, L.: Road Traffic Regulations (2018).

[293] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), a: p. 66.

[294] Bagschik, G. et al.: Ontology based Scene Creation (2018), p. 1816-1817.

[295] Glatzki, F. et al.: Behavioral Attributes for a Behavior-Semantic Scenery Description (2021), p. 669-670.

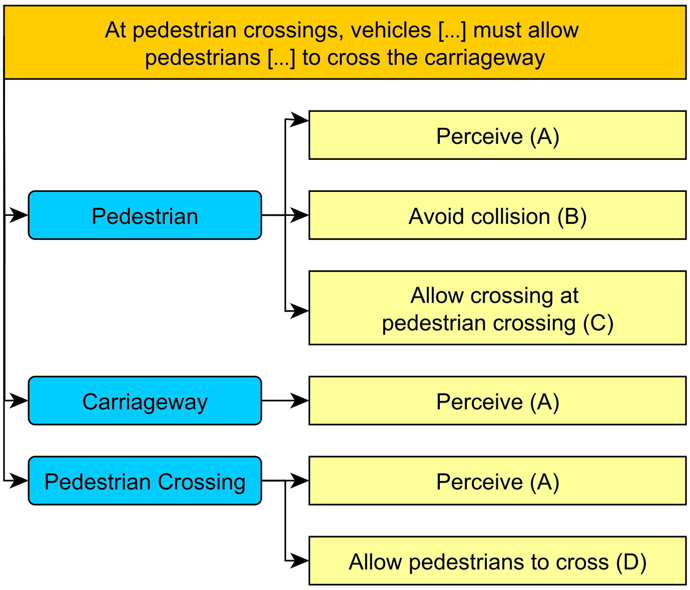

Figure 5-1: Example for process of grouping behavioral requirements by object categories. Uppercase letters are introduced as shorthand for visualization in Fig. 5-2.[298]

As first step, all behavioral requirements which potentially apply to the ego vehicle are extracted from the text. Some of these requirements depend on preconditions relating to the situational context. Since these preconditions include other objects and their categories, they are also recorded. This also includes requirements pertaining to the internal behavior of the ego vehicle, which considers the internal states of the system[296]. A notable example is if attention is required in a specific situation. Some requirements are not explicitly stated, but implicitly included in the text. Examples here include collision avoidance or perceiving specific objects. Other requirements only apply to specific relations between objects or states or attributes which change over time. However, these are neglected here since only object categories are considered. Any category mentioned in combination with a relation or an attribute simply receives the implicit requirement of perceiving it. An example for spatial relations are queues of vehicles or how lanes are arranged. Frequently referenced attributes include traffic light signals or flashing lights.

The process is visualized for an example in Fig. 5-1. The original requirement is that "At pedestrian crossings, vehicles [...] must allow pedestrians [...] to cross the carriageway"[297]. This requirement is decomposed into its components relating to different object categories. The carriageway provides an example of an implicit requirement to perceive an object. Adherence to the behavioral requirement is only possible if the carriageway is perceived. However, no further behavioral requirements apply to the carriageway itself. As a result, a list of object categories is obtained. Each object category is associated with a potentially non-unique set of behavioral requirements.

[296] Nolte, M. et al.: Towards a skill- and ability-based development process (2017), p. 3.

[297] Böttcher, L.: Road Traffic Regulations (2018), Sec.26 (1).

[298] Mori, K. T. et al.: Systematic Classification Requirements (2023) Fig. 1 © 2023 IEEE.

Extracting Categories

The list of categories identified thus far is not limited to legal aspects. Firstly, it contains categories such as vehicles and buses which overlap. Secondly, some categories may even have identical behavioral requirements. This section is however only concerned with objects which require distinction from a legal perspective.

To resolve these ambiguities, two steps are applied. Firstly, categories which are a subcategory of other categories are identified using the hyponyms in WordNet[299]. Any behavioral requirement which applies to the higher level category is also appended to the respective subcategories. Secondly, any categories possessing the same legal requirements are merged. This yields categories which are represented by a unique set of behavioral requirements. In many cases, a set of requirements only applies to a single or few object categories. To simplify notation in these cases, the name of the respective object categories is used to identify a node instead of the lengthy requirements.

Structuring Categories

Once the categories have been obtained, the objective is to structure them hierarchically as visualized for an example in Fig. 5-2. Each of the nodes visualizes a set of requirements, each of which is abbreviated by a single uppercase letter. The root node is given by an empty set {}, which is a subset of all other sets of requirements. Practically speaking, this indicates any objects for which no behavioral requirements exist.

Starting from this node, new nodes are added iteratively based on the requirements. The process begins with the sets containing the lowest number of requirements. Accordingly, the second node added to Fig. 5-2 is the orange node with one requirement {A}. This is followed by the two yellow nodes {A,B} and {B,C} which each have two requirements. For each new set, all subsets already present in the tree are identified. These subsets represent potential parent node candidates. Since {B,C} only has one subset {} as potential candidate, it is directly added to this node. The node {A,B} has the two potential parent node candidates {} and {A}. Among these candidates, those candidates whose requirements are subsets of other candidates are removed. In the example, this means that {} is removed as parent node candidate since it is a subset of {A}. This corresponds to choosing the lowest level in the hierarchy or the most specific subclass. The procedure described thus far is sufficient to add all colored nodes in Fig. 5-2 to the structure.

However, it is also possible that multiple candidates are still present at this stage. An example is provided by the white node {A,B,C}. In this case, both {A,B} and {B,C} remain as parent node candidates. To resolve the ambiguity, the severity of infractions is considered. For this purpose, the intersections between the new node and each parent node candidate are obtained. For the example, the intersections are {A,B} and {B,C}, respectively. Next, the penalty as listed in the

[299] Miller, G. A.: WordNet: a lexical database for English (1995).

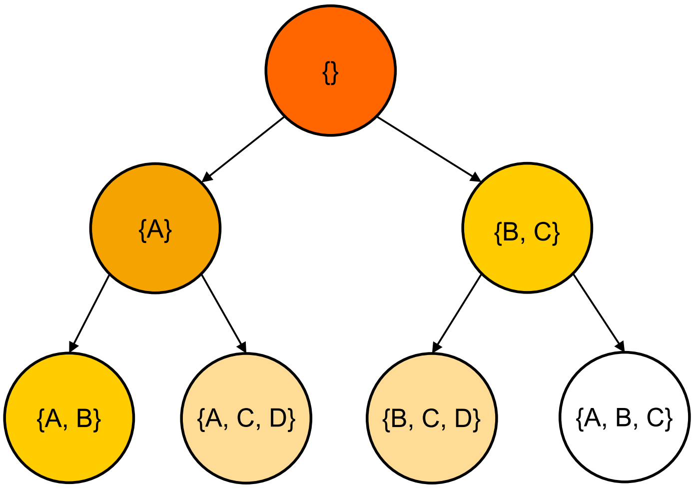

Figure 5-2: Example for iteratively adding nodes to structure. Stronger saturation and lower brightness of the colors indicate nodes added earlier.[301]

penalty catalogue regulation (BKatV)[300] for failing to meet a requirement is identified. For each set of behavioral requirements, the most severe penalty is considered. Finally, the new node is added to the parent node with the most severe penalty. This approach considers the importance of requirements based on the severity of infractions. Assuming that infringing upon {C} is penalized more severely than infringing upon {A}, this means that the node {A,B,C} is added to {B,C}.

5.2.2 Safety Structure

This section considers the aspect of safety in addition to the previously defined legal categories. Here, collision risk is considered with respect to collision severity [302] and likelihood [303,304] as defined in the ISO 26262[305]. Both aspects are used to further subdivide the categories already present from the legal structure if applicable.

Collision Severity

Road regulations only demand that collisions be avoided without providing a distinction of different collision types.[306] However, the severity can be judged similarly to the previous section by considering infractions. The penalties for noncompliance with laws are stated in the penalty

---

[300] Bundesministerium für Justiz und Verbraucherschutz: Bußgeldkatalog-Verordnung (14.03.2013).

[301] Modified from Mori, K. T. et al.: Systematic Classification Requirements (2023), Fig. 2 © 2023 IEEE.

[302] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 4.

[303] Nilsson, G.: Traffic Safety Dimensions (2004), p. 61-67.

[304] Lefèvre, S. et al.: A survey on motion prediction and risk assessment (2014), p. 11.

[305] ISO/TC 22/SC 32: ISO 26262-1 Functional Safety (12/2018), p. 15-24.

[306] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), p. 1.

catalogue regulation[307] in accordance with other legal texts[308,309]. Surveying the documents shows an increasing penalty from simple infractions over hindering other traffic participants to causing damage to property. More severe penalties occur if negligence leads to bodily harm. German law considers such a case to be a criminal offence.[310] Similar ideas of distinguishing property damage from various degrees of injury are found in road safety reports from Germany[311] as well as the United Kingdom[312]. This justifies distinguishing different outcomes of accidents in this work.

Accidents with injuries occur less frequently for cars than for other road users.[312] One reason is the fact that the occupants are protected by larger crush zones.[313] Less protected traffic participants including pedestrians, cyclists and motorcylists[312] are often summarized as vulnerable road users (VRUs)[314]. When considering AD, injury may also occur for occupants of the vehicle. Therefore, a distinction of accident severity is also required for other obstacles besides traffic participants. Different types of static objects show different severities in the case of run-off-roadway accidents.[315] Similar considerations apply for the severity of accidents with different animal species.[316] However, predicting the exact severity of an accident is difficult under real-world settings including perception uncertainties. The difficulties in modeling risk is also acknowledged in prior work which incorporates ethical considerations into trajectory planning.[317]

Therefore, this work only provides a coarse distinction between static obstacles. Here, harmful objects are distinguished without a strict formal approach based on their rigidity and size. In addition, VRU are distinguished from other road users as in existing literature.

Collision Likelihood

One factor known to influence the likelihood of an accident[318] as well as the severity[319,320] is the velocity. While the possible range of velocity is influenced by the class[321], the velocity is a distinct and dynamic property. Therefore, the velocity is not considered for the categorization. Furthermore, the accident likelihood is influenced by various types of unexpected behaviors.

[307] Bundesministerium für Justiz und Verbraucherschutz: Bußgeldkatalog-Verordnung (14.03.2013).

[308] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013).

[309] Bundesministerium für Justiz und Verbraucherschutz: Straßenverkehrsgesetz: StVG (05.03.2003).

[310] Bundesministerium für Justiz und Verbraucherschutz: Strafgesetzbuch: StGB (13.11.1998), p. 118.

[311] Statistisches Bundesamt: Verkehr: Verkehrsunfälle April (2022), p. 3-4.

[312] Department of Transport: Reported road causalties Great Britain (2021).

[313] Richards, D. C.: Relationship between Speed and Risk of Fatal Injury (2010), p. 26.

[314] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 4.

[315] Lee, J.; Mannering, F.: Analysis of roadside accident frequency (1999), p. 60.

[316] Huijser, M. P. et al.: Wildlife-Vehicle Collision Reduction Study: Report to Congress (2008), p. 4-5.

[317] Geisslinger, M. et al.: An ethical trajectory planning algorithm (2023), p. 140-141.

[318] Nilsson, G.: Traffic Safety Dimensions (2004), p. 61-67.

[319] Han, Y. et al.: Effects of vehicle impact velocity (2012), p. 508, 516.

[320] Richards, D. C.: Relationship between Speed and Risk of Fatal Injury (2010), p. 24-25.

[321] Dietmayer, K.: Predicting of Machine Perception for Automated Driving (2016), p. 409-413.

These may relate to simply unexpected behaviors as well as infractions.[322] In addition, some objects are capable of sudden changes in direction or speed of their movement. These classes pose inherent difficulties for behavior prediction.[323]

This work considers the potentially erratic movements of animals both as road users and as obstacles. Especially for malicious behavior of other objects, an accident may be unavoidable.[324] Therefore, collision likelihood is considered secondary to collision severity. Accordingly, severity is placed higher in the hierarchy.

5.2.3 Human Perception Structure

As defined in the common principles, this section attempts to improve interpretability and consider the human baseline. Therefore, the previously defined categories are supplemented with categories considering human perception.

The first obvious choice for additional categories are those categories explicitly mentioned within the legal text[325]. This refers to the categories which were initially omitted when eliciting the legal requirements. Nevertheless, these perceptual categories are now reinserted since they are more interpretable than sets of requirements. While the behavioral requirements are explicitly stated, the safety categorization is coarsely estimated. It should be noted that including perceptual categories may lead to the presence of overlapping categories such as taxis and cars.

Another addition following the categories found in legal texts are dataset classes. This work includes datasets for the two tasks of object detection[326,327,328,329,330] and semantic segmentation[328,329,331,332] for the context of driving. The categories are generally selected by the benchmarks based on aspects related to perception. Examples include the frequency of occurrence[333], diversity, coverage[329a] and visual similarity[334]. To associate the categories with the correct nodes in the hierarchy, the synonyms and hypernyms from WordNet[335] are leveraged once more.

[322] Lefèvre, S. et al.: A survey on motion prediction and risk assessment (2014), p. 11.

[323] Ahmed, S. et al.: Pedestrian and Cyclist Detection and Intent Estimation (2019), p. 21.

[324] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017), p. 6.

[325] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013).

[326] Sun, P. et al.: Scalability in Perception (2020).

[327] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012).

[328] Geyer, J. et al.: A2D2: Audi Autonomous Driving Dataset (2020).

[329] Yu, F. et al.: BDD100K: A Diverse Driving Dataset (2020), a: p. 2636.

[330] Caesar, H. et al.: nuScenes (2020).

[331] Huang, X. et al.: The ApolloScape Open Dataset (2020).

[332] Neuhold, G. et al.: The Mapillary Vistas Dataset (2017).

[333] Cordts, M. et al.: The Cityscapes Dataset (2016), p. 3214-3215.

[334] Ertler, C. et al.: The Mapillary Traffic Sign Dataset (2020), p. 4-5.

[335] Miller, G. A.: WordNet: a lexical database for English (1995).

5.3 Results

This section shows the results obtained from the application of the method presented in the previous section. The resulting hierarchical classification structure is shown in Fig. 5-3 and Fig. 5-4. The structure is split at the “avoid collision” node for better visibility.

5.3.1 Legal Structure

The yellow and orange nodes in Fig. 5-3 and Fig. 5-4 represent nodes obtained from considering legal requirements. Overall, 39 categories are obtained in this manner. These categories are distributed across seven levels in the hierarchy if the category “no regulation” is included. Note that neglecting the green category “6 VRU” in Fig. 5-4 means that the yellow categories from level 8 move up to level 7 for purpose of counting the levels of the hierarchy.

By design, each node represents a unique set of legal requirements which apply to the object category. For instance, the yellow node “4 Stopping prohibited” indicates areas where stopping is prohibited which also implies the requirement that these areas must be perceived. Category names indicating specific objects such as “bus” are also used in the hierarchy. However, these are merely introduced for notational convenience and still refer to the unique set of requirements which apply to the respective category. Collision relevant categories and non collision relevant objects each comprise approximately half of the categories. Examples for non collision relevant categories include locations while collision relevant objects include both static and dynamic objects. Orange nodes for the categories “traffic signs/markings” and “traffic installations” indicate incomplete sections. However, it should be noted that these subcategories are well specified and documented in the road regulations.[336] Therefore, adding these categories to the proposed structure is considered straightforward. Furthermore, the recognition of traffic signs is generally considered as a separate task which is solved with dedicated systems.[337,338,339] In addition, the distinction of different traffic signs is not directly related to the task of collision avoidance which is emphasized in this work. Therefore, the subcategories of the incomplete sections are not further considered in this work for the sake of brevity.

[336] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), p. 38-81.

[337] Ertler, C. et al.: The Mapillary Traffic Sign Dataset (2020), p. 8-9.

[338] Chen, J. et al.: A real-time and high-precision method for small traffic-signs recognition (2022), p. 2233-2234.

[339] Karthika, R.; Parameswaran, L.: Traffic Sign Recognition from Road Scenes (2022), p. 351-354.

5.3.2 Safety Structure

Including safety considerations adds six nodes to the classification structure, increasing the number of categories to 45. The categories from the safety structure are depicted as green nodes in Fig. 5-4. The limited number of new categories owes to the fact that many differences are already considered by the legal structure.

Nevertheless, considering safety provides a contribution to the overall structure. Firstly, obstacles are coarsely distinguished based on the possibility of an injury in the event of a collision. Additionally, erratic movements of animals are considered. Animals may occur as road users or vehicles in the case of riders or animal drawn carriages. In addition, wild animals are treated as obstacles and are distinguished accordingly. Due to inclusion of the "6 VRU" category, the depth of the hierarchy increases to eight levels.

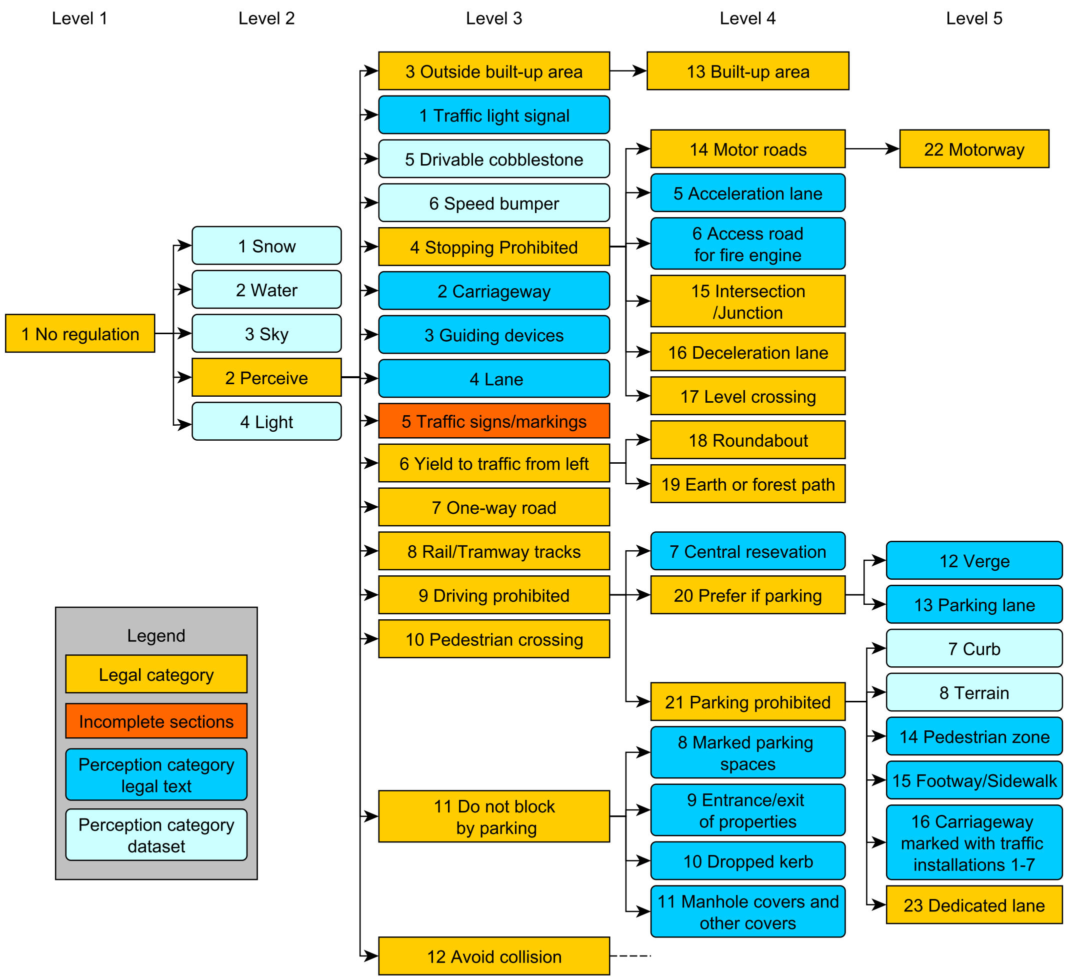

Figure 5-3: Classification hierarchy for non collision relevant categories which is continued for collision relevant objects in Fig. 5-4. [340]

[340] Modified from: Mori, K. T. et al.: Systematic Classification Requirements (2023), Fig. 3 © 2023 IEEE.

5.3.3 Human Perception Structure

Blue nodes in Fig. 5-3 and Fig. 5-4 indicate categories originating from considering interpretability and the human baseline for perception. Dark blue categories such as “17 Tunnel” indicate categories which are adopted from the legal text. Nodes depicted in light blue such as “13 Vegetation” are categories which are taken from datasets.

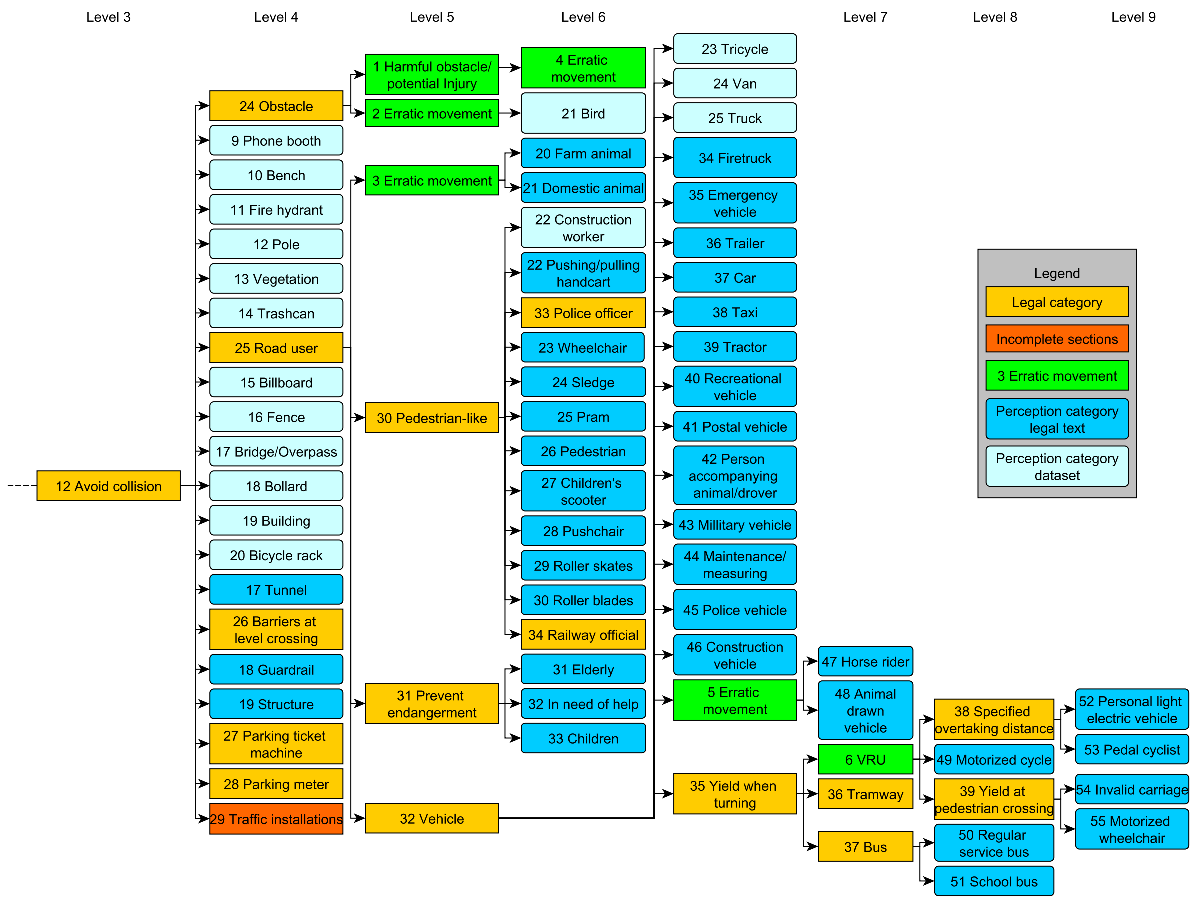

Figure 5-4: Classification hierarchy for collision relevant objects. Continuation of the classification hierarchy for collision relevant categories from Fig. 5-3. Note that the vehicle categories were moved to the right for visualization purposes. Therefore, their visualized horizontal position does not directly correspond to their depth in the hierarchy.[341]

The legal text contributes 55 additional categories. Among these categories, 16 are not collision relevant while 39 are collision relevant. This indicates that road regulations emphasize collision relevant objects, in particular road users. Datasets contribute another 25 categories which are not mentioned in road regulations, resulting in 80 categories based on human perception. Adding these categories to the structure increases the overall number of categories to 125. The fact that datasets provide additional complementary categories to road regulations underlines the utility of including these dataset categories. However, it also shows that existing perception datasets are concerned with many classes which are not included in road regulations. This particularly applies

[341] Modified from: Mori, K. T. et al.: Systematic Classification Requirements (2023), Fig. 4 © 2023 IEEE.

to the static collision relevant objects, where datasets add the most categories. Of the additional dataset classes, five relate to segmentation and 20 relate to detection. This indicates that while categories which are part of the object detection are more frequent, some classes do require the task of segmentation. Eight of the dataset categories belong to the non collision relevant objects, while 17 more categories are collision relevant. This shows that datasets and road regulations put a similar emphasis on collision relevant objects.

5.4 Conclusion

In this chapter, a method to answer the RQ1 regarding classification was proposed. It was demonstrated that it is indeed possible to systematically identify categories and structure them in a hierarchical manner. The common principle of interpretability is reflected within the method itself by providing a structured approach. The remaining common principles of legal and safety requirements as well as the human baseline were explicitly considered in the three steps of the method.

By applying road regulations and the behavioral requirements contained therein, 39 categories with behavioral requirements were obtained. The application of safety requirements yielded additional six categories. Finally, the application of the human perception baseline yielded another 55 categories which also improved the interpretability. In each of the steps, collision relevant objects comprised more than half of the categories. The final 125 categories were structured hierarchically across nine levels. It is observed that the legal and safety categories at higher levels in the hierarchy show little agreement with dataset categories. However, the latter were grouped in accordance with the higher level categories. A full discussion of the classification results along with the corresponding RQ1 is provided in section 9.3.1. This allows considering the validation results presented in chapter 8 in the discussion.

6 Relevance

This chapter considers RQ 1.3 pertaining to the relevance of objects:

Is it possible to systematically identify objects relevant for detection based only on information contained in the object list while considering the safety in driving context?

Two previous publications coauthored by the author[342,343] already included the contents of this chapter which are therefore presented with little modification. The content also appears in the work by Storms[344] created concurrently with this work. Firstly, an abstract method to derive relevance is defined. This method is then applied to urban traffic and the corresponding results are presented. Finally, a conclusion of the relevance chapter is provided. The discussion of these results is postponed until section 9.3.2 to allow a joint discussion including the validation results presented in chapter 8.

6.1 Abstract Method

This section outlines the abstract method applied to the definition of relevance. An overview is provided in Fig. 6-1.

Relevance is first conceptually defined in order to proceed. Literature distinguishes different related terms among which situational relevance or utility is considered in this work. Utility is

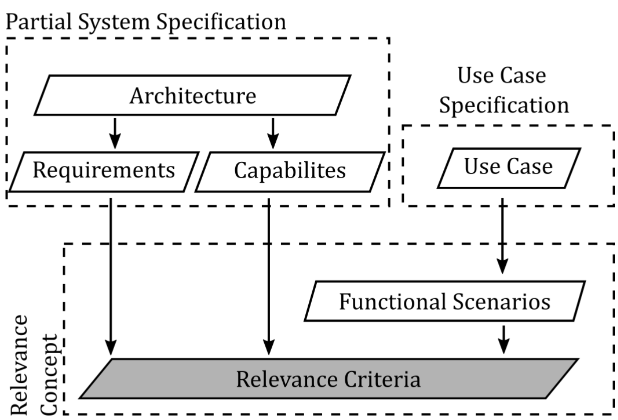

Figure 6-1: Overview of the proposed relevance method with its key outputs.[342a]

[342] Mori, K. et al.: Conservative Estimation of Perception Relevance (2023), a: Fig. 1 © 2023 IEEE.

[343] Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023).

[344] Storms, K.: Context Aware Data Reduction (2023).

only clearly defined if a specific task is provided.[345] For simplicity, this work continues to refer to this concept with the term relevance. Based on this conceptual understanding, the definition of the task is required. This includes a partial specification of the system and the use case to which it is applied. To ensure interpretability and simplicity, the use case is decomposed into functional scenarios. Considering the uncertainties yields the outer bounds of relevance for each functional scenario.

6.1.1 Partial System Specification

Since relevance depends on the task and system, a partial system specification is required for the specification of relevance. Note that the objective is to avoid fully specifying the downstream planner. Accordingly, the system specification is limited to high-level aspects regarding system requirements and capabilities.

System Requirements

While the downstream planner is not fully specified, it is assumed that any valid planner fulfills certain requirements. Similar to the detection task, the common principles of legal and safety requirements also apply to the planner.

Firstly, the planner must provide actions which conform to legal requirements:

REQ1: *The actions must adhere to applicable legal restrictions.*

In addition, the action can only be performed if it can be executed by the downstream act module. Physical limitations provide a definitive bound for the possible trajectories.

REQ2: *The actions must adhere to physical limitations.*

These requirements so far only apply to the planning task and not for detection. Establishing the connection between these behavioral requirements and the relevance for detection is discussed in section 6.1.3.

System Capabilities

Having defined the high-level system requirements, a definition of the system capabilities is required. The underlying assumption is that any valid planner provides information regarding these high-level capabilities. The planner therefore provides guarantees upon its abilities, which can be relied upon during perception testing. In this work, the system latency and acceleration limits are considered.

The effects of latency on the system is depicted in Fig. 6-2. It is assumed that an event occurs at an initial point in time denoted by t_0. This event is first perceived by the system, followed by

[345] Cosijn, E.; Ingwersen, P.: Dimensions of relevance (2000), p. 537-540.

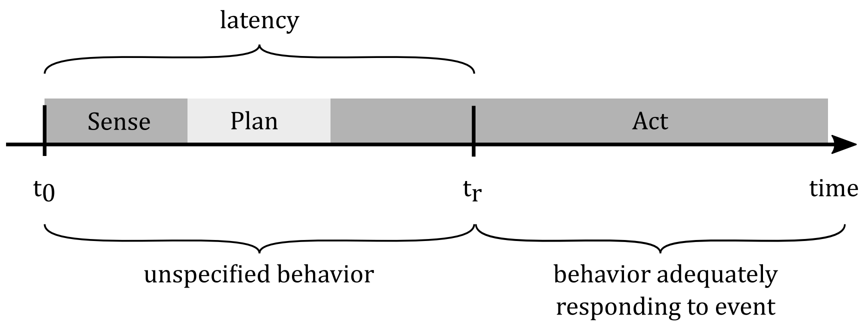

Figure 6-2: Initially, the ego behavior is unspecified with respect to a sudden event. An adequate response to the sudden event is only guaranteed after the latency.[346]

planning a new trajectory. Until the actuators actually execute the new trajectory as event response, a certain amount of time passes. This time is the system latency or the reaction time which is denoted as t_r. For any system, this latency is larger than zero. Therefore, it is unreasonable to demand that a response to an event already occurs during the reaction time. Instead, the system is expected to provide a guarantee on the latency after which a correct response to an event is provided. This latency must be considered by the trajectory planning in order to ensure that sufficient time is available for the system to react even if sudden events occur. An example is maintaining sufficient distance to a leading vehicle in case the leading vehicle suddenly brakes. Accordingly, this latency is a parameter which is required for the partial specification of the system capabilities.

The available actions are mainly defined by the available accelerations of the ego vehicle. Limitations on this acceleration may result from outer circumstances such as ice on the road. Other limitations result from factors owing to the system specification such as limited power of the engine. In order to maintain interpretability and simplicity, two accelerations are distinguished in this work:

- minimum guaranteed braking deceleration
- minimum guaranteed acceleration

The guaranteed acceleration is available for longitudinal acceleration as well as for lateral acceleration when steering. All accelerations are limited by the available friction on the road surface. Additional limitations on the longitudinal acceleration typically result from the limited power of the vehicle.[347] For lateral acceleration, human preference typically provides the limiting factor.[348] All parameters defined in this section are later applied within the relevance method application.

[346] Modified from Mori, K. et al.: Conservative Estimation of Perception Relevance (2023), Fig. 2 © 2023 IEEE.

[347] Bokare, P. S.; Maurya, A. K.: Acceleration-Deceleration Behaviour (2017), p. 4742.

[348] Bertolazzi, E. et al.: Supporting Drivers in Keeping Safe Speed and Safe Distance (2010), p. 530.

6.1.2 Use Case specification

For the proposed relevance method, a use case is required. This use case implicitly defines the operating environment, situations and requirements. Typical perception datasets focus on the urban domain. Therefore, this work also applies the relevance method to the urban domain. Note that the scope of this chapter is limited to object detection and collision avoidance. It does not include other traffic rules which may be present in an urban environment.

6.1.3 Relevance Concept

Once both the system and the use case are specified, a concept to specify relevance is defined. This section first provides an overview over the challenges such a concept faces. This is followed by the proposed solution including an overview and more detailed explanations of its two steps.

Defining relevance for detection is faced with two difficulties. While high-level behavioral requirements are specified, they currently lack specificity and awareness of situational context. Furthermore, they relate to behavior and are thus not directly applicable to detection.

In addition, uncertainties are present in any given situation. Firstly, the future trajectories of traffic participants including all other objects are inherently uncertain. The ego trajectory is also uncertain from the perspective of the detection. The reason is that the ego trajectory is only planned after obtaining the information from the detection. Additionally, uncertainties also persist with respect to the static environment. Since only the object list is considered as interface, no reliable information on road and lane geometry is available. This also includes information on applicable traffic rules such as right of way. Furthermore, low-level information such as local friction coefficients is also unavailable.

Overview

While the current behavioral requirements are still high-level, further specification is possible. Behavioral requirements are generally dependent on the situational context of traffic. Therefore, a use case decomposition is applied to yield interpretable functional scenarios. Each functional scenario considers a specific driving context, for which specific behavioral requirements are obtained. However, it is still necessary to account for uncertainties to derive relevance for each functional scenario.

Considering uncertainties allows transferring behavioral requirements to object detection requirements. The reasoning is that an object requires detection if it incurs a behavioral requirement for the ego vehicle. In this case, an object is considered relevant for detection. However, uncertainties are present regarding the future object trajectory, the future ego trajectory and the static environment. This means that based on these uncertainties, a range of scenarios is possible. An object is considered relevant if a behavioral requirement for the ego may occur within this range of possible scenarios. Therefore, uncertainties are accounted for by relying on worst-case assumptions. For instance, traffic rules such as right of way are not specified in the object list. Thus, the conservative worst-case assumption is that the object of interest (OOI) has the right of way. Accordingly, the presence of a behavioral requirement under worst case assumptions indicates that an object is relevant for object detection.

Use Case Decomposition

The objective of the use case decomposition is to identify interpretable functional scenarios. This allows for the subsequent specification of relevance for each scenario.

The object list provided by the detection contains multiple objects. However, this work follows Topan et al.[349] in focusing on pairwise interactions between a dynamic OOI and the ego vehicle. One notable exception is the case where the ego vehicle enters the opposite lane to pass an object on its own lane. This case is discussed separately in a later section. Generally, considering pairwise interactions reduces the complexity. While additional objects may further restrict the behavior of the OOI, ignoring these restrictions overestimates the possible behaviors of the OOI. Therefore, the estimates are conservative and thus aligned with the common principle of safety. The object pair consisting of the ego vehicle and the OOI is represented by a set of parameters including attributes such as location or velocity.

Similar attempts to decompose a use case into functional scenarios are found in prior work.[350,351,352,353] This work identifies potentially applicable scenarios based on a formalized approach. Scenarios are distinguished with equations using the parameter sets of the respective object pair at the current point in time. As previously mentioned, uncertainties regarding the static environment as well as future behaviors of the OOI and the ego vehicle are present. Therefore, different scenarios such as stopping or merging may follow the same initial configuration. Thus, any potentially applicable scenario is evaluated. Since potential scenarios are not mutually exclusive, multiple hypothetical scenarios may be considered.

The evaluation of relevance for each of the potential scenarios is described in the following section. After evaluating all potential scenarios, the results are aggregated to yield the final relevance result. Within this work, a superposition approach similar to Schönemann et al.[353a] is applied. If any of the hypothetical scenarios consider an object to be relevant, the object is considered relevant.

[349] Topan, S. et al.: Interaction-Dynamics-Aware Perception Zones (2022), p. 1207.

[350] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017).

[351] PEGASUS Project: PEGASUS Method: An Overview (2019).

[352] Philipp, R. et al.: Systematization of Relevant Road Users (2022).

[353] Schönemann, V. et al.: Scenario-Based Functional Safety (2019), a: p. 347.

Relevance for Functional Scenarios

In this section, the abstract method to define relevance for a functional scenario is outlined. The approach is based on a formal specification of behavioral requirements using simplifications and worst case assumptions.

The behavioral assumptions and requirements are elicited in accordance with the common principles of legal and safety requirements. Firstly, German road regulations[354] are considered to yield behavioral requirements for the ego vehicle. These behavioral requirements are in many cases too abstract to directly yield parametrized equations. In such cases, further specification and interpretation is provided by worst case assumptions. For the ego vehicle, a valid response to an event is only demanded after the system latency. During the latency of the ego vehicle, worst-case behavior is assumed. For the OOI, adherence to behavioral requirements is not assumed. The reason is that many objects such as children and animals may not comply with traffic rules. In addition, the object list does not include information regarding right of way. This means that the conservative assumption is to assume that the OOI has the right of way. Therefore, the OOI is assumed to follow worst case behavior throughout the scenario.

To specify the behavior of the OOI and the ego vehicle, trajectories are represented as parametrized equations. For simplicity, several assumptions are adopted from the worst time to collision metric[355]. Objects including the ego vehicle are generally treated as point masses. However, their size is accounted for by considering the points to have a radius. The action space of all objects including the ego vehicle for the worst case assumptions is considered using Kamm's circle. This provides a comprehensive model which overestimates the actions available for the worst case. Kamm's circle considers the isotropic maximum acceleration $a_{\max}$. The available acceleration is assumed to be independent of the direction.[356] While additional consideration of kinematic constraints is possible, this conservative overestimation is preferred due to its simplicity.

With these assumptions, both the behavior of the OOI and the ego vehicle are specified. This includes worst case assumptions for the OOI and the ego during its latency. Furthermore, the valid reaction of the ego vehicle to an event is also formalized. Given this specification, relevance for a given functional scenario is conceptualized as follows:

> *All objects that potentially change the set of viable trajectories are relevant for solving the combined planning/detection task.*

With this, the concept of relevance and an abstract method to derive it are sufficiently specified. The next section shows the application of the method for the specified system and use case.

---

[354] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013).

[355] Wachenfeld, W. et al.: The worst-time-to-collision metric (2016), p. 730-731.

[356] Schmidt, C.: Fahrstrategien zur Unfallvermeidung (2013), p. 13-15.

6.2 Method Application

This section provides the application of the abstract method outlined in the previous section. The objective is to demonstrate the practical approach for the context of urban driving.

6.2.1 Use Case Decomposition

As previously described, the use case is first decomposed into different functional scenarios. Functional scenarios are distinguished by equations which provide a formal distinction based on object parameters. It should be noted that while the examples mostly refer to vehicles as OOI, the results apply generally. For instance, the same scenario may describe the interaction with a vehicle at an intersection or the interaction with a pedestrian at a pedestrian crossing.

The first question when formalizing the distinction is which coordinate system to apply. Previous popular works such as RSS[357] rely on lane-based coordinates for the scenario descriptions. Contrary to previous works[357,358,359], lane information is assumed to be unavailable in this work which relies solely on object lists. In absence of lane information, the remaining options are vehicle coordinates and polar coordinates. Using vehicle coordinates with longitudinal and lateral direction is not suitable for the decomposition. Vehicle coordinates imply a linear motion assumption or a straight road, which may not be the case. Therefore, polar coordinates are favored for the decomposition. In addition, polar coordinates directly relate to collision safety. For a collision to occur, the radial distance must approach zero.

A visualization is provided in Fig. 6-3 for the case that vehicle velocity and lane direction do not coincide. In the visualization, (A) depicts a pair of objects as included in the object list with velocities v and a distance d. The blue vehicle represents the ego vehicle while the red vehicle represents the OOI. The second part of the image (B) demonstrates how the scenario is transformed into polar coordinates. In polar coordinates, the radial direction denoted by the index r is provided by the line connecting the two vehicles. The tangential direction denoted by the index t is perpendicular to the connecting line. All distances and velocities are considered relative to this radial and tangential direction. The angle of the ego vehicle heading with respect to the connecting line is denoted with α. Practically speaking, this coordinate system represents a hypothetical lane connecting both vehicles. It should be noted that the velocities do not coincide with the direction of the hypothetical lane representing the radial direction. Simultaneously, the object geometry is simplified by introducing the sizes s as radius for each objects. Finally, this yields the simplified model depicted in (C). Using this model allows formalizing behaviors for each scenario as described in section 6.2.2.

[357] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017) , p. 12.

[358] Philipp, R. et al.: Systematization of Relevant Road Users (2022).

[359] Schönemann, V. et al.: Scenario-Based Functional Safety (2019).

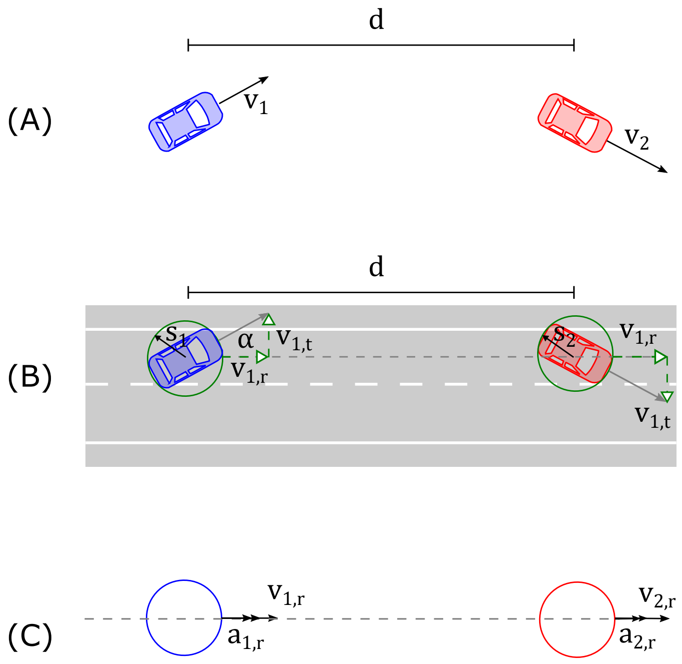

Figure 6-3: Development of a simplified environment model for a radial scenario. Shown for scenario R.TA. (A) Object information (B) simplifications through hypothetical road model (C) Simplified model.[360]

Before describing each scenario, the different functional scenarios are first distinguished. In polar coordinates, the scenario distinction is based on distance, radial and tangential velocity. First, different scenarios considering the interaction in radial direction are discussed. Next, the criteria to distinguish different tangential scenarios are derived.

Radial Scenarios

This section discusses the distinction of different scenarios with respect to their radial interaction. Relevant notation, the criteria to distinguish the scenarios and some simplifications used in later section are introduced.

The formal description to distinguish functional scenarios applies the following notation which is maintained throughout this work. Location vectors of objects are denoted as $\vec{r}_\mathrm{i}$. The index i refers to the respective object where 1 is the ego vehicle and 2 is the OOI. If multiple indices are present, the object is denoted by the first index. The last index indicates a state, where an index 0 denotes the initial state. Analogously, the velocities are denoted as $\vec{v}_\mathrm{i}$. The two vehicles are connected by the distance vector $\vec{d}$ which is:

$$\vec{d} = \vec{r}_2 - \vec{r}_1 \tag{6-1}$$

The distinction between the scenarios is drawn according to the relative radial motion of the

[360] Modified from: Mori, K. et al.: Conservative Estimation of Perception Relevance (2023), Fig. 3 © 2023 IEEE.

objects. Every scenario is indicated by a three letter abbreviation. The first letter distinguishes tangential and radial scenarios. The first letter after the period indicates if the ego vehicle is moving towards or away from the OOI. The second letter provides the same information for the OOI with respect to the ego. Accordingly, the four radial scenarios are:

- ego moving towards OOI, OOI moving away from ego (R.TA):

$$\vec{d}_0 \cdot \vec{v}_{1,0} \geq 0 \cap \vec{d}_0 \cdot \vec{v}_{2,0} \geq 0 \tag{6-2}$$

- ego moving away from OOI, OOI moving towards ego (R.AT):

$$\vec{d}_0 \cdot \vec{v}_{1,0} \leq 0 \cap \vec{d}_0 \cdot \vec{v}_{2,0} \leq 0 \tag{6-3}$$

- both vehicles moving towards each other (R.TT):

$$\vec{d}_0 \cdot \vec{v}_{1,0} > 0 \cap \vec{d}_0 \cdot \vec{v}_{2,0} < 0 \tag{6-4}$$

- both vehicles moving away from each other (R.AA):

$$\vec{d}_0 \cdot \vec{v}_{1,0} < 0 \cap \vec{d}_0 \cdot \vec{v}_{2,0} > 0 \tag{6-5}$$

As will be shown in later sections, the R.TT scenario encompasses two different cases. The regular R.TT scenario considers two objects simply moving towards each other. However, there is an additional R.TT' scenario which requires consideration. R.TT' is the case where the ego vehicle is on the opposite lane in order to pass an object on its own lane. While passing dynamic objects is not required for urban driving, passing static objects is required to fulfill the driving task. To distinguish this case, a three-wise interaction is required unlike for the other scenarios. Therefore, the first index is expanded with the option 3, which denotes the static object the ego passes. The R.TT' scenario is applied if all following conditions are met:

- Ego moving towards static object: $(\vec{r}_{3,0} - \vec{r}_{1,0}) \cdot \vec{v}_{1,0} > 0$
- OOI moving towards static object: $(\vec{r}_{3,0} - \vec{r}_{2,0}) \cdot \vec{v}_{2,0} > 0$
- OOI located behind static object: $(\vec{r}_{3,0} - \vec{r}_{1,0}) \cdot (\vec{r}_{3,0} - \vec{r}_{2,0}) < 0$

Finally, the following simplified assumptions common to all radial scenarios are introduced. The radial scenarios are described by a one-dimensional model as depicted in Fig. 6-3. To derive the model, all attributes are projected onto the connecting line between the two objects. During this projection, conservative estimates are applied. The radial distance between the ego and the OOI is:

$$d = |\vec{d}| \tag{6-6}$$

If the road is not straight, the actual path travelled by the objects may be longer. Nevertheless, the distance is the shortest possible path and therefore provides a conservative estimate. When projecting the velocities, the tangential component is neglected. This is again conservative since a tangential velocity component may prevent a potential collision.

$$v_{\mathrm{i,r}} = |\vec{v}_{\mathrm{i,r}}| = \left| \vec{v}_{\mathrm{i}} \cdot \frac{\vec{d}}{|\vec{d}|} \right| \tag{6-7}$$

The second index generally denotes the coordinate or direction, if it is present. In this case the index r refers to the radial direction. The worst case behavior generally assumes the maximum physically possible accelerations. When projecting to the radial direction, the available braking acceleration is reduced. The reason is that part of the friction may be required for a tangential acceleration. Practically speaking, this is the case if the vehicle is required to steer on a curved road. Therefore, only a reduced braking acceleration in radial direction $a_{\mathrm{i,r,b}}$ is available:

$$a_{\mathrm{i,r,b}} = cos(\alpha) \cdot a_{\mathrm{i,b}} = \frac{v_{\mathrm{i,r}}}{v_{\mathrm{i}}} \cdot a_{\mathrm{i,b}} \tag{6-8}$$

The last index b refers to the state of braking. Besides the braking acceleration, the guaranteed accelerations are also projected. As previously discussed in section 6.1.1, the longitudinal and lateral acceleration are limited by the vehicle power[361] and human preference[362], respectively. Since neither guaranteed acceleration is limited by road friction, they are not reduced so that $a_{\mathrm{i,r,g}} = a_{\mathrm{i,g}}$.

Tangential Scenarios

Similar to the radial case, the tangential interaction also provides different options. This section introduces their distinction as well as the special case of passing objects.

The scenario abbreviations introduced by the radial scenarios are expanded. The X signifies that the direction of the radial ego velocity is not relevant. Two cases of tangential interaction are distinguished:

- OOI moving away from ego (T.XA):

$$\vec{d}_0 \cdot \vec{v}_{2,0} \geq 0 \tag{6-9}$$

- OOI moving towards ego (T.XT):

$$\vec{d}_0 \cdot \vec{v}_{2,0} < 0 \tag{6-10}$$

361 Bokare, P. S.; Maurya, A. K.: Acceleration-Deceleration Behaviour (2017), p. 4742.

362 Bertolazzi, E. et al.: Supporting Drivers in Keeping Safe Speed and Safe Distance (2010), p. 530.

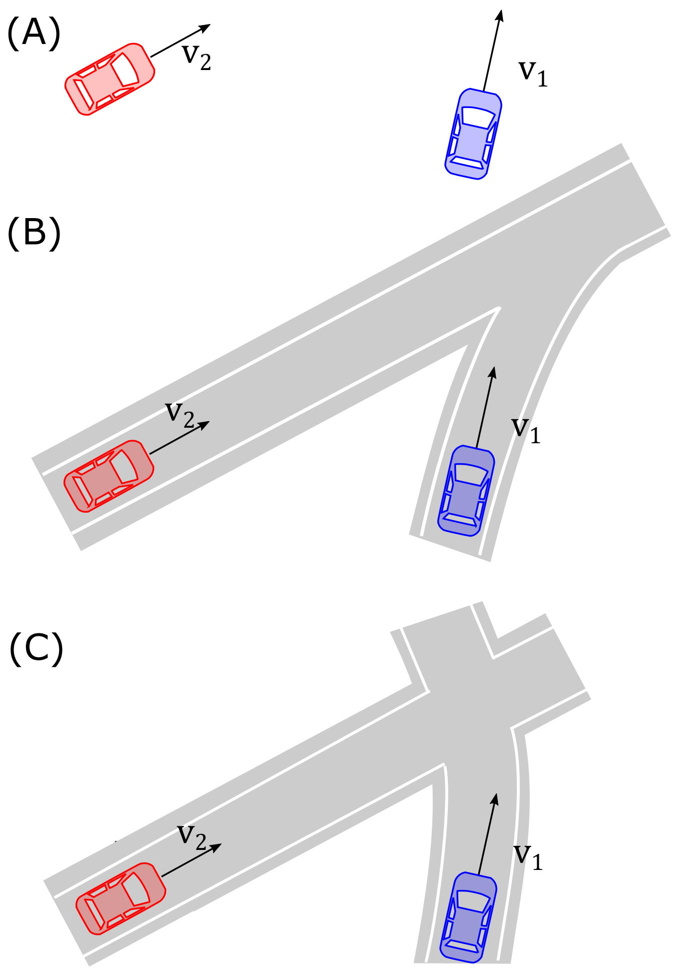

Figure 6-4: Different variants of the tangential scenario T.XT for the same initial object configuration (A) Object information (B) Potential merging scenario (C) Potential intersection scenario[363]

Tangential interactions do not require any further consideration if the OOI is moving away from the ego. Due to the separate consideration of radial interactions, this does however not necessarily mean that the object is irrelevant. T.XT is the case where the OOI is moving towards the ego vehicle. In this case, merging or intersection scenarios are possible. An example is depicted in Fig. 6-4. As shown in the visualization, the initial object configuration (A) is ambiguous regarding the road layout and the resulting scenario. Examples for possible scenarios are a merging procedure shown in (B) and an intersection scenario depicted in (C).

6.2.2 Relevance for Functional Scenarios

This section shows the application of the previously presented method to derive relevance for a functional scenario. Each functional scenario is treated separately to derive equations to distinguish relevant objects.

[363] Mori, K. et al.: Conservative Estimation of Perception Relevance (2023), Fig. 4 © 2023 IEEE.

R.TA: Ego moving towards OOI, OOI moving away from ego

An illustrative description of the R.TA scenario is the ego following the OOI. The corresponding behavioral requirement is explicitly stated in the road regulation StVO as follows:[364]

> REQ2.1: *The ego vehicle shall be able to brake to halt behind a vehicle in front to avoid a collision in the event that the front car suddenly brakes.*

The RSS model for formal planning similarly demands that other objects should not be hit from behind.[365] Formalizing collision avoidance in terms of the minimum distance d_{min} yields:

$$d_{\mathrm{min}} > 0 \tag{6-11}$$

Within the one-dimensional model, the vehicle behavior is fully specified by the radial acceleration. In accordance with worst case assumptions, the OOI performs a full brake. Therefore, its acceleration is directed towards the ego vehicle throughout the scenario. During its latency, the ego accelerates towards the OOI as worst case. After the latency, a valid breaking reaction of the ego with the specified $a_{1,\mathrm{b}}$ is performed.

The position of either vehicle is described by the following equations for the position r_{i} and velocity v_{i}. Constant acceleration is assumed, which is negative for the case of braking.

$$r_{\mathrm{i,r}} = r_{\mathrm{i,r,0}} + v_{\mathrm{i,r,0}}\ t + \frac{1}{2} a_{\mathrm{i,r,0}} t^2 \tag{6-12}$$

$$v_{\mathrm{i,r}} = v_{\mathrm{i,r,0}} + a_{\mathrm{i,r,0}}\ t \tag{6-13}$$

The first index again denotes the vehicle. The second index r refers to the radial component, while the third index 0 refers to the initial state. For both vehicles, the corresponding radial braking distance $r_{\mathrm{i,r,b}}$ is:

$$r_{\mathrm{i,r,b}} = \frac{v_{\mathrm{i,r}}^2}{2a_{\mathrm{i,r,b}}} \tag{6-14}$$

With this, it is possible to describe the ego braking procedure. First, the ego accelerates during its reaction time t_{r} and then performs a full brake until it stops. Considering the initial acceleration with (6-12) as well as the braking distance in (6-14) by leveraging (6-13), the final ego position is:

$$\begin{aligned} r_{1,\mathrm{r,s}} = r_{1,\mathrm{r},0} + v_{1,\mathrm{r},0}\ t_{1,\mathrm{r}} + \frac{1}{2} a_{1,\mathrm{r},0} t_{1,\mathrm{r}}^2 \\ + \frac{(v_{1,\mathrm{r},0} + t_{1,\mathrm{r}} a_{1,\mathrm{r},0})^2}{2a_{1,\mathrm{r,b}}} \end{aligned} \tag{6-15}$$

[364] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), p. 3.

[365] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017), p. 6-7.

The minimal distance occurs when both vehicles have stopped. Subtracting (6-14) from the braking distance of the OOI as specified in (6-12) provides the minimal distance. Furthermore, the object sizes s_i are considered by subtracting them similar to Wachenfeld et al.[366] The final equation includes the worst case assumptions, the radial components and the intial requirement from (6-11):

$$\begin{aligned} 0 < d_\mathrm{min} = d_0 - s_1 - s_2 + \frac{v_{2,\mathrm{r},0}^2}{2a_\mathrm{max}} - v_{1,\mathrm{r},0}\, t_{1,\mathrm{r}} \\ - \frac{1}{2}a_\mathrm{max}t_{1,\mathrm{r}}^2 - \frac{(v_{1,\mathrm{r},0} + t_{1,\mathrm{r}}a_\mathrm{max})^2}{2a_{1,\mathrm{r},\mathrm{b}}} \end{aligned} \tag{6-16}$$

Satisfying this condition means that even for worst case behavior, no collision occurs. Conversely, violating this conditions means that certain ego actions are not available if a collision is to be avoided. Therefore, if the condition is violated, the other object is relevant.

R.AT: ego moving away from OOI, OOI moving towards ego

This scenario basically represents the reversal of the situation described in the previous section. Accordingly, this time the ego vehicle is being followed by the OOI. The corresponding requirement obtained from the road regulation is:[367]

> REQ2.2: *other vehicles should not be unnecessarily impeded.*

Impeding is a term that requires further interpretation according to the safety principle. Within this work, impeding is considered to mean that the OOI receives additional behavioral requirements originating from the ego vehicle.

This case is identical to the previous R.TA scenario with exchanged roles. Accordingly, the requirement for the OOI is to avoid collisions with the ego vehicle. Two different cases are distinguished in the following. The first is the R.AT+ scenario where the ego vehicle is already travelling with an adequate speed. The other is the R.AT- scenario in which the ego vehicles speed is below the desired or adequate speed.

R.AT+: ego with desired speed moving away from OOI, OOI moving towards ego

This section considers the scenario where the ego velocity is already adequately high. Under this assumption, REQ2.2 is specified with the following sub-requirement:

> REQ2.3: *The ego vehicle may not restrict the actions of the following vehicle by unnecessarily braking.*

[366] Wachenfeld, W. et al.: The worst-time-to-collision metric (2016), p. 730-731.

[367] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), p. 1.

The corresponding equation is obtained by directly reusing the equation from the previous section. However, the roles of the vehicles and the corresponding indices are exchanged:

$$\begin{aligned} 0 < d_{\min} = d_0 - s_1 - s_2 + \frac{v_{1,\mathrm{r},0}^2}{2a_{\max}} - v_{2,\mathrm{r},0}\, t_{2,\mathrm{r}} \\ - \frac{1}{2}a_{\max}t_{2,\mathrm{r}}^2 - \frac{(v_{2,\mathrm{r},0} + t_{2,\mathrm{r}}a_{\max})^2}{2a_{2,\mathrm{r,b}}} \end{aligned} \tag{6-17}$$

Exchanging the roles means that the reaction time $t_{2,\mathrm{r}}$ and the braking acceleration $a_{2,\mathrm{r,b}}$ both correspond to the OOI. The interpretation of this equation also changes slightly. In this case, it signifies the cases where the actions available to the OOI are not restricted by a potential full braking of the ego. Violating this equation conversely means that the actions of the OOI are restricted by the ego. This means the ego is hindering the other object for which a valid reason is required. Accordingly, the OOI is relevant in this case.

R.AT-: ego with less than desired speed moving away from OOI, OOI moving towards ego

In the R.AT- scenario, the initial ego velocity is below the desired velocity. A common example of this scenario is a lane change onto a faster lane. Therefore, the ego vehicle is required to accelerate until it reaches a desired adequate speed of $v_{1,\mathrm{r,d}}$. From that point onward, the scenario transitions to the R.AT+ scenario.

Before this transition to the R.AT+ scenario, the REQ2.2 is interpreted as follows:

> REQ2.4: *The ego vehicle may not restrict the actions of the following vehicle by having insufficient speed.*

As previously mentioned, this scenario typically occurs after a lane change. Therefore, the ego latency is not considered here, but is instead considered at the beginning of the lane change. Thus, the ego vehicle simply accelerates until it reaches its desired speed. During the whole scenario, the OOI performs its maximum acceleration towards the ego vehicle.

As previously indicated, the R.AT- scenario ends with a transition to the R.AT+ scenario when the ego vehicle reaches its desired speed. The point in time where the desired speed is reached denoted with the index d is calculated as:

$$t_{\mathrm{d}} = \frac{v_{1,\mathrm{r,d}} - v_{1,\mathrm{r},0}}{a_{1,\mathrm{r,g}}} \tag{6-18}$$

Generally speaking, it is difficult to ascertain a suitable desired speed. One possible option is to utilize the speed limit v_{lim}. However, since speed infractions may occur, it may be more reasonable to additionally consider the initial velocity of the other vehicle. To maintain conservativity, the maximum of the two values is used:

$$v_{1,\mathrm{r,d}} = max\left\{v_{\mathrm{lim}}, v_{2,0}\right\} \tag{6-19}$$

It is in principle possible that the OOI accelerates beyond its initial velocity during the scenario. However, the OOI has a higher worst case acceleration a_{max} than the acceleration available to the ego $a_{1,\mathrm{g}}$. Therefore, assuming an accelerating OOI fails to provide an upper boundary to the desired speed of the ego vehicle. In addition, it appears unreasonable to demand that the ego vehicle may not impede an OOI which is accelerating beyond both its initial velocity and the speed limit. While further discussion and validation of such assumptions regarding behavioral requirements is required, this work considers above equation (6-19). It should be noted that for the later implementation, the speed limits are unknown. Therefore, the implementation instead assumes that $v_{1,\mathrm{r,d}} = v_{2,0}$ for practical purposes. Since the OOI accelerates faster than the ego vehicle, the distance decreases throughout the R.AT- scenario prior to the transition to R.AT+. The distance between the two vehicles is calculated by inserting the assumptions of worst case behavior into (6-12) to yield:

$$d = d_0 - s_1 - s_2 + (v_{1,\mathrm{r},0} - v_{2,\mathrm{r},0})t + \frac{1}{2}(a_{1,\mathrm{r,g}} - a_{\mathrm{max}})t^2 \tag{6-20}$$

The critical point in time for a collision occurs when the distance becomes minimal. Since the distance continues to decrease until the ego reaches its desired velocity, the minimal distance occurs at t_{d}. The corresponding velocity of the OOI at t_{d} is obtained by inserting (6-18) into (6-13):

$$v_{2,\mathrm{r,d}} = v_{2,\mathrm{r},0} + a_{\mathrm{max}} t_{\mathrm{d}} \tag{6-21}$$

The distance between the vehicles at t_{d} is obtained by inserting (6-18) into (6-20). Then, the scenario is described by modifying (6-17) from the R.AT+ scenario. Firstly, the initial distance $(d_0 - s_1 - s_2)$ is substituted by the resulting distance calculated for t_{d}. The corresponding velocity at this time is the desired velocity $v_{1,\mathrm{r,d}}$ which replaces the initial ego velocity $v_{1,\mathrm{r},0}$. Finally, the initial velocity of the OOI $v_{2,\mathrm{r},0}$ is substituted by (6-21). The resulting equation is:

$$\begin{aligned} 0 < d_{\mathrm{min}} = d(t = t_{\mathrm{d}}) + \frac{v_{1,\mathrm{r,d}}^2}{2a_{\mathrm{max}}} - v_{2,\mathrm{r,d}}\, t_{2,\mathrm{r}} \\ - \frac{1}{2} a_{\mathrm{max}} t_{2,\mathrm{r}}^2 - \frac{(v_{2,\mathrm{r,d}} + t_{2,\mathrm{r}} a_{\mathrm{max}})^2}{2a_{2,\mathrm{r,b}}} \end{aligned} \tag{6-22}$$

Both the reaction time $t_{2,\mathrm{r}}$ and the guaranteed braking acceleration $a_{2,\mathrm{r,b}}$ correspond to the OOI. Similar to the previous scenarios, objects which violate this requirement are relevant.

R.TT: both vehicles moving towards each other

The R.TT scenario occurs if two objects move towards each other either in longitudinal or lateral direction. The former case primarily occurs with vehicles driving on the other side of the road. On the other hand, the latter case occurs during merging as well as in intersections.

One possible assumption for the case of two vehicles moving towards each other is both vehicles reacting correctly. This assumption is explicitly introduced by RSS[368]. German road regulations also implicitly contain this assumption in the demand that stopping is required within half of the visible distance.[369a] However, these approaches do not align with the worst case assumptions applied in this work. Instead, the legal requirement to avoid harming or endangering others[369b] is used as basis. In cases where an accident is unavoidable, an attempt to minimize harm and damage is required. Naturally, the ego vehicle can only minimize its own speed to reduce its own contribution to the accident. However, no behavioral assumptions for the other vehicle are available. The resulting requirement is:

> REQ2.5: *The ego vehicle shall brake to a standstill before the other vehicle collides with it.*

The worst case behavior is thus formalized as follows. The worst case is the OOI accelerating towards the ego vehicle with the maximum physically possible acceleration. Similar to the R.TA scenario, the ego accelerates towards the other vehicle during the latency. After the latency, it brakes to a full stop. Since collisions require the distance to be zero, the requirement for the minimum distance is:

$$d_{\mathrm{min}} > 0 \tag{6-23}$$

The ego braking under worst case assumptions is already formalized in (6-14). The equation for the velocity during the latency and after the ego reaction are given by the equations:

$$v_{1,\mathrm{r}} = v_{1,0} + a_{\mathrm{max}} t \quad \text{for} \quad t \leq t_{\mathrm{r}} \tag{6-24}$$

$$v_{1,\mathrm{r}} = v_{1,\mathrm{r,b}} - a_{1,\mathrm{r,b}}(t - t_{\mathrm{r}}) \quad \text{for} \quad t \geq t_{\mathrm{r}} \tag{6-25}$$

In above equation, the entry $v_{1,\mathrm{r,b}}$ denotes the ego velocity when it begins braking after its latency. The time which the ego requires for braking $t_{1,\mathrm{b}}$ is derived by inserting (6-24) into (6-25). Requiring that the ego comes to a stop $v_{1,\mathrm{r}} = 0$ and solving for the braking time yields:

$$t_{1,\mathrm{b}} = t_{1,\mathrm{r}} + \frac{(v_{1,\mathrm{r},0} + t_{1,\mathrm{r}}\, a_{\mathrm{max}})}{a_{1,\mathrm{r,b}}} \tag{6-26}$$

Utilizing (6-12) with negative velocity and acceleration yields the location of the OOI. Considering the ego stopping distance (6-15), the vehicle sizes, the requirement (6-23) and inserting worst

[368] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017), p. 6-7.

[369] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), a: p. 2, b: p. 1.

case assumptions finally yields:

$$
\begin{aligned}
0 < d_{\min} = d_0 - s_1 - s_2 - v_{1,\mathrm{r},0}\, t_{1,\mathrm{r}} \\
- \frac{1}{2} a_{\max} t_{1,\mathrm{r}}^2 - \frac{(v_{1,\mathrm{r},0} + t_{1,\mathrm{r}} a_{\max})^2}{2 a_{1,\mathrm{r,b}}} \\
- v_{2,\mathrm{r},0} t_{1,\mathrm{b}} - \frac{1}{2} a_{\max} t_{1,\mathrm{b}}
\end{aligned}
\tag{6-27}
$$

In this equation, velocities are considered positive if they are pointing towards the OOI. Any OOI violating this equation is a candidate for potential collisions and therefore relevant.

R.TT': opposing OOI while passing a static object

As previously indicated, the R.TT' scenario considers passing a static object. This means that a three-wise interaction between static object, ego vehicle and the OOI on the opposite lane is required. The values 1, 2 and 3 for the first index denote the ego vehicle, the opposing OOI and the static object, respectively. For simplicity, this scenario is only applied to OOI which are not considered relevant according to the R.TT scenario.

To model the scenario, different interactions are considered. Firstly, the static object is required to be sufficiently close to be relevant to the ego vehicle for passing to be realistic. Then, the opposing vehicle is considered during the passing procedure. The passing itself is modeled as the three phases of lateral movement onto the opposite lane, longitudinal acceleration and another lateral movement onto its own lane. An overview over the procedure as well as variables is provided in Fig. 6-5. The velocities of the ego vehicle and the opposing OOI are projected onto the radial

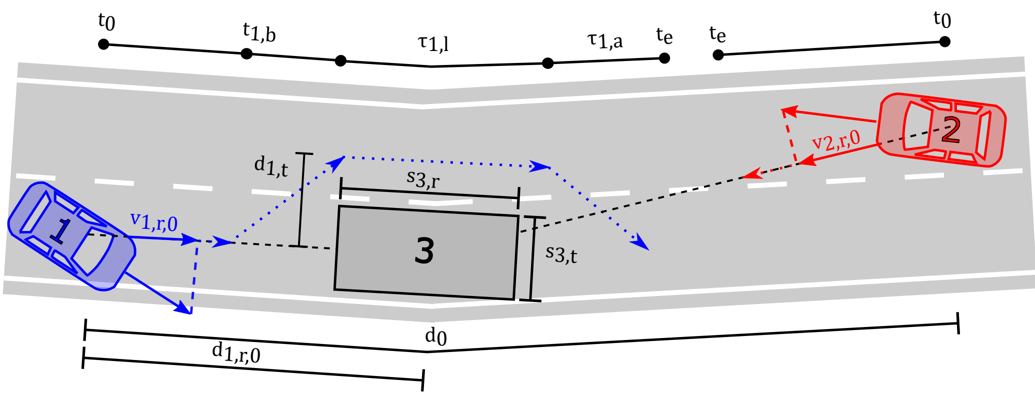

Figure 6-5: Sequence and variables for the R.TT' scenario.[370]

direction provided by the connecting line to the static object. Static objects include any objects that narrow the road in such a manner that the ego vehicle must use the opposing lane to pass it. This includes objects such as parked vehicles, bus stops and construction works. Especially for the latter case, the length of the obstacle is significantly larger than the width. Therefore, the static object is modeled with two size dimensions, $s_{3,\mathrm{r}}$ in radial and $s_{3,\mathrm{t}}$ in tangential direction

[370] Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023), Figure 1.

relative to the ego vehicle. For the lateral movement, the ego vehicle travels a lateral distance of $d_{1,\mathrm{t}}$. The different points in time t and the different time spans τ are discussed in the following.

To model the scenario, the static object is considered as OOI as a first step. The static object must be relevant according to either R.TT or R.TA which are equivalent for static objects. Arbitrarily selecting R.TA yields the requirement for the distance to the static object $d_{1,\mathrm{r},0}$

$$0 < d_{1,\mathrm{r,min}} = d_{1,\mathrm{r},0} - s_1 - s_{3,\mathrm{r}} - v_{1,\mathrm{r},0}\, t_{1,\mathrm{r}} + \frac{1}{2} a_{\max} t_{1,\mathrm{r}}^2 - \frac{(v_{1,\mathrm{r},0} + t_{1,\mathrm{r}} a_{\max})^2}{2 a_{1,\mathrm{r,b}}} \tag{6-28}$$

A static object violating this requirement is considered relevant. Other objects on the opposite lane are only considered if a relevant static object is present.

If a static object is present, it is assumed the ego vehicle passes it based on worst case assumptions. Therefore, the ego vehicle brakes during its latency. The radial and tangential direction denoted with the indices r and t are defined with respect to the connecting line of the ego and the static object. This results in a distance covered $d_{1,\mathrm{r,b}}$ and the corresponding velocity $v_{1,\mathrm{r,b}}$:

$$d_{1,\mathrm{r,b}} = v_{1,\mathrm{r},0} t_{1,\mathrm{b}} - \frac{1}{2} a_{\max} t_{1,\mathrm{b}}^2 \tag{6-29}$$

$$v_{1,\mathrm{r,b}} = v_{1,\mathrm{r},0} - a_{\max} t_{1,\mathrm{b}} \tag{6-30}$$

The braking maneuver terminates either when the ego vehicle comes to a full stop or at the reaction time. The end time of this braking maneuver $t_{1,\mathrm{b}}$ is therefore described as follows:

$$t_{1,\mathrm{b}} = min\left\{ t_{1,\mathrm{r}}\ ,\ \frac{v_{1,\mathrm{r},0}}{a_{\max}} \right\} \tag{6-31}$$

After the reaction time, the first lateral movement onto the opposite lane is initiated. The lateral movement must cover half of the ego size as well as half of the static objects size for the ego to avoid collision with the static object. It should be noted that $s_{3,\mathrm{t}}$ may also include a safety margin which is maintained to the static object when passing it.

$$d_{1,\mathrm{t}} = \frac{s_1}{2} + \frac{s_{3,\mathrm{t}}}{2} \tag{6-32}$$

During this lateral movement, the ego vehicle first steers to initiate a lateral movement. This is followed by another steering which is a lateral deceleration so that the lateral velocity is zero again after completing the lateral movement. Since both the lateral acceleration and lateral deceleration are steering maneuvers, the situation is symmetrical. Therefore, the time span $\tau_{1,\mathrm{l}}$ required for the lateral movement can be calculated considering only the first half of acceleration. The acceleration procedure covers half of the overall lateral distance in half of the time required for the entire lateral movement. Considering the fact that the initial lateral location and lateral velocity of the ego are zero, the corresponding equation for the lateral location is:

$$\frac{1}{2}d_{1,\mathrm{t}} = \frac{1}{2}a_{1,\mathrm{g}}\left(\frac{\tau_{1,\mathrm{l}}}{2}\right)^2 \tag{6-33}$$

Solving this equation for the lateral movement time $\tau_{1,\mathrm{l}}$ yields:

$$\tau_{1,\mathrm{l}} = \sqrt{\frac{2(s_{3,\mathrm{t}} + s_1)}{a_{1,\mathrm{g}}}} \tag{6-34}$$

This lateral motion is followed by an acceleration with the guaranteed acceleration denoted by the index a. Accelerating for a duration of $\tau_{1,\mathrm{a}}$ leads to a velocity $v_{1,\mathrm{a}}$:

$$v_{1,\mathrm{r,a}} = v_{1,\mathrm{r,b}} + a_{1,\mathrm{g}}\tau_{1,\mathrm{a}} \tag{6-35}$$

The radial distance $d_{1,\mathrm{a}}$ travelled during the longitudinal acceleration is the distance required to pass the static object. This is the sum of the length of the ego vehicle and the length of the static object. Accordingly, the radial distance travelled is:

$$d_{1,\mathrm{a}} = s_1 + s_{\mathrm{s,r}} = v_{1,\mathrm{r,b}}\tau_{1,\mathrm{a}} + \frac{1}{2}a_{1,\mathrm{g}}\tau_{1,\mathrm{a}}^2 \tag{6-36}$$

Solving for the time $\tau_{1,\mathrm{a}}$ required to pass the static object results in the equation:

$$\tau_{1,\mathrm{a}} = \frac{-v_{1,\mathrm{r,b}} + \sqrt{2a_{1,\mathrm{g}}(2s_1 + 2s_{3,\mathrm{r}}) + v_{1,\mathrm{r,b}}^2}}{a_{1,\mathrm{g}}} \tag{6-37}$$

After the the acceleration, the second lateral movement occurs. This lateral movement leads to the end of the scenario which is denoted with the index e. The second lateral movement of the ego vehicle requires a time span of $\tau_{1,\mathrm{l}}$. This is equal to the first lateral movement due to the symmetry of the situation. The overall radial distance $d_{1,\mathrm{r,e}}$ travelled by the ego vehicle during the passing is a sum of all individual maneuvers:

$$d_{1,\mathrm{r,e}} = d_{1,\mathrm{r,b}} + v_{1,\mathrm{r,b}}\tau_{1,\mathrm{l}} + d_{1,\mathrm{a}} + v_{1,\mathrm{r,a}}\tau_{1,\mathrm{l}} \tag{6-38}$$

Throughout the passing scenario, the vehicle on the opposite lane acts in accordance with worst case assumptions. Therefore, it performs a maximum acceleration towards the ego vehicle until the end of the passing. The corresponding distance $d_{2,\mathrm{r,e}}$ which is travelled, the final velocity $v_{2,\mathrm{r,e}}$ and the time t_e required are:

$$d_{2,\mathrm{r,e}} = v_{2,\mathrm{r,0}}t_\mathrm{e} + \frac{1}{2}a_{\max}t_\mathrm{e}^2 \tag{6-39}$$

$$v_{2,\mathrm{r,e}} = v_{2,\mathrm{r,0}} + a_{\max}t_\mathrm{e} \tag{6-40}$$

$$t_\mathrm{e} = t_{1,\mathrm{r}} + 2\tau_{1,\mathrm{l}} + \tau_{1,\mathrm{a}} \tag{6-41}$$

These equations provide the positions and velocities of both vehicles when the ego vehicle completes the passing. At this point in time, the ego vehicle and the opposite vehicle are treated in accordance with the R.TT scenario. However, the initial values of the R.TT scenario as utilized in (6-43) are replaced by the values at the moment where passing is completed. The resulting criterion is:

$$
\begin{aligned}
0 < d_{\mathrm{min}} = (d_0 - d_{1,\mathrm{r,e}} - d_{2,\mathrm{r,e}}) - s_1 - s_2 - v_{1,\mathrm{r,a}}\, t_{1,\mathrm{r}} \\
- \frac{1}{2} a_{\mathrm{max}} t_{1,\mathrm{r}}^2 - \frac{(v_{1,\mathrm{r,a}} + t_{1,\mathrm{r}} a_{\mathrm{max}})^2}{2 a_{1,\mathrm{r,b}}} - v_{2,\mathrm{r,e}} t_{1,\mathrm{b}} - \frac{1}{2} a_{\mathrm{max}} t_{1,\mathrm{b}}
\end{aligned}
\tag{6-42}
$$

Violating this equation means that the opposite vehicle is relevant.

However, inserting exemplary values into this equation results in distances in the hundreds of meters. These values surpass dataset annotation ranges by at least one order of magnitude. Furthermore, passing an object only requires consideration if there is an ego intention to pass. If no intention to pass a static object while entering the opposite lane is present, evaluation is not required. However, this ego intention is unavailable on contemporary datasets. Therefore, this scenario is neglected for the practical implementation used in the results. The results of this section are shown for the sake of completeness within the urban domain.

R.AA: both vehicles moving away from each other

Similar to the previous scenario, this scenario describes objects moving away from each other in either longitudinal or lateral direction. This may occur for vehicles travelling on the opposite lane as well as for cut out maneuvers or diverging lanes.

For lateral movement such as during cut out maneuvers, the OOI may reverse its direction of movement. In this case, braking by the ego is required. For this case, the results of the previous section 6.2.2 are applied. Therefore, the requirement is identical to the previous result obtained from REQ2.5. The only difference is the fact that the velocity vectors are pointing away from each other. Accordingly, the radial velocity components in (6-43) are negative. For purpose of completeness, the equation is restated as:

$$
\begin{aligned}
0 < d_{\mathrm{min}} = d_0 - s_1 - s_2 - v_{1,\mathrm{r},0}\, t_{1,\mathrm{r}} \\
- \frac{1}{2} a_{\mathrm{max}} t_{1,\mathrm{r}}^2 - \frac{(v_{1,\mathrm{r},0} + t_{1,\mathrm{r}} a_{\mathrm{max}})^2}{2 a_{1,\mathrm{r,b}}} \\
- v_{2,\mathrm{r},0} t_{1,\mathrm{b}} - \frac{1}{2} a_{\mathrm{max}} t_{1,\mathrm{b}}
\end{aligned}
\tag{6-43}
$$

Again, a violation of the requirement indicates the possibility of a collision which means the object is relevant.

T.XT: OOI moving towards ego

The T.XT scenario applies in cases where the ego vehicle merges in front of the OOI. As previously, the ego vehicle is required to avoid unnecessarily impeding the other vehicle[371] as stated in the behavioral requirement REQ2.2. This situation may occur for merging on adjacent lanes as well as for intersections.

In an intersection, two different potential maneuvers are available as depicted in Fig. 6-6. The ego vehicle can either pass through the intersection (A) or turn to merge onto the lane upon which the OOI is travelling (B). If the ego vehicle enters the intersection, merging onto the OOIs lane is the worst case maneuver. The reason is that the ego vehicle is not only required to leave the intersection, but also to accelerate to an adequate speed. A collision in the merging maneuver is only avoided if this adequate velocity is reached sufficiently fast. Therefore, only the maneuver (B) is considered in the following.

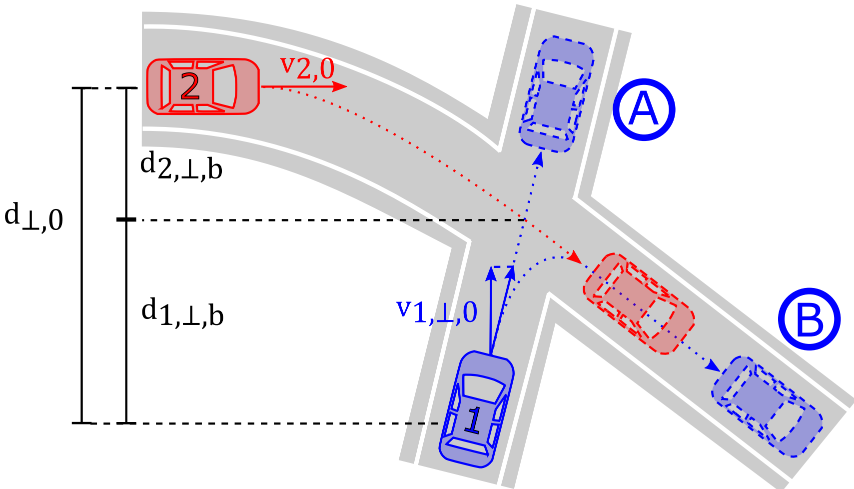

Figure 6-6: Maneuvers in the intersection and variables in the T.XT scenario.[372]

Assuming the maneuver (B), the whole process is decomposed into two different cases. Firstly, the ego vehicle may brake in order to avoid entering the intersection, thus avoiding a collision. The option of braking is only available if there is sufficient distance between the intersection and the ego vehicle. If entering the intersection is unavoidable, a merging maneuver is considered. However, the location of the intersection is assumed to be unknown within this work. Therefore, both braking and merging as potential scenario developments require consideration in order to develop the worst case assumptions. For the coordinate system, the OOI is used to define lateral and longitudinal direction. As shown in the following, this provides a suitable reference to define the worst case intersections.

The description of the scenario consists of two parts. Firstly, the case where the ego vehicle brakes before the intersection is discussed. This evaluates if a hypothetical intersection exists for

[371] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), p. 1.

[372] Modified from Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023), Fig. 2.

which the ego can no longer avoid entering the intersection by braking. Further consideration of merging is only required if such an intersection exists. A visualization is provided in Fig. 6-6. The distance required by the ego vehicle to brake to a full stop is:

$$d_{1,\perp,\mathrm{b}} = v_{1,\perp,0} t_{1,\mathrm{r}} + \frac{v_{1,\perp,0}^2}{2a_{1,\perp,\mathrm{b}}} \tag{6-44}$$

Since the location of the intersection is assumed to be unknown, a worst case is considered. The worst case to determine if braking is possible is where the intersection is as close as possible to the ego vehicle. However, the OOI is still required to be able to reach the intersection. To this end, a lateral movement of the OOI towards the ego vehicle is assumed. Accordingly, the worst case intersection occurs where the OOI performs its maximum acceleration in lateral direction towards the ego vehicle. The lateral distance traveled $d_{2,\perp}$ during the braking time of the ego vehicle $t_{1,\mathrm{b}}$ is:

$$d_{2,\perp,\mathrm{b}} = \frac{1}{2} a_{\max} t_{1,\mathrm{b}}^2 \tag{6-45}$$

$$t_{1,\mathrm{b}} = t_{1,\mathrm{r}} + \frac{v_{1,\perp,0} + t_{1,\mathrm{r}} a_{\max}}{a_{1,\mathrm{b}}} \tag{6-46}$$

If the movements of both vehicles exceed the original lateral distance $d_{\perp,0}$, braking no longer avoids entering the intersection:

$$d_{\perp,0} < d_{1,\perp,\mathrm{b}} + d_{2,\perp,\mathrm{b}} = v_{1,\perp,0} t_{\mathrm{r}} + \frac{v_{1,\perp,0}^2}{2a_{1,\perp,\mathrm{b}}} + \frac{1}{2} a_{\max} \left[t_{1,\mathrm{r}} + \frac{v_{1,\perp,0} + t_{1,\mathrm{r}} a_{\max}}{a_{1,\mathrm{b}}} \right]^2 \tag{6-47}$$

Therefore, if this equation is fulfilled, further consideration of the a merging scenario within the intersection is required.

This section describes a merging process in which the ego vehicle merges onto the lane upon which the OOI is travelling. A visualization of the scenario is provided in Fig. 6-7. The symbols and the notation in the image are referenced and explained throughout this section. The worst case for this merging scenario differs from the assumptions in the previous section. In the previous section, assumptions considered what the worst case is for a merging to happen. This section describes the worst case merging already presupposing that a merging happens. Conservatively overestimating the merging scenarios in this way simplifies the following formalization.

The worst case intersection for the merging procedure is provided by the current direction of motion of the OOI. Therefore, the motion of the OOI serves as reference coordinate system with a longitudinal and lateral component. Whenever the words “longitudinal” and “lateral” are applied in this section, they refer to the coordinate system of the OOI. Since the OOI requires no acceleration for lateral movement, it uses its full acceleration to accelerate towards the ego vehicle. The ego movement is divided into two steps. First, a purely lateral movement into the intersection is assumed. Afterwards, the ego vehicle performs a longitudinal acceleration. The conditions for the ego vehicle to not impede the OOI before and after reaching the target speed during the

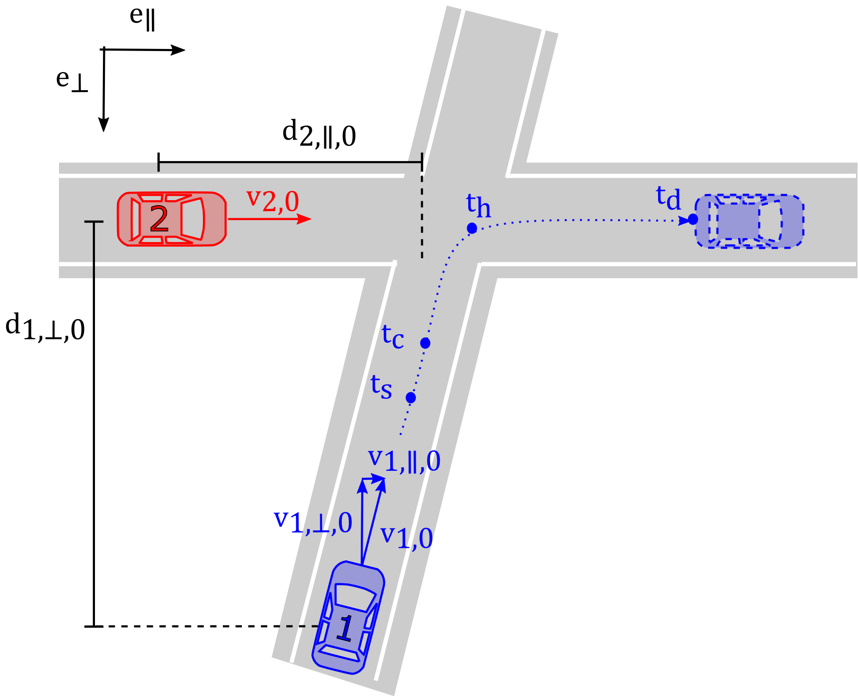

Figure 6-7: Worst-case intersection and variables in the T.XT scenario.[373]

longitudinal acceleration are formalized in the R.AT- and R.AT+ scenario, respectively. This model provides a conservative estimate since simultaneous acceleration in longitudinal and lateral direction may shorten the duration of merging. The coordinate system is provided by two unit vectors which are defined with respect to the direction of motion of the OOI. The first longitudinal or parallel unit vector $\vec{e}_{\|}$ is obtained by normalizing velocity vector of the OOI $\vec{v}_2$. It consists of the two components $x_{\|}$ and $y_{\|}$:

$$\vec{e}_{\|} = \begin{bmatrix} x_{\|} \\ y_{\|} \end{bmatrix} = \frac{\vec{v}_2}{|\vec{v}_2|} \tag{6-48}$$

The second lateral unit vector $\vec{e}_{\perp}$ is perpendicular to the parallel unit vector. It consists of the two components $x_{\perp}$ and $y_{\perp}$ which are calculated by exchanging the two vector components of the parallel unit vector:

$$\vec{e}_{\perp} = \begin{bmatrix} x_{\perp} \\ y_{\perp} \end{bmatrix} = \begin{bmatrix} y_{\|} \\ -x_{\|} \end{bmatrix} \tag{6-49}$$

The ego velocity is divided into the two velocity components in lateral direction $v_{1,\perp}$ and longitudinal direction $v_{1,\|}$ relative to the OOI as:

$$v_{1,\perp} = \vec{v}_1 \cdot \vec{e}_{\perp} \quad \text{and} \quad v_{1,\|} = \vec{v}_1 \cdot \vec{e}_{\|} \tag{6-50}$$

The lateral movement of the ego vehicle is modeled with worst case assumptions. In the case

[373] This image was created together with Kai Storms and also appears in his work[374] created concurrently.

[374] Storms, K.: Context Aware Data Reduction (2023).

of the merging, the worst case occurs if the ego vehicle requires a lot of time. Therefore, the worst case for the lateral movement is a lateral deceleration of the ego vehicle. However, a lateral movement away from the intersection is not possible since this case is excluded by (6-47). Within this constraint, worst case assumptions are applied within the reaction time t_r. During the reaction time, the lateral velocity and position are described by:

$$v_{1,\perp} = v_{1,\perp,0} + a_\mathrm{max} \cdot t \tag{6-51}$$

$$r_{1,\perp} = r_{1,\perp,0} + v_{1,\perp,0} \cdot t + \frac{1}{2} a_\mathrm{max} t^2 \tag{6-52}$$

Since movements away from the intersection were previously excluded, the lateral velocity is required to be negative. Limiting the lateral velocity to a maximum of $v_{1,\perp} = 0$ is achieved by inserting the following expression in above equations:

$$t_\mathrm{1,s} = min\left\{ t_\mathrm{1,r}\ ,\ -\frac{v_{1,\perp,0}}{a_\mathrm{max}} \right\} \tag{6-53}$$

The worst case action during the reaction time is followed by a lateral movement. A substitution $t' = t - t_\mathrm{1,r}$ is applied in the following equations for brevity. The start of the lateral movement is indicated by the index s. The point in time where the ego vehicle changes its acceleration to begin decelerating in lateral direction is denoted by c. Decelerating is required for the ego vehicle to stay upon the lane of the OOI. Both the lateral acceleration and the lateral deceleration use the guaranteed acceleration $a_\mathrm{1,g}$. While lateral deceleration may be higher, the conservative assumption is that lateral deceleration is achieved by steering. The velocities for both phases of lateral movement are:

$$v_{1,\perp} = v_{1,\perp,\mathrm{s}} - a_\mathrm{1,g} \cdot t' \quad \text{for} \quad t' \leq t'_\mathrm{c} \tag{6-54}$$

$$v_{1,\perp} = v_{1,\perp,\mathrm{c}} + a_\mathrm{1,g} \cdot (t' - t'_\mathrm{c}) \quad \text{for} \quad t' > t'_\mathrm{c} \tag{6-55}$$

During the reaction time where $t' \leq t'_\mathrm{c}$, the location of the ego vehicle is:

$$r_{1,\perp} = r_{1,\perp,\mathrm{s}} + v_{1,\perp,\mathrm{s}} \cdot t' - \frac{1}{2} a_\mathrm{1,g} t'^2 \tag{6-56}$$

The location after the ego latency for $t' > t'_\mathrm{c}$ is:

$$r_{1,\perp} = r_{1,\perp,\mathrm{c}} + v_{1,\perp,\mathrm{c}} \cdot (t' - t'_\mathrm{c}) + \frac{1}{2} a_\mathrm{1,g} (t' - t'_\mathrm{c})^2 \tag{6-57}$$

Finally, the ego vehicle arrives on the the lane of the OOI after the halting time t'_h and stays there. These conditions after t'_h can be formalized as follows:

$$v_{1,\perp,\mathrm{h}} = v_{1,\perp}(t = t'_\mathrm{h}) = 0 \tag{6-58}$$

$$r_{1,\perp,\mathrm{h}} = r_{1,\perp}(t = t'_\mathrm{h}) = 0 \tag{6-59}$$

The time which is required for the lateral movement is defined through an equation system. As first step, the two equations (6-54) and (6-55) are inserted into (6-58):

$$0 = v_{1,\perp,\mathrm{h}} = v_{1,\perp,\mathrm{s}} + a_{1,\mathrm{g}} \cdot (t'_\mathrm{h} - 2t'_\mathrm{c}) \tag{6-60}$$

Combining (6-51), (6-52), (6-57) as well as (6-59) yields:

$$\begin{aligned} 0 = {} & r_{1,\perp,\mathrm{s}} + v_{1,\perp,\mathrm{s}} \cdot t'_\mathrm{c} - \frac{1}{2} a_{1,\mathrm{g}} {t'_\mathrm{c}}^2 \\ & + (v_{1,\perp,\mathrm{s}} - a_{1,\mathrm{g}} \cdot t_\mathrm{c}) \cdot (t'_\mathrm{h} - t'_\mathrm{c}) \quad + \frac{1}{2} a_{1,\mathrm{g}} (t'_\mathrm{h} - t'_\mathrm{c})^2 \end{aligned} \tag{6-61}$$

Solving the two equations (6-60) and (6-61) for t'_c and t'_h provides the time required for the lateral movement:

$$t'_\mathrm{c} = \frac{v_{1,\perp,\mathrm{s}}}{a_\mathrm{max}} + \sqrt{\frac{r_{1,\perp,\mathrm{s}}}{a_{1,\perp,\mathrm{g}}}} \tag{6-62}$$

$$t'_\mathrm{h} = t'_\mathrm{c} - \frac{v_{1,\perp,\mathrm{s}}}{a_{1,\perp,\mathrm{g}}} \tag{6-63}$$

Throughout the lateral movement, the ego vehicles continues moving in longitudinal direction with constant velocity. The lateral movement is followed by an acceleration similar to R.AT-. However, the longitudinal velocity components from the equations for the T.XT scenario are applied instead of the previously applied radial components. Furthermore, the time to reach the desired velocity t'_d is modified from (6-18) to include reaction time and lateral movement:

$$t'_\mathrm{d} = t_\mathrm{r} + t'_\mathrm{h} + \frac{v_{1,\parallel,\mathrm{d}} - v_{1,\parallel,0}}{a_{1,\parallel,\mathrm{g}}} \tag{6-64}$$

As for the R.AT- scenario, the desired velocity is estimated based on the initial speed of the OOI for practical reasons so that $v_{1,\parallel,\mathrm{d}} = v_{2,0}$. Inserting these changes into (6-21) and (6-20) yields

$$v'_{2,\mathrm{d}} = v_{2,0} + a_\mathrm{max} t'_\mathrm{d} \tag{6-65}$$

$$\begin{aligned} d(t = t'_\mathrm{d}) = {} & r_{1,\parallel,0} - s_1 - s_2 + (v_{1,\parallel,0} - v_{2,0}) t'_\mathrm{d} \\ & + \frac{1}{2} (a_{1,\parallel,\mathrm{g}} - a_\mathrm{max}) {t'_\mathrm{d}}^2 \end{aligned} \tag{6-66}$$

By substituting t_d and $v_{2,\mathrm{r},\mathrm{d}}$ in (6-22) with these equations, the final requirement is:

$$\begin{aligned} 0 < d_\mathrm{min} = {} & d(t = t'_\mathrm{d}) + \frac{v_{1,\parallel,\mathrm{d}}^2}{2a_\mathrm{max}} - v'_{2,\mathrm{d}}\, t_\mathrm{r} \\ & - \frac{1}{2} a_\mathrm{max} t_\mathrm{r}^2 - \frac{(v'_{2,\mathrm{d}} + t_\mathrm{r} a_\mathrm{max})^2}{2a_{1,\parallel,\mathrm{g}}} \end{aligned} \tag{6-67}$$

An object violating this requirement may be impeded by the ego vehicle during a merging and is therefore relevant.

6.3 Results

In this section, the applied method from the previous section is implemented on a public dataset for the urban domain. The results are visualized to offer interpretability and plausibilization. Firstly, visual BEV images show exemplary results for urban scenes including applicable scenarios and their respective relevance. Secondly, the distributions of relevant objects across distances are analyzed.

6.3.1 Implementation

Results are presented for the popular nuScenes dataset[375] which includes object list annotations. More specifically, this work utilizes standard validation split of the nuScenes motion prediction task. As shown in later sections, the motion prediction task offers certain advantages for the validation of the results. The prediction task defines a subset of all annotated objects for which trajectories are predicted. For each of these objects, the corresponding surrounding objects used as input for the prediction are loaded. The relevance criteria of this work are then applied to these input objects.

The equations presented in previous sections contain parameters. Practical implementation and application of these equations requires specific parameter values. All values are selected to be applicable to common, realistic conditions as encountered in traffic. However, finding parameters which are applicable under all substances requires additional argumentation. The maximum acceleration is limited by the friction available on the road surface. Following previous work[376,377], a maximum acceleration of $a_{\mathrm{max}} = 10\ \frac{m}{s^2}$ is selected. For the braking deceleration, a previous study focusing on braking decelerations during emergency stops are applied. The braking deceleration is chosen to be $a_{\mathrm{b}} = 7\ \frac{m}{s^2}$ in accordance with minimum decelerations obtained even for wet road surfaces.[378] For the positive longitudinal acceleration, the value differs according to the type of vehicle as well as the velocity. Based on the results of Bokare and Maurya[379], an available acceleration of $a_{\mathrm{g}} = 0.5\ \frac{m}{s^2}$ is assumed. While lateral accelerations range between $1 - 2\ \frac{m}{s^2}$, the minimum system specification of this work models the ego vehicle with a single parameter $a_{1,\mathrm{g}}$ for both the longitudinal and lateral acceleration. Therefore, this parameter for the guaranteed acceleration is limited by the longitudinal direction. Another parameter of the vehicles is the reaction time. For human drivers, a common reaction time for a surprise intrusion is approximately 1.5 s.[380] For simplicity, these parameters are applied to all of the vehicles involved.

[375] Caesar, H. et al.: nuScenes (2020).

[376] Althoff, M.; Magdici, S.: Set-Based Prediction of Traffic Participants (2016), p. 197.

[377] Wachenfeld, W. et al.: The worst-time-to-collision metric (2016), p. 732.

[378] Greibe, P.: Braking distance, friction and behaviour (2007), p. 41-42.

[379] Bokare, P. S.; Maurya, A. K.: Acceleration-Deceleration Behaviour (2017), p. 4741-4743.

[380] Green, M.: How Long Does It Take to Stop? (2000), p. 213.

6.3.2 Relevance Visualization

The relevance method is applied to all objects which are part of the nuScenes motion prediction task. Results show exemplary visual results for the disambiguation and the relevance.

The disambiguation results show which functional scenario is applied to which objects. A visualization of an exemplary road scene in BEV is shown in Fig. 6-8. Each color indicates a functional scenario as distinguished by this work.

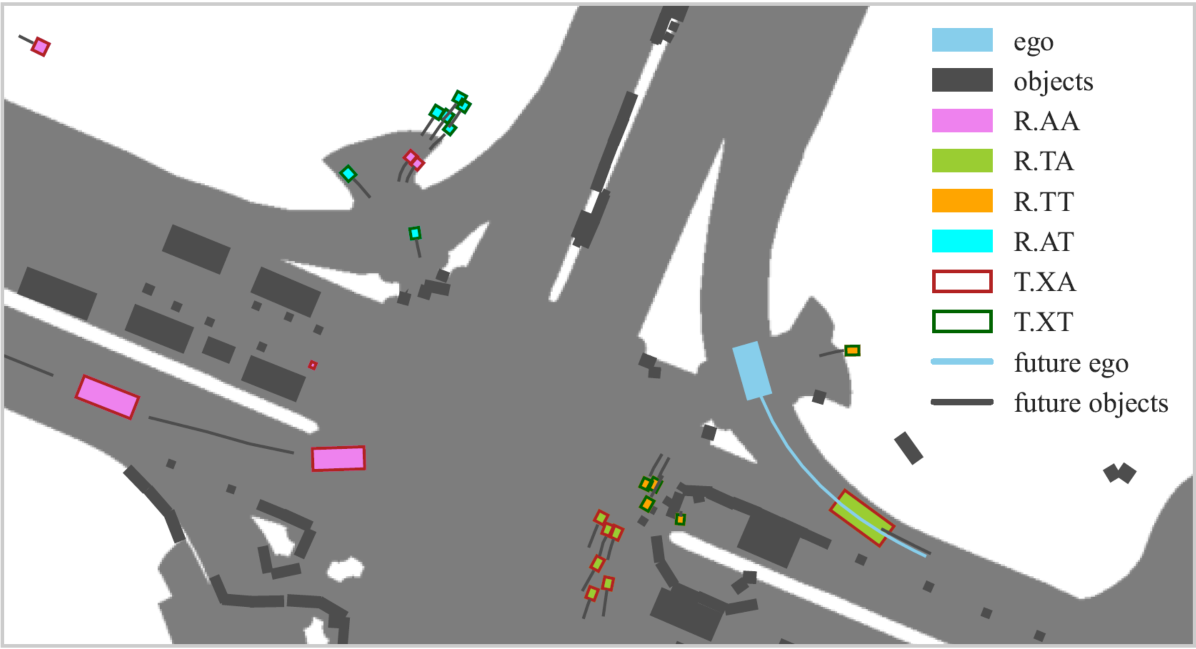

Figure 6-8: Visualization of applicable functional scenarios on nuScenes dataset. Since radial and tangential scenarios may overlap, tangential scenarios are indicated by edge colors while radial scenarios are indicated by fill colors of the rectangles.[381]

Since the traffic scene is recorded in Singapore, vehicles drive on the left side of the road. The ego vehicle and its future trajectory are depicted in light blue. The relevance of all other objects is evaluated with respect to this ego vehicle. Other objects are visualized as rectangles of different color. Firstly, all objects and their future trajectories are plotted in dark gray while dynamic objects received colored boxes. Therefore, the dark gray boxes visible in Fig. 6-8 show static objects. As previously indicated, this work leverages the functions provided by nuScenes to load objects surrounding the ego vehicle as input for predictions. While the results of this work are also applicable to static objects, these functions only load the dynamic objects. For consistency with the existing dataset pipeline as well as with the prediction results which will be shown in later chapters, only dynamic objects are evaluated in this section. Therefore, the scenario disambiguation is only applied to the dynamic objects displayed as colored rectangles in the image. Each color represents a different type of functional scenario. The fill color distinguishes

[381] This image was created together with Kai Storms and also appears in his work[382] created concurrently.

[382] Storms, K.: Context Aware Data Reduction (2023).

different radial scenarios while the edge color distinguishes different tangential scenarios. For instance, the magenta fill color indicates objects for which the R.AA scenario is considered. As shown in the image, the R.AA scenario is applied if the ego vehicle and an object move apart even if they are not moving in exact opposite directions. Objects for which the R.TA scenario is considered are depicted with a light green fill color. One example is the vehicle which is travelling in front of the ego vehicle on the same lane. However, the R.TA scenario also applies to the pedestrians visible as smaller rectangles in the lower part of the image. Similarly, orange and cyan fill colors indicate objects for which the R.TT and the R.AT scenario are applied, respectively. The tangential scenarios are distinguished according to the edge color of the rectangles. Objects for which the radial velocity component is pointing away from the ego vehicle are considered according to the T.XA scenario indicated by the red edge. The T.XT scenario indicated by the dark green edge applies to objects for which the radial velocity component is pointing towards the ego vehicle.

For the exemplary visualization, examples for each of the functional scenarios are shown. It is observed that even seemingly different constellations and traffic participants may be treated according to the same functional scenario as in the case of the R.TA scenario. In particular, this means that not all occurrences of a scenario represent the archetypal case. For example, pedestrians in different locations on the intersection are all evaluated according to T.XT. The reason is that neither the class information nor the road layout included in the image are considered for the relevance evaluation. Therefore, it is possible that the hypothetical scenarios considered are not always practically possible on the real road geometry. However, this is expected as the relevance method provides a conservative overestimation.

The relevance results in Fig. 6-9 show which objects are considered relevant without distinction of different functional scenarios. As in the previous visualization, the ego vehicle is depicted in light blue while static objects are presented as dark gray rectangles. The relevance evaluation is once again only conducted for dynamic objects or agents which are loaded as prediction input by the nuScenes functions. These dynamic agents are represented by the rectangles colored in red and blue. Red rectangles represent relevant agents while blue rectangles represent irrelevant agents. It should be noted that the distinction between static and dynamic objects is based on a velocity calculated by a differentiation of bounding box locations. However, inaccuracies in the location annotations of the objects may result in velocities larger than zero even for static object categories. An example of this is also visible in the upper right part of the image. The long rows of objects in the center of the street are static concrete barriers. One of these barriers is depicted in red since it is considered dynamic based on an inaccurate velocity annotation. However, these inaccuracies do not affect the results of this work since the relevance method is equally applicable to static and dynamic objects.

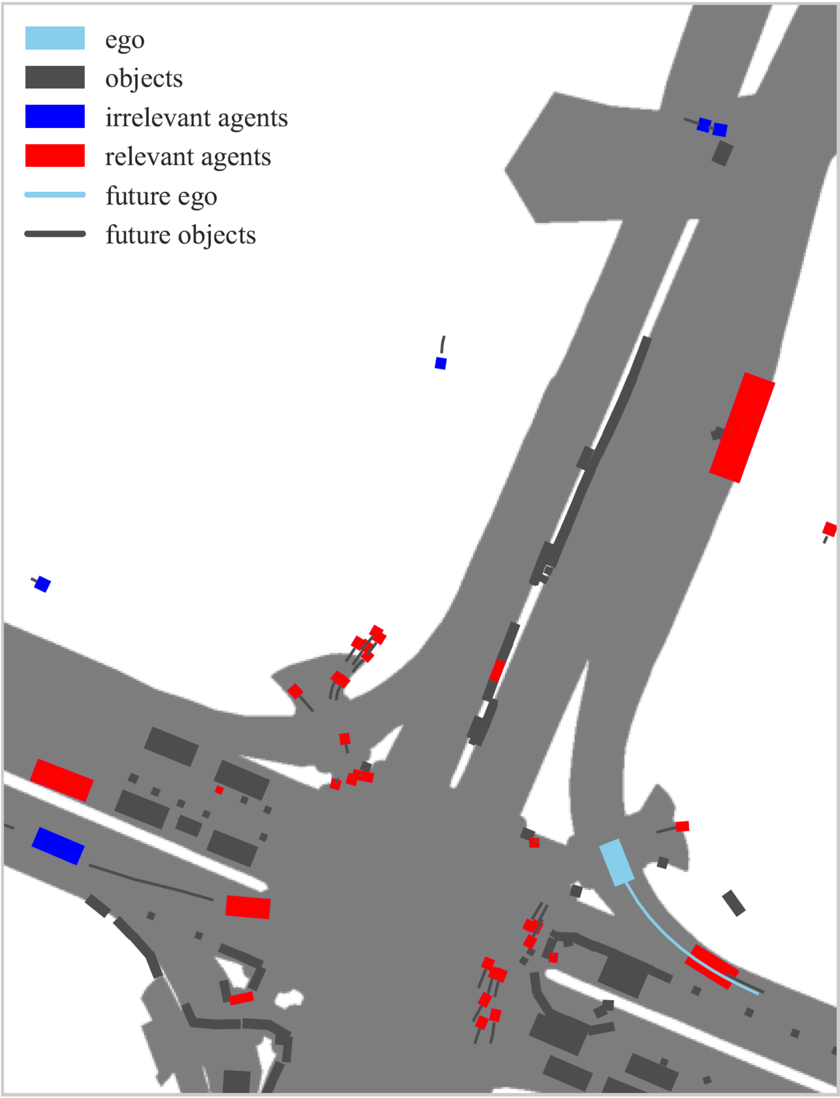

Figure 6-9: Visualizations of object relevance on an exemplary scene from nuScenes.[383]

Overall, it is observed that most of the agents depicted in the visualization are considered relevant. In particular, agents in the vicinity of the ego vehicle are all considered relevant. Irrelevant objects occur at large distances from the ego vehicle, often additionally moving away from the ego vehicle. The fact that objects moving towards the ego vehicle and nearby objects are considered relevant is plausible. While only few objects are considered irrelevant, this is expected due to the conservative design of the relevance criteria.

[383]This image was created together with Kai Storms and also appears in his work[384]created concurrently.

[384]Storms, K.: Context Aware Data Reduction (2023).

6.3.3 Distance Distributions

While the exemplary visual results provide some insight into the relevance criteria, this section analyzes the whole dataset. The objective is to provide a quantitative analysis of the relevance results. To this end, the number and distribution of relevant objects across distance are considered.

Firstly, an analysis is conducted to quantify the number of objects considered irrelevant. In accordance with the conservative approach of this work, only 10% of objects are considered irrelevant. This also agrees with the qualitative results presented in the previous section.

Besides the number of the relevant objects, the spatial distribution of the objects is considered in the following. This serves to further analyze the qualitative observation from the previous section that irrelevant objects only occur at large distances. For this purpose, the distributions of objects over the distance from the ego vehicle are presented as ECDFs. A visual overview is provided in Fig. 6-10. The image depicts the distance distributions of different types of objects as different ECDFs. Each curve shows the proportion of objects below a specific distance value. For example, the cyan curve for the objects considered relevant according to the R.AT scenario reaches a value of approximately 50% at a distance of 25 m. This means that the median for the distance of objects relevant according to the R.AT scenario is 25 m.

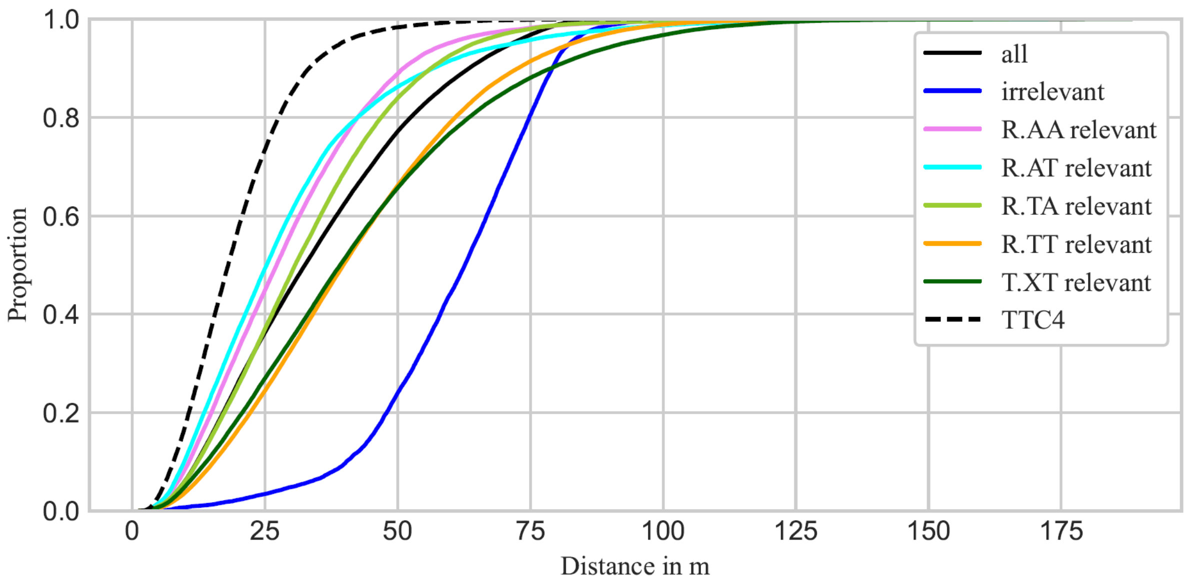

Figure 6-10: ECDFs of distances for all objects and objects which are considered relevant according to different criteria. The distribution for a typical suggested TTC is provided as reference.[385]

[385]This image was created together with Kai Storms and also appears in his work[386]created concurrently.

[386]Storms, K.: Context Aware Data Reduction (2023).

The distance distribution of all dynamic objects is depicted by the solid black line while the irrelevant objects are shown by the dark blue line. It is observed that irrelevant objects mainly occur at distances above 50 m. Furthermore, the distribution of irrelevant objects is further to the right than for all other distributions. This means that irrelevant objects occur at larger distances than relevant objects. Therefore, the quantitative results corroborate the qualitative observation from the previous section. However, the distance distributions also show significant overlap in distance between relevant and irrelevant objects. Since the applicable scenarios and the velocities influence relevance, there is no single distance threshold that separates relevant from irrelevant objects. The other distributions show the distances for objects considered relevant by each of the functional scenarios. Overall, the results for different scenarios appear visually similar. All scenarios show relevant objects across a range of distances. The highest distance values are observed for the T.XT scenario which yields relevant objects even for distances above 100 m.

Since the distance distributions are difficult to interpret, a TTC threshold as a common criticality metric is provided as reference. The TTC thresholds vary for different sources and scenarios, but is generally chosen to be within 4 s.[387] Accordingly, an arbitrary value of 4 s is chosen for this work. Another ECDF for the distance values is shown as dashed black line in Fig. 6-10 for all objects which fall within the selected TTC threshold. The dashed line is further on the left than the other distributions, indicating that TTCs below the threshold occur at smaller distances. Conversely, this means that the relevance results of this work consider objects to be relevant at higher distances than when using a TTC threshold. Again, this is in line with the objective of utilizing conservative assumptions for the relevance method of this work.

[387] Mahmud, S. S. et al.: Application of proximal surrogate indicators (2017), p. 155-156.

6.4 Conclusion

This chapter showed the proposed approach to identify relevant objects for the task of object detection. A structured method which first specified the system as well as the use case and then derived relevance was presented.

Relevance was hereby defined based on behavioral requirements for the context of driving. Behavioral requirements were obtained by considering the two common principles of legal and safety requirements. The driving task was decomposed into different interpretable functional scenarios, for which relevance was then derived. The relevance concept only requires the object list and a minimum system specification. Most importantly, no information regarding intentions, future behaviors or road geometry is required. Instead, these factors were considered as uncertainties for which conservative estimates were provided. Applying the relevance method yielded interpretable equations for each functional scenario which distinguish relevant from irrelevant objects. Since relevance is fully specified by the common principles of interpretability, legal and safety requirements, consideration of the human baseline was not required.

Visualizations of the results were provided for the different functional scenarios as well as for the resulting relevance. In each case, the visualizations showed that the results appear plausible. As expected due to the conservative relevance method, only 10% of the objects are considered irrelevant. While irrelevant objects mainly occur at larger distances, distance alone is not a sufficient criterion to distinguish irrelevant objects. Relevance results are more conservative than a TTC threshold of 4 s obtained from literature[387] with some scenario types yielding relevant objects beyond 100 m. The implications of these relevance results are presented along with a full discussion in section 9.3.2. There, a joint discussion including the validation results obtained in chapter 8 is provided to answer the RQ2.

7 Attributes

Once the relevant object classes and instances are identified, the question is which requirements apply for their attributes. The corresponding research question is RQ 1.3:

Is it possible to define requirements for the detected object attributes considering the safety of the driving task?

Most of the content in this chapter was previously included in a publication coauthored by the author[388]. However, this chapter adapts and extends the previously published results. Firstly, the overall method to derive attribute requirements is outlined. This is followed by the application of the method for different attributes. Next, results for applying these criteria to the detection output of popular detection baselines are presented. Finally, a conclusion of this chapter is provided. As with previous chapters, the joint discussion is presented in chapter 9.

7.1 Method

This section outlines the approach to derive accuracy requirements for different attributes considered in the object detection task. Overall, road regulations and safety norms do not provide specific requirements including thresholds for object detection. Therefore, the remaining common principles of interpretability and human baseline are emphasized in this section.

In order to provide interpretable criteria, different attributes are considered separately in this work. Attributes are limited to those commonly evaluated on perception benchmarks. While this list may not be exhaustive, it is expected to provide a reasonable baseline which considers the driving context. For completeness, both the object detection and the related tracking task are considered and reconciled. For each of these attributes, the human performance is estimated. In accordance with safety considerations, a conservative estimate of the upper bound on human performance is provided. If there is uncertainty regarding the human performance, the highest performance is used as requirement. This ensures that any perception system which fulfils the requirement is at least as good as the average human. Furthermore, errors for some attributes differ with respect to their safety consequences. For instance, underestimating a distance may not be critical while overestimating the object distance may lead to an accident. In this case, the unsafe error direction is prohibited while the other errors are allowed within human performance bounds. All errors refer to L1 distances of the GT value and the value of the detection. The criteria are formulated as simple analytic formulas which remain fully interpretable.

[388] Mori, K. T.; Peters, S.: SHARD (2023) © 2023 IEEE.

7.2 Method Application

This section presents results for applying the high-level method described in the previous section. The attributes are structured into different aspects. Firstly, the process of associating the detections with the GT objects is outlined. Next, anything related to tracking or the temporal aspects of the detection task is considered. This is followed by the requirements for attributes which are considered relevant for collisions. Finally, further attributes which are not considered collision relevant are discussed. Each aspect contains different interpretable attributes for which requirements are developed.

7.2.1 Association

When evaluating object detection, the first step is associating detected objects with GT annotations[389] which is discussed in the following. Attributes relating to the association include classification, confidence, the choice of reference point as well as the procedure and thresholds used for the association.

The first question is if the common paradigm of location based association of objects is reasonable. Studies on human perception mechanisms indicate that visual working memory indeed uses an object-centric representation including various corresponding features.[390] Furthermore, evidence suggest that features are bound together by a location map.[391] While errors during the combination of features are possible,[392,393] these problems in in binding features typically only occur under specific conditions such as very brief presentations[394]. Therefore, the general process of associating integrated object hypotheses is supported by the human baseline.

Once an association between detections and GT is defined, this also defines FNs and FPs.[389] Assuming a pairwise matching, detected objects which correspond to a GT object are considered a TP. On the other hand, GT objects which are not matched with a detection are considered a FN. Conversely, a detection without corresponding GT object is a FP.[395] However, the details of the association or matching procedure require further discussion.

[389] Hoss, M. et al.: A Review of Testing Object-Based Environment Perception (2022), p. 229.
[390] Luck, S. J.; Vogel, E. K.: The capacity of visual working memory (1997), p. 279-280.
[391] Vul, E. et al.: The structure of illusory conjunctions (2020), p. 550-551.
[392] Treisman, A.; Schmidt, H.: Illusory conjunctions in the perception of objects (1982), p. 138-139.
[393] Bulakowski, P. F. et al.: Independent coding of object motion and position (2007), p. 815-816.
[394] Wolfe, J. M.; Cave, K. R.: Evidence for a Binding Problem in Human Vision (1999), p. 16.
[395] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3358.

Association and Classification

Most commonly, the association procedure is performed individually for each class.[396] This section challenges this paradigm and discusses whether the class of an object should be considered for association.

For instance, safely avoiding obstacles requires their detection irrespective of their class. Class-agnostic detection may therefore be better suited for obstacles.[397] Human perception mechanisms also appear to integrate various different features into object hypotheses.[398] Additionally, it is necessary to consider the fact that classes may be unknown or difficult to separate clearly.[399] Similar concepts are found in different object detectors which leverage a class-agnostic objectness property.[400,401,402,403,404] Considering these factors, an association irrespective of the predicted class is favored in this work.

Confidence

One attribute which is often contained in an object list is the confidence score. This score is commonly provided by object detectors and also relied upon by the typical mAP metric.[405]

However, there is no clear analogy for confidence scores in human perception. In fact, typical dataset annotations implicitly acknowledge this, since the human annotations do not include confidence scores. In addition, the usage of confidence scores is inconsistent between tasks with semantic segmentation not providing confidence.[406] Rather than requiring confidence scores, a single object list without confidence scores is assumed as detection output within this work. Failures are considered with respect to this object list. If confidence scores are present, there are various conceivable methods how to obtain a single object list. Following previous work,[407,408] the confidence is optimized with regard to the final metric. This approach utilizes the optimal operating point of the detector for the proposed metrics.

---

[396] Liu, L. et al.: Deep Learning for Generic Object Detection: A Survey (2020), p. 271.

[397] Jaiswal, A. et al.: Class-agnostic Object Detection (2021), p. 919-920.

[398] Vul, E. et al.: The structure of illusory conjunctions (2020), p. 550-551.

[399] Chan, R. et al.: SegmentMeIfYouCan (2021), p. 1-2.

[400] Wu, C. et al.: A hierarchical loss (2019), p. 13-14.

[401] Singh, B. et al.: R-FCN-3000 at 30fps: Decoupling Detection and Classification (2018), p. 1087-1088.

[402] Redmon, J.; Farhadi, A.: YOLOv3: An Incremental Improvement (2018), p. 1-2.

[403] Yin, T. et al.: Center-based 3D Object Detection and Tracking (2021), p. 11782.

[404] Yang, Z. et al.: IPOD: Intensive Point-based Object Detector for Point Cloud (2018), p. 3.

[405] Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010), p. 313.

[406] Kirillov, A. et al.: Panoptic Segmentation (2019), p. 9398.

[407] Oksuz, K. et al.: One Metric to Measure them All (2020), p. 9452-9453.

[408] Guo, Y. et al.: CS-R-FCN: Cross-supervised Learning for Large-Scale Object Detection (2020), p. 1.

Object Reference

When associating detected and GT objects, an object reference is required. The reference may either utilize the entire bounding box representation of the object or a single reference point.

Considering the whole object as done by IoU couples location, size and orientation.[409] This coupling is undesired since it reduces interpretability. In addition, accurately estimating the size of an object may be difficult.[410] Therefore, a reference point is used as object reference in this work. Generally, different potential reference points are available. While it is possible to use object centers as reference points[409], the minimum distance to the other vehicle is the safety relevant attribute[411]. The difference between the two possible reference points is visualized for an exemplary following scenario in Fig. 7-1.

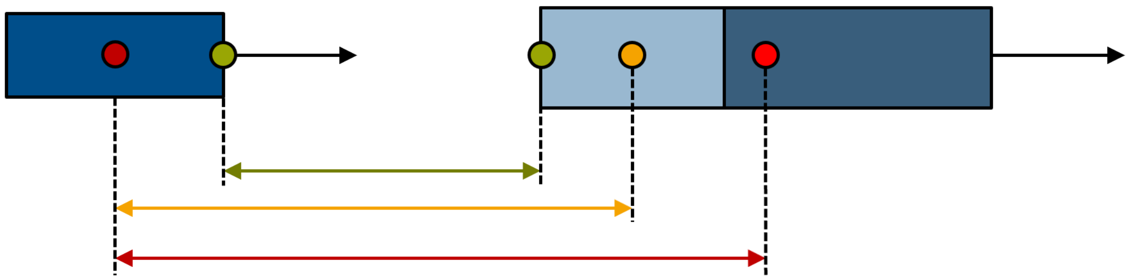

Figure 7-1: Difference between using closest point (green) and center point (yellow/red) as reference.[412]

For different vehicle lengths, the center points shown in yellow and red differ. However, the closest point depicted in green remains the same. The closest point corresponds to the available distance during an emergency brake. Therefore, the closest point better reflects the principle of safety as also argued in literature[413] and in the previous chapter 6 relating to relevance. In addition, objects such as humans have a geometry which is not closely approximated by a rectangular bounding box. For a pedestrians whose arms are both stretched forward, the bounding box center may be outside the pedestrian. In such cases, the closest point provides better interpretability than the center of an arbitrarily defined rectangular bounding box. Using the closest point decouples the reference point estimation from the size estimation. Unlike the center point, the closest point is on the object surface which is visible in sensor data.[414,415,416] Therefore, surface points such

409 Caesar, H. et al.: nuScenes (2020), p. 11622.

410 Wang, Y. et al.: Train in Germany, Test in The USA (2020), p. 11714-11715.

411 Zhao, H. et al.: Suraksha: A Quantitative AV Safety Evaluation Framework (2021), p. 36.

412 Mori, K. T.; Peters, S.: SHARD (2023), Fig. 1 © 2023 IEEE.

413 Wachenfeld, W. et al.: The worst-time-to-collision metric (2016), p. 730-731.

414 Qi, C. R. et al.: Deep Hough Voting for 3D Object Detection in Point Clouds (2019), p. 2.

415 Chen, Q. et al.: Every View Counts (2020), p. 68-70.

416 Wang, Y. et al.: 1st Place Solutions for Waymo Open Dataset Challenges - 2D and 3D Tracking (2020), p. 2-5.

as object face centers[417] or closest corners[418] are more suitable for the localization of objects. Accordingly, the closest point is selected as reference in this work due to its advantages regarding interpretation, detection and safety.

Association Procedure

Within this section, the procedure of associating two object hypotheses is discussed in detail. The questions answered are which metrics and methods to use for the association.

As previously indicated, the human analogy suggests a location based association.[419] Common datasets also use location for the association of objects.[420,421,422] Therefore, this approach is also followed in this work. Common metrics for location based association are 3D IoU as applied by KITTI[420] and center point distance as applied by nuScenes[421]. As argued in the section 7.2.1 regarding a reference point, IoU entangles the location and the size estimation. Therefore, the Euclidean distance of the reference points is utilized. Note that the reference point applied in this work is the point closest to the ego vehicle, not the center point. As presented in Fig. 7-2, the matching is based on the closest distance d of the GT reference point to the perceived object.

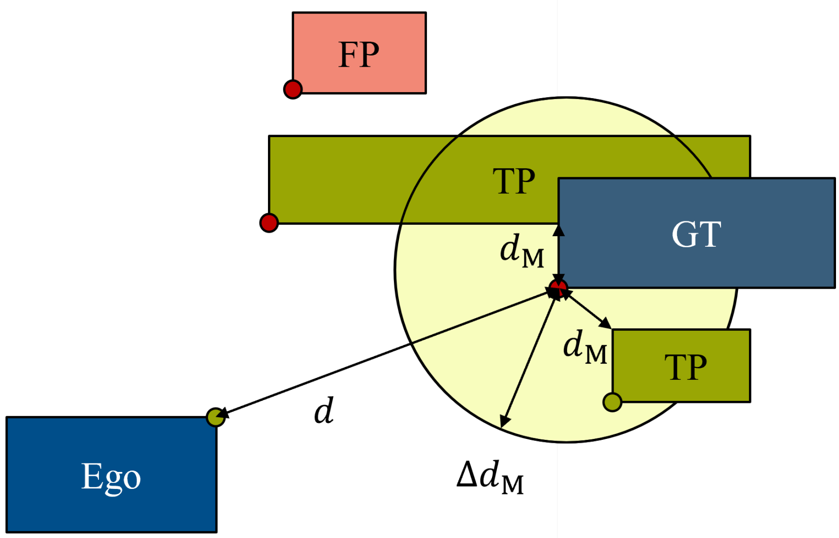

Figure 7-2: Association based on circular distance threshold r_{M} around the closest point of GT. Green objects are potential matching candidates while red objects are beyond the matching threshold.[423]

[417] Yin, T. et al.: Center-based 3D Object Detection and Tracking (2021), p. 11782.

[418] Yang, B. et al.: Auto4D: Learning to Label 4D Objects (2021), p. 4.

[419] Vul, E. et al.: The structure of illusory conjunctions (2020), p. 550-551.

[420] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3358.

[421] Caesar, H. et al.: nuScenes (2020), p. 11622.

[422] Sun, P. et al.: Scalability in Perception (2020), p. 2447.

[423] Mori, K. T.; Peters, S.: SHARD (2023), Fig. 2 © 2023 IEEE.

Once the metric for association is selected, an algorithm for defining the optimal matching is required. Optimization can either be performed by iteratively associating in a greedy manner or by performing global optimization with bipartite matching using the Hungarian algorithm.[424] Greedy optimization is applied by KITTI[420] and nuScenes[421] while the Waymo Open Dataset uses the Hungarian algorithm[422]. Previous work observed that for tracking, the greedy algorithm obtains better results for distance based matching. Outliers were suggested as potential reason for this outcome.[424] In order to verify this hypothesis, three matching procedures including a modified Hungarian algorithm in addition to the greedy and the Hungarian algorithm are implemented. The modified Hungarian algorithm is based on the observation that matching uses a discrete association threshold. For distances larger than the threshold, optimization does not yield meaningful matching results and is therefore detrimental. Therefore, this work proposes to set all distances in the cost matrix which are larger than the association threshold to an equally large value. The value is chosen to be higher than typically occurring distances to ensure that unmatched objects incur a high cost. As shown in the results in section 7.3.2, the modified global optimization and the greedy matching show similar performance. Both methods outperform the standard global optimization. For simplicity, the greedy matching is applied for all other results in this work.

Association Threshold

In addition to the metric, association also requires a maximum threshold. Generally, the border between the criteria used for association and the localization criteria is fuzzy.[425] Nevertheless, this principle corresponds well to existing object detectors.[426] In addition, treating existence and localization separately conforms to the principle of interpretability.

Since the distinction between existence and localization is essentially arbitrary, a simple circular distance threshold is proposed. The approach is visualized in Fig. 7-2. The threshold is chosen to be compatible with the localization criteria of this work. Contrary to prior work, the permissible location errors depend on the direction of the error in this work. Therefore, a circular threshold based on the most lenient localization threshold is selected. For the purpose of this section, the results of the later section 7.2.3 are preempted. Here, only the results are applied while details regarding the method of obtaining these results are left for section 7.2.3. As will be shown later, the most lenient criterion is applied for the distance d between the GT object and the ego vehicle. The corresponding distance accuracy requirement is $0.15 \cdot d$ which scales with the distance. Accordingly, the circular threshold for the matching is chosen as $\Delta d_{\mathrm{M}} = 0.15 \cdot d$ which also scales with the distance. However, scaling with the distance introduces very strict requirements at small distances. To avoid overly severe requirements for matching, a minimum permissible offset is introduced. The value for this minimum permissible object is chosen as 2 m in accordance with

[424] Pang, Z. et al.: SimpleTrack (2023), p. 5.

[425] Hoss, M. et al.: A Review of Testing Object-Based Environment Perception (2022), p. 229-230.

[426] Zhou, X. et al.: Probabilistic two-stage detection (2021), p. 1-3.

prior work.[427,428] To summarize, the radius for matching is defined as $\Delta d_{\mathrm{M}} = \max(0.15 \cdot d, 2\,\mathrm{m})$, where d is the egocentric distance of the object. A detected object is associated with a GT object if the matching distance $d_{\mathrm{M}} < \Delta d_{\mathrm{M}}$.

Note that both matching and localization failures are present in the final evaluation. Therefore, their arbitrary distinction introduced here is inconsequential for the overall number of failures.

7.2.2 Tracking

The tracking task typically imposes additional requirements when compared to the detection task. In this section, all aspects related to the task of tracking are explicitly discussed. This includes identifier switches as well as the temporal distribution of failures.

Identifier Switches

The popular CLEAR MOT metrics[429] as well as other tracking metrics[430] also identify the so called identifier switch. An identifier switch is a failure mode unique to tracking across sequences, where a correct detection receives the wrong identifier.[431] While this unique identification may be required for surveillance tasks, it is not obvious if it is required for the task of driving.

Experiments regarding human perception show that interruptions may cause large changes to remain unnoticed. Examples for such interruptions include saccades, blank images, mud splashes and cuts or pans in motion pictures.[432] Notably, this phenomenon occurs in real-world settings and is not negated by attention.[433] Overall, these findings demonstrate that humans are unable to reliably perform unique identification of objects. Since humans are able to drive despite these shortcomings, unique identification is likely not required for the driving task. Therefore, no unique identification is required in this work and identifier switches are not evaluated.

Tracking Accuracy

Another aspect which is typically only considered for tracking is the temporal sequence of failures. While detection typically neglects the aspect of detection times, the aspect is safety relevant for driving.[434] However, no requirements are available in literature without presupposing a given frame rate or annotation frequency.

---

[427] Caesar, H. et al.: nuScenes (2020), p. 11622.

[428] Ge, R. et al.: AFDet: Anchor Free One Stage 3D Object Detection (2020), p. 6-7.

[429] Bernardin, K.; Stiefelhagen, R.: Evaluating Multiple Object Tracking Performance (2008), p. 2-5.

[430] Song, B. et al.: A Stochastic Graph Evolution Framework for Robust Multi-target Tracking (2010), p. 615.

[431] Yamaguchi, K. et al.: Who are you with and where are you going? (2011), p. 1345-1352.

[432] Simons, D. J.: Current Approaches to Change Blindness (2000), p.1-2.

[433] Simons, D. J.; Levin, D. T.: Failure to detect changes to people during a real-world interaction (1998), p. 646-648.

[434] Volk, G. et al.: A Comprehensive Safety Metric (2020), p. 3.

Even for the tracking task, the temporal sequence is not always considered. For instance, it is neglected entirely by the popular MOTA metric.[435] Previous metrics conceived different alternatives for including temporal requirements. For instance, the detection time can be considered explicitly by using it to re-weight the perception scores.[436a, 434] An alternative is directly evaluating the time required for detection.[437, 436b] Another option is to implicitly consider the temporal sequence by evaluating the completeness of a track.[438] While different metrics are available, safety requires a direct evaluation of the detection time. Therefore, detection is required within a specified maximum detection time in this work.

The maximum detection time is specified by considering the human baseline. Perception times are distinct from reaction times, but can be measured directly using event-related potentials. The perception time for the recognition of the presence of different object categories is approximately 150 ms.[439,440] Related tasks such as detecting a change in a geometric constellation requires approximately 200 ms.[441] The time required to perceive motion depends on the type of motion and ranges between 160-200 ms.[442] Overall, the human baseline indicates that an object may be perceived as fast as in 150 ms.

Therefore, failures due to detection delays are permitted within the first 150 ms in this work. This includes FNs immediately after an object appears and FPs immediately after an object disappears. Within a track, no FN or FP failures are permitted since no equivalent is identified for human perception. While the number of objects tracked by humans is limited[443,444], this is not considered in this work. Since it is difficult to determine which objects humans are tracking, it is conservatively assumed all relevant objects require detection. It should also be noted that current dataset annotation frequencies are lower than this human perception latency. This means that no failures are permissible for the annotated frames.

7.2.3 Collision Relevant Attributes

This section considers the collision relevant attributes of an object. More specifically, localization and velocity are each treated separately. Localization is split into the two aspects of distance to the ego vehicle and the angular positions due to correspondence with human perception. Velocity

---

[435] Bernardin, K.; Stiefelhagen, R.: Evaluating Multiple Object Tracking Performance (2008), p. 5.

[436] Kim, K.-Y. et al.: Real-Time Performance Evaluation Metrics (2016), a: p. 177, b: p. 175-176.

[437] Caesar, H. et al.: nuScenes (2020), p. 11623.

[438] Li, Y. et al.: Learning to associate: HybridBoosted multi-target tracker for crowded scene (2019), p. 2958.

[439] Thorpe, S. et al.: Speed of processing in the human visual system (1996), p. 520.

[440] van Rullen, R.; Thorpe, S. J.: The time course of visual processing (2001), p. 456-459.

[441] Koivisto, M.; Revonsuo, A.: An ERP study of change detection (2003), p. 424-427.

[442] Kuba, M. et al.: Motion-onset VEPs: characteristics, methods, and diagnostic use (2007), p. 189-192.

[443] Pylyshyn, Z. W.; Storm, R. W.: Tracking multiple independent targets (1988), p. 182-185.

[444] Alvarez, G. A.; Franconeri, S. L.: How many objects can you track? (2007), p. 1-4.

is split into radial and angular velocity due to safety considerations. Angular velocity determines the possibility, while the radial velocity determines the timing of a collision. Note that size and angle are not part of the collision relevant attributes due to choosing the closest point as reference.

Distance

As argued with regards to the reference point, the distance of the closest point corresponds best to driving safety.

Distance errors are well studied in perception literature. Experiments generally utilize static setups in open spaces, many of which are unrelated to the context of driving. Participants are paying full attention and receive sufficient time to produce their estimates. Distance errors consist of random and systematic error components which also depend on the distance.[445] Generally, egocentric distance between an object and the observer and exocentric distance between two objects are distinguished. Driving generally requires egocentric distance estimation. However, the human estimation error is similar for both egocentric and exocentric distances.[446a] Therefore, available information for exocentric distance information is also incorporated in this work. Analyzing errors for interobject distances presented by Levin and Haber[446b] shows that the maximum of the standard deviation and mean errors is always $> 15\%$. Daum and Hecht observe distance estimation errors of $> 25\%$ distances beyond 54 m.[445] Moeller et al. show the influence of the driving context for distances under 20 m. Here, even higher relative errors of approximately 40% are observed.[447] In a roadside environment, an analysis of the results by Strauss and Carnahan[448] yields errors of 23%. If annotators are directly asked to label image depth from monocular image, relative errors of $> 20\%$ are obtained.[449] Overall, human distance estimations are no more accurate than 15% relative distance error.

Conservative estimations allow underestimating the distance. Overestimating the distance is not permissible, since this may cause accidents. Considering both the human baseline and conservative estimates yields:

$$\Delta d = d_{\mathrm{GT}} - d_{\mathrm{PRED}} \leq 0.15 \cdot d \tag{7-1}$$

If accidents are avoided, the distance to an object always remains larger than zero. This ensures that the requirement never requires an error of zero.

[445] Daum, O. S.; Hecht, H.: Distance estimation in vista space (2009), p. 1130-1136.
[446] Levin, C. A.; Haber, R. N.: Perceived interobject distance (1993), a: p. 253, b: p. 259.
[447] Moeller, B. et al.: What a car does to your perception (2016), p. 783-785.
[448] Strauss, M.; Carnahan, J.: Distance Estimation Error in a Roadway Setting (2009), p. 249-253.
[449] Wu, Y. et al.: Size-to-depth (2018), p. 1-2.

Angular Position

Positions can be evaluated as either lateral distances[450,451] or as viewing angles. Since the human performance is best quantified in terms of the visual angle[452], visual angles are used in this work.

Results from literature indicate that the human accuracy is unaffected by different viewing conditions including fixed gaze, fixed head and free head.[453a] Since the gaze and head orientation are unknown on public dataset, the heading of the car is used in this work. Due to the negligible effect of viewing conditions this is considered feasible. Furthermore, results are consistent across different angular estimation methods such as numeric estimates and nonverbal measures.[453b] Therefore, available data is considered regardless of the estimation method. In order to ensure conservative estimates, the object point with the smallest visual angle is used as reference. Note that this point may differ from the closest point used for distance estimates. This difference may be significant especially when considering large objects. Using the smallest angle is conservative since smaller angles are typically closer to the future ego trajectory. Accordingly, smaller visual angles correspond to a closer proximity to potential accidents. Generally, human visual angle estimation exhibits both systematic and random errors. Across a range of visual angles between $3° - 80°$, both horizontal and vertical random errors are $> 5°$.[453c,454]

To ensure conservative estimates, the visual angle may be underestimated, but not overestimated. Considering the human baseline, a constant permissible error of $5°$ is selected. For simplicity, this error is applied for both azimuth and elevation regardless of the value of the angle. Fig. 7-3 provides a visual representation of the angular errors.

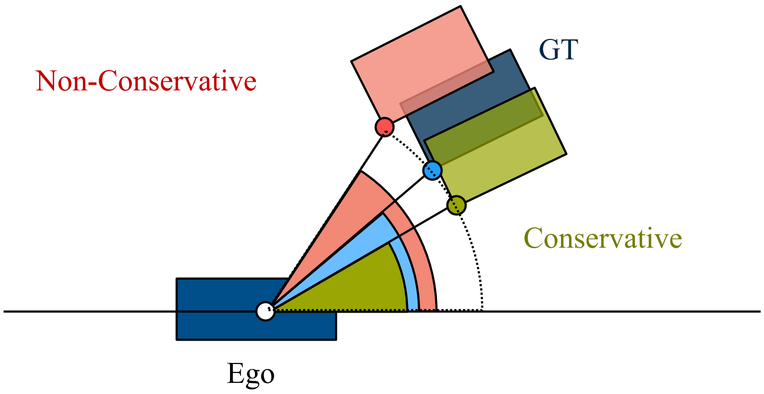

Figure 7-3: Angular position referencing ego heading and conservative estimates.[455]

[450] Sontges, S. et al.: Worst-case Analysis of the Time-To-React Using Reachable Sets (2018), p. 1891.

[451] Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017), p. 10-11.

[452] Levin, C. A.; Haber, R. N.: Perceived interobject distance (1993), p. 252-253.

[453] Li, Z.; Durgin, F. H.: Perceived azimuth direction is exaggerated (2016), a: p. 6-7, b: p. 7, c: p. 7-15.

[454] Higashiyama, A.: Anisotropic perception of visual angle (1992), p. 225-226.

[455] Mori, K. T.; Peters, S.: SHARD (2023), Fig. 3 © 2023 IEEE.

Radial Velocity

Radial velocity is considered based on the TTC in this work. One reason is the correspondence of the TTC to the proximity of a collision. In addition, humans also estimate the TTC through the expansion rate of the image rather than estimating speed.[456]

When considering human estimation performance, interceptive movements show high temporal precision.[457] However, this accuracy is based on continuous adjustments to the estimate rather than an initial accurate estimate.[458] Therefore, this accuracy does not transfer to the driving context where the objective is avoiding collisions. For a following scenario, estimations inside the moving car show errors of at least 30%.[459,460] Other experiments show standard deviations of at least 10% for TTC estimations with different velocities.[461] For velocities approaching zero, humans exhibit a threshold for perceiving looming which can be quantified as $\mathrm{iTTC}_{\mathrm{low}} = 0.2\frac{1}{\mathrm{s}}$ for naturalistic conditions such as braking.[462]

This inverse time to collision (iTTC) is defined as $1/\mathrm{TTC} = \mathrm{iTTC} = v/d$. The reason for applying the iTTC is to avoid unbounded values for the TTC which occur for velocities of zero. Since the distance never reaches zero as long as accidents are avoided, the values of the iTTC remain bounded. For calculation of the iTTC, the distance of the GT object rather than the detected object is used. This allows separating the velocity estimation error from the distance error. The data from literature[459,461] yields human iTTC estimation errors of $> 10\%$. Conservative estimates demand that the iTTC is never underestimated. This requirement is applicable for positive and negative radial velocities. Adding the human perception threshold $\mathrm{iTTC}_{\mathrm{low}}$ to the relative errors and including the requirement of a conservative estimate yields the permissible error:

$$\Delta\mathrm{iTTC} \leq 10\% \cdot \mathrm{iTTC} + \mathrm{iTTC}_{\mathrm{low}} \tag{7-2}$$

Angular Velocity

The angular velocity indicates the possibility of a collision. Collisions are only possible for sufficiently low angular velocities.

Compared to the previously discussed attributes, angular velocity estimation experiments are generally conducted under less naturalistic conditions. Experimental setups involve estimating velocities of gratings or points on a screen. Human performance for estimating angular velocity

[456] Yan, J.-J. et al.: Visual processing of the impending collision (2011), p. 2,19.

[457] Brenner, E.; Smeets, J. B. J.: How people achieve their amazing temporal precision in interception (2015), p. 16-17.

[458] Brenner, E.; Smeets, J. B. J.: Continuous visual control of interception (2011), p. 490-491.

[459] Cavallo, V.; Laurent, M.: Visual Information and Skill Level in Time-To-Collision Estimation (1988), p. 626.

[460] Kiefer, R. J. et al.: Time-to-collision judgments under realistic driving conditions (2006), p. 339-342.

[461] Hoffmann, E. R.; Mortimer, R. G.: Drivers' estimates of time to collision (1994), p. 514-516.

[462] Markkula, G. et al.: A farewell to brake reaction times? (2016), p. 17-18.

depends on factors such as luminance[463] or velocity[464]. Nevertheless, errors above 5% are observed for various settings.[463,464,465,466] For angular velocities near zero, human perception shows a threshold for motion detection at $\dot{\Theta}_{\text{low}} = 0.03°/s$ for various settings.[467,468]

Since low angular velocities indicate collisions, underestimation is permissible while overestimation is not. Considering this conservative estimate as well as the human accuracy, the permissible error for angular velocities is:

$$\Delta\dot{\Theta} \leq 5\% \cdot \dot{\Theta} + \dot{\Theta}_{\text{low}} \tag{7-3}$$

7.2.4 Non Collision Relevant Attributes

Collision avoidance is generally possible when considering the attributes discussed previously. However, the attributes of size and orientation are typically considered by benchmarks[469,470] and may provide additional useful cues. An additional aspect is the object class which was previously only discussed in relation to existence.

Size

Object sizes are typically considered as length, width and height of a bounding box. While size is a relatively constant object property which may be inferred using priors[471], this may lead to degraded performance in other domains[472].

Experimental conditions for size estimation are diverse and include naturalistic settings in open spaces, monocular images and monitors with disparity available. However, the results show general agreement across different experimental conditions. The accuracy of size estimation does not differ significantly between different distant ranges such as 0-50 m and 50-100 m.[471] This phenomenon known as size constancy is shown for distance ranges of up to 700 m and depends on the availability of context and distance cues.[473]. In naturalistic driving conditions these cues are generally available. Studies show relative errors of above 6% for screens with disparity

[463] Takeuchi, T.; Valois, K. K.: Velocity discrimination in scotopic vision (2000), p. 2012-2017.
[464] Bruyn, B.; Orban, G. A.: Human velocity and direction discrimination (1988), p. 1327.
[465] Chen, Y. et al.: The precision of velocity discrimination across spatial frequency (1998), p. 1330-1331.
[466] Haarmeier, T.; Thier, P.: Detection of speed changes during pursuit eye movements (2006), p. 345-346.
[467] Snowden, R. J.; Kavanagh, E.: Motion Perception in the Ageing Visual System (2006), p. 11-12.
[468] Murakami, I.: Correlations between fixation stability and visual motion sensitivity (2004), p. 754.
[469] Sun, P. et al.: Scalability in Perception (2020), p. 2447.
[470] Caesar, H. et al.: nuScenes (2020), p. 11622-11623.
[471] Haber, R. N.; Levin, C. A.: The independence of size perception and distance perception (2001), p. 1140-1141.
[472] Wang, Y. et al.: Train in Germany, Test in The USA (2020), p. 11715-11717.
[473] Zalevski, A. et al.: Size Estimation with Night Vision Goggles (2001), p. 2-4.

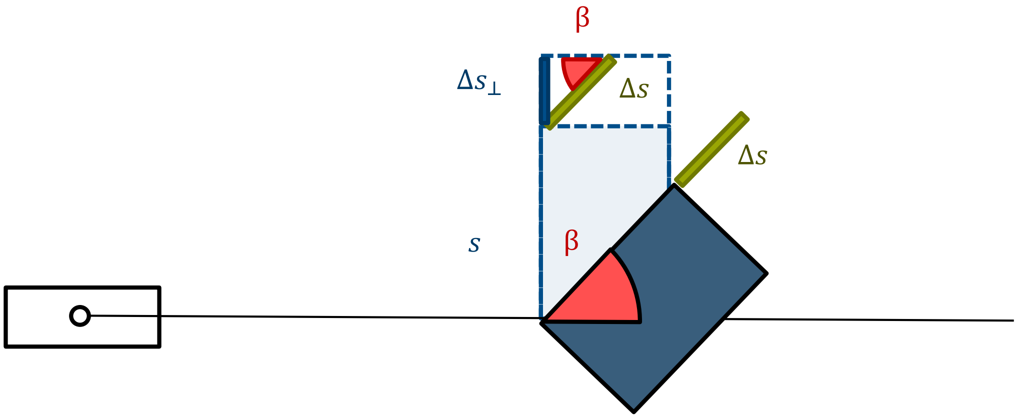

Figure 7-4: Size error in perpendicular direction and projection according to orientation.

information[474] and 8% for estimating indoor object size from monocular images[475]. In outdoor settings, relative errors of above 5% are obtained for an unfamiliar object[476]. The estimation error is generally lower for familiar objects that show little variance in size[471]. All aforementioned studies consider height or size perpendicular to the line of sight of the observer. For objects in space, results from judging aspect ratios of an L-shape on the ground show the ratio of true and perceived aspect ratio to be < 0.8[477], suggesting lower accuracy in depth direction.

To account for this fact, the permissible size error is first quantified for the perpendicular case as $\Delta s_\perp$. Summarizing the literature introduced above[474,475,476] yields a human perpendicular size error of $\Delta s_\perp > 5\% \cdot s$. Since the accuracy in depth direction is lower[477], the size error $\Delta s_\perp$ is projected according to the orientation of the object as shown in Fig. 7-4. Exclusively evaluating this projected error means that the permissible errors in radial direction are not bounded. However, from section 7.2.3 it is known that human errors of 15% occur when estimating distances in radial direction. Therefore, it is assumed that the same threshold of 15% is also applicable to bound permissible size errors in radial direction. Accordingly, the size error is evaluated using the projected perpendicular size error which is additionally bounded in radial direction. Conservative estimates prohibit underestimation of size, while the permissible overestimation is:

$$\Delta s \leq \min\left(\frac{\Delta s_\perp}{\sin(\beta)}, 15\% \cdot s\right) \tag{7-4}$$

[474] Mckee, S. P.; Welch, L.: The precision of size constancy (1992), p. 1457.

[475] Wu, Y. et al.: Size-to-depth (2018), p. 2.

[476] Gilinsky, A. S.: The Effect of Attitude upon the Perception of Size (1955), p. 180.

[477] Loomis, J. M.; Philbeck, J. W.: Is the anisotropy of perceived 3-D shape invariant across scale? (1999), p. 400.

Orientation

Orientation estimation provides useful cues in addition to motion such as distinguishing a driving and reversing car[478] or for pedestrians in standstill[479]. Existing requirements on datasets are inconsistent and range between $20° - 45°$ depending on the class.[480,481]

Generally, little information about human performance is available in literature. However, a small-scale study on amazon mechanical turk workers for human orientation estimation from monocular images shows a standard deviation of approximately $9°$.[482] Therefore, this requirement is adopted and $\pm 9°$ orientation error is used as requirement. Since no contrary information is available, this requirement is extended beyond the yaw angle reported in the studies to also include roll and pitch angle. While the practical impact of this assumption is not fully clear, it should be noted that large roll and pitch angles are uncommon in traffic scenarios.

Class

Classification differs from the other attributes since it has been previously treated in chapter 5. There, all required object categories are defined. Classification is evaluated as a separate attribute according to the required object categories. However, common object detection datasets do not offer sufficiently fine-grained annotations. Therefore, results regarding the attributes in this section are presented using the predefined dataset classes.

7.2.5 Summary

In this section, an overview across the results obtained from the method and its application is provided.

So far, the entire pipeline for evaluating object detection has been reviewed. The detection task is decomposed to distinguish different interpretable aspects. For each aspect, the common principles of conservative estimations and the human baseline are referenced. This yields a metric and a threshold defining acceptable performance. These criteria define different types of perception failures. All failures are designed in such a way that they typically do not occur for human performance.

[478] You, R.; Kwon, J.-W.: VoNet (2016), p. 196.

[479] Yu, D. et al.: Continuous Pedestrian Orientation Estimation (2019), p. 1.

[480] Braun, M. et al.: EuroCity Persons (2019), p. 1848.

[481] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3357.

[482] Hara, K.; Chellappa, R.: Growing Regression Tree Forests (2017), p. 306-308.

An overall evaluation of a detection algorithm is performed simply by counting the occurrence of failures. However, simply considering the number of occurrences neglects the number of scenes or objects. Therefore, a normalization with the number of GT objects similar to MOTA[483] is applied to enhance interpretability. Since each attributes defines clear failures separately, no weighting between different attributes is required.

7.3 Results

In this section, results for the practical application of the criteria developed in previous sections are shown. First, the implementation used to derive results is presented. This is followed by an overview over the evaluation results. The effect of different matching strategies on the evaluation is shown as first result. Subsequently, the prevalence of different failures, comparisons of different detection algorithms and the potential of sensor fusion are each considered separately.

7.3.1 Implementation

Performing a perception evaluation requires evaluation data as well as detection algorithms.

The public nuScenes dataset[484] is used as source for evaluation data. Like other datasets, nuScenes only offers a limited annotation frequency and a limited number of classes. However, the number of classes is higher than for other 3D object detection datasets. More specifically, the evaluation procedure is implemented for the standard nuScenes validation split. To obtain popular detection baselines, the mmDetection3D framework[485] is leveraged. The objective in this work is to show exemplary evaluation results for common baselines. Accordingly, the detectors PointPillars[486] and CenterPoint[487] for lidar as well as the detector FCOS3D[488] for camera are selected. Backbones are selected arbitrarily since the effect on mAP is within a few percentage points. The focus of this section is to obtain a general estimate of the detection performance on the proposed requirements. Therefore, detailed choices of backbones are inconsequential. For PointPillars, the SECFPN backbone is used while FCOS3D leverages a ResNet101 w/ DCN backbone. CenterPoint is applied with a SECFPN backbone using voxels with a size of 0.1 m and with circular non-maximum suppression. All baselines apply default settings and pretrained weights downloaded from mmDetection3D[489].

[483] Bernardin, K.; Stiefelhagen, R.: Evaluating Multiple Object Tracking Performance (2008), p. 2-5.

[484] Caesar, H. et al.: nuScenes (2020).

[485] MMDetection3D Contributors: MMDetection3D (2020).

[486] Lang, A. H. et al.: PointPillars: Fast Encoders for Object Detection from Point Clouds (2019).

[487] Yin, T. et al.: Center-based 3D Object Detection and Tracking (2021).

[488] Wang, T. et al.: FCOS3D (2021).

[489] MMDetection3D Contributors: MMDetection3D (2020).

7.3.2 Matching

As previously indicated, the association procedure may apply different methods for matching. In this section, the corresponding results are presented for the mini split of the nuScenes dataset[484].

To consider matching failures, both FP and FN failures are summed up and normalized with the number of GT objects. Results in Fig. 7-5 show the prevalence of failures for different association procedures for different algorithms. While the performance for the different detectors differs, the conclusions for the matching strategy are identical. First, it is verified that applying the standard Hungarian algorithm does indeed yield higher failure rates when compared with a greedy matching strategy. Secondly, the custom modification of the Hungarian algorithm which considers the outliers performs similar to the greedy matching. This indicates that the outliers indeed account for the difference, which is eliminated by modifying the cost matrix. With this custom modification, both matching strategies perform equally. Due to its simplicity, greedy matching is applied for the remainder of this work.

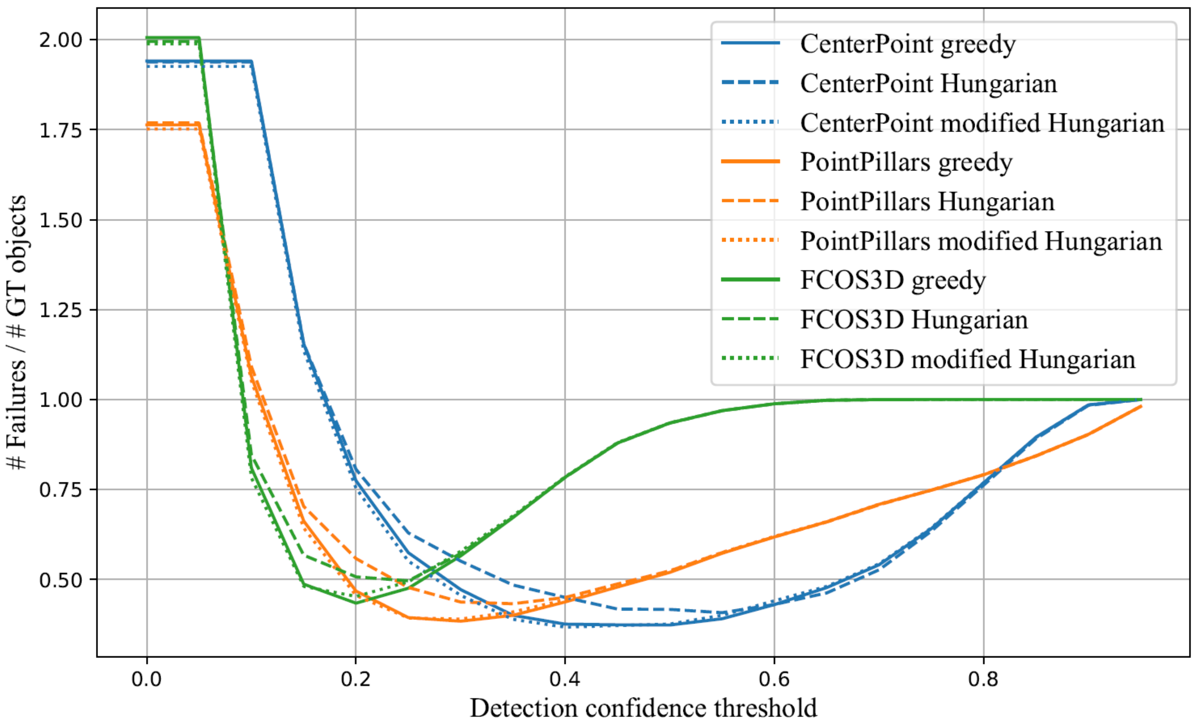

Figure 7-5: Comparison of detection performance for different matching algorithms.

7.3.3 Prevalence of Failure Types

This section focuses on the commonalities observed in different object detection baselines. For brevity, only the results of CenterPoint[487] are shown.

One noteworthy observation is that the detection results are generally not conservative. This means that for each attribute, approximately half of the estimates exhibit a corresponding failure.

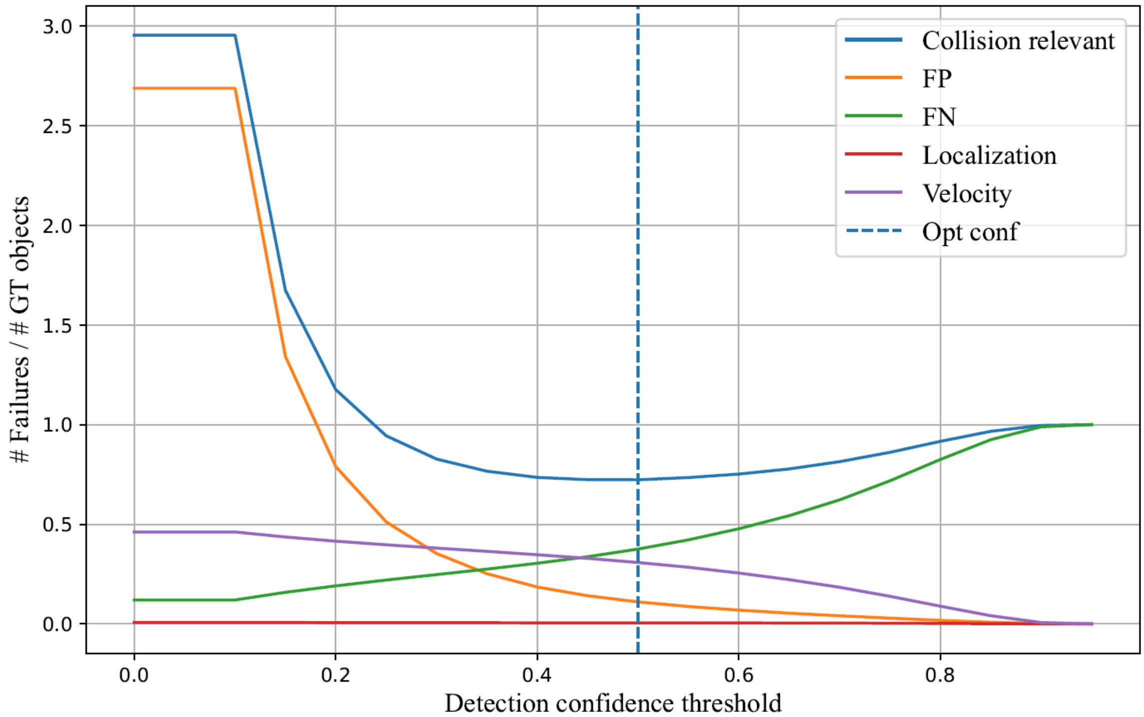

Figure 7-6: Frequency and optimum of failures of the CenterPoint detector normalized with the number of GT objects for collision relevant attributes. Unlike metrics based on recall and precision, lower is better.[490]

Considering multiple attributes, this means that almost every object shows perception failures owing to nonconservative estimations.

However, besides conservative estimates, detection accuracy is also of practical relevance. Therefore, accuracy is evaluated separately by neglecting failures due to lack of conservativity. Results for different failure types of the detector are depicted in Fig. 7-6 for different confidence thresholds. The most frequent collision relevant failures occur for association. TP failures for attributes such as velocity and location of correctly matched objects are less frequent. However, among the different TP failures for collision relevant attributes, velocity failures are the most frequent. Only considering accuracy for the optimal confidence thresholds yields 0.72 collision relevant failures per GT box on average for CenterPoint. Further failures for the non collision relevant attributes occur as depicted in Fig. 7-7. Similar results are obtained for the other detectors with FCOS3D at 0.82 and Pointpillars at 0.76 collision relevant failures per GT object.

[490] Modified from Mori, K. T.; Peters, S.: SHARD (2023), Fig. 6 © 2023 IEEE.

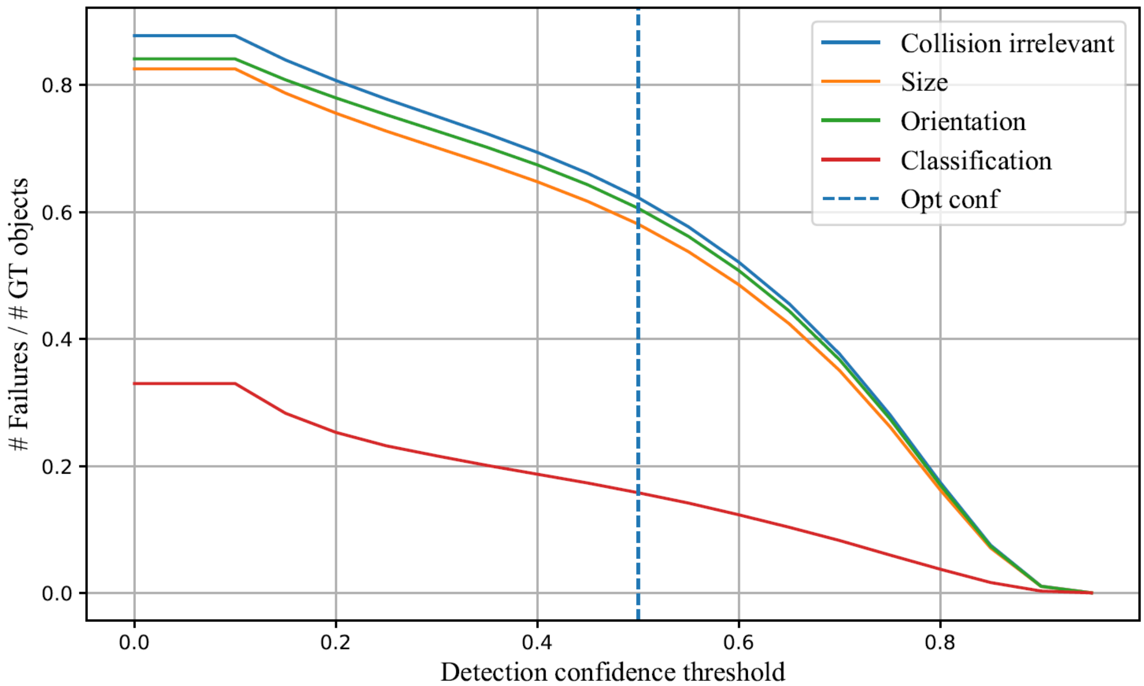

Figure 7-7: Failure frequency of the CenterPoint detector normalized with the number of GT objects for non collision relevant attributes. The optimum confidence threshold is provided for the collision relevant failures.

7.3.4 Different Detectors

The general trends shown in the previous section are similar between detectors. However, this section emphasizes the remaining differences between different detection architectures.

For this section, matching and velocity failures are considered. Both failure types are selected since the previous section showed them to be the dominant failure types.

Different baselines consisting of two lidar detectors and one camera detector are compared in Fig. 7-8. As indicated by the dashed lines in the image, the optimal confidence thresholds differ between detectors. The overall distribution of failures is similar between different detectors. Nevertheless, differences regarding the number of failures as well as their distribution over different confidence thresholds prevail.

7.3.5 Ideal Fusion and Uncorrelated Fusion

It is commonly assumed that the fusion of different detectors and modalities improves performance. This section explicitly tests this assumption. As in the previous section, results are presented for matching and localization failures.

The objective in this section is to provide an upper bound for the potential of sensor fusion. In addition, state of the art fusion architectures show high variability and are in many cases complex. Therefore, a hypothetical ideal fusion is constructed in this section. This fusion operates on the

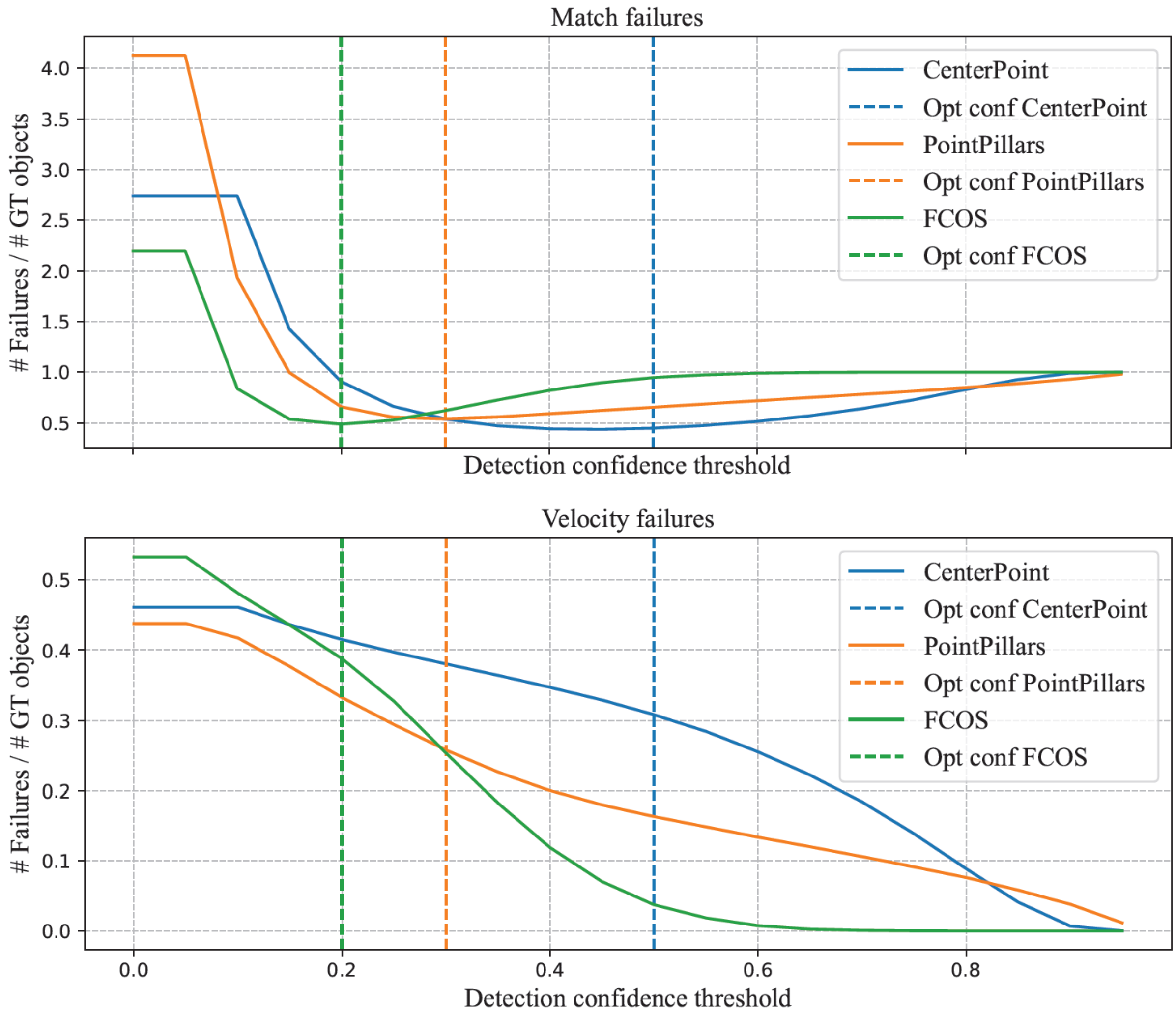

Figure 7-8: Frequency and optimum of matching and localization failures for different detectors normalized with the number of GT objects.[491]

two object list outputs from two detectors and constructs a single object list as output. Basically, the assumption is that a correct output is generated if either detector is correct. Accordingly, false negatives are only output if they are present in both input object lists. Similarly, false positives are also only output if present in both object lists. This fusion effectively assumes knowledge of the GT to generate results. It therefore inevitably overestimates the performance which can be practically achieved. However, it does provide an upper performance bound.

A common factor which reduces fusion performance are potential correlations of failures. In order to achieve the optimum redundancy, two detectors should not exhibit correlations of their failures. Therefore, a second fusion reference is calculated under the assumption of zero correlation. If no correlation is present, the final failure likelihood can be calculated by multiplication of the two failure likelihoods of both detectors. This uncorrelated fusion also represents an idealized fusion.

[491] Modified from Mori, K. T.; Peters, S.: SHARD (2023), Fig. 7 © 2023 IEEE.

Both fusion approaches are compared with the baseline performance in Fig. 7-9. It is observed that the ideal fusion achieves performance improvements over the baseline. The effect is most pronounced for the most frequent matching failures. However, the ideal fusion of three detectors fails to yield substantial additional performance gains. For lower failures probabilities, the ideal fusion shows higher failure likelihoods than the uncorrelated fusion.

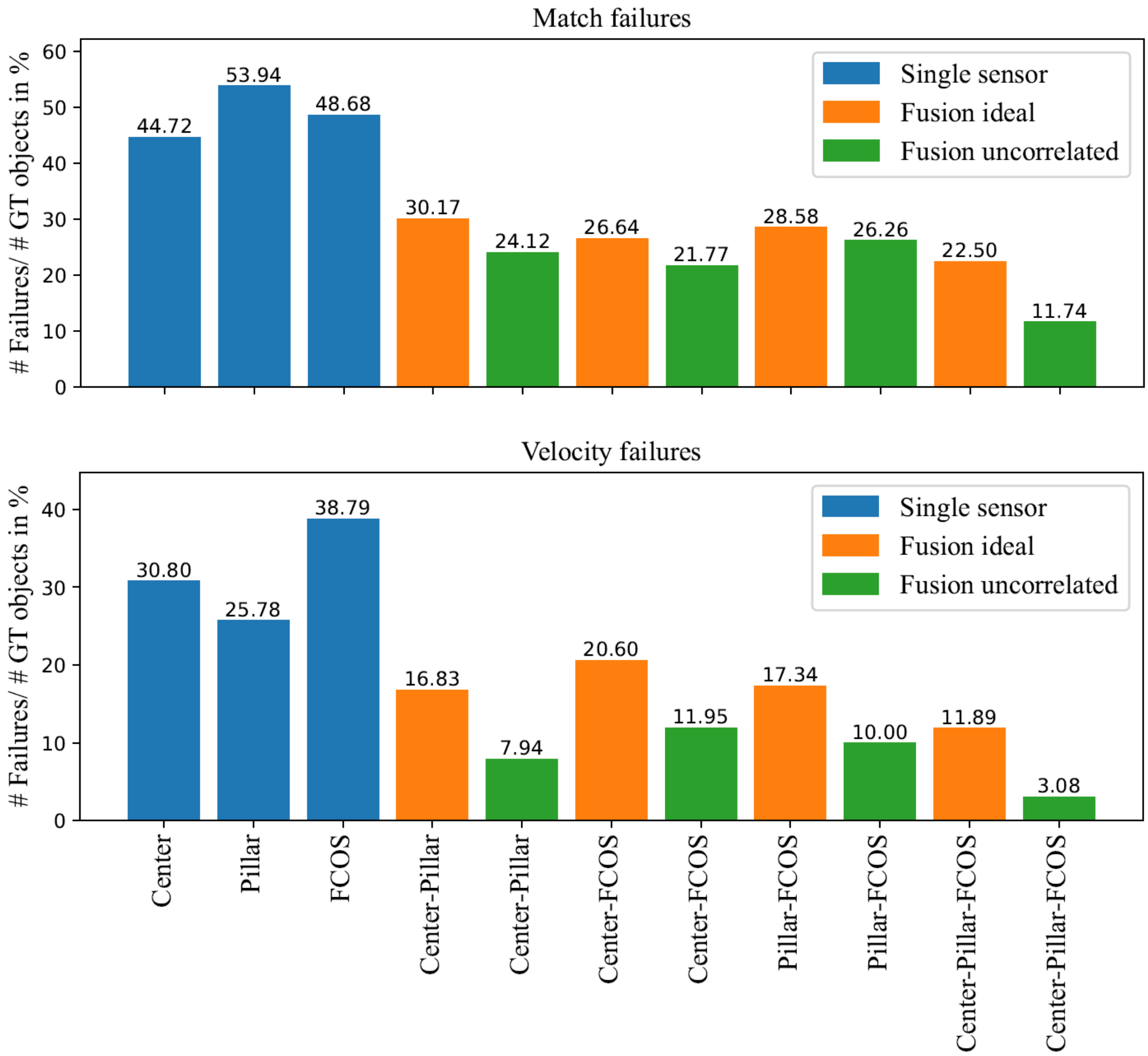

Figure 7-9: Frequency of matching failures normalized the number of GT objects and localization failures normalized with TP boxes of three detectors for the optimum confidence threshold compared with an ideal fusion and with a hypothetical case with no failure correlation.[492]

[492] Modified from Mori, K. T.; Peters, S.: SHARD (2023), Fig. 8 © 2023 IEEE.

7.4 Conclusion

This chapter presented a method to define requirements for attributes of detected objects. Based on the common principle of interpretability, different attributes each received a separate requirement.

It was found that legal and safety requirements are insufficient to define requirements for attributes. Therefore, this chapter relied heavily on considerations of the human detection performance. The entire object detection evaluation pipeline was reconsidered while considering safety and the human baseline. It was found that greedy optimization is a suitable strategy for matching. Matching is performed using the closest point as reference without consideration of object class. The matching threshold is a point distance which scales with the egocentric distance. While temporal requirements of detection within 150 ms exist, no requirements for unique identification or tracking of objects over time were identified. Therefore, the attributes are evaluated separately and for every frame. For interpretability, the collision relevant attributes and non collision relevant attributes are distinguished. The former include distance, angular position, radial velocity and angular velocity while the latter include attributes such as size, orientation and class. In the case of each attribute, suitable requirements from literature were identified.

Applying the novel requirements to existing detectors showed that conservative estimates are not reflected in the detection results. However, even when neglecting conservativity, frequent failures ranging from 0.72 to 0.82 failures of collision relevant attributes per GT object occurred. These results were similar for all three object detectors studied in this chapter. Furthermore, an idealized sensor fusion was considered to estimate an upper performance bound of fusion. Results showed that such an idealized fusion does demonstrate potential benefits over single detectors. However, even the idealized fusion exhibits higher failure rates than when calculating theoretical failure rates for zero correlation.

A full discussion of the attribute results along with the corresponding RQ3 is provided in section 9.3.3 under consideration of the validation results presented in chapter 8.

8 Validation

So far, results were based on an argumentation and then partly plausibilized. However, the open question remains how to validate object detection requirements. Therefore, the following RQ4 is addressed in this chapter:

How can detection requirements be validated with respect to the safety of the driving task?

The first section presents an overview of the validation method. This is followed by an application of the validation method to the relevance and attribute results. Finally, a conclusion of this relevance chapter is provided. The content of this chapter was in parts published in previous publications[493,494] coauthored by the author. The content also appears in the work by Storms[495] created concurrently with this work.

8.1 Method

In this section, the method for validating the detection requirements is presented. First, a preliminary discussion introduces the basis of the method. Next, an overview of the proposed validation method is provided. Finally, this approach is formally specified.

8.1.1 Preliminaries

As noted in section 3.4, limited attention has been directed to the validation of detection metrics. Furthermore, there is no generally accepted method available in literature.

Previous approaches for detection criteria are generally either based on an argumentation or on a specific implementation of a downstream planner. However, previous methods fail to reconcile these two complementary approaches. The proposed method therefore aims to combine these approaches. Once again, the common principle of the human baseline is applied. This objective is also present in previous planners[496] as well as the planner-centric metric Planning Kullback-Leibler Divergence (PKL)[497]. However, the objectives for a path planner typically include additional requirements. Objectives regarding infractions, mission goals[498] and comfort lead to ambiguous

[493] Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023).

[494] Mori, K. T.; Peters, S.: SHARD (2023) © 2023 IEEE.

[495] Storms, K.: Context Aware Data Reduction (2023).

[496] Bansal, M. et al.: ChauffeurNet (2018), p. 8-12.

[497] Philion, J.; Fidler, S.: Lift, Splat, Shoot (2020), p. 14057-14058.

[498] Dosovitskiy, A. et al.: CARLA: An Open Urban Driving Simulator (2017), p. 6

planning objectives. In addition, the ambiguity in objectives also complicates the definition of consistent evaluation metrics.[499]

The idea presented in this work is to adapt the method presented by PKL[500] and apply it to validate the previously developed requirements. However, one core limitation of planning based approaches is the lack of validity of the planner itself.[501] Therefore, modifications are introduced in this work to disambiguate the objective and allow a clear performance evaluation.

8.1.2 Overview

Following the preliminary discussion, the overall concept of the validation method is presented in this section.

In accordance with the common principle of the human baseline, the objective is to determine the influence of perception failures on the human driving task. An ideal study of this kind may be conducted in a driving simulator. While this approach directly studies the influence on humans, it incurs substantial cost. Therefore, the proposed method utilizes a motion prediction algorithm to approximate the human driving behavior instead. Formulating the objective as motion prediction of all vehicles including the ego vehicle instead of path planning offers a number of advantages. Firstly, motion prediction can be evaluated open-loop on recorded data in an offline manner. Therefore, real world recordings from public datasets can be applied directly. This avoids the need for closed-loop simulations as in previous work[502,503,504], for which validity is difficult to achieve. The prediction performance is explicitly quantified in this work and used as part of the validation. This is in contrast to previous work[500], which does not present a method to deal with potential errors in the planner[501]. The issue of potential planning errors is exacerbated by the ambiguities in the planning objectives.[505] By contrast, the objective of the motion prediction is unambiguous. Another difference to prior work is that for the proposed method, no full AD pipeline is created. Instead, the motion prediction acts as proxy for human behavior.

The basic idea of the validation considers the fact that the pass/fail criteria define boundaries for acceptable errors. For example, removing irrelevant objects or slightly offsetting the location within the permitted bounds is considered acceptable. Validity with respect to the human baseline means that the acceptable perception errors do not influence human driving behavior. The motion prediction is used as proxy for this human behavior. Therefore, manipulating the object list within the acceptance criteria should not influence the prediction performance. Conversely, a

[499] Caesar, H. et al.: nuPlan: A closed-loop ML-based planning benchmark (2021), p. 4.

[500] Philion, J. et al.: Learning to Evaluate Perception (2020).

[501] Mao, J. et al.: 3D Object Detection for Autonomous Driving (2023), p. 1914.

[502] Henze, F. et al.: Admissible Uncertainty Bounds for Planning Algorithms (2021), p. 3120-3130.

[503] Jha, S. et al.: Watch Out for the Safety-Threatening Actors (2022), p. 6-9

[504] Philipp, R. et al.: Accuracy Requirements for Environmental Perception (2021), p. 131-133.

[505] Guo, Y. et al.: The efficacy of Neural Planning Metrics (2020), p. 2.

difference in prediction performance indicates a criterion is not valid. By subjecting the object list to different manipulations, each criterion is validated separately.

8.1.3 Formal Specification

The previous description is high-level and thus does not provide sufficient formal specification of the approach. Therefore, this section introduces the metrics and mathematical methods applied to specify prediction performance and to perform comparisons.

Prediction errors are measured with the common average distance error (ADE) metric for the top ten trajectories used by the nuScenes motion prediction benchmark[506]. For different inputs, the global prediction performance is evaluated across the whole dataset. This global evaluation avoids effects resulting from local performance issues and non-deterministic predictions. Contrary to previous approaches locally evaluating scenarios, this procedure acknowledges the inherent uncertainty of predicting future trajectories. The global error is represented as ECDF of local prediction errors per object. To account for potential non-deterministic predictions, predictions are performed multiple times for each input. Finally, the sets of ECDFs for different inputs are compared. Accordingly, a criterion is considered valid if the global error distribution remains unchanged. The equivalence of the error distribution is evaluated with a statistical test. The Cramer-von Mises test for two empirical distributions[507] is applied. This test yields a p-value reflecting the confidence that both ECDFs are samples drawn from the same underlying distribution. Notably, this confidence reflects the distribution similarity as well as the sample size. Pairwise comparisons are performed for multiple runs with the same input as well as for different inputs. All p-values are plotted as a box plot to obtain a visual representation. While it is possible to introduce threshold values for p, the suitable choice for such a threshold is debated.[508] Instead, this work leverages the comparison of two GT inputs as reference for valid p-values. Therefore, p-values for other inputs are compared with the GT-GT case instead of introducing thresholds.

8.2 Results

In this section, results for the application of the validation method are presented. Firstly, the implementation used for the application of the validation is briefly described. Next, a verification of the validation procedure itself is provided. This is followed by a validation of the criteria proposed earlier in this work. However, it should be noted that the classification results of this work are not validated. The reason is that publicly available datasets do not provide annotations for the proposed categories.

[506] nuScenes: nuScenes prediction task: Leaderboard (2020).

[507] Anderson, T. W.: On the Distribution of the Two-Sample Cramer-von Mises Criterion (1962), p. 1148.

[508] Benjamin, D. J. et al.: Redefine statistical significance (2018), p. 6.

8.2.1 Implementation

To evaluate the practical utility of the proposed method, an implementation on a public dataset is provided.

As for previous results, the standard validation split of the nuScenes dataset[509] is utilized. The prediction network is selected from the nuScenes motion prediction leaderboard[510]. An algorithm with a publicly available implementation is selected among the top entries at the time of implementation. Therefore, the PGP algorithm[511] with its corresponding implementation[512] is selected and applied with pretrained weights and default settings. The prediction takes inputs from a square region of interest reaching from [-20 m, 80 m] in longitudinal and [-50 m, 50 m] in lateral direction. Within this region, lane information, humans and vehicles are provided to the network. This default region of interest is maintained in all experiments. All experiments are performed ten times with the same input.

8.2.2 Verification

This section presents different results which verify the proposed validation procedure. Firstly, experimental results are presented for global and local evaluation. This is followed by boxplots comparing the regular input with verification inputs.

Global Versus Local Evaluation

In this section, results of comparing results locally for each object and frame as well as comparing globally across the whole dataset are presented. Results are displayed in Fig. 8-1. The evaluation is run for multiple runs with the same GT input. The dashed lines show pairwise comparisons of multiple runs according to the local evaluation. Due to nondeterministic predictions of the PGP algorithm, significant noise averaging 0.33 m is observed even for identical GT inputs. The corresponding prediction error as used by the global evaluation is shown by the bold lines. Prediction performance is explicitly shown by the distribution, averaging at 0.96 m. Despite the local noise in predictions, the global error distributions are barely visually distinguishable in the plot. This means that the global prediction error distribution is robust to noise resulting from nondeterministic predictions.

[509] Caesar, H. et al.: nuScenes (2020).

[510] nuScenes: nuScenes prediction task: Leaderboard (2020).

[511] Deo, N. et al.: Multimodal Trajectory Prediction Conditioned on Lane-Graph Traversals (2022).

[512] nachiket92: PGP: Multimodal Trajectory Prediction Conditioned on Lane-Graph Traversals (2022).

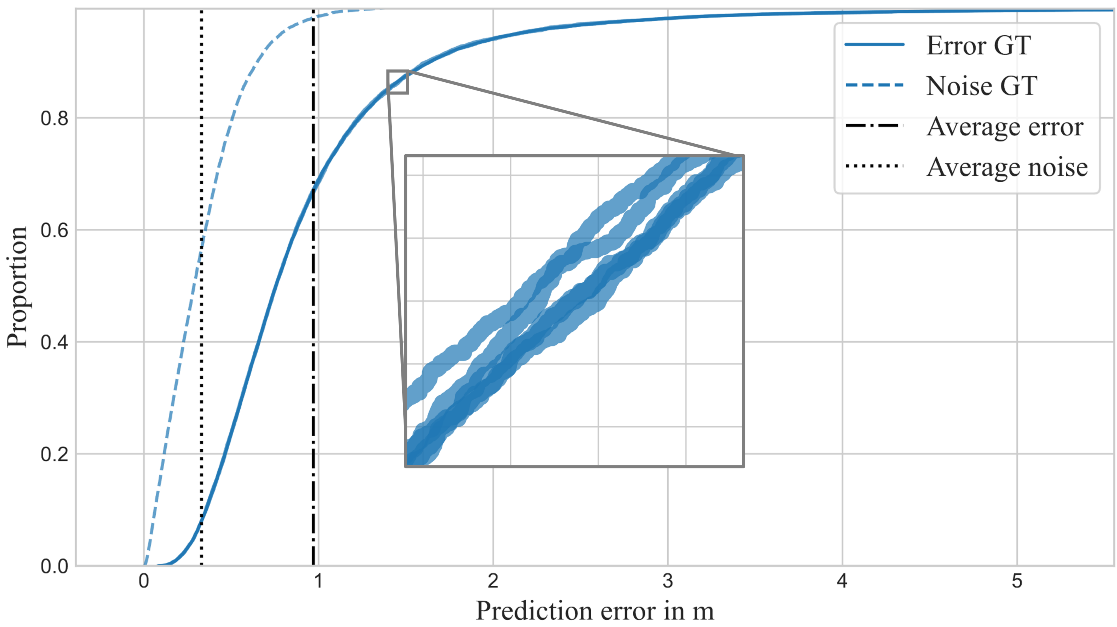

Figure 8-1: ECDFs of prediction errors as ADE for top ten trajectories for different inputs and noise for multiple runs with same GT inputs.[513]

Verification Inputs

For the verification of the validation procedure, artificial inputs for verification purposes are constructed. The idea is to construct inputs which are implausible and invalid to ensure that the validation procedure uncovers them.

The first two verification inputs considers the relevance of objects. Verification is abbreviated with “V” with an additional number to distinguish different verification inputs. The second letter abbreviates relevance and distance as “R” and “D”, respectively. The first input denoted as VR1 simply removes all objects in a scene. This ensures that any relevant objects present in the scene are definitely removed. However, this does not ensure that the validation is also able to uncover subtler invalid inputs. Most importantly, removing fewer objects may be difficult to uncover for the validation procedure. Therefore, the second verification input VR2 is constructed to selectively remove relevant objects. For this purpose, all vehicles within 2 m from the heading axis of the ego vehicle are removed in VR2. In this case, approximately 5% ob the objects are removed. This value is chosen to be lower than the number of objects filtered out by the relevance criteria put forth in this work. It should be noted that both prediction inputs still include lane information in addition to object information. This is necessary since preliminary experiments showed the prediction performance severely deteriorates in absence of lane information.

[513] Modified from: Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023), Fig. 3.

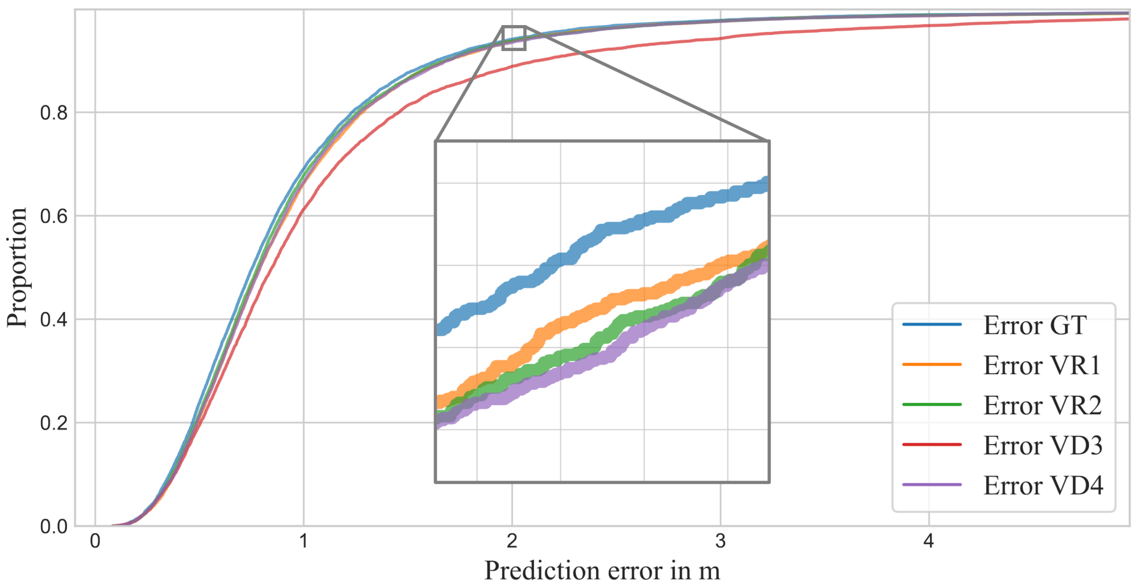

Figure 8-2: ECDFs of prediction errors as ADE for top ten trajectories for different verification inputs.[514]

Besides manipulating the existence of objects, manipulating attributes is also studied. For this purpose, two additional verification inputs utilize large distance errors. The third verification input VD3 halves the distances while the fourth verification input VD4 doubles the distance. It should be noted that increasing the distance can lead to objects being removed since they move outside of the region of interest applied to the prediction input.

Results for the verification are displayed visually in two different ways. Firstly, the error distributions as ECDFs are shown in Fig. 8-2. Note that only one of ten runs is depicted in the graph for better visualization. Overall, it is observed that the error distributions are visually similar. This is true even for the condition VR1 where all other objects are removed from the input. However, enlarging the image indicates differences between the verification inputs and the GT input. In addition to a visual comparison, the results of the statistical tests are shown as boxplot in Fig. 8-3.

The y-axis depicting the p-values is log scaled. The leftmost box shows the comparison of different runs using the same GT input (GT-GT). This case exhibits high p-values with limited variance due to noise. All remaining boxes except the second column indicate comparisons of verification inputs with the GT input. It is observed that p-values are orders of magnitude smaller with increased variance. The low p-values indicate a high confidence that the error distributions are in fact different for the verification inputs. Most importantly, the p-values differ from the GT-GT case. Therefore, the validation procedure considers the verification inputs invalid. Since the verification inputs are invalid by design, this is the desired behavior. Thus, the verification of the validation procedure is successful.

[514] Modified from Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023), Fig. 3.

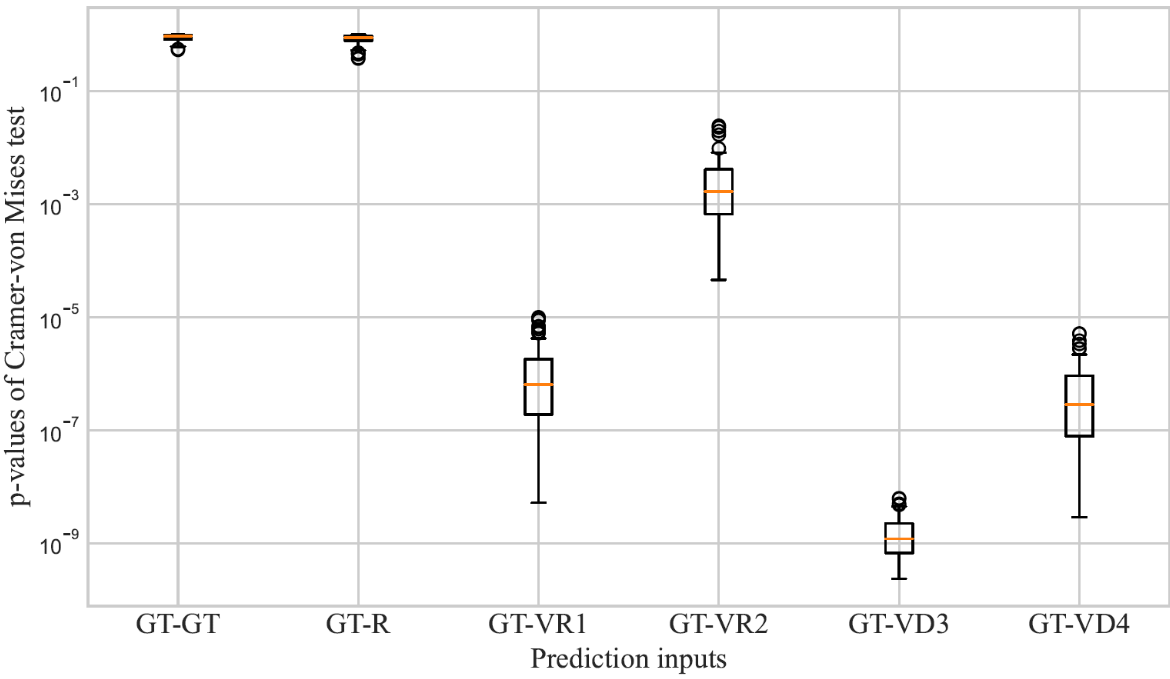

Figure 8-3: Boxplot of p-values testing for equality of distribution for pairwise combinations.[515]

8.2.3 Relevance

After verifying the validation procedure, the validation is applied to the requirements developed in this work. Since dataset annotations are insufficient to evaluate the classification criteria of this work, only relevance and attribute requirements are validated. This section focuses on the relevance results.

In this section, the validation procedure is applied to the relevance criteria of this work. It should once more be noted that the input of the prediction network already uses a geometric region which also filters objects. This filtering is maintained when evaluating the relevance criteria of this work for consistency. The proposed relevance criteria are only additionally applied within the predefined region of interest. In this case, approximately 10% of the objects are additionally filtered out by the proposed relevance criteria. The visual results of the error distribution are presented as red line in Fig. 8-4.

As previously, only one of multiple runs is shown for this visualization. The same verification results as in the previous Fig. 8-2 are depicted in gray for reference. It is observed that the results for the proposed relevance criteria appear very similar to the GT input. Qualitatively, the two distributions appear indistinguishable, even for the enlarged portion of the image. This result supports the relevance criteria of this work, since it indicates that the prediction performance is unaffected. The quantitative results of statistically testing for equality are depicted in the left part of Fig. 8-3.

[515] Modified from Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023), Fig. 4.

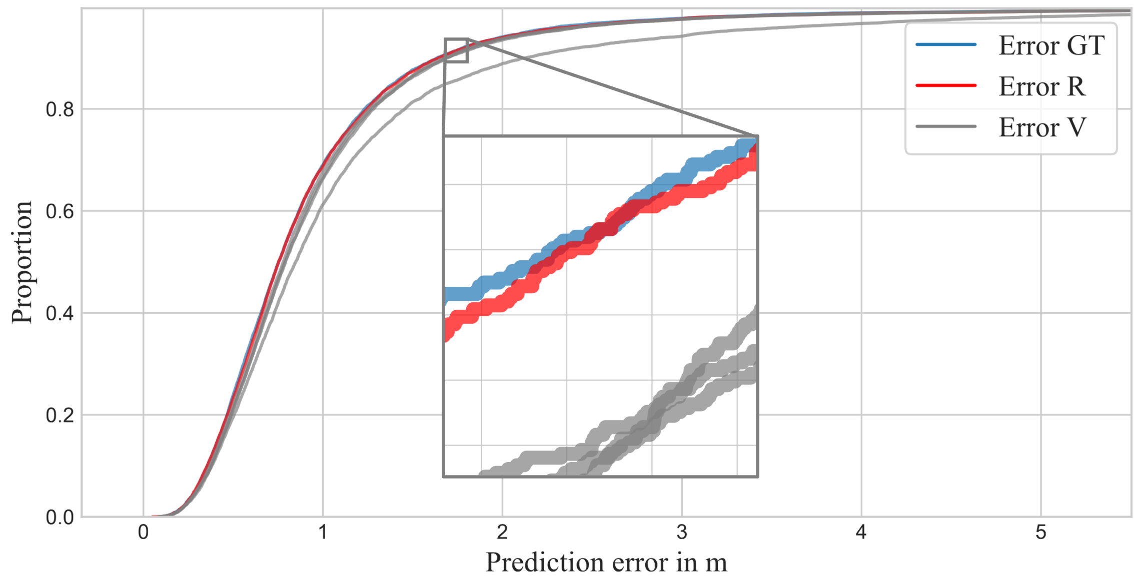

Figure 8-4: ECDFs of prediction errors as ADE for top ten trajectories for verification inputs and relevant inputs.[516]

Comparing relevant inputs with GT inputs (GT-R) shows similarly high p-values as for identical GT inputs (GT-GT). This occurs despite the fact that (GT-R) filters out more objects than (GT-VR2). Therefore, the prediction performance is not affected by the relevance criteria proposed in this work. Since the falsification of the relevance criteria failed, this lends support to the proposed relevance criteria.

8.2.4 Attributes

As next step, the same validation procedure is also applied to the attribute criteria.

For each attribute, the requirement defines the permissible error. Each attribute is considered separately by modifying the GT attribute with the maximum permissible error. This modification is applied to all input objects in a scene. In this section, the focus lies upon the criteria for location. The reason is the simple manipulation of the prediction input which is less trivial for other attributes. Different assumptions regarding localization requirements are validated separately in the following. The first part focuses on the distinction of different localization error types such as distance and azimuth angle. This is followed by considering the direction or conservativity of errors. Finally, the magnitude or thresholds of the errors are scrutinized. Previous results showed that differences in error distributions are difficult to visually distinguish. Therefore, the following evaluation only visualizes the boxplots which resolve differences more accurately.

[516] Modified from Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023), Fig. 3.

Different Types of Error

In this work, the pass/fail criteria assert that different types of localization error such as distance and angle should be evaluated separately. Therefore, a validation of this hypothesis is provided in this section.

To this end, the locations of the objects are subjected to manipulation in radial distance direction and in azimuth direction by changing the azimuth angle. For both types of error, an equal error magnitude is selected in accordance with the largest nuScenes threshold of 4 m[517]. The result of the comparison is depicted in Fig. 8-5. Distance errors and angular errors clearly exhibit different p-values with p-values for angular errors being orders of magnitude smaller. The different impact on the prediction indicates that distinguishing different types of error is valid. Furthermore, the results also show that for the threshold of 4 m, different values than for the GT input are observed. This means that depending on the direction of the error, the largest threshold of 4 m applied by the nuScenes dataset already impacts the prediction performance. Accordingly, this threshold is falsified by the validation procedure.

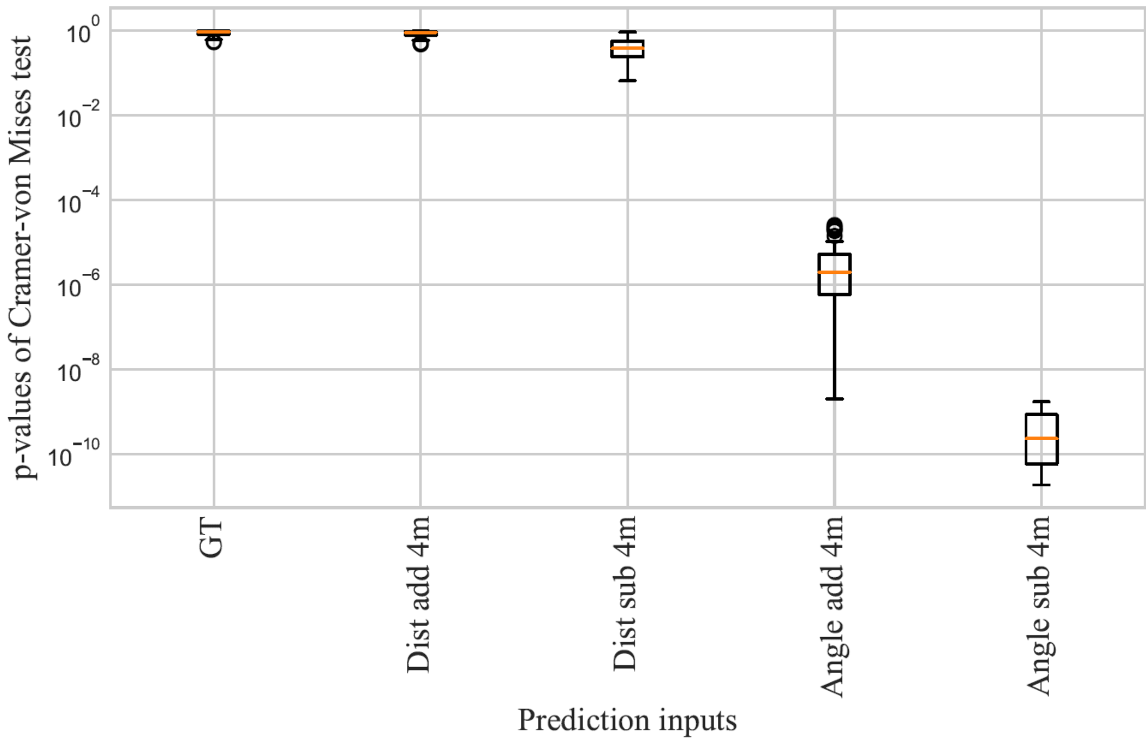

Figure 8-5: Boxplot of log-scaled p-values for comparing prediction error distributions of different inputs with different error types and directions with the GT input.[518]

[517] Caesar, H. et al.: nuScenes (2020), p. 11622.

[518] Modified from Mori, K. T.; Peters, S.: SHARD (2023), Fig. 4 © 2023 IEEE.

Direction of Error

In addition to different types of error, conservativity is demanded by the requirements developed in this work. Therefore, results for applying the same error magnitude in different directions are evaluated.

Results are also depicted in Fig. 8-5 for both distance and azimuth errors. In either case, the p-values observed differ between different error directions. However, the effect is more pronounced for angular errors which differ by orders of magnitude. It should be noted that the ADE metric in this work only considers the Euclidean distance of points. The influence of the direction occurs despite this symmetry of the evaluation metric. This indicates that the prediction is impacted differently for different error directions. However, the results do not support conservativity. In fact, the non-conservative estimate of overestimating distances exhibits similar p-values as the GT input. This indicates that non-conservative estimates are valid for human behavior despite conflicting with safety principles previously introduced in this work.

Error Scale and Magnitude

After studying the impact of different types and error directions, the permissible thresholds remain as open question. In this work, thresholds scaled with distance are proposed which are validated in the following.

When comparing thresholds scaling with distance with constant thresholds, a fair comparison requires equal error budgets. Therefore, the fixed errors are chosen to be equal to distance scaled errors when averaged across the whole dataset. This yields an average distance error of 5.29 m which corresponds to the distance error of 15%. For an azimuth error of $5°$, an average location error of 3.07 m is obtained. As previously, results are depicted in boxplots in Fig. 8-6. Constant errors exhibit lower p-values than their respective counterparts which scale with distance. For the distance errors, p-values range in the same range as for GT inputs for distance scaled errors and adding constant errors. One exception is subtracting a constant error which exhibits lower p-values. For the angular errors, all p-values are lower than for GT inputs by orders of magnitude. Overall, the p-values for constant errors are smaller than the p-values for distance scaled errors. This means that the statistical evidence supporting an influence of the errors on the prediction is stronger for constant errors. Conversely, the influence of distance scaled errors on the prediction performance is less pronounced. Therefore, the validation supports scaling errors with distance. In addition, the distance threshold selected in this work is supported despite being more lenient than common dataset thresholds. However, the thresholds for angular error are not supported.

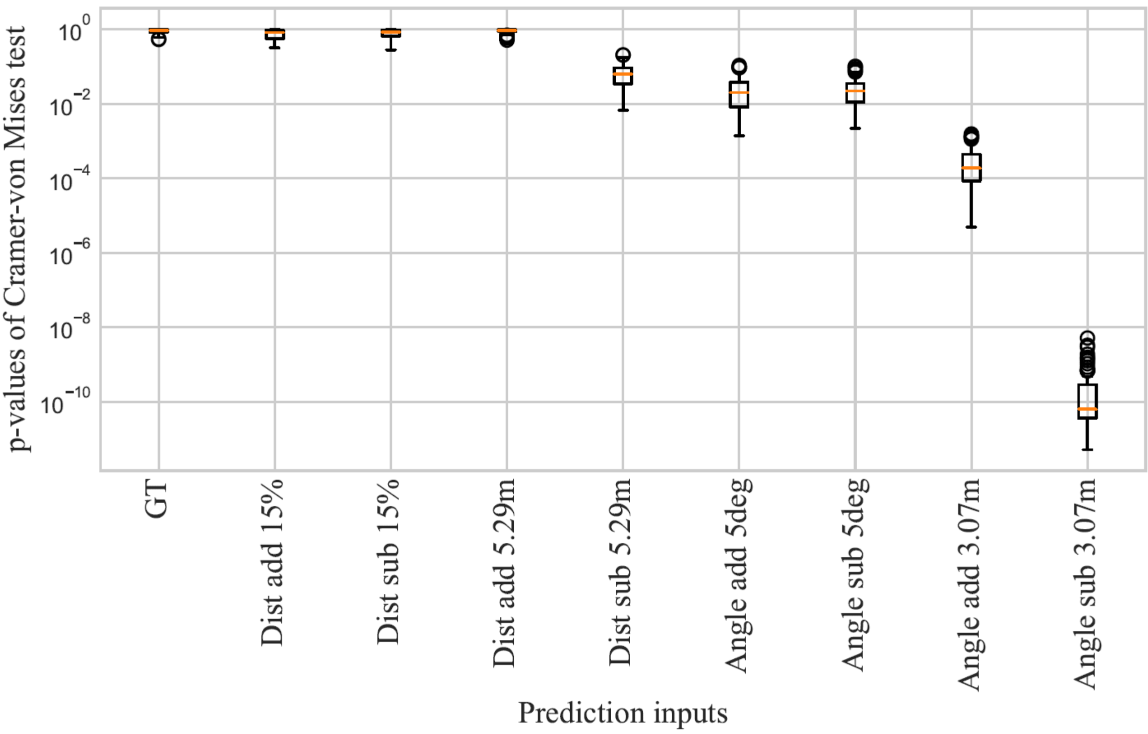

Figure 8-6: Boxplot of log-scaled p-values for comparing prediction error distributions of different inputs with different error scales and magnitudes with the GT input.[519]

8.3 Conclusion

In this chapter, a novel method was presented to validate requirements for object detection.

The approach leverages a DNN for motion prediction. Since the prediction provides a proxy for human driving behavior, it directly reflects the common principle of the human baseline. Contrary to prior methods based on neural planners, prediction has unambiguous task objectives which can be evaluated open-loop. However, the common principles of legal and safety requirements are only implicitly considered to the degree that they are reflected in the human driven trajectories of the training data. The validation introduces modifications to the object list used as input for the prediction. If the modification within the threshold specified by the requirement does not impact the prediction, the requirement is considered valid. In order to account for the inherent uncertainties of prediction, a global prediction evaluation across the whole dataset was proposed. Whether the prediction is impacted was evaluated by a statistical test comparing the global error distribution. While the DNN itself is not interpretable, the validation procedure still reflects the common principle of interpretability if applied to interpretable requirements.

[519] Modified from Mori, K. T.; Peters, S.: SHARD (2023), Fig. 5 © 2023 IEEE.

The validation method was only applied to the relevance and localization requirements of this work due to the availability of annotations and suitable motion prediction networks. A verification using artificially constructed inputs showed that the validation procedure successfully identifies invalid input modifications. The subsequent application of the validation method to the relevance results supported the proposed criteria. Regarding the localization requirements, the validation supported the approach of this work to distinguish different types and directions of errors. Conversely, the center distance of 4 m applied by the nuScenes dataset[520] was shown to be invalid. Furthermore, the validation also supported scaling the error thresholds with the egocentric distance as performed in this work. However, the conservative estimates applied in this work were not supported. The distance error threshold of this work was supported while the angular error threshold was not.

A full discussion of the validation results along with the corresponding RQ4 is provided in section 9.2 following the discussion of the overall method of this work.

[520]Caesar, H. et al.: nuScenes (2020), p. 11622.

9 Discussion

The joint discussion of the methodology and results of the previous chapters is presented in the following. An overview of the structure of the discussion is provided in Fig. 9-1.

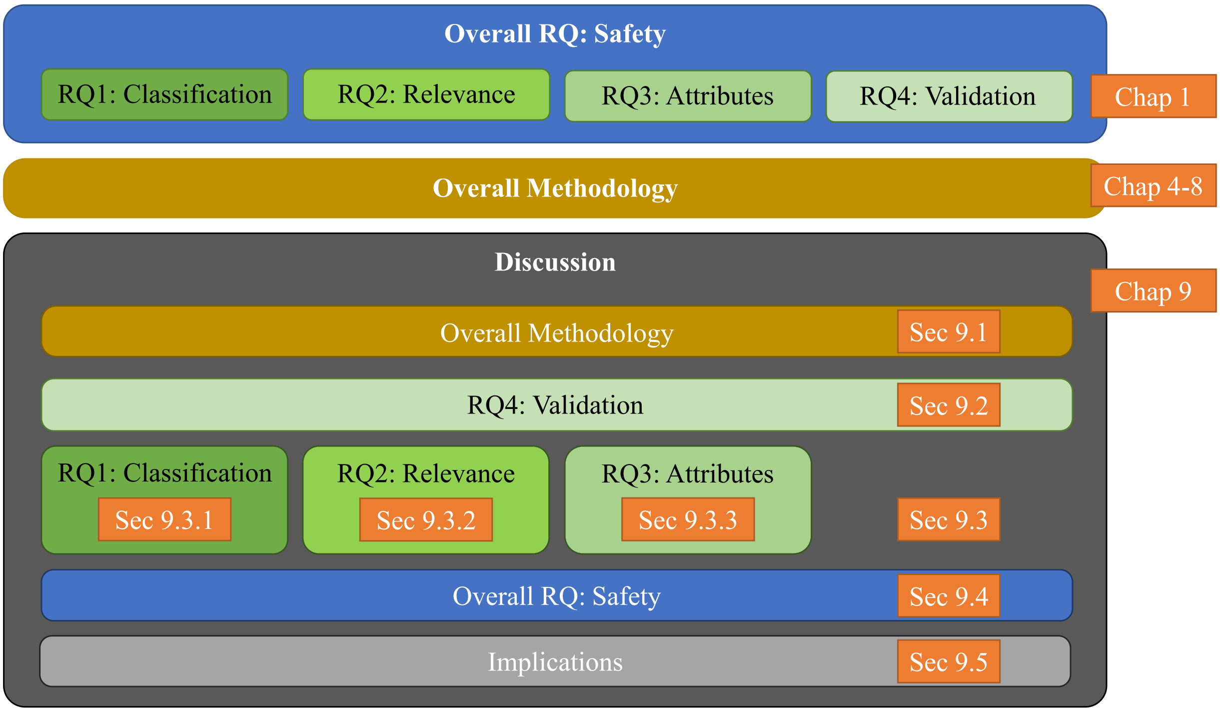

Figure 9-1: Overview of the structure of the discussion.

While this chapter is also structured according to the RQs, it occasionally deviates from the sequence of previous chapters. Firstly, the overall methodology introduced in chapter 4 is discussed. However, the discussion of the overall RQ regarding safety is postponed until the other RQs 1-4 are discussed. Among the other RQs, the discussion of RQ 4 regarding the validation of detection requirements is presented first. The reason is that consideration of the validity of the results is required for the discussion of the other RQs. Therefore, RQs 1-3 are each discussed separately in the sections following the validation. Having treated the overall methodology as well as the other RQs, the overall RQ regarding safety is discussed. With this, all RQs regarding requirements for different aspects of the object detection task are answered. Finally, the implications which these novel requirements bear on data acquisition and object detection are discussed.

9.1 Overall Methodology

In order to answer the overall RQ, an overall methodology was proposed in chapter 4 of this work. The overall RQ was decomposed into four different RQs which treat different aspects separately. However, common principles were introduced and applied to each of the RQs.

In this section, the overall methodology of this work is discussed. Firstly, an overview is provided regarding how the common principles are reflected in each method and result presented in chapters 5-8. This is followed by a discussion of the high-level concept of eliciting and validating requirements. Next, limitations of the overall approach are presented. Finally, this section is concluded with a brief summary. Note that answering the overall RQ is postponed until section 9.4 after all the other RQs have been answered.

9.1.1 Common Principles

The common principles of this work consist of interpretability, legal, safety and human aspects. In this section, the overall methodology and the results from chapters 5-8 are discussed with respect to these common principles.

Interpretability is reflected in the results in different ways. Firstly, the results show that it is possible to consider a perception module separately from the downstream planner. All requirements are interpretable on the detection interface without requiring a full specification of the planner. The assumptions regarding the planner are mostly limited to assuming conformity with road regulations and respecting safety requirements. Only the relevance results require additional information regarding the planner. However, even in this case the information is limited to high-level attributes such as latency and available accelerations. Secondly, the presented approach emphasizes interpretability by developing requirements for different aspects of the object detection task. Furthermore, different attributes are also decomposed and considered separately. Overall, the results of this work show that interpretability is feasible for detection requirements.

Legal requirements mainly refer to road regulations, which impose specific behavioral requirements for automated vehicles. This principle finds direct application in the development of classification requirements. In addition, it provides the basis for the behavioral requirements regarding relevance. Compliance with legal requirements must be ensured in any case for the driving task. However, behavioral requirements do not directly provide detection requirements. Therefore, further substantiation is required. The results of this work show that applying the remaining common principles sufficiently bridges this gap.

Safety requirements are implicitly included in legal requirements. As with legal requirements, safety is only well-defined with respect to safety outcomes on a behavioral level. For classification, a coarse categorization based on existing traffic studies was introduced to refine sub-categories of obstacles. While safety outcomes require consideration of the future, the corresponding uncertainty is successfully addressed using worst-case assumptions. This principle is applicable when formalizing possible trajectories as equations for relevance estimation. The safety principle is essentially sufficient to formally specify relevance. Safety is also incorporated into attribute requirements when defining the object reference. Furthermore, conservative estimates are demanded for various attributes to incorporate safety. However, classification, the parametrization of the relevance model and attribute requirements still require further specification from an additional principle.

The common principle of the human baseline is frequently applied throughout this work. It is applied for all aspects which are insufficiently specified when considering road regulations and safety aspects. Firstly, they are applied to the classification requirements to increase interpretability. While the classification requirements are applicable without this consideration, it improves the practical application for typical labeling approaches. For the relevance estimation, the human baseline is applied for the parametrization of the minimum planning specifications. Both system latency and available accelerations are chosen to match their human counterparts. The most prolific use of the human baseline is found in the attribute requirements. Various aspects considering association, temporal aspects of tracking, the choice of attributes and the thresholds to define requirements are based on human performance. The validation approach also heavily relies on the human baseline for driving behavior.

Overall, the proposed solution includes the common principles for each aspect. The different research questions merely differ regarding the degree to which each principle is applied. This indicates that the selected principles generalize across different aspects and are practical. Combining the common principles is sufficient to develop the individual methods for each of the aspects considered in this work.

9.1.2 Overall Methodology

This section only contains the discussion of the overall methodology followed in this work. Discussion on the details of different aspects, the validation and the relation to the overall RQ is left for later sections.

In this work, an argumentative approach to develop the detection requirements is proposed. Most notably, the results show that it is possible to derive detection requirements based on an argumentation. All requirements consider the context of the driving task while retaining interpretability. Since they are based on broad common principles, they generalize across different scenarios in a large-scale public dataset. When evaluating a detection component, these requirements are independent of the specific downstream implementation. Furthermore, no complex neural network is required for the application of the evaluation. Nevertheless, a neural network is applied for the validation as is explained in the following. This offers benefits regarding the availability of an implementation[521] as well as regarding computational effort when applying the requirements. In addition, the approach avoids the problems associated with the lack of robustness which neural networks typically exhibit[522,523]. However, these requirements based on an argumentation require further validation by a different approach.

For validation, this work proposes to leverage an approach based on a DNN for prediction. Predicting human trajectories refers to the human baseline instead of simply using a specific implementation of a planner. Utilizing a motion prediction component also circumvents the ambiguities of the planning task. Standard motion prediction metrics are utilized to evaluate the prediction component. Results empirically verify the presence of noise in the non-deterministic predictions used in this work. However, it is important to note that future trajectories are inherently uncertain. Any single GT trajectory is therefore considered to be a sample drawn from an underlying unknown distribution. Any local prediction performance contains the scenario context as well as a stochastic component. For an unknown distribution, these two components cannot be distinguished. This means that local stability of trajectories as applied by PKL[524] is not meaningful. Accordingly, the prediction evaluation is applied globally across an entire dataset in this work. The sufficiently large sample size ensures the results are not influenced by local stochastic effects. Furthermore, considering the error distribution across a large number of samples is also robust to the local performance issues that DNNs may exhibit. It should be noted that any observed differences in predictions cannot be explained due to the black-box nature of DNNs. Therefore, this approach is only used to validate interpretable requirements obtained from another source.

This validation approach effectively reconciles the analytic requirements with a DNN-based approach. Firstly, it is demonstrated that obtaining simple interpretable requirements based on an argumentative approach is feasible. Furthermore, it is possible to validate these requirements with a context-aware motion prediction DNN. The requirements are thus substantiated by evidence from two independent sources. The agreement of two different sources allows gaining confidence in the results.

9.1.3 Limitations

Despite the encouraging results, the proposed methodology naturally also has limitations.

The methodology relies upon the German StVO[525] as exemplary road regulation as a basis for legal requirements. Therefore, no definitive conclusion regarding the transfer to other road regulations can be drawn. For the classification, specific terminology and regulations of the German StVO

[521] Wolf, M. et al.: Safety-Aware Metric for People Detection (2021), p. 2760.

[522] Houben, S. et al.: Inspect, Understand, Overcome (2022), p. 6.

[523] Willers, O. et al.: Safety Concerns and Mitigation Approaches (2020), p. 341.

[524] Philion, J.; Fidler, S.: Lift, Splat, Shoot (2020), p. 14057-14058.

[525] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013).

structure the final result. The relevance results also rely on the road regulations, but mainly on implicitly specified safety objectives. Accordingly, the specifics of the German StVO are weakly expressed in the final result. It can be observed that other traffic regulations such as the California Driver's Handbook[526] and the Japanese Road Traffic Act[527] include similar safety objectives regarding collisions. Thus, a transfer of the relevance results to other regions and road regulations seems plausible. This is also supported by the validation of the relevance results on the nuScenes dataset, which is recorded in Singapore and Boston[528]. However, this aspect requires further validation. In addition, a direct transfer of the classification requirements seems less likely. If differences between regulations emerge, it is currently not known if general requirements encompassing multiple road regulations are possible.

The results for this work are aimed at the urban domain. It should be noted that the results do not include any assumptions which are specific to the urban domain. Nevertheless, the results and validation results are only obtained for one urban dataset. Study on other datasets is warranted to ensure that no overfitting to the dataset content is present. Most importantly, the dataset only includes normal driving situations. This is of note when considering the relevance methodology which is also designed for driving maneuvers at the physical limits or in hazardous situations. However, since these extreme maneuvers in proximity of accidents are not contained in existing public datasets, they cannot be validated. More generally, the data-based validation is limited by available datasets. Relevant limitations for this work include the number of classes and available annotation ranges.

Overall, the limitations center around the two aspects of the generality of the legal framework and the available data for testing. The transfer to other legal frameworks and other data cannot be guaranteed. Nevertheless, the methodology is in principle applicable to these cases, warranting further investigation.

9.1.4 Summary

In this section, a brief summary for the discussion concerning the overall methodology is provided.

Firstly, the overall RQ was decomposed into different aspects. For each of these aspects, a combination of the common principles was utilized to derive an answer. It was shown that the suggested approach is feasible to develop object detection requirements in driving context. The methodology is based on an argumentation incorporating interpretability, legal and safety requirements as well as the human baseline. The methodology was only investigated for a single legal framework and a single dataset. While a transfer to other regulations and datasets is plausible, this remains an open question. Nevertheless, the interpretable requirements were

[526] State of California Department of Motor Vehicles: Autonomous Vehicle Testing Permit Holders (2022).

[527] Ministry of Justice: Road Traffic Act (1960).

[528] Caesar, H. et al.: nuScenes (2020), p. 11620.

successfully validated. The validation relies upon the complementary approach of leveraging a DNN component in the form of a motion prediction. Thus, the interpretable requirements are supported by an independent validation procedure. To the knowledge of the author, this is the first reconciliation of this kind which generalizes across a whole large-scale perception dataset. Before answering the overall RQ, the following sections first discuss the other RQs 1-4.

9.2 Validation

In this section, the validation method as well as its verification are discussed. This allows a joint discussion of the proposed object detection requirements and their validity in the following sections. Finally, the conclusion of this section provides an answer to the RQ4.

9.2.1 Validation Method

The high-level aspects of the validation method relating to the overall methodology of this work were previously discussed in section 9.1.2. Rather, the focus here lies on more detailed aspects of the specific choices regarding the validation method.

Prediction Network

The validation method of this work utilizes a neural network for motion prediction. One natural question is which requirements apply to the motion prediction network. As opposed to PKL which develops a custom network[529], no specific requirements apply. The method is applicable without modification for any motion prediction which outputs a set of potential trajectories. This flexibility is exemplified for the motion prediction applied in this work which provides discrete, non-deterministic prediction trajectories.

As discussed in section 3.3.2, one potential problem with such an approach is the validity of the DNN itself. The proposed approach has the advantage that the performance is explicitly quantified as distribution of prediction performances. For the presently applied prediction, the global performance exhibits an average error of 0.96 m. For this performance, it is possible to apply accuracy requirements for the prediction quality. However, there are currently no such requirements available. This means that it is currently difficult to ascertain if the prediction performance of an algorithm is sufficient. Nevertheless, the prediction performance at minimum provides a ranking between two different prediction algorithms. Higher prediction accuracy should be favored if two algorithms are available, especially in the case of disagreements. For this work, a state-of-the-art prediction network at the time of implementation was selected. Essentially, this means that for currently available prediction quality, a falsification of the object detection

[529] Philion, J.; Fidler, S.: Lift, Splat, Shoot (2020), p. 14057-14058.

requirements was not possible. Results may differ for other prediction algorithms, especially if more accurate algorithms become available in the future. However, the transfer of the results to other available predictions was not explicitly tested and is left for future work.

To sum up, the validation method is flexible with respect to the motion prediction network. Since prediction performance is explicitly evaluated, the accuracy of the motion prediction is explicitly considered as part of the validation procedure.

Application

Where possible, the validation method is applied to the requirements of this work. However, not all requirements can be validated by the proposed method due to limitations of the available data and the prediction network. This includes the classification requirements as well as some of the attribute requirements. Therefore, the validation method is only applied to the relevance results and to the localization requirements for the attributes. Details are discussed in later sections.

Limitations

As already indicated in the previous sections, the validation procedure is limited by the available data and the prediction network. The validation of the classification is not possible since current datasets do not provide annotations for the classes suggested in this work. In addition, the prediction network only processes a limited number of classes.

For the validation of the relevance, no critical situations are evaluated due to their absence in the dataset. Furthermore, the validation only evaluates false negatives where relevant objects are erroneously removed. Among those objects declared relevant, the validation cannot distinguish truly relevant objects from false positives which are in fact irrelevant. This is in line with the conservative estimates of this work. However, improving and testing the specificity of the relevance remains an open question left for later work. Another limitation of the relevance is that it is currently only evaluated within the fixed region of interest of the prediction network.

Regarding attributes, only location is validated since the validation is trivially applicable. For the velocity errors, location and velocity must be consistently manipulated to achieve physically plausible trajectories. How to perform this manipulation is currently unclear. Furthermore, the validation is limited by the inputs considered by the prediction implementation. Attributes such as size or orientation are currently not considered by typical prediction components. Therefore, validating these attributes requires prediction algorithms which leverage these attributes with sufficient sensitivity.

Yet another limitation is the fact that safety outcomes are not explicitly considered. The prediction component only allows an evaluation of the influence on human trajectories. More specifically, the validation approach evaluates whether a change in the input is likely to yield changes in human driven trajectories. This means that safety is only implicitly considered in the results to the degree that is reflected in human driving behavior. In particular, the results in section 8.2.4 show that conservative estimates are not supported. This means that a nonconservative estimate of attributes such as distance does not lead to a change in human driving behavior. However, this does not necessarily mean that this unchanged driving behavior selected by the human driving behavior is also safe. Furthermore, the significance of such nonconservative estimates likely also depends on the safety margins which the downstream planner adopts. How to determine if a given behavior is safe and how to integrate this into the validation procedure remains an open question.

Overall, most of the current limitations do not concern the validation method but the current implementation. Given more diverse datasets including critical situations and predictions which consider more different types of input is likely to alleviate these issues. Nevertheless, the validation method is already applicable for the relevance and localization requirements.

9.2.2 Verification

Since a novel validation method is presented, it is necessary to provide a verification. In this section, a discussion of the verification experiments performed in this work is provided.

The verification is attempted with four different inputs. Each input is constructed to contain implausible modifications to the object list. Therefore, each modification provides an invalid input to the prediction network. Invalidity is based on intuitive notions such as the importance of the existence of objects in front of the vehicle or the importance of large distance errors. Firstly, the results show that the impact on prediction as visualized by the distribution functions is visually difficult to distinguish. The reason is that the prediction network also relies on lane information. Most likely, many driving situations are influenced by lane geometry and driver preference more than by other traffic participants. Despite the visually subtle differences, the proposed statistical analysis is able to clearly resolve these. The resulting p-values plotted as boxplots successfully identify all verification inputs as invalid. For the relevance, it is verified that the validation method is able to resolve small differences. Even the filtering of 5% of objects in one verification scenario is falsified. This indicates that the validation method is sufficiently sensitive to small numbers of objects. The number of 5% is considered because the relevance method removes 10% of the input objects.

However, it is currently not clear what the minimum differences are which the proposed validation method can resolve. Firstly, it is unclear how large the difference in p-values must be to consider an input invalid. However, it should be noted that most of the results of this work achieve orders of magnitude difference in p-values. Therefore, the necessity to argue with minute differences does not arise. Secondly, determining the sensitivity is complicated by the fact that the corresponding verification inputs are difficult to define. Since the current verification inputs represent large variations of the input, judging them as invalid is possible. This becomes increasingly difficult as the invalidity of the verification inputs becomes more subtle. Further improvements to the validation with regards to sensitivity may be possible. Creating subsets of data with specific corner cases may better resolve small differences in perception performance.

While the exact sensitivity of the validation method is unknown, it is sufficient for the verification inputs of this work. Overall, the verification of the validation is successful and supports the proposed validation method.

9.2.3 Conclusion

The key points from the discussion regarding the validation procedure are summarized in this section. Finally, an answer to RQ4 is provided.

Overall, the validation method leveraged a motion prediction algorithm to assess the impact of detection errors. The approach is agnostic to the design of the prediction network. Explicit consideration of prediction performance allows comparing different motion prediction algorithms. Evaluating globally across an entire dataset avoided difficulties arising from nondeterminism and local performance issues that DNNs may exhibit. Since the impact on the prediction performance was often subtle, a statistical evaluation was required. While the novel method currently still possesses some limitations, future datasets and prediction algorithms are likely to alleviate these issues. The only remaining principle limitation of the proposed validation approach is the lack of explicit consideration of safety outcomes in the prediction task. The validation only considers safety implicitly through the impact on human driven trajectories. At the same time, it was not explicitly evaluated if the impact or lack thereof is actually safe. Nevertheless, the validation is already sufficiently practical to validate the relevance and the localization requirements. Testing the validation with artificially constructed implausible inputs successfully verified the validation procedure.

To conclude the discussion of the validation, RQ4 is considered:

How can detection requirements be validated with respect to the safety of the driving task?

This work showed that a validation relying on a motion prediction network trained on human behavior is possible. Potential errors and nondeterminism in the motion prediction DNN required a global evaluation with statistical methods. Since the validation method is not interpretable in itself, it is most suitable for application to simple interpretable requirements.

9.3 Object Detection Requirements

In this section, results for the different types of requirements are discussed separately. Classification, relevance and attributes are each considered in the following. Each section begins with a general discussion of the method and results regarding each aspect. Furthermore, a comparison with prior approaches such as those found in datasets is provided. This is followed by additional content specific to each aspect and the limitations of the results. Each section concludes with an answer to the respective RQ 1-3. Finally, this section is concluded with a summary and answer to the overall RQ.

9.3.1 Classification

This section presents a general discussion followed by a comparison with datasets which represent the common approach prior to this work. After discussing the severity of misclassifications, some limitations of this work are outlined. Finally, a conclusion regarding the RQ1 is provided.

General Discussion

In this work, the common principles are applied to obtain requirements for classification. This section provides general discussion of the classification method including the validity. Finally, the research question is discussed.

To derive the required categories, each of the common principles is applied in a structured manner. These requirements are represented by a hierarchy of classes which must be distinguished for the task of driving. The minimum number of categories is defined by the legal categories. One benefit of these abstract categories is that they are applicable to any object which is encountered. This is conceptually similar to the task of detecting out-of-distribution objects or obstacles. For these obstacles, the exact type is typically not further classified.[530,531,532,533,534] Furthermore, the proposed requirements provide additional categories encoding safety. Accident statistics may be used to further substantiate the division according to severity and likelihood of collisions. However, this is left for future work. Finally, human perception is considered to yield interpretable categories which humans can readily perceive.

Current publicly available datasets do not provide annotations for the classification requirements of this work. Accordingly, it is at present not possible to apply the validation method of this work. While further validation is required, the classes are substantiated by an explicit argumentation incorporating the safety of the driving task.

Comparison with Datasets

A comparison shows that the classification structure proposed in this work deviates from existing dataset categories.

[530] Pinggera, P. et al.: Lost and Found (2016), p. 1102.

[531] Geyer, J. et al.: A2D2: Audi Autonomous Driving Dataset (2020), p. 7.

[532] Blum, H. et al.: Fishyscapes (2019), p. 2405-2406.

[533] Chan, R. et al.: SegmentMeIfYouCan (2021), p. 3.

[534] Metzger, K. A. et al.: A Fine-Grained Dataset (2021), p. 7894-7899.

As shown in section 5.3.2, this work identifies 45 classes originating from legal and safety requirements. Furthermore, section 5.3.3 introduces 80 additional categories based on human perception, which increases the overall number of categories to 125. The number of categories is therefore higher than in typical driving datasets such as Mapillary Vistas[535] with 66 segmentation classes or nuScenes with 23 detection classes[536]. Since traffic signs, markings and traffic installations are neglected thus far, the final number of categories is expected to be higher. For instance, the Mapillary traffic sign detection dataset distinguishes 313 different traffic signs.[537]

Neither the dataset classes nor the results of this work are validated. However, the dataset categories are apparently arbitrary and fail to provide any argumentation for their selection. Accordingly, legal and safety requirements are not considered in the dataset class definition. It is therefore likely that the explicit consideration for classification improves the safety of the driving task.

Severity of Misclassification

One question regarding the classification structure is if the severity of class confusions is encoded.

It should be noted that the edges of the graph containing the categories differ in their meaning. Edges may imply a distinction due to legal, safety or perception requirements. Since these different aspects are difficult to compare, severity is only captured to a limited degree. Even for different edges relating to legal categories, the severity may differ. A higher distance in the tree corresponds to a larger number of legal requirements which are transgressed. However, the severity of the transgression may differ depending on which node is considered.

Even if severity is not fully captured in the hierarchy, it may be possible to identify conservative class estimates. For such classes, the conservative estimates may be selected if uncertainty is present for the classification. In case of the legal categories, potential candidates are legal categories which only have a single legal category as child node. Within the entire hierarchy, only four such cases including motorways and built-up areas are present. However, in case of the provided examples, a confusion may lead to a severe misjudgement regarding the permissible velocity. Since the resulting potential harm to other traffic participants is not acceptable, only two conservative estimates remain. This shows that most misclassifications are critical since they indicate a transgression of rules. It should be noted that a confusion of different perceptual categories within the same legal and safety category is acceptable. An example visible in the classification structure from section 5.3.3 are the perceptual categories “Taxi” and “Truck” which are visually distinct. However, both classes belong to the legal category “vehicle” so that the same traffic rules apply for interaction with both of these perceptual categories. Regardless, the

[535] Neuhold, G. et al.: The Mapillary Vistas Dataset (2017), p. 5002-5003.

[536] Caesar, H. et al.: nuScenes (2020), p. 11621.

[537] Ertler, C. et al.: The Mapillary Traffic Sign Dataset (2020), p. 6.

fact that conservative class estimates are rare remains. Therefore, hierarchical losses may not be required for adequate classification.

Limitations

The current classification hierarchy still has limitations regarding specific classes, attributes and relations.

For the legal structure, different sub-classes for traffic signs, markings and traffic installations are currently not distinguished. However, their recognition is typically considered separately[538,539] and is not related to the task of collision avoidance considered in this work. Furthermore, these categories are well specified in road regulations[540] and are in parts also included in previous datasets[541,542]. While an integration of such classes into the present structure is therefore likely straightforward, this is left for future work.

However, another aspect limiting traffic signs, markings and traffic installations is present. These categories rely on relations and attributes, which are not present in the current structure. For the task of AD, attributes depending on the object category are relevant.[543] An example is the state of a traffic light which is available in the DriveU Traffic Light Datastet.[544] Furthermore, the relations between different objects and attributes may also be relevant. An example is the interaction between traffic signs, lights and police officers.[545] One previous work incorporates interactive attributes between objects into a pre-existing dataset.[543] Other works formalize semantic and spatial relations for objects, the road and infrastructure elements.[546,547] Yet another approach is the behavioral semantic scenery description which directly considers the behavioral space.[548] Overall, different approaches to consider attributes and relations are available. While an integration into the proposed ontology may be possible, this is left for future work.

Conclusion

Regarding the classification it was demonstrated that it is possible to obtain classification requirements in a structured manner. The novel categories differ substantially from existing datasets, hindering the validation of the proposed categories. Nevertheless, the proposed classification

[538] Ertler, C. et al.: The Mapillary Traffic Sign Dataset (2020), p. 8-9.
[539] Chen, J. et al.: A real-time and high-precision method for small traffic-signs recognition (2022).
[540] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), p. 38-81.
[541] Huang, Y.; Chen, Y.: Autonomous Driving with Deep Learning (2020), p. 2706-2707.
[542] Ertler, C. et al.: The Mapillary Traffic Sign Dataset (2020), p. 6.
[543] Metwaly, K. M. et al.: CAR - Cityscapes Attributes Recognition (2021), p. 2-3.
[544] Fregin, A. et al.: The DriveU Traffic Light Dataset (2018), p. 3379.
[545] Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013), p. 20-21.
[546] Buechel, M. et al.: Ontology-based traffic scene modeling (2017), p. 1473-1474.
[547] Karimi, A.; Duggirala, P. S.: Formalizing traffic rules for uncontrolled intersections (2020), p. 44-45.
[548] Glatzki, F. et al.: Behavioral Attributes for a Behavior-Semantic Scenery Description (2021), p. 669-670.

requirements provide an argumentation incorporating legal and safety requirements. Despite the hierarchical structure, severity of misclassifications is only encoded to a limited degree. Furthermore, only few categories allow conservative class estimates. The classification hierarchy is currently limited with respect to traffic signs, markings and installations as well as relations and attributes. While an integration into the classification structure may be possible, this is left for future work.

Finally, as a conclusion to the discussion of the classification requirements, RQ1 is considered:

Is it possible to systematically identify categories which must be distinguished to safely perform the driving task?

Currently, the answer to this question provided by this work is a tentative yes. The results demonstrate that it is indeed possible to provide a structured approach to identify categories. Both legal requirements and safety were explicitly considered in an interpretable way. However, the current lack of validation prevents providing a more definitive answer. While a validation with the method proposed in this work is conceptually possible, it requires dataset annotations reflecting the categories of this work. Furthermore, prediction networks which leverage such class information are required.

9.3.2 Relevance

In this section, a discussion of the relevance requirements is presented. First, a general discussion including the validity is provided. This is followed by a comparison with common datasets and the limitations of the results. Finally, this section is concluded with a discussion of RQ2.

General Discussion

The proposed approach first defines the concept of relevance specifically for the task of object detection. Furthermore, relevance is developed based on an understanding of behavioral requirements in driving context. By considering uncertainties and a minimum system specification, the safety of the driving task is taken into account. Overall, the approach is successful in identifying relevant objects. The approach is conservative by design, designed to exclude objects which can confidently be declared irrelevant.

As discussed for the results in section 6.3, 10% of the objects are filtered out by the relevance criteria of this work. Nevertheless, the validation procedure indicates that the prediction is unaffected if irrelevant objects are discarded. The p-values between verification inputs and the proposed relevance method differ by orders of magnitude. Furthermore, the p-values for all inputs show similar boxplots as when comparing with only the relevant inputs. Therefore, the validation supports the proposed relevance criteria. This validity is observed even though the validation method is sufficiently sensitive to clearly distinguish the verification scenario which removes 5%

of the objects. Validity is only ensured with respect to the conservative assumptions, meaning that false negatives are avoided.

The distance distributions of different scenarios shown in section 6.3.3 are generally similar. However, the R.TT and the T.XT scenario exhibit the largest distances. All scenarios are more conservative when compared to a TTC threshold of 4 s obtained from literature[549]. This is expected since the relevance is constructed in a conservative manner. Furthermore, the distances of relevant objects are large. Depending on the scenario type, distance values above 100 m are obtained. This fact is further discussed in the following section regarding the comparison with datasets.

Comparison with Datasets

In this work, relevance is considered as an attribute of an object. This conceptually differs from common datasets which use heuristics including distance and visibility.

As shown in the previous sections, relevant objects at large distances of above 100 m are observed. This indicates that common distance thresholds ranging between 30 m[550] and 75 m[551] are insufficient. Overall, more explicit consideration of the annotation range of datasets is required. It should however be noted that relevance as conceptualized in this work is not a geometric region but a property of an object. Transferring the results to determine a relevant region requires assumptions regarding the attributes of the objects. However, how to argue these assumptions is at present unclear. Therefore, this question is left for future work.

The fact that relevance is a property of the object also has implications beyond the annotation distance. A noteworthy fact is that objects may be relevant despite not being visible in the sensor data. Naturally, this case does not occur in datasets which annotate based on visibility in sensor data. Nevertheless, it is possible that this typical annotation approach fails to include objects which may be relevant to the ego vehicle. The implications of the relevance concept of this work for sensor setups and annotation procedures of datasets are discussed in section 9.5.

Limitations

Despite the encouraging results, the relevance results exhibit some limitations.

Firstly, only 10% of objects can be excluded from the data based on relevance. It is currently not known if the specificity of the relevance model can be improved while maintaining validity. Furthermore, no approaches to evaluate specificity are available. The current reference to the annotated objects is limited by the annotation procedure. Furthermore, relevance is conceptualized as the possibility of a collision. It does not determine which objects directly influence the behavior

[549] Mahmud, S. S. et al.: Application of proximal surrogate indicators (2017), p. 155-156.

[550] nuScenes: nuScenes Detection Task: Leaderboard (2020).

[551] Sun, P. et al.: Scalability in Perception (2020), p. 2447.

in a given situation. Defining and improving specificity is likely to deliver a more complete picture of relevance.

Besides the general limitations to the validity, the relevance model also includes parameters. These parameters are required for application of the relevance model to real world data. The parameter values selected in this work are validated together with the relevance method. However, the parameter values of reaction time and braking accelerations may differ in some situations. For instance, icy or oily road surfaces may reduce the available braking acceleration. Further argumentation and validation is required to select parameter values under such conditions.

Conclusion

The relevance criteria of this work are successful in providing a conservative estimate of relevant objects. Leveraging a novel validation methodology, the relevance results were shown to be valid. An analysis of relevant objects revealed that these occur at large distances of above 100 m. In addition to the large ranges, even objects which are not visible in sensor data may be considered relevant. This indicates that more explicit consideration of relevance by datasets is required. Limitations of this work include the limited number of 10% of objects considered irrelevant.

Finally, a conclusion is presented for RQ2 regarding relevance:

Is it possible to systematically identify objects relevant for detection based only on information contained in the object list while considering the safety in driving context?

The results obtained obtained in this work are sufficient to answer the question with yes. The proposed relevance method exclusively relies on information contained in the object list. At the same time, the approach presented a systematic method to identify objects relevant for detection. The safety of the driving context was explicitly considered in the argumentative approach. Furthermore, the driving context is also present in the validation procedure which successfully confirms the relevance requirements.

9.3.3 Attributes

In this section, requirements for the attributes of objects are discussed. After an initial general discussion considering both the validity and the initial RQ, a comparison with existing detection metrics is performed. This is followed by a performance analysis of standalone detectors as well as the idealized fusion algorithms. Next, limitations of the proposed attribute requirements are presented. Finally, a conclusion regarding RQ3 is provided.

General Discussion

The method proposed in this work is able to derive requirements for different attributes. Different attributes are each treated separately. Human detection errors from different literature sources show sufficient agreement to provide simple requirements. Plausible requirements are obtained for the attributes of localization, velocity, size and orientation. In addition, the whole evaluation pipeline is reconsidered including the matching operation. Furthermore, aspects of the tracking task are also considered for the requirements. The result are interpretable requirements which allow independent evaluation of different attributes. All requirements are substantiated by an argumentation based on the common principles of safety and human performance.

The validation procedure is applied to the localization requirements in this work. Different assumptions of this work are validated successfully. Firstly, the distinction between angular and distance errors is valid according to its impact on the prediction. The direction or sign of the error also exhibits an impact on the prediction. While distinguishing the direction of the error is supported, conservativity is not. Rather, nonconservative errors appear to have a smaller impact on the prediction. One reason for this result is most likely the failure of the validation method to consider safety outcomes. The validation uses a prediction which quantifies the impact on a human driven trajectory. However, this does not necessarily mean that the impact or lack thereof on the human trajectory is also safe. Another potential reason is the fact that the full impact of conservativity is only observed in near-accidents which are not included in the dataset. The validation further supports to scale location errors with the distance from the ego vehicle for both angle and distance. For the distance, the proposed distance threshold of 15% error is confirmed by the validation. This indicates that the general approach of using human errors from literature other than for driving context is a feasible approach. Furthermore, it supports the process of using interpretable error thresholds as simple upper bound on human detection performance. The angular error threshold of $5°$ is not supported by the results. It appears that the angular error thresholds suggested by the prediction are more restrictive than previously proposed. One potential reason is that the angular errors found in literature provided no data below $3°$. If the error thresholds are smaller for small angles, this may be sufficient to explain the validation results. Another possible reason is that humans rely upon relative distances to points of reference such as the lane. In this case, an estimation of absolute angle as assumed previously in this work may not be an adequate perceptual model. Further research is required to investigate the reason and develop a valid alternative requirement.

Comparison with Previous Metrics

The proposed requirements provide distinct requirements for different attributes. This is in contrast to typical evaluation procedures which tend to emphasize metrics aggregating different attributes. However, standard evaluation pipelines use association thresholds based on localization requirements. The nuScenes dataset considers a GT object to be correctly matched if an object detection is within a specified radius from the GT object center. While this general approach is retained for interpretability, the threshold is modified for this work. For evaluation, the nuScenes dataset applies different radii for matching, among which the most permissive radius is 4 m.[552a] However,

[552] Caesar, H. et al.: nuScenes (2020), a: p. 11622, b: p. 11621.

the results in section 8.2.4 show this constant radius to be invalid. Similarly, various assumptions which typical detection and detection evaluation pipelines follow are falsified. The validation results clearly show that different types and directions of error require separate consideration. Despite the higher validity of the results of this work, requirements are more permissive. This means that higher average errors in direction of the radial distance are shown to be valid than commonly assumed. Potentially, previous benchmarks have overemphasized precise localization in depth direction. Furthermore, this work derives thresholds for attributes other than localization. These other attributes have so far received comparatively less attention by benchmark evaluation metrics.

Analyzing human performance leads to the requirement of detecting an object within 150 ms. This is comparatively strict when considering existing datasets. For nuScenes, the time between two annotated frames is 500 ms.[409b] Therefore, it is not possible to study or to meet the temporal requirement of 150 ms on this dataset. Studying such temporal requirements requires higher annotation frequencies. While the maximum annotation frequency is also limited by the scan rate of 3D sensors such lidar and radar, datasets with annotation frequencies of 100 ms are available[553,554].

The importance of classification is diminished for the suggested evaluation procedure compared to mAP which considers it a prerequisite for matching. However, aspects related to classification can only be tested to a limited degree due to limitations in available annotated data. Even though classification is not required for matching, it is still part of the requirements. The classification required for the driving task was discussed in section 9.3.1. For current object detectors, it is observed that perception errors are common even when classification is no longer required for matching. One potential reason for this observation is an overfitting of current detectors to prevalent detection metrics.

Detection Performance

The detection baselines evaluated in this work exhibit frequent failures. On average, FCOS3D obtains 0.82, Pointpillars obtains 0.76 and CenterPoint obtains 0.72 collision relevant failures per GT object. Unlike other performance metrics, lower is better for the failure likelihoods defined within this work. The numbers mean that multiple failures are present in every frame. Matching failures are the dominant failure type despite more lenient matching thresholds than commonly applied. Furthermore, all detectors disregard conservativity, which may pose safety risks. Overall, the object detection requirements are not met by current detectors. This provides evidence for the absence of safety considerations not only in common evaluation protocols, but also in detectors.

In addition to an evaluation using the novel requirements, the performance is also discussed regarding mAP. The mAP values obtained by the detectors are 32% for FCOS3D, 34% for

[553] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3355.

[554] Sun, P. et al.: Scalability in Perception (2020), p. 2448.

PointPillars and 56% for CenterPoint[555]. The ranking between the different detectors is identical for mAP and the proposed requirements. It is possible to compare the discrepancies betweeen the metrics if linear scaling assumed. In this case, the performance gap between PointPillars and CenterPoint is larger on mAP than for the requirements of this work. Potentially, this observation is explained by the assumption that CenterPoint provides a superior fine-grained localization. Conversely, the performance gap between PointPillars and FCOS3D is larger in this work than when evaluating mAP. A possible explanation is the fact that classification is not among the collision relevant requirements of this work. Camera images generally contain richer semantic information compared to lidar.[556,557] Assuming that FCOS3D is capable of leveraging this information, this may explain why better performance is shown for mAP which emphasizes classification.

The present work proposes to evaluate for a single optimum confidence threshold. Each detector has a different optimum confidence score for the evaluation. It has previously been observed that confidence scores are not sufficiently calibrated.[558,559] However, this calibration of confidence scores is not a prerequisite for the proposed requirements. Contrary to metrics such as mAP, confidence scores can be discarded entirely. Only a single object list is required as evaluation input. How to leverage this fact along with other specifics of the proposed requirements for detectors is a question left for future work.

Fusion Performance

To study correlation in fusion, two ideal fusions are proposed in this work. The proposed ideal fusion has access to GT information in a late fusion scheme. While this is not a realistic assumption for practical application, it provides an upper bound of fusion performance. This fusion is compared with the uncorrelated case. For this mathematical model, failure likelihoods are directly calculated based on the assumption of zero correlation. The ideal fusion generally shows potential for large improvements over the single detector baselines. However, it is likely that these performance gains cannot be fully realized for practical fusion algorithms. Nevertheless, even this ideal fusion shows higher failure likelihood than the uncorrelated case for likelihoods below 10%. This means that even an idealized fusion is still affected by correlation of failures. Such correlations can originate from common causes for certain failures.[560] In the present case, factors such as occlusion or size may similarly affect different detectors. Furthermore, other external factors such as adverse weather may further increase these correlations. The results

[555] MMDetection3D Contributors: MMDetection3D (2020)

[556] Fei, J. et al.: SemanticVoxels (2020), p. 185.

[557] Liu, Z. et al.: BEVFusion (2023), p. 2774.

[558] Willers, O. et al.: Safety Concerns and Mitigation Approaches (2020), p. 341.

[559] Kato, Y.; Kato, S.: A Conditional Confidence Calibration Method (2022), p. 1835.

[560] Stapelberg, R. F.: Handbook of Reliability (2009), p. 621-623.

show that naively assuming correlation to be absent is not justified. Similarly, prior work has also shown substantial correlations may occur for more theoretical settings.[561]

One common assumption is that utilizing different modalities is able to minimize the correlation of errors. However, no substantial difference between combinations of the same or of different modalities are observed. These results are of course limited by the camera and lidar algorithms studied in this work. Therefore, this particular aspect requires further investigation. Another potential for reduction of common causes is provided by utilizing other sensor types. Potential candidates of other sensors which are currently understudied in public datasets include frequency modulated continuous wave lidar, high-resolution radar or thermal cameras.

Limitations

While the attribute requirements are based on an argumentation and are partially validated, limitations are still present.

Firstly, not all attributes for which requirements are put forth are in fact validated. Specifically, the velocity, size and orientation requirements are not validated in this work. Furthermore, this work proposes a unified consideration of tracking and detection. However, these aspects are currently insufficiently considered in datasets. Since the annotation frequencies of current datasets are insufficient, the validation method is not applicable.

While the localization requirements were validated, not all requirements could be confirmed. The validation procedure does support to distinguish positive from negative errors. However, the conservative property is not validated. A potential reason is the symmetric distance metric used to evaluate the prediction. Therefore, safety outcomes resulting from these deviations are not considered.

Conclusion

The results of this work showed that it is possible to quantify human detection for different attributes separately. Furthermore, the detection evaluation pipeline was reconsidered regarding the matching operation and the reconciliation with the tracking task. The resulting requirements showed disagreement with dataset metrics. However, the validation procedure showed that requirements currently applied by datasets are invalid. At the same time, proposals of this work such as separating different error types and directions as well as scaling thresholds with distance were supported. While the distance threshold of this work was validated, the conservative estimates and the angular threshold were not. Overall, the requirements of this work are more valid despite being more permissive than previous requirements. Applying these novel requirements to existing detectors showed that these exhibit frequent failures. Failures persisted even for an idealized fusion, demonstrating that errors are correlated even for different detectors and modalities. This result questions the capability of sensor setups relying on fusion to fully alleviate detection failures.

[561] Gottschalk, H. et al.: Does Redundancy in AI Perception Systems Help? (2022), p. 100-101.

Finally, RQ3 regarding attributes is considered:

Is it possible to define requirements for the detected object attributes considering legal and safety requirements for the driving task?

With the results of this work, the initial RQ is answered with a tentative yes. The results demonstrate that it is possible to define requirements for the detected object attributes. Legal and safety requirements were explicitly addressed in the method. Furthermore, the context of driving task was also included into the validation method. While the possibility of obtaining valid attribute requirements was demonstrated, further refinement and validation is required to extend the results to all collision relevant attributes.

9.4 Overall Research Question: Safety

In this section, a conclusion of the discussion regarding all object detection requirements of this work is presented to answer the overall RQ regarding safety.

Overall, it was shown that this work successfully derives requirements for classification, relevance and attributes. This shows that it is feasible to obtain fully interpretable requirements based on an argumentation. The argumentation is rooted in the common principles of legal and safety requirements as well as the human baseline. A comparison with datasets showed substantial disagreement across different aspects. However, for some of the requirements, a validation on public datasets was possible. In these cases, many of the assumptions present in dataset metrics were falsified. At the same time, the validation results largely supported the requirements proposed in this work.

Reviewing the RQs 1-3, each question was at least tentatively answered positively in section 9.3. Furthermore, RQ 1.3 was already positively answered in section 9.2.3. This means that presently, all RQs relating to individual aspects of the overall question are sufficiently answered. Thus, it is now possible to consider the overall RQ:

Is it possible to specify requirements for 3D object detection which consider the safety of the driving task without fully specifying a downstream planning module?

Summing up the answers for the RQs 1-4, the answer for the overall RQ is therefore also positive. The results of this work showed that it is indeed possible to derive requirements for 3D object detection. All requirements consider the safety of the driving task both in the argumentative approach and in the validation. In addition, the full specification of a downstream planning module was avoided. Instead, the proposed method leveraged the human baseline for the argumentation as well as for the validation. With this, all the research questions of this work are successfully answered.

9.5 Implications

So far, the discussion mainly focused on the methodology and the requirements pertaining to the RQs of this work. However, the resulting requirements obtained as answer to the RQs may also impact other related domains. The aspects of data annotation, sensor setups and algorithm development are each discussed in turn.

9.5.1 Implications for Data Annotation

Current detection pipelines generally leverage large scale data for training and testing. In fact, the data samples provide the requirements for data-based detection components. Therefore, the novel general detection requirements require changes in the creation of datasets. Both the classification and relevance require consideration during annotation.

The classification results showed that the proposed classes differ substantially from the classes commonly found in datasets. This shows that the safety of the driving task requires more explicit consideration. When defining dataset classes, an argumentation should be presented instead of selecting classes arbitrarily. Among the proposed classes, some classes such as locations are not boxable objects which are considered by object detection. Therefore, it may be necessary to treat these classes as a segmentation task. Such a segmentation may be performed as 3D semantic segmentation of lidar[562] or as segmentation of a birds-eye-view map segmentation[563]. The classes proposed in this work may be directly applied for future labeling purposes.

The relevance estimation impacts the objects which are required to be contained in the GT of a dataset. It should be noted that this aspect cannot be explicitly studied on current dataset since only objects contained in the GT can be evaluated. Nevertheless, the relevance of this work is conceptualized as property of an object with specific attributes. This indicates that simple heuristics such as circular regions of interest[564,565] are insufficient. Any attempt to define a region of interest requires assumptions regarding the object attributes. The author urges future dataset creators to explicitly define and substantiate any such assumptions. Another notable aspect is that relevant objects occur across the whole distance range of annotations. This implies that current annotation ranges may be insufficient. Furthermore, objects are considered relevant regardless of their visibility. This means that annotation procedures which solely rely on onboard vehicle sensors are not able to capture relevant occluded objects.

Similar to the classification requirements, attribute requirements may be directly applied. As done for the results in this work, the evaluation with existing annotations is possible. Since this is done

[562] Huang, Y.; Chen, Y.: Autonomous Driving with Deep Learning (2020), p. 2707-2709.

[563] Caesar, H. et al.: nuScenes (2020), p. 11621.

[564] Sun, P. et al.: Scalability in Perception (2020), p. 2447.

[565] nuScenes: nuScenes Detection Task: Leaderboard (2020).

at evaluation time, no further modifications during annotation are required. Conversely, it may be possible to simplify the data annotation when applying the novel requirements. For instance, the collision relevant attributes only require the closest point of an object. The annotation of such a point may be performed substantially faster than the annotation of the entire tight fitting bounding box. However, the temporal requirements elicited in this work indicate that current annotation frequencies may not always be sufficient. Annotations every 100 ms as already provided by some datasets[566,567] or even higher frequencies may be required to study and meet the temporal requirement of detection within 150 ms.

Overall, the new requirements proposed in this work require reconsidering dataset annotation. While the classes may be directly applicable, the relevance criteria may require substantial modifications to the annotation procedure by leveraging additional external sensors.

9.5.2 Implications for Sensor Setups

The previous section considered the annotation procedure. However, this step is typically preceded by the selection of a sensor setup for the dataset. In this section, the implications of the proposed requirements on the design of these sensor setups are discussed.

As discussed in previous section, relevance is conceptualized as property of the objects independent of visibility. Therefore, occluded or distant objects may be relevant for the ego vehicle. In real-world datasets, obtaining accurate annotations for occluded areas may be difficult. One option to alleviate this issue may be to utilize sensors external to the vehicle. While these approaches are not currently applied in large-scale sensor datasets aiming at 3D object detection, different approaches are put forth in literature. Datasets for various purposes have already explored leveraging additional vehicles[568], drones[569,570,571] or infrastructure sensors[572,573,574]. Incorporating these ideas into 3D detection datasets may be required to ensure that all relevant objects are annotated. An alternative may be to modify the environment representation. Approaches using occupancy grids allow to explicitly distinguish free areas from unobserved areas.[575] Applying conservative assumptions for potential objects in the occluded region as suggested by previous works[576,577] may be feasible.

[566] Geiger, A. et al.: Are we ready for Autonomous Driving? (2012), p. 3355.

[567] Sun, P. et al.: Scalability in Perception (2020), p. 2448.

[568] Xu, R. et al.: V2V4Real (2023).

[569] Krajewski, R. et al.: The highD Dataset (2018).

[570] Breuer, A. et al.: openDD: A Large-Scale Roundabout Drone Dataset (2020).

[571] Lu, D. et al.: CAROM Air (2023).

[572] Yu, H. et al.: DAIR-V2X (2022).

[573] Hetzel, M. et al.: The IMPTC Dataset (2023).

[574] Tang, Z. et al.: CityFlow (2019).

[575] van Kempen, R. et al.: Combined Registration and Fusion (2023), p.2.

[576] Poncelet, R. et al.: Safe Geometric Speed Planning (2020).

[577] Orzechowski, P. F. et al.: Tackling Occlusions & Limited Sensor Range (2018).

As indicated in the previous section, the temporal requirement of 150 ms proposed by this work may require an annotation frequency which is not present for all datasets. However, the annotation procedure typically leverages 3D sensors such as lidar and radar which provide a scan every 100 ms. Accordingly, the maximum annotation frequency is currently also limited by the maximum sensor frequency. For the appearance of an object, one scan every 100 ms means that the object needs to be detected immediately in the first scan to meet the temporal requirement of 150 ms. However, it is currently common to aggregate information from multiple frames for tasks such as detection[578,579,580] or tracking[581,582,583]. Therefore, it may be necessary to consider sensors with higher frame rates in order to meet the novel temporal requirements.

In addition, the failure rates of current detection algorithms were shown to be high for the requirements of this work. Furthermore, camera and lidar showed non-negligible correlations even for an idealized fusion. These findings may necessitate large modifications to existing sensor setups. Different options may be considered for future setups to potentially improve redundancy. Firstly, redundancy may be increased by adding other sensor modalities such as radar, FMCW lidar and thermal cameras. Furthermore, redundancy may be improved by increasing the number of redundant sensors. These approaches are both supported by the fact that many companies aiming at AD already apply additional modalities[584,585] and sensors[586,587,588,589]. However, it is currently not clear if and to what degree these modifications improve the redundancy of a sensor setup. Therefore, further research in this direction is required.

Overall, the findings of this work regarding relevance and redundancy in fusion both suggest deficiencies of existing sensor setups. Therefore, future sensor setups may require modifications to better address these issues.

[578] Wang, Y. et al.: Train in Germany, Test in The USA (2020), p. 4.

[579] Chen, Q. et al.: Every View Counts (2020), p. 4.

[580] Laddha, A. et al.: RV-FuseNet (2021), p. 7060-7061.

[581] Pang, Z. et al.: SimpleTrack (2023), p. 2.

[582] Weng, X. et al.: 3D Multi-Object Tracking (2020), p. 10361.

[583] Chiu, H. et al.: Probabilistic 3D Multi-Object Tracking for Autonomous Driving (2021), p. 14229.

[584] Aurora: Voluntary Safety Self-Assessment (2021), p. 17-18.

[585] nuro: Delivering Safety (2021), p. 9-10.

[586] ApolloAuto: apollo (2021).

[587] General Motors: Self Driving Safety Report (2018), p. 7.

[588] Pony.ai: Our Approach to Safety: Pony.ai Safety Report (2020) p. 7.

[589] Zoox: Safety Innovation at Zoox (2018), p. 13-15.

9.5.3 Implications for Detectors

Current object detection algorithms are typically designed for the prevalent benchmarks and their requirements. Therefore, different requirements may also require reconsidering detection architectures and training procedures.

It has been shown that the driving task requires a higher number of classes than present in current datasets. For object detection in 2D image detection, the influence of the number of classes is better studied. Some detection architectures such as class-specific convolutional heads in single-stage detectors are not suitable for higher numbers of categories regarding computational efficiency.[590,591] Furthermore, large-scale classification over high numbers of categories often requires dedicated architectures which differ from standard detectors. Previous work suggests separate classification heads[592,593] or even separate expert models[594,595]. Other works propose to modify the training losses to improve upon the typical softmax classification.[596,597] The stricter classification requirement may also impact sensor fusion. While current benchmarks[598,599] are dominated by lidar-based methods, camera is considered to be better suited for classification[600]. Therefore, stricter classification requirements may increase the benefit of incorporating information from camera. This is also supported by the fact that lidar-camera fusion is more present on the nuScenes leaderboard[598.] where more classes are distinguished than for Waymo[601,602a]. Overall, results from 2D detection imply that detection architectures for 3D detection may require modifications when scaling to a larger number of classes.

As indicated by the relevance criteria, relevant objects may occur at large ranges beyond 100 m. For current object detectors, it is known that their performance deteriorates at larger distances. Even on the Waymo Open dataset with denser lidar points[602b], detectors exhibit lower performance at higher distances[602c]. Furthermore, 3D object detection at large distances has been shown to improve with dedicated architectures.[603] Combined, this evidence indicates that object detectors may require modifications to detect objects at larger distances than currently evaluated upon.

[590] Zhou, X. et al.: Probabilistic two-stage detection (2021), p. 2.

[591] Singh, B. et al.: R-FCN-3000 at 30fps: Decoupling Detection and Classification (2018), p. 1083-1084.

[592] Ertler, C. et al.: The Mapillary Traffic Sign Dataset (2020), p. 8-9.

[593] Zhu, B. et al.: Class-balanced Grouping and Sampling for Point Cloud 3D Object Detection (2019), p. 4.

[594] Liu, Y. et al.: 1st Place Solutions for OpenImage2019 (2020), p. 3-4.

[595] Niitani, Y. et al.: Team PFDet's Methods for Open Images Challenge 2019 (2019), p. 2.

[596] Redmon, J.; Farhadi, A.: YOLO9000: Better, Faster, Stronger (2017) p. 6522-6524.

[597] Guo, Y. et al.: The efficacy of Neural Planning Metrics (2020), p. 2555.

[598] nuScenes: nuScenes prediction task: Leaderboard (2020).

[599] Waymo LLC: 3D Detection: Waymo Open Dataset: Challenge 1 (2019-2022).

[600] Feng, D. et al.: Deep Multi-Modal Object Detection and Semantic Segmentation (2020), p. 1342-1343.

[601] Caesar, H. et al.: nuScenes (2020), p.11621.

[602] Sun, P. et al.: Scalability in Perception (2020), a: p. 2447, b: p. 2444, c: p. 2450.

[603] Zhang, H. et al.: Faraway-Frustum (2021), p. 2646-2647.

The attribute requirements proposed in this work may also impact the detection architectures. As discussed in previous sections, this work proposes a temporal requirement of detecting an object within 150 ms. However, it is currently common to aggregate information from multiple frames for detection[604,605,606] or tracking[607,608,609]. Therefore, the requirements indicate that either sensors with higher frame rates are required or that aggregating multiple frames is not feasible. Furthermore, section 7.3.3 showed that matching failures are the dominant failure mode for the more permissive localization requirements of this work. In addition, velocity requirements which previously received comparatively less attention are proposed. Therefore, novel architectures may be required to meet the novel requirements which emphasize different attributes than prior benchmarks.

Overall, detection architectures are optimized to perform well on existing benchmarks. Evidence from detection on 2D images and lidar detectors implies that architectural modifications may be required to satisfy the novel requirements. However, compelling evidence on the task of 3D detection with the proposed requirements is still lacking and thus requires further research.

9.5.4 Summary of Implications

In this section, a concise summary of the implications regarding different aspects of object detection pipelines is presented.

The requirements developed in this work challenge existing object detection pipelines with respect to annotation, sensor setups and detectors. Firstly, the classification and relevance criteria are shown to be incompatible with common datasets. Applying the classification requirements of this work to dataset annotation is straightforward. However, the relevance criteria indicate that additional external sensors are required. Furthermore, the attribute requirements and the current fusion performance also influence sensor setups. Ensuring sufficient redundancy between sensors is likely only possible with more sensors and more different sensor modalities. Thus, fully applying the proposed requirements and their validation requires significant modification to datasets. In addition, existing detectors are optimized for performance on prevalent dataset metrics. Evidence suggests that meeting the proposed requirements also requires modification to the detection architectures.

[604] Wang, Y. et al.: Train in Germany, Test in The USA (2020), p. 4.

[605] Chen, Q. et al.: Every View Counts (2020), p. 4.

[606] Laddha, A. et al.: RV-FuseNet (2021), p. 7060-7061.

[607] Pang, Z. et al.: SimpleTrack (2023), p. 2.

[608] Weng, X. et al.: 3D Multi-Object Tracking (2020), p. 10361.

[609] Chiu, H. et al.: Probabilistic 3D Multi-Object Tracking for Autonomous Driving (2021), p. 14229.

9.6 Discussion Summary

With this, the findings and results of this work have been fully discussed. For the sake of brevity and clarity, the key discussion points are summarized in this section.

In this work, the overall RQ regarding safety requirements in object detection was decomposed into different aspects. The overall approach applied common principles of interpretability, legal and safety requirements as well as human performance. Application of the proposed methodology successfully yielded simple and interpretable detection requirements. The novel validation approach reconciled these requirements with a motion prediction DNN in a general way.

RQ4 was answered by proposing a novel method for validation. The method applied a motion prediction DNN, thus avoiding the difficulties of a full specification of a downstream planner. Furthermore, prediction performance was incorporated to assess the reliability of the prediction component. The validation is limited with respect to the available data and the prediction implementation. This means that at present, not all requirements of this work can be validated. Nevertheless, the validation was first successfully verified and then applied to the relevance and localization requirements.

Encouraging results were obtained by the proposed methodology, which successfully derives detection requirements. Interpretable requirements were obtained for classification, relevance and attributes of relevant objects. Thus, the initial RQs 1-3 were successfully answered. In addition to the argumentation, the encouraging results from the validation also support the requirements. The assumptions and requirements of this work were largely shown to be valid, contrary to those of datasets. Combining all of the aforementioned results allowed providing a positive answer to the overall RQ. While individual requirements may necessitate further modification and validation, this does not impact the overall conclusion. It was shown that considering the safety of the driving task is possible when deriving and validating detection requirements. Neither the application of the methodology nor the validation required fully specifying the downstream planner.

The novel requirements also impact the object detection pipelines including the dataset creation and the algorithm development. Applying the classification and attribute requirements to dataset annotation and evaluation is straightforward. However, results regarding relevance and redundancy uncovered limitations of present annotation procedures and sensor setups. Most likely, additional external sensors and additional vehicle sensors are required to fully test the novel requirements. Furthermore, the available evidence suggests that modifications to current object detection architectures are required to meet the proposed requirements.

10 Conclusion and outlook

In this chapter, a conclusion is presented summarizing the key findings of this work. Finally, the outlook regarding future research is discussed.

10.1 Conclusion

This work considered the field of AD which is expected to yield gains in efficiency, comfort and safety of traffic. For the task of AD, a reliable perception of the environment is required. More specifically, this work considered the task of 3D object detection which aims at predicting 3D bounding boxes for predefined object classes. However, testing object detection with respect to safety in driving context remains challenging.

One difficulty in dealing with perception functions in general is the difficulty of specifying the perception task. Therefore, the testing of perception functions such as object detection generally relies on datasets. These datasets use large-scale human annotated data which provide samples used for training and testing algorithms. Performance is generally measured by performance metrics provided by the datasets. However, current metrics emphasize average performance without considering the safety in the context of the driving task. Furthermore, no clear requirements including thresholds are provided.

Based on this observation, the objective of this work was to provide these requirements which consider the safety of the driving task. The requirements were decomposed into several interpretable components. The first aspect is which object categories to consider for the object detection. Once the categories were defined, it was necessary to consider the relevance of objects. For example, an object may no longer be relevant at very large distances. Finally, requirements were necessary to define how accurately different attributes such as location and velocity must be detected for a relevant object.

The methodology followed in this work was to consider each of the research questions separately. For each RQ, a unique method was developed and applied to answer it. However, the overall RQ was considered in each of the methods by defining common principles which apply to all research questions. Firstly, interpretability was demanded to increase understanding and acceptance. Secondly, all requirements must conform to road regulations and other applicable legal requirements. Since road regulations insufficiently specify safety, it was necessary to additionally explicitly consider safety outcomes. Furthermore, neither legal nor safety aspects fully specify the task of perception. Therefore, the human baseline was leveraged to specify any aspects not defined by any of the previous principles.

For each of the aspects of classification, relevance and attributes, a method to specify requirements was presented. For classification, the common principles were applied sequentially to obtain a hierarchical structure. For the relevance of objects, the concept was based on behavioral requirements. The idea is that the ego vehicle should be able to comply with all behavioral requirements even under worst-case conditions. Using a minimum specification of the driving system and incorporating uncertainties, potentially collision relevant objects were identified. Finally, different attribute requirements were considered separately to allow for interpretability. For each attribute, the corresponding human detection performance was derived from literature and repurposed as requirement. Furthermore, the entire object detection pipeline including matching and temporal aspects was reconsidered and aligned with the common principles. Overall, simple and interpretable requirements were obtained for all aspects. The requirements are substantiated by an argumentation and comply with the common principles.

Another question for such object detection requirements is how these can be validated. For this purpose, a novel validation method was presented in this work. The idea is to provide two different inputs to a DNN for motion prediction. Since the network was trained on human trajectories, it acted as proxy for human driving behaviors. Doing so allowed to estimate if a certain change in the object list influences human driving behavior. It was shown that the prediction is affected by uncertainties and local performance issues. However, by considering the change in performance distribution over the whole dataset, these issues were successfully circumvented. Successful verification experiments showed the approach to be practical in identifying invalid changes to an object list. Applying the approach to the requirements of this work fully supported the relevance requirements and largely supported the localization requirements.

Overall, requirements for various aspects of the detection task were successfully defined. The resulting requirements are simple, interpretable and substantiated by an argumentation. In addition, they were reconciled with the complementary approach of leveraging a context aware DNN for validation. However, the requirements showed substantial disagreement with existing evaluation pipelines. Furthermore, existing object detectors showed frequent detection failures. In addition, failures persisted even for an idealized fusion, demonstrating that errors are correlated even for different detectors and modalities. This result questions the capability of sensor setups relying on fusion to fully avoid detection failures. The results indicate that changes are required in both detection and the corresponding evaluation to better incorporate the safety of the driving task. Despite the advantages of the novel requirements, the results currently possess some limitations. Firstly, the classification requirements lack traffic signs, attributes and object relations. For the relevance results, a method to evaluate and validate specificity may be required. Regarding the attributes, additional investigation is required with regards to the human angular accuracy and conservativity. Furthermore, the validation procedure is currently limited by the available datasets and the available prediction networks.

To sum up, this work achieves encouraging results towards the definition of requirements for 3D object detection. Results showed that obtaining interpretable and valid criteria is possible. However, substantial changes to evaluation of object detection evaluation are required to fully consider safety requirements in the context of driving.

10.2 Outlook

The outlook regarding future research focuses on two distinct aspects. One is refining the requirements and their validation while the other is to develop detectors and sensor setups which better fulfill the requirements.

Further study is required to refine the object detection requirements of this work. Firstly, effort is required to gather data which allows testing the boundaries of validity for the requirements. Particularly the relevance testing is likely to benefit from critical driving situations. In addition, datasets which provide the classification labels defined in this work are required to test the requirements and their validity. Furthermore, investigation of different prediction architectures is recommended. Designing architectures specifically for the purpose of validation may allow validating object attributes which could not be validated within this work.

Furthermore, the novel requirements are likely to impact datasets, sensor setups and detectors used to fulfil the requirements. Literature indicates that specifically the larger amount of classes is likely to require modifications to common 3D detection architectures. In addition, the relevance and attribute requirements differ especially with regards to the spatial distribution of the permissible error. Designing and evaluating detectors specifically designed for the novel requirements may improve the failure rates. Furthermore, it may be necessary to modify sensor setups of datasets. By incorporating additional sensors and sensor modalities, the aspects of relevance and redundancy may be studied in more detail. These investigations are also likely to yield improvements to sensor setups with respect to the requirements.

Overall, the author hopes that this work encourages the explicit consideration and improvement of safety for future object detection.

Bibliography

Ahmed, S. et al.: Pedestrian and Cyclist Detection and Intent Estimation (2019)
Ahmed, Sarfraz; Huda, M. Nazmul; Rajbhandari, Sujan; Saha, Chitta; Elshaw, Mark; Kanarachos, Stratis: Pedestrian and Cyclist Detection and Intent Estimation for Autonomous Vehicles: A Survey, in: Applied Sciences, Vol. 9, p. 38, 2019

Althoff, M.: Reachability Analysis and its Application (2010)
Althoff, Matthias: Reachability Analysis and its Application to the Safety Assessment of Autonomous Cars, Dissertation, Technische Universität München, 2010

Althoff, M. et al.: Set-Based Prediction of Traffic Participants (2016)
Althoff, Matthias; Magdici, Silvia: Set-Based Prediction of Traffic Participants on Arbitrary Road Networks, in: IEEE Transactions on Intelligent Vehicles, Vol. 1, pp. 187–202, 2016

Alvarez, G. A. et al.: How many objects can you track? (2007)
Alvarez, G. A.; Franconeri, S. L.: How many objects can you track? Evidence for a resource-limited attentive tracking mechanism, in: Journal of vision, Vol. 7, pp. 1–10, 2007

Amersbach, C. T.: Functional Decomposition Approach (2020)
Amersbach, C. T.: Functional Decomposition Approach - Reducing the Safety Validation Effort for Highly Automated Driving, Dissertation, Technische Universität Darmstadt, 2020

Anderson, T. W.: On the Distribution of the Two-Sample Cramer-von Mises Criterion (1962)
Anderson, T. W.: On the Distribution of the Two-Sample Cramer-von Mises Criterion, in: The Annals of Mathematical Statistics, Vol. 33, pp. 1148–1159, 1962

ApolloAuto: apollo (2021)
ApolloAuto: apollo: README.md, URL: https://github.com/ApolloAuto/apollo, 2021

Ashmore, R. et al.: Assuring the Machine Learning Lifecycle (2022)
Ashmore, R.; Calinescu, R.; Paterson, C.: Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges, in: ACM Computing Surveys, Vol. 54, pp. 1–39, 2022

Aurora: Voluntary Safety Self-Assessment (2021)
Aurora: Voluntary Safety Self-Assessment, URL: https://assets.ctfassets.net/8byw6jksp7h2/12klTCAz1K43qbYuyZQUxe/b9eb91087815ecc1e54ff3a80e9ef580/Aurora_VSSA_2021_Final.pdf, 2021

Bagloee, S. A. et al.: Autonomous vehicles (2016)
Bagloee, Saeed Asadi; Tavana, Madjid; Asadi, Mohsen; Oliver, Tracey: Autonomous vehicles: challenges, opportunities, and future implications for transportation policies, in: Journal of Modern Transportation, Vol. 24, pp. 284–303, 2016

Bagschik, G. et al.: Ontology based Scene Creation (2018)
Bagschik, Gerrit; Menzel, Till; Maurer, Markus: Ontology based Scene Creation for the Development of Automated Vehicles, in: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 1813–1820, 2018

Bai, X. et al.: TransFusion (2022)
Bai, Xuyang; Hu, Zeyu; Zhu, Xinge; Huang, Qingqiu; Chen, Yilun; Fu, Hongbo; Tai, Chiew-Lan: TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers, in: (Hrsg.): 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022

Bansal, A. et al.: Verifiable Obstacle Detection (2022)
Bansal, A.; Kim, H.; Yu, S.; Li, B.; Hovakimyan, N.; Caccamo, M.; Sha, L. R.: Verifiable Obstacle Detection, in: (Hrsg.): 2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE), 2022

Bansal, A. et al.: Risk Ranked Recall (2021)
Bansal, A.; Singh, J.; Verucchi, M.; Caccamo, M.; Sha, L. R.: Risk Ranked Recall: Collision Safety Metric for Object Detection Systems in Autonomous Vehicles, in: (Hrsg.): 2021 10th Mediterranean Conference on Embedded Computing (MECO), 2021

Bansal, M. et al.: ChauffeurNet (2018)
Bansal, Mayank; Krizhevsky, Alex; Ogale, Abhijit: ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst, in: ArXiv, Vol. abs/1812.03079, 2018

Benjamin, D. J. et al.: Redefine statistical significance (2018)
Benjamin, Daniel J.; Berger, James O.; Johannesson, Magnus; Nosek, Brian A.; Wagenmakers, E-J; Berk, Richard; Bollen, Kenneth A.; Brembs, Björn; Brown, Lawrence; Camerer, Colin; Cesarini, David; Chambers, Christopher D.; Clyde, Merlise; Cook, Thomas D.; Boeck, Paul de; Dienes, Zoltan; Dreber, Anna; Easwaran, Kenny; Efferson, Charles; Fehr, Ernst; Fidler, Fiona; Field, Andy P.; Forster, Malcolm; George, Edward I.; Gonzalez, Richard; Goodman, Steven; Green, Edwin; Green, Donald P.; Greenwald, Anthony G.; Hadfield, Jarrod D.; Hedges, Larry V.; Held, Leonhard; Hua Ho, Teck; Hoijtink, Herbert; Hruschka, Daniel J.; Imai, Kosuke; Imbens, Guido; Ioannidis, John P. A.; Jeon, Minjeong; Jones, James Holland; Kirchler, Michael; Laibson, David; List, John; Little, Roderick; Lupia, Arthur; Machery, Edouard; Maxwell, Scott E.; McCarthy, Michael; Moore, Don A.; Morgan, Stephen L.; Munafó, Marcus; Nakagawa, Shinichi; Nyhan, Brendan; Parker, Timothy H.; Pericchi, Luis; Perugini, Marco; Rouder, Jeff; Rousseau, Judith; Savalei, Victoria; Schönbrodt, Felix D.; Sellke, Thomas; Sinclair, Betsy; Tingley, Dustin; van Zandt, Trisha; Vazire, Simine; Watts, Duncan J.; Winship, Christopher; Wolpert, Robert L.; Xie, Yu; Young, Cristobal; Zinman, Jonathan; Johnson, Valen E.: Redefine statistical significance, in: Nature human behaviour, Vol. 2, pp. 6–10, 2018

Bernardin, K. et al.: Evaluating Multiple Object Tracking Performance (2008)
Bernardin, K.; Stiefelhagen, R.: Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics, in: EURASIP Journal on Image and Video Processing, Vol. 2008, pp. 1–10, 2008

Bertolazzi, E. et al.: Supporting Drivers in Keeping Safe Speed and Safe Distance (2010)
Bertolazzi, Enrico; Biral, Francesco; Da Lio, Mauro; Saroldi, Andrea; Tango, Fabio: Supporting Drivers in Keeping Safe Speed and Safe Distance: The SASPENCE Subproject Within the European Framework Programme 6 Integrating Project PReVENT, in: IEEE Transactions on Intelligent Transportation Systems, Vol. 11, pp. 525–538, 2010

Bevandic, P. et al.: Multi-domain semantic segmentation with overlapping labels (2022)
Bevandic, Petra; Orsic, Marin; Grubisic, Ivan; Saric, Josip; Segvic, Sinisa: Multi-domain semantic segmentation with overlapping labels, in: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2422–2431, 2022

Blum, H. et al.: Fishyscapes (2019)
Blum, Hermann; Sarlin, Paul-Edouard; Nieto, Juan; Siegwart, Roland; Cadena, Cesar: Fishyscapes: A Benchmark for Safe Semantic Segmentation in Autonomous Driving, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 2403–2412, 2019

Bokare, P. S. et al.: Acceleration-Deceleration Behaviour (2017)
Bokare, P. S.; Maurya, A. K.: Acceleration-Deceleration Behaviour of Various Vehicle Types, in: Transportation Research Procedia, Vol. 25, pp. 4733–4749, 2017

Böttcher, L.: Road Traffic Regulations (2018)
Böttcher, Lorenz: Road Traffic Regulations (Straßenverkehrs-Ordnung, StVO) with Annexes, URL: https://germanlawarchive.iuscomp.org/?p=1290, 2018

Braun, M. et al.: EuroCity Persons (2019)
Braun, M.; Krebs, S.; Flohr, F.; Gavrila, D.: EuroCity Persons: A Novel Benchmark for Person Detection in Traffic Scenes, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 41, pp. 1844–1861, 2019

Brenner, E. et al.: Continuous visual control of interception (2011)
Brenner, E.; Smeets, J. B. J.: Continuous visual control of interception, in: Human Movement Science, Vol. 30, pp. 475–494, 2011

Brenner, E. et al.: How people achieve their amazing temporal precision in interception (2015)
Brenner, E.; Smeets, J. B. J.: How people achieve their amazing temporal precision in interception, in: Journal of vision, Vol. 15, pp. 1–21, 2015

Breuer, A. et al.: openDD: A Large-Scale Roundabout Drone Dataset (2020)
Breuer, Antonia; Termohlen, Jan-Aike; Homoceanu, Silviu; Fingscheidt, Tim: openDD: A Large-Scale Roundabout Drone Dataset, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6, 2020

Bruyn, B. et al.: Human velocity and direction discrimination (1988)
Bruyn, B.; Orban, G. A.: Human velocity and direction discrimination measured with random dot patterns, in: Vision Research, Vol. 28, pp. 1323–1335, 1988

Buechel, M. et al.: Ontology-based traffic scene modeling (2017)
Buechel, M.; Hinz, G.; Ruehl, F.; Schroth, H.; Gyoeri, C.; Knoll, A.: Ontology-based traffic scene modeling, traffic regulations dependent situational awareness and decision-making for automated vehicles, in: 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 1471–1476, 2017

Bulakowski, P. F. et al.: Independent coding of object motion and position (2007)
Bulakowski, P. F.; Koldewyn, K.; Whitney, D.: Independent coding of object motion and position revealed by distinct contingent aftereffects, in: Vision Research, Vol. 47, pp. 810–817, 2007

Bundesministerium der Justiz: AFGBV (24.06.2022)
Bundesministerium der Justiz: Verordnung zur Genehmigung und zum Betrieb von Kraftfahrzeugen mit autonomer Fahrfunktion in festgelegten Betriebsbereichen: Autonome-Fahrzeuge-Genehmigungs-und-Betriebs-Verordnung - AFGBV, URL: https://www.gesetze-im-internet.de/afgbv/AFGBV.pdf, 24.06.2022

Bundesministerium für Justiz und Verbraucherschutz: Strafgesetzbuch: StGB (13.11.1998)
Bundesministerium für Justiz und Verbraucherschutz: Strafgesetzbuch: StGB, URL: https://www.gesetze-im-internet.de/stgb/StGB.pdf, 13.11.1998

Bundesministerium für Justiz und Verbraucherschutz: StVO (06.03.2013)
Bundesministerium für Justiz und Verbraucherschutz: Straßenverkehrs-Ordnung: StVO, URL: https://www.gesetze-im-internet.de/stvo_2013/StVO.pdf, 06.03.2013

Bundesministerium für Justiz und Verbraucherschutz: Straßenverkehrsgesetz: StVG (05.03.2003)
Bundesministerium für Justiz und Verbraucherschutz: Straßenverkehrsgesetz: StVG, URL: https://www.gesetze-im-internet.de/stvg/STVG.pdf, 05.03.2003

Bundesministerium für Justiz und Verbraucherschutz: Bußgeldkatalog-Verordnung (14.03.2013)
Bundesministerium für Justiz und Verbraucherschutz: Verordnung über die Erteilung einer Verwarnung, Regelsätze für Geldbußen und die Anordnung eines Fahrverbotes wegen Ordnungswidrigkeiten im Straßenverkehr (Bußgeldkatalog-Verordnung - BKatV): BKatV, URL: https://www.gesetze-im-internet.de/bkatv_2013/BKatV.pdf, 14.03.2013

Caesar, H. et al.: nuScenes (2020)
Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. Erin; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O.: nuScenes: A multimodal dataset for autonomous driving, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11618–11628, 2020

Caesar, H. et al.: nuPlan: A closed-loop ML-based planning benchmark (2021)
Caesar, Holger; Kabzan, Juraj; Tan, Kok Seang; Fong, Whye Kit; Wolff, Eric M.; Lang, Alex; Fletcher, Luke; Beijbom, Oscar; Omari, Sammy: nuPlan: A closed-loop ML-based planning benchmark for autonomous vehicles, in: ArXiv, Vol. abs/2106.11810, 2021

Caesar, H. et al.: COCO-Stuff (2018)
Caesar, Holger; Uijlings, Jasper; Ferrari, Vittorio: COCO-Stuff: Thing and Stuff Classes in Context, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018

Cavallo, V. et al.: Visual Information and Skill Level in Time-To-Collision Estimation (1988)
Cavallo, V.; Laurent, M.: Visual Information and Skill Level in Time-To-Collision Estimation, in: Perception, Vol. 17, pp. 623–632, 1988

Chan, R. et al.: SegmentMeIfYouCan (2021)
Chan, R.; Lis, K.; Uhlemeyer, S.; Blum, H.; Honari, S.; Siegwart, R.; Salzmann, M.; Fua, P.; Rottmann, M.: SegmentMeIfYouCan: A Benchmark for Anomaly Segmentation, in: (Hrsg.): 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks, 2021

Chang, M. et al.: Argoverse (2019)
Chang, M.; Lambert, J.; Sangkloy, P.; Singh, J.; Bak, S.; Hartnett, A.; Wang, d.; Carr, P.; Lucey, S.; Ramanan, D.; Hays, J.: Argoverse: 3D Tracking and Forecasting with Rich Maps, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8740–8749, 2019

Chapman, P. et al.: Visual Search of Driving Situations (1998)
Chapman, Peter; Underwood, Geoffrey: Visual Search of Driving Situations: Danger and Experience, in: Perception, Vol. 27, pp. 951–964, 1998

Chen, J. et al.: A real-time and high-precision method for small traffic-signs recognition (2022)
Chen, Junzhou; Jia, Kunkun; Chen, Wenquan; Lv, Zhihan; Zhang, Ronghui: A real-time and high-precision method for small traffic-signs recognition, in: Neural Computing and Applications, Vol. 34, pp. 2233–2245, 2022

Chen, Q. et al.: Every View Counts (2020)
Chen, Q.; Sun, L.; Cheung, E.; Yuille, A.: Every View Counts: Cross-View Consistency in 3D Object Detection with Hybrid-Cylindrical-Spherical Voxelization, in: Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F.; Lin, H. (Hrsg.): Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 2020

Chen, Y. et al.: The precision of velocity discrimination across spatial frequency (1998)
Chen, Y.; Bedell, H. E.; Frishman, L. J.: The precision of velocity discrimination across spatial frequency, in: Perception & psychophysics, Vol. 60, pp. 1329–1336, 1998

Chiu, H. et al.: Probabilistic 3D Multi-Object Tracking for Autonomous Driving (2021)
Chiu, H.; Li, J.; Prioletti, A.; Bohg, J.: Probabilistic 3D Multi-Object Tracking for Autonomous Driving, in: (Hrsg.): 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021

Cho, K. et al.: Deep Predictive Autonomous Driving (2019)
Cho, K.; Ha, T.; Lee, G.; Oh, S.: Deep Predictive Autonomous Driving Using Multi-Agent Joint Trajectory Prediction and Traffic Rules, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2076–2081, 2019

COCO Consortium: COCO Common Objects in Context: Detection Evaluation (2015)
COCO Consortium: COCO Common Objects in Context: Detection Evaluation, URL: https://cocodataset.org/#detection-eval, 2015

Cordts, M. et al.: The Cityscapes Dataset (2016)
Cordts, Marius; Omran, Mohamed; Ramos, Sebastian; Rehfeld, Timo; Enzweiler, Markus; Benenson, Rodrigo; Franke, Uwe; Roth, Stefan; Schiele, Bernt: The Cityscapes Dataset for Semantic Urban Scene Understanding, in: Proceedings, 29th IEEE Conference on Computer Vision and Pattern Recognition : 26 June-1 July 2016, Las Vegas, Nevada, pp. 3213–3223, 2016

Cosijn, E. et al.: Dimensions of relevance (2000)
Cosijn, Erica; Ingwersen, Peter: Dimensions of relevance, in: Information Processing & Management, Vol. 36, pp. 533–550, 2000

D'Amour, A. et al.: Underspecification Presents Challenges (2022)
D'Amour, A.; Heller, K. A.; Moldovan, D.; Adlam, B.; Alipanahi, B.; Beutel, A.; Chen, C.; Deaton, J.; Eisenstein, J.; Hoffman, M. D.; Hormozdiari, F.; Houlsby, N.; Hou, S.; Jerfel, G.; Karthikesalingam, A.; Lucic, M.; Ma, Y.-A.; McLean, C. Y.; Mincu, D.; Mitani, A.; Montanari, A.; Nado, Z.; Natarajan, V.; Nielson, C.; Osborne, T. F.; Raman, R.; Ramasamy, K.; Sayres, R.; Schrouff, J.; Seneviratne, M.; Sequeira, S.; Suresh, H.; Veitch, V.; Vladymyrov, M.; Wang, X.; Webster, K.; Yadlowsky, S.; Yun, T.; Zhai, X.; Sculley, D.: Underspecification Presents Challenges for Credibility in Modern Machine Learning, in: Journal of Machine Learning Research, Vol. 23, pp. 1–61, 2022

Daum, O. S. et al.: Distance estimation in vista space (2009)
Daum, O. S.; Hecht, H.: Distance estimation in vista space, in: Attention, perception & psychophysics, Vol. 71, pp. 1127–1137, 2009

Deng, B. et al.: Revisiting 3D Object Detection From an Egocentric Perspective (2021)
Deng, B.; Qi, C.; Najibi, M.; Funkhouser, T. A.; Zhou, Y.; Anguelov, D.: Revisiting 3D Object Detection From an Egocentric Perspective, in: Advances in Neural Information Processing Systems 34 (NeurIPS 2021), pp. 1–14, 2021

Deng, J. et al.: ImageNet: A large-scale hierarchical image database (2009)
Deng, Jia; Dong, Wei; Socher, Richard; Li, Li-Jia; Li, Kai; Fei-Fei, Li: ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, 2009

Deo, N. et al.: Multimodal Trajectory Prediction Conditioned on Lane-Graph Traversals (2022)
Deo, Nachiket; Wolff, Eric; Beijbom, Oscar: Multimodal Trajectory Prediction Conditioned on Lane-Graph Traversals, in: Proceedings of the 5th Conference on Robot Learning, vol. 164, pp. 203–212, 2022

Department of Transport: Reported road causalties Great Britain (2021)
Department of Transport: Reported road causalties Great Britain, annual report: 2020: National statistics, URL: https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2020/reported-road-casualties-great-britain-annual-report-2020#overall-casualties, 2021

Dietmayer, K.: Predicting of Machine Perception for Automated Driving (2016)
Dietmayer, Klaus: Predicting of Machine Perception for Automated Driving, in: Maurer, Markus; Gerdes, J. Christian; Lenz, Barbara; Winner, Hermann (Hrsg.): Autonomous Driving: Technical, Legal and Social Aspects, Springer Berlin Heidelberg, 2016

Dosovitskiy, A. et al.: CARLA: An Open Urban Driving Simulator (2017)
Dosovitskiy, Alexey; Ros, German; Codevilla, Felipe; Lopez, Antonio; Koltun, Vladlen: CARLA: An Open Urban Driving Simulator, in: ArXiv, Vol. 1711, 2017

Ertler, C. et al.: The Mapillary Traffic Sign Dataset (2020)
Ertler, Christian; Mislej, Jerneja; Ollmann, Tobias; Porzi, Lorenzo; Neuhold, Gerhard; Kuang, Yubin: The Mapillary Traffic Sign Dataset for Detection and Classification on a Global Scale, in: (Hrsg.): Computer Vision - ECCV 2020, 2020

Ethics Comission: Automated and Connected Driving: Report (2017)
Ethics Comission: Automated and Connected Driving: Report, URL: https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile, 2017

Ettinger, S. M. et al.: The Waymo Open Motion Dataset (2021)
Ettinger, Scott M.; Cheng, Shuyang; Caine, Benjamin; Liu, Chenxi; Zhao, Hang; Pradhan, Sabeek; Chai, Yuning; Sapp, Benjamin; Qi, C.; Zhou, Yin; Yang, Zoey; Chouard, Aurelien; Sun, Pei; Ngiam, Jiquan; Vasudevan, Vijay; McCauley, Alexander; Shlens, Jonathon; Anguelov, Drago: Large Scale Interactive Motion Forecasting for Autonomous Driving : The Waymo Open Motion Dataset, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9690–9699, 2021

European Commission: Annexes to Regulation (EU) 2019/2144 (2022)
European Commission: ANNEXES to the Commission Implementing Regulation laying down rules for the application of Regulation (EU) 2019/2144 of the European Parliament and of the Council as regards uniform procedures and technical specifications for the type-approval of the automated driving system (ADS) of fully automated vehicles: Draft act: Annex 2, URL: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12152-Automated-cars-technical-specifications_en, 2022

European Commission: Implementing Regulation (EU) 2022/1426 (05.07.2022)
European Commission: Commission Implementing Regulation (EU) 2022/1426, URL: http://data.europa.eu/eli/reg_impl/2022/1426/oj, 05.07.2022

European Commission: Artificial Intelligence Act (2021)
European Commission: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206, 2021

Everingham, M. et al.: The Pascal Visual Object Classes (VOC) Challenge (2010)
Everingham, M.; van Gool, L.; Williams, C. K. I.; Winn, J.; Zisserman, A.: The Pascal Visual Object Classes (VOC) Challenge, in: International Journal of Computer Vision, Vol. 88, pp. 303–338, 2010

Fei, J. et al.: SemanticVoxels (2020)
Fei, J.; Chen, W.; Heidenreich, P.; Wirges, S.; Stiller, C.: SemanticVoxels: Sequential Fusion for 3D Pedestrian Detection using LiDAR Point Cloud and Semantic Segmentation, in: (Hrsg.): 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), 2020

Feng, D. et al.: Deep Multi-Modal Object Detection and Semantic Segmentation (2020)
Feng, D.; Haase-Schütz, C.; Rosenbaum, L.; Hertlein, H.; Gläser, C.; Timm, F.; Wiesbeck, W.; Dietmayer, K.: Deep Multi-Modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges, in: IEEE Transactions on Intelligent Transportation Systems, pp. 1341–1360, 2020

Fong, W. K. et al.: Panoptic nuScenes (2022)
Fong, Whye Kit; Mohan, Rohit; Hurtado, Juana Valeria; Zhou, Lubing; Caesar, Holger; Beijbom, Oscar; Valada, Abhinav: Panoptic nuScenes: A Large-Scale Benchmark for LiDAR Panoptic Segmentation and Tracking, in: IEEE Robotics and Automation Letters, Vol. 7, pp. 3795–3802, 2022

Fregin, A. et al.: The DriveU Traffic Light Dataset (2018)
Fregin, Andreas; Muller, Julian; Krebel, Ulrich; Dietmayer, Klaus: The DriveU Traffic Light Dataset: Introduction and Comparison with Existing Datasets, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 3376–3383, 2018

Ge, R. et al.: AFDet: Anchor Free One Stage 3D Object Detection (2020)
Ge, R.; Ding, Z.; Hu, Y.; Wang, Y.; Chen, S.; Huang, L.; Li, Y.: AFDet: Anchor Free One Stage 3D Object Detection, in: ArXiv, Vol. 2006.12671, 2020

Geiger, A. et al.: Are we ready for Autonomous Driving? (2012)
Geiger, A.; Lenz, P.; Urtasun, R.: Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3354–3361, 2012

Geisslinger, M. et al.: An ethical trajectory planning algorithm (2023)
Geisslinger, Maximilian; Poszler, Franziska; Lienkamp, Markus: An ethical trajectory planning algorithm for autonomous vehicles, in: Nature Machine Intelligence, Vol. 5, pp. 137–144, 2023

General Motors: Self Driving Safety Report (2018)
General Motors: Self Driving Safety Report, URL: https://www.gm.com/content/dam/company/docs/us/en/gmcom/gmsafetyreport.pdf, 2018

Geyer, J. et al.: A2D2: Audi Autonomous Driving Dataset (2020)
Geyer, Jakob; Kassahun, Yohannes; Mahmudi, Mentar; Ricou, Xavier; Durgesh, Rupesh; Chung, Andrew S.; Hauswald, Lorenz; Pham, Viet Hoang; Mühlegg, Maximilian; Dorn, Sebastian; Fernandez, Tiffany; Jänicke, Martin; Mirashi, Sudesh; Savani, Chiragkumar; Sturm, Martin; Vorobiov, Oleksandr; Oelker, Martin; Garreis, Sebastian; Schuberth, Peter: A2D2: Audi Autonomous Driving Dataset, in: ArXiv, Vol. abs/2004.06320, 2020

Gilinsky, A. S.: The Effect of Attitude upon the Perception of Size (1955)
Gilinsky, A. S.: The Effect of Attitude upon the Perception of Size, in: The American Journal of Psychology, Vol. 68, p. 173, 1955

Glatzki, F. et al.: Behavioral Attributes for a Behavior-Semantic Scenery Description (2021)
Glatzki, Felix; Lippert, Moritz; Winner, Hermann: Behavioral Attributes for a Behavior-Semantic Scenery Description (BSSD) for the Development of Automated Driving Functions, in: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), pp. 667–672, 2021

Gopinath, D. et al.: Property Inference for Deep Neural Networks (2019)
Gopinath, Divya; Converse, Hayes; Pasareanu, Corina S.; Taly, Ankur: Property Inference for Deep Neural Networks, in: 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 797–809, 2019

Gottschalk, H. et al.: Does Redundancy in AI Perception Systems Help? (2022)
Gottschalk, H.; Rottmann, M.; Saltagic, M.: Does Redundancy in AI Perception Systems Help to Test for Super-Human Automated Driving Performance?, in: Fingscheidt, Tim; Gottschalk, Hanno; Houben, Sebastian (Hrsg.): Deep Neural Networks and Data for Automated Driving: Robustness, Uncertainty Quantification, and Insights Towards Safety, Springer International Publishing, 2022

Green, M.: How Long Does It Take to Stop? (2000)
Green, Marc: How Long Does It Take to Stop? Methodological Analysis of Driver Perception-Brake Times, in: Transportation Human Factors, Vol. 2, pp. 195–216, 2000

Greibe, P.: Braking distance, friction and behaviour (2007)
Greibe, Poul: Braking distance, friction and behaviour: Findings, analyses and recommendations based on braking trials, URL: https://www.trafitec.dk/sites/default/files/publications/braking%20distance%20-%20friction%20and%20driver%20behaviour.pdf, 2007

Guo, Y. et al.: The efficacy of Neural Planning Metrics (2020)
Guo, Y.; Caesar, H.; Beijbom, O.; Philion, J.; Fidler, S.: The efficacy of Neural Planning Metrics: A meta-analysis of PKL on nuScenes, in: ArXiv, Vol. 2010.09350, 2020

Guo, Y. et al.: CS-R-FCN: Cross-supervised Learning for Large-Scale Object Detection (2020)
Guo, Ye; Li, Yali; Wang, Shengjin: CS-R-FCN: Cross-supervised Learning for Large-Scale Object Detection, in: (Hrsg.): ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020

Haarmeier, T. et al.: Detection of speed changes during pursuit eye movements (2006)
Haarmeier, T.; Thier, P.: Detection of speed changes during pursuit eye movements, in: Experimental brain research, Vol. 170, pp. 345–357, 2006

Haber, R. N. et al.: The independence of size perception and distance perception (2001)
Haber, R. N.; Levin, C. A.: The independence of size perception and distance perception, in: Perception & psychophysics, Vol. 63, pp. 1140–1152, 2001

Hallerbach, S. et al.: Simulation-Based Identification of Critical Scenarios (2018)
Hallerbach, Sven; Xia, Yiqun; Eberle, Ulrich; Koester, Frank: Simulation-Based Identification of Critical Scenarios for Cooperative and Automated Vehicles, in: SAE International Journal of Connected and Automated Vehicles, Vol. 1, pp. 93–106, 2018

Han, Y. et al.: Effects of vehicle impact velocity (2012)
Han, Yong; Yang, Jikuang; Mizuno, Koji; Matsui, Yasuhiro: Effects of vehicle impact velocity, vehicle front-end shapes on pedestrian injury risk, in: Traffic injury prevention, Vol. 13, pp. 507–518, 2012

Hara, K. et al.: Growing Regression Tree Forests (2017)
Hara, K.; Chellappa, R.: Growing Regression Tree Forests by Classification for Continuous Object Pose Estimation, in: International Journal of Computer Vision, Vol. 122, pp. 292–312, 2017

Hendy, N. et al.: FISHING Net (2020)
Hendy, Noureldin; Sloan, Cooper; Tian, F.; Duan, Pengfei; Charchut, Nick; Xie, Y.; Wang, C.; Philbin, J.: FISHING Net: Future Inference of Semantic Heatmaps In Grids, in: ArXiv, Vol. abs/2006.09917, 2020

Henze, F. et al.: Admissible Uncertainty Bounds for Planning Algorithms (2021)
Henze, F.; Fabender, D.; Stiller, C.: Identifying Admissible Uncertainty Bounds for the Input of Planning Algorithms, in: IEEE Transactions on Intelligent Vehicles, pp. 3129–3143, 2021

Hetzel, M. et al.: The IMPTC Dataset (2023)
Hetzel, Manuel; Reichert, Hannes; Reitberger, Günther; Fuchs, Erich; Doll, Konrad; Sick, Bernhard: The IMPTC Dataset: An Infrastructural Multi-Person Trajectory and Context Dataset, in: 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1–7, 2023

Higashiyama, A.: Anisotropic perception of visual angle (1992)
Higashiyama, A.: Anisotropic perception of visual angle: implications for the horizontal-vertical illusion, overconstancy of size, and the moon illusion, in: Perception & psychophysics, Vol. 51, pp. 218–230, 1992

Hoffmann, E. R. et al.: Drivers' estimates of time to collision (1994)
Hoffmann, E. R.; Mortimer, R. G.: Drivers' estimates of time to collision, in: Accident Analysis & Prevention, Vol. 26, pp. 511–520, 1994

Hoss, M. et al.: A Review of Testing Object-Based Environment Perception (2022)
Hoss, M.; Scholtes, M.; Eckstein, L.: A Review of Testing Object-Based Environment Perception for Safe Automated Driving, in: Automotive Innovation, Vol. 5, pp. 223–250, 2022

Houben, S. et al.: Inspect, Understand, Overcome (2022)
Houben, Sebastian; Abrecht, Stephanie; Akila, Maram; Bär, Andreas; Brockherde, Felix; Feifel, Patrick; Fingscheidt, Tim; Gannamaneni, Sujan Sai; Ghobadi, Seyed Eghbal; Hammam, Ahmed; Haselhoff, Anselm; Hauser, Felix; Heinzemann, Christian; Hoffmann, Marco; Kapoor, Nikhil; Kappel, Falk; Klingner, Marvin; Kronenberger, Jan; Küppers, Fabian; Löhdefink, Jonas; Mlynarski, Michael; Mock, Michael; Mualla, Firas; Pavlitskaya, Svetlana; Poretschkin, Maximilian; Pohl, Alexander; Ravi-Kumar, Varun; Rosenzweig, Julia; Rottmann, Matthias; Rüping, Stefan; Sämann, Timo; Schneider, Jan David; Schulz, Elena; Schwalbe, Gesina; Sicking, Joachim; Srivastava, Toshika; Varghese, Serin; Weber, Michael; Wirkert, Sebastian; Wirtz, Tim; Woehrle, Matthias: Inspect, Understand, Overcome: A Survey of Practical Methods for AI Safety, in: Fingscheidt, Tim; Gottschalk, Hanno; Houben, Sebastian (Hrsg.): Deep Neural Networks and Data for Automated Driving, Springer Cham, 2022

Houston, J. et al.: One Thousand and One Hours (2020)
Houston, John; Zuidhof, Guido; Bergamini, Luca; Ye, Yawei; Chen, Long; Jain, Ashesh; Omari, Sammy; Iglovikov, Vladimir; Ondruska, Peter: One Thousand and One Hours: Self-driving Motion Prediction Dataset, in: ArXiv, Vol. abs/2006.14480, 2020

Hu, Y. et al.: Scenario-Transferable Semantic Graph Reasoning (2022)
Hu, Yeping; Zhan, Wei; Tomizuka, Masayoshi: Scenario-Transferable Semantic Graph Reasoning for Interaction-Aware Probabilistic Prediction, in: IEEE Transactions on Intelligent Transportation Systems, Vol. 23, pp. 23212–23230, 2022

Huang, X. et al.: The ApolloScape Open Dataset (2020)
Huang, Xinyu; Wang, Peng; Cheng, Xinjing; Zhou, Dingfu; Geng, Qichuan; Yang, Ruigang: The ApolloScape Open Dataset for Autonomous Driving and Its Application, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, pp. 2702–2719, 2020

Huang, Y. et al.: A Survey on Trajectory-Prediction Methods for Autonomous Driving (2022)
Huang, Yanjun; Du, Jiatong; Yang, Ziru; Zhou, Zewei; Zhang, Lin; Chen, Hong: A Survey on Trajectory-Prediction Methods for Autonomous Driving, in: IEEE Transactions on Intelligent Vehicles, Vol. 7, pp. 652–674, 2022

Huang, Y. et al.: Autonomous Driving with Deep Learning (2020)
Huang, Yu; Chen, Y.: Survey of State-of-Art Autonomous Driving Technologies with Deep Learning, in: (Hrsg.): 2020 IEEE 20th International Conference, 2020

Huijser, M. P. et al.: Wildlife-Vehicle Collision Reduction Study: Report to Congress (2008)
Huijser, M. P.; McGowen, P.; Fuller, J.; Hardy, A.; Kociolek, A.; Clevenger, A. P.; Smith, D.; Ament, R.: Wildlife-Vehicle Collision Reduction Study: Report to Congress, URL: https://www.fhwa.dot.gov/publications/research/safety/08034/02.cfm, 2008

Ilievski, M. et al.: Design Space of Behaviour Planning for Autonomous Driving (2019)
Ilievski, Marko; Sedwards, Sean; Gaurav, Ashish; Balakrishnan, A.; Sarkar, Atrisha; Lee, J.; Bouchard, F.; Iaco, R. D.; Czarnecki, K.: Design Space of Behaviour Planning for Autonomous Driving, in: ArXiv, Vol. abs/1908.07931, 2019

ISO/TC 22/SC 32: ISO 26262-1 Functional Safety (12/2018)
ISO/TC 22/SC 32: Road Vehicles - Functional Safety: Part 1: Vocabulary, 12/2018

ISO/TC 22/SC 32: ISO/PAS 21448 SOTIF (06/2022)
ISO/TC 22/SC 32: Road vehicles - Safety of the intended functionality, URL: https://www.iso.org/standard/77490.html, 06/2022

ISO/TC22/SC32/WG14: ISO PAS 8800 Road Vehicles (2021)
ISO/TC22/SC32/WG14: ISO PAS 8800 Road Vehicles - Safety and Artificial Intelligence, URL: https://unece.org/transport/documents/2021/09/informal-documents/iso-iso-pas-8800-road-vehicles-safety-and-artificial, 2021

Ivanovic, B. et al.: Injecting Planning-Awareness (2022)
Ivanovic, B.; Pavone, M.: Injecting Planning-Awareness into Prediction and Detection Evaluation, in: (Hrsg.): 2022 IEEE Intelligent Vehicles Symposium (IV), 2022

Jaiswal, A. et al.: Class-agnostic Object Detection (2021)
Jaiswal, A.; Wu, Y.; Natarajan, P.; Natarajan, P.: Class-agnostic Object Detection, in: (Hrsg.): 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021

Jha, S. et al.: Watch Out for the Safety-Threatening Actors (2022)
Jha, Saurabh; Cui, Shengkun; Kalbarczyk, Zbigniew; Iyer, Ravishankar K.: Watch Out for the Safety-Threatening Actors: Proactively Mitigating Safety Hazards, in: ArXiv, 2022

Junietz, P. et al.: Criticality Metric for Safety Validation (2018)
Junietz, P.; Bonakdar, F.; Klamann, B.; Winner, H.: Criticality Metric for the Safety Validation of Automated Driving using Model Predictive Trajectory Optimization, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 60–65, 2018

Junietz, P. et al.: Macroscopic Safety Requirements (2019)
Junietz, P.; Steininger, U.; Winner, H.: Macroscopic Safety Requirements for Highly Automated Driving, in: Transportation Research Record, Vol. 2673, pp. 1–14, 2019

Junietz, P. et al.: Evaluation of Different Approaches to Address Safety Validation (2018)
Junietz, P.; Wachenfeld, W.; Klonecki, K.; Winner, H.: Evaluation of Different Approaches to Address Safety Validation of Automated Driving, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 491–496, 2018

Kalra, N. et al.: Driving to Safety (2016)
Kalra, Nidhi; Paddock, Susan M.: Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?, in: Transportation Research Part A: Policy and Practice, Vol. 94, pp. 182–193, 2016

Kang, Y. et al.: Test Your Self-Driving Algorithm (2019)
Kang, Y.; Yin, H.; Berger, C.: Test Your Self-Driving Algorithm: An Overview of Publicly Available Driving Datasets and Virtual Testing Environments, in: IEEE Transactions on Intelligent Vehicles, Vol. 4, pp. 171–185, 2019

Karimi, A. et al.: Formalizing traffic rules for uncontrolled intersections (2020)
Karimi, A.; Duggirala, P. S.: Formalizing traffic rules for uncontrolled intersections, in: 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS), pp. 41–50, 2020

Karthika, R. et al.: Traffic Sign Recognition from Road Scenes (2022)
Karthika, R.; Parameswaran, Latha: A Novel Convolutional Neural Network Based Architecture for Object Detection and Recognition with an Application to Traffic Sign Recognition from Road Scenes, in: Pattern Recognition and Image Analysis, Vol. 32, pp. 351–362, 2022

Kato, Y. et al.: A Conditional Confidence Calibration Method (2022)
Kato, Y.; Kato, S.: A Conditional Confidence Calibration Method for 3D Point Cloud Object Detection, in: (Hrsg.): 2022 IEEE Intelligent Vehicles Symposium, 2022

Kiefer, R. J. et al.: Time-to-collision judgments under realistic driving conditions (2006)
Kiefer, R. J.; Flannagan, C. A.; Jerome, C. J.: Time-to-collision judgments under realistic driving conditions, in: Human factors, Vol. 48, pp. 334–345, 2006

Kim, K.-Y. et al.: Real-Time Performance Evaluation Metrics (2016)
Kim, K.-Y.; Kim, Y.; Park, J.; Kim, Y.-S.: Real-Time Performance Evaluation Metrics for Object Detection and Tracking of Intelligent Video Surveillance Systems, in: Asia Pacific Journal of Contemporary Education and Communication Technology, pp. 173–179, 2016

Kirillov, A. et al.: Panoptic Segmentation (2019)
Kirillov, Alexander; He, Kaiming; Girshick, Ross; Rother, Carsten; Dollár, Piotr: Panoptic Segmentation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9396–9405, 2019

Klamann, B. et al.: Defining Pass-/Fail-Criteria for Particular Tests (2019)
Klamann, Björn; Lippert, Moritz; Amersbach, Christian; Winner, H.: Defining Pass-/Fail-Criteria for Particular Tests of Automated Driving Functions, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 169–174, 2019

Klueck, F. et al.: Using Ontologies for Test Suites Generation (2018)
Klueck, F.; Li, Y.; Nica, M.; Tao, J.; Wotawa, F.: Using Ontologies for Test Suites Generation for Automated and Autonomous Driving Functions, in: 2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 118–123, 2018

Kobs, K. et al.: SimLoss: Class Similarities in Cross Entropy (2020)
Kobs, Konstantin; Steininger, M.; Zehe, Albin; Lautenschlager, Florian; Hotho, A.: SimLoss: Class Similarities in Cross Entropy, in: Foundations of Intelligent Systems, pp. 431–439, 2020

Koivisto, M. et al.: An ERP study of change detection (2003)
Koivisto, M.; Revonsuo, A.: An ERP study of change detection, change blindness, and visual awareness, in: Psychophysiology, Vol. 40, pp. 423–429, 2003

Krajewski, R. et al.: The highD Dataset (2018)
Krajewski, Robert; Bock, Julian; Kloeker, Laurent; Eckstein, Lutz: The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2118–2125, 2018

Kuba, M. et al.: Motion-onset VEPs: characteristics, methods, and diagnostic use (2007)
Kuba, M.; Kubová, Z.; Kremlácek, J.; Langrová, J.: Motion-onset VEPs: characteristics, methods, and diagnostic use, in: Vision Research, Vol. 47, pp. 189–202, 2007

Kuznetsova, A. et al.: The Open Images Dataset V4 (2020)
Kuznetsova, Alina; Rom, Hassan; Alldrin, Neil; Uijlings, Jasper; Krasin, Ivan; Pont-Tuset, Jordi; Kamali, Shahab; Popov, Stefan; Malloci, Matteo; Kolesnikov, Alexander; Duerig, Tom; Ferrari, Vittorio: The Open Images Dataset V4, in: International Journal of Computer Vision, Vol. 128, pp. 1956–1981, 2020

Laddha, A. et al.: RV-FuseNet (2021)
Laddha, A.; Gautam, Shivam; Meyer, Gregory P.; Vallespi-Gonzalez, Carlos: RV-FuseNet: Range View based Fusion of Time-Series LiDAR Data for Joint 3D Object Detection and Motion Forecasting, in: (Hrsg.): 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021

Lambert, J. et al.: MSeg: A Composite Dataset (2020)
Lambert, J.; Liu, Z.; Sener, O.; Hays, J.; Koltun, V.: MSeg: A Composite Dataset for Multi-Domain Semantic Segmentation, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2876–2885, 2020

Lang, A. H. et al.: PointPillars: Fast Encoders for Object Detection from Point Clouds (2019)
Lang, A. H.; Vora, S.; Caesar, H.; Zhou, L.; Yang, J.; Beijbom, O.: PointPillars: Fast Encoders for Object Detection from Point Clouds, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 12697–12705, 2019

Le Mero, L. et al.: Survey on Imitation Learning Techniques (2022)
Le Mero, Luc; Yi, Dewei; Dianati, Mehrdad; Mouzakitis, Alexandros: A Survey on Imitation Learning Techniques for End-to-End Autonomous Vehicles, in: IEEE Transactions on Intelligent Transportation Systems, Vol. 23, pp. 14128–14147, 2022

Lee, J. et al.: Analysis of roadside accident frequency (1999)
Lee, Jinsun; Mannering, Fred: Analysis of roadside accident frequency and severity and roadside safety management, URL: https://depts.washington.edu/trac/bulkdisk/pdf/475.1.pdf, 1999

Lefèvre, S. et al.: A survey on motion prediction and risk assessment (2014)
Lefèvre, Stéphanie; Vasquez, Dizan; Laugier, Christian: A survey on motion prediction and risk assessment for intelligent vehicles, in: ROBOMECH Journal, Vol. 1, pp. 1–14, 2014

Levin, C. A. et al.: Perceived interobject distance (1993)
Levin, C. A.; Haber, R. N.: Visual angle as a determinant of perceived interobject distance, in: Perception & psychophysics, Vol. 54, pp. 250–259, 1993

Li, G. et al.: Visual Saliency Based on Multiscale Deep Features (2015)
Li, Guanbin; Yu, Yizhou: Visual Saliency Based on Multiscale Deep Features, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5455–5463, 2015

Li, Y. et al.: Learning to associate: HybridBoosted multi-target tracker for crowded scene (2019)
Li, Y.; Huang, C.; Nevatia, R.: Learning to associate: HybridBoosted multi-target tracker for crowded scene, in: (Hrsg.): 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019

Li, Z. et al.: Perceived azimuth direction is exaggerated (2016)
Li, Z.; Durgin, F. H.: Perceived azimuth direction is exaggerated: Converging evidence from explicit and implicit measures, in: Journal of vision, Vol. 16, pp. 1–19, 2016

Lin, T.-Y. et al.: Microsoft COCO: Common Objects in Context (2014)
Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C. L.; Dollár, P.: Microsoft COCO: Common Objects in Context, in: Computer Vision – ECCV 2014, pp. 740–755, 2014

Liu, L. et al.: Deep Learning for Generic Object Detection: A Survey (2020)
Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M.: Deep Learning for Generic Object Detection: A Survey, in: International Journal of Computer Vision, Vol. 128, pp. 261–318, 2020

Liu, P. et al.: How Safe Is Safe Enough for Self-Driving Vehicles? (2019)
Liu, Peng; Yang, Run; Xu, Zhigang: How Safe Is Safe Enough for Self-Driving Vehicles?, in: Risk analysis : an official publication of the Society for Risk Analysis, Vol. 39, pp. 315–325, 2019

Liu, X.-F. et al.: Reinforced Wasserstein Training (2020)
Liu, Xiao-Feng; Zhang, Yimeng; Liu, Xiongchang; Bai, Song; Li, Site; You, J.: Reinforced Wasserstein Training for Severity-Aware Semantic Segmentation in Autonomous Driving, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), vol. abs/2008.04751, pp. 12563–12572, 2020

Liu, Y. et al.: 1st Place Solutions for OpenImage2019 (2020)
Liu, Yu; Song, Guanglu; Zang, Yuhang; Gao, Yan; Xie, Enze; Yan, Junjie; Loy, Chen Change; Wang, Xiaogang: 1st Place Solutions for OpenImage2019 – Object Detection and Instance Segmentation, in: ArXiv, Vol. abs/2003.07557, 2020

Liu, Z. et al.: BEVFusion (2023)
Liu, Z.; Tang, H.; Amini, A.; Yang, X.; Mao, H.; Rus, D.; Han, S.: BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation, in: (Hrsg.): 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023

Loomis, J. M. et al.: Is the anisotropy of perceived 3-D shape invariant across scale? (1999)
Loomis, J. M.; Philbeck, J. W.: Is the anisotropy of perceived 3-D shape invariant across scale?, in: Perception & psychophysics, Vol. 61, pp. 397–402, 1999

Lu, D. et al.: CAROM Air (2023)
Lu, Duo; Eaton, Eric; Weg, Matt; Wang, Wei; Como, Steven; Wishart, Jeffrey; Yu, Hongbin; Yang, Yezhou: CAROM Air - Vehicle Localization and Traffic Scene Reconstruction from Aerial Videos, in: 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 10666–10673, 2023

Luck, S. J. et al.: The capacity of visual working memory (1997)
Luck, S. J.; Vogel, E. K.: The capacity of visual working memory for features and conjunctions, in: Nature, Vol. 390, pp. 279–281, 1997

Luiten, J. et al.: HOTA: A Higher Order Metric for Evaluating Multi-object Tracking (2021)
Luiten, J.; Osep, A.; Dendorfer, P.; Torr, P.; Geiger, A.; Leal-Taixé, L.; Leibe, B.: HOTA: A Higher Order Metric for Evaluating Multi-object Tracking, in: International Journal of Computer Vision, Vol. 129, pp. 548–578, 2021

Luo, W. et al.: Multiple Object Tracking: A Literature Review (2021)
Luo, W.; Xing, J.; Milan, A.; Zhang, X.; Liu, W.; Zhao, X.; Kim, T.-K.: Multiple Object Tracking: A Literature Review, in: Artificial Intelligence, Vol. 293, pp. 1–23, 2021

Lyssenko, M. et al.: Towards Safety-Aware Pedestrian Detection (2022)
Lyssenko, M.; Gladisch, C.; Heinzemann, C.; Woehrle, M.; Triebel, R.: Towards Safety-Aware Pedestrian Detection in Autonomous Systems, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 293–300, 2022

Mahmud, S. S. et al.: Application of proximal surrogate indicators (2017)
Mahmud, S. SohelM.; Ferreira, Luis; Hoque, Md. Shamsul; Tavassoli, Ahmad: Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs, in: IATSS Research, Vol. 41, pp. 153–163, 2017

Makrigiorgos, A. et al.: Human Visual Attention Prediction (2019)
Makrigiorgos, Alexander; Shafti, Ali; Harston, Alex; Gérard, Julien; Faisal, A.: Human Visual Attention Prediction Boosts Learning & Performance of Autonomous Driving Agents, in: ArXiv, Vol. abs/1909.05003, 2019

Mao, J. et al.: One Million Scenes for Autonomous Driving: ONCE Dataset (2021)
Mao, J.; Niu, M.; Jiang, C.; Liang, H.; Chen, J.; Liang, X.; Li, Y.; Ye, C.; Zhang, W.; Li, Z.; Yu, J.; Xu, H.; Xu, C.: One Million Scenes for Autonomous Driving: ONCE Dataset, in: ArXiv, pp. 1–21, 2021

Mao, J. et al.: 3D Object Detection for Autonomous Driving (2023)
Mao, Jiageng; Shi, Shaoshuai; Wang, Xiaogang; Li, Hongsheng: 3D Object Detection for Autonomous Driving: A Comprehensive Survey, in: International Journal of Computer Vision, pp. 1909–1963, 2023

Markkula, G. et al.: A farewell to brake reaction times? (2016)
Markkula, G.; Engström, J.; Lodin, J.; Bärgman, J.; Victor, T.: A farewell to brake reaction times? Kinematics-dependent brake response in naturalistic rear-end emergencies, in: Accident; analysis and prevention, Vol. 95, pp. 209–226, 2016

Mckee, S. P. et al.: The precision of size constancy (1992)
Mckee, S. P.; Welch, L.: The precision of size constancy, in: Vision Research, Vol. 32, pp. 1447–1460, 1992

Mehrabi, M. et al.: Bounds on the Approximation Power (2018)
Mehrabi, Mohammad; Tchamkerten, Aslan; Yousefi, Mansoor I.: Bounds on the Approximation Power of Feedforward Neural Networks, in: ArXiv, Vol. abs/1806.11416, 2018

Meletis, P. et al.: Training on Multiple Heterogeneous Datasets (2018)
Meletis, Panagiotis; Dubbelman, Gijs: Training of Convolutional Networks on Multiple Heterogeneous Datasets for Street Scene Semantic Segmentation, in: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 1045–1050, 2018

Metwaly, K. M. et al.: CAR - Cityscapes Attributes Recognition (2021)
Metwaly, Kareem M.; Kim, Aerin; Branson, Elliot; Monga, Vishal: CAR - Cityscapes Attributes Recognition A Multi-category Attributes Dataset for Autonomous Vehicles, in: ArXiv, Vol. abs/2111.08243, 2021

Metzger, K. A. et al.: A Fine-Grained Dataset (2021)
Metzger, Kai A.; Mortimer, Peter; Wuensche, H.: A Fine-Grained Dataset and its Efficient Semantic Segmentation for Unstructured Driving Scenarios, in: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 7892–7899, 2021

Miller, G. A.: WordNet: a lexical database for English (1995)
Miller, George A.: WordNet: a lexical database for English, in: Communications of the ACM, Vol. 38, pp. 39–41, 1995

Ministry of Justice: Road Traffic Act (1960)
Ministry of Justice: Road Traffic Act, URL: https://www.japaneselawtranslation.go.jp/en/laws/view/2962/en, 1960

MMDetection3D Contributors: MMDetection3D (2020)
MMDetection3D Contributors: MMDetection3D: OpenMMLab next-generation platform for general 3D object detection, URL: https://github.com/open-mmlab/mmdetection3d, 2020

Moeller, B. et al.: What a car does to your perception (2016)
Moeller, B.; Zoppke, H.; Frings, C.: What a car does to your perception: Distance evaluations differ from within and outside of a car, in: Psychonomic Bulletin & Review, Vol. 23, pp. 781–788, 2016

Moghadam, M. et al.: A Hierarchical Architecture (2019)
Moghadam, Majid; Elkaim, Gabriel Hugh: A Hierarchical Architecture for Sequential Decision-Making in Autonomous Driving using Deep Reinforcement Learning, in: ArXiv, Vol. abs/1906.08464, 2019

Mohta, A. et al.: Investigating the Effect of Sensor Modalities (2021)
Mohta, Abhishek; Chou, Fang-Chieh; Becker, B. C.; Vallespi-Gonzalez, Carlos; Djuric, Nemanja: Investigating the Effect of Sensor Modalities in Multi-Sensor Detection-Prediction Models, in: ArXiv, Vol. abs/2101.03279, 2021

Mori, K. et al.: Conservative Estimation of Perception Relevance (2023)
Mori, Ken; Storms, Kai; Peters, Steven: Conservative Estimation of Perception Relevance of Dynamic Objects for Safe Trajectories in Automotive Scenarios, in: 2023 IEEE International Conference on Mobility, Operations, Services and Technologies (MOST), pp. 83–95, 2023

Mori, K. T. et al.: Systematic Classification Requirements (2023)
Mori, Ken T.; Brown, Trent; Peters, Steven: Towards Establishing Systematic Classification Requirements for Automated Driving, in: 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1–8, 2023

Mori, K. T. et al.: The Inadequacy of Discrete Scenarios (2022)
Mori, Ken T.; Liang, Xu; Elster, Lukas; Peters, Steven: The Inadequacy of Discrete Scenarios in Assessing Deep Neural Networks, in: IEEE Access, Vol. 10, pp. 118236–118242, 2022

Mori, K. T. et al.: SHARD (2023)
Mori, Ken T.; Peters, Steven: SHARD: Safety and Human Performance Analysis for Requirements in Detection, in: IEEE Transactions on Intelligent Vehicles, pp. 1–12, 2023

Murakami, I.: Correlations between fixation stability and visual motion sensitivity (2004)
Murakami, I.: Correlations between fixation stability and visual motion sensitivity, in: Vision Research, Vol. 44, pp. 751–761, 2004

nachiket92: PGP: Multimodal Trajectory Prediction Conditioned on Lane-Graph Traversals (2022)
nachiket92: PGP: Multimodal Trajectory Prediction Conditioned on Lane-Graph Traversals, URL: https://github.com/nachiket92/PGP, 2022

Neuhold, G. et al.: The Mapillary Vistas Dataset (2017)
Neuhold, Gerhard; Ollmann, Tobias; Bulò, Samuel Rota; Kontschieder, Peter: The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 5000–5009, 2017

Niitani, Y. et al.: Team PFDet's Methods for Open Images Challenge 2019 (2019)
Niitani, Yusuke; Ogawa, Toru; Suzuki, Shuji; Akiba, Takuya; Kerola, Tommi; Ozaki, Kohei; Sano, Shotaro: Team PFDet's Methods for Open Images Challenge 2019, in: ArXiv, Vol. abs/1910.11534, 2019

Nilsson, G.: Traffic Safety Dimensions (2004)
Nilsson, G.: Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety, Doctoral Thesis, Lund Institute of Technology, 2004

Nolte, M. et al.: Towards a skill- and ability-based development process (2017)
Nolte, Marcus; Bagschik, Gerrit; Jatzkowski, Inga; Stolte, Torben; Reschka, Andreas; Maurer, Markus: Towards a skill- and ability-based development process for self-aware automated road vehicles, in: (Hrsg.): 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017

nuro: Delivering Safety (2021)
nuro: Delivering Safety, URL: https://nuro.sfo3.digitaloceanspaces.com/nuro-vssa-2022.pdf?mtime=20220111152245&focal=none, 2021

nuScenes: nuScenes Detection Task: Leaderboard (2020)
nuScenes: nuScenes Detection Task: Leaderboard, URL: https://www.nuscenes.org/object-detection?externalData=all&mapData=all&modalities=Any, 2020

nuScenes: nuScenes prediction task: Leaderboard (2020)
nuScenes: nuScenes prediction task: Leaderboard, URL: https://www.nuscenes.org/prediction, 2020

Oksuz, K. et al.: One Metric to Measure them All (2020)
Oksuz, K.; Cam, B. C.; Kalkan, S.; Akbas, E.: One Metric to Measure them All: Localisation Recall Precision (LRP) for Evaluating Visual Detection Tasks, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 44, pp. 9446–9463, 2020

Orzechowski, P. F. et al.: Tackling Occlusions & Limited Sensor Range (2018)
Orzechowski, Piotr Franciszek; Meyer, Annika; Lauer, Martin: Tackling Occlusions & Limited Sensor Range with Set-based Safety Verification, in: (Hrsg.): 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018

Palazzi, A. et al.: Predicting the Driver's Focus of Attention (2018)
Palazzi, Andrea; Abati, Davide; Calderara, Simone; Solera, Francesco; Cucchiara, Rita: Predicting the Driver's Focus of Attention: the DR(eye)VE Project, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 41, pp. 1270–1733, 2018

Pan, B. et al.: Cross-view Semantic Segmentation (2020)
Pan, Bowen; Sun, Jiankai; Leung, Ho Yin Tiga; Andonian, Alex; Zhou, Bolei: Cross-view Semantic Segmentation for Sensing Surroundings, in: IEEE Robotics and Automation Letters, Vol. 5, pp. 4867–4873, 2020

Pang, Z. et al.: SimpleTrack (2023)
Pang, Z.; Li, Z.; Wang, N.: SimpleTrack: Understanding and Rethinking 3D Multi-object Tracking, in: (Hrsg.): Computer Vision - ECCV 2022 Workshops, 2023

Patil, A. et al.: The H3D Dataset (2019)
Patil, Abhishek; Malla, Srikanth; Gang, Haiming; Chen, Yi-Ting: The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes, in: IEEE International Conference on Robotics and Automation (ICRA), pp. 9552–9557, 2019

PEGASUS Project: PEGASUS Method: An Overview (2019)
PEGASUS Project: PEGASUS Method: An Overview, URL: https://www.pegasusprojekt.de/files/tmpl/Pegasus-Abschlussveranstaltung/PEGASUS-Gesamtmethode.pdf, 2019

Pham, Q.-H. et al.: A 3D Dataset (2020)
Pham, Q.-H.; Sevestre, P.; Pahwa, R. S.; Zhan, H.; Pang, C. H.; Chen, Y.; Mustafa, A.; Chandrasekhar, V.; Lin, J.: A 3D Dataset: Towards Autonomous Driving in Challenging Environments, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 2267–2273, 2020

Philion, J. et al.: Learning to Evaluate Perception (2020)
Philion, J.; Kar, A.; Fidler, S.: Learning to Evaluate Perception Models Using Planner-Centric Metrics, in: (Hrsg.): 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020

Philion, J. et al.: Lift, Splat, Shoot (2020)
Philion, Jonah; Fidler, S.: Lift, Splat, Shoot: Encoding Images From Arbitrary Camera Rigs by Implicitly Unprojecting to 3D, in: Vedaldi, Andrea (Hrsg.): Computer vision - ECCV 2020, Springer, 2020

Philipp, R. et al.: Accuracy Requirements for Environmental Perception (2021)
Philipp, R.; Qian, H.; Hartjen, L.; Schuldt, F.; Howar, F.: Simulation-Based Elicitation of Accuracy Requirements for the Environmental Perception of Autonomous Vehicles, in: Leveraging Applications of Formal Methods, Verification and Validation, pp. 129–145, 2021

Philipp, R. et al.: Systematization of Relevant Road Users (2022)
Philipp, R.; Rehbein, J.; Grun, F.; Hartjen, L.; Zhu, Z.; Schuldt, F.; Howar, F.: Systematization of Relevant Road Users for the Evaluation of Autonomous Vehicle Perception, in: (Hrsg.): 2022 IEEE International Systems Conference (SysCon), 2022

Piazzoni, A. et al.: Modeling Perception Errors (2020)
Piazzoni, Andrea; Cherian, Jim; Slavík, Martin; Dauwels, Justin: Modeling Perception Errors towards Robust Decision Making in Autonomous Vehicles, in: International Joint Conference on Artificial Intelligence, pp. 3494–3500, 2020

Pinggera, P. et al.: Lost and Found (2016)
Pinggera, Peter; Ramos, Sebastian; Gehrig, Stefan; Franke, Uwe; Rother, Carsten; Mester, Rudolf: Lost and Found: detecting small road hazards for self-driving vehicles, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1099–1106, 2016

Poncelet, R. et al.: Safe Geometric Speed Planning (2020)
Poncelet, Renaud; Verroust-Blondet, Anne; Nashashibi, F.: Safe Geometric Speed Planning Approach for Autonomous Driving through Occluded Intersections, in: 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), pp. 393–399, 2020

Pony.ai: Our Approach to Safety: Pony.ai Safety Report (2020)
Pony.ai: Our Approach to Safety: Pony.ai Safety Report, URL: https://static.cdn.xiaomazhixing.com/file/1652065422385/ed5046fc-395a-46cf-a471-1728658f5001/Pony.ai%20safety%20report%20(Mar%2031%202022%20edit).pdf, 2020

Pylyshyn, Z. W. et al.: Tracking multiple independent targets (1988)
Pylyshyn, Z. W.; Storm, R. W.: Tracking multiple independent targets: evidence for a parallel tracking mechanism, in: Spatial Vision, Vol. 3, pp. 179–197, 1988

Qi, C. R. et al.: Deep Hough Voting for 3D Object Detection in Point Clouds (2019)
Qi, C. R.; Litany, O.; He, K.; Guibas, L. J.: Deep Hough Voting for 3D Object Detection in Point Clouds, in: (Hrsg.): 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019

Qi, C. R. et al.: Offboard 3D Object Detection (2021)
Qi, C. R.; Zhou, Y.; Najibi, M.; Sun, P.; Vo, K.; Deng, B.; Anguelov, D.: Offboard 3D Object Detection from Point Cloud Sequences, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 6130–6140, 2021

Redmon, J. et al.: YOLOv3: An Incremental Improvement (2018)
Redmon, J.; Farhadi, A.: YOLOv3: An Incremental Improvement, in: ArXiv, Vol. 1804.02767, 2018

Redmon, J. et al.: YOLO9000: Better, Faster, Stronger (2017)
Redmon, Joseph; Farhadi, Ali: YOLO9000: Better, Faster, Stronger, in: (Hrsg.): 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017

Richards, D. C.: Relationship between Speed and Risk of Fatal Injury (2010)
Richards, D. C.: Relationship between Speed and Risk of Fatal Injury: Pedestrians and Car Occupants, URL: https://nacto.org/docs/usdg/relationship_between_speed_risk_fatal_injury_pedestrians_and_car_occupants_richards.pdf, 2010

Riedmaier, S. et al.: Survey on Scenario-Based Safety Assessment (2020)
Riedmaier, Stefan; Ponn, Thomas; Ludwig, Dieter; Schick, Bernhard; Diermeyer, Frank: Survey on Scenario-Based Safety Assessment of Automated Vehicles, in: IEEE Access, Vol. 8, pp. 87456–87477, 2020

On-Road Automated Driving (ORAD) committee: Taxonomy and Definitions (30.04.2021)
On-Road Automated Driving (ORAD) committee: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, 30.04.2021

Russakovsky, O. et al.: ImageNet Large Scale Visual Recognition Challenge (2015)
Russakovsky, Olga; Deng, Jia; Su, Hao; Krause, Jonathan; Satheesh, Sanjeev; Ma, Sean; Huang, Zhiheng; Karpathy, Andrej; Khosla, Aditya; Bernstein, Michael; Berg, Alexander C.; Fei-Fei, Li: ImageNet Large Scale Visual Recognition Challenge, in: International Journal of Computer Vision, Vol. 115, pp. 211–252, 2015

Sadat, A. et al.: Perceive, Predict, and Plan (2020)
Sadat, A.; Casas, S.; Ren, Mengye; Wu, X.; Dhawan, Pranaab; Urtasun, R.: Perceive, Predict, and Plan: Safe Motion Planning Through Interpretable Semantic Representations, in: ArXiv, 2020

Schmidt, C.: Fahrstrategien zur Unfallvermeidung (2013)
Schmidt, Christian: Fahrstrategien zur Unfallvermeidung im Straßenverkehr für Einzel- und Mehrobjektszenarien, 2013

Schönemann, V. et al.: Scenario-Based Functional Safety (2019)
Schönemann, Valerij; Winner, Hermann; Glock, Thomas; Otten, Stefan; Sax, Eric; Boeddeker, Bert; Verhaeg, Geert; Tronci, Fabrizio; Padilla, Gustavo G.: Scenario-Based Functional Safety for Automated Driving on the Example of Valet Parking, in: Advances in Information and Communication Networks, pp. 53–64, 2019

Schwalbe, G. et al.: Structuring the Safety Argumentation for Perception (2020)
Schwalbe, Gesina; Knie, Bernhard; Sämann, Timo; Dobberphul, Timo; Gauerhof, Lydia; Raafatnia, Shervin; Rocco, Vittorio: Structuring the Safety Argumentation for Deep Neural Network Based Perception in Automotive Applications, in: Casimiro, António (Hrsg.): Computer safety, reliability, and security, Springer, 2020

Shalev-Shwartz, S. et al.: On a Formal Model of Safe and Scalable Self-driving Cars (2017)
Shalev-Shwartz, S.; Shammah, S.; Shashua, A.: On a Formal Model of Safe and Scalable Self-driving Cars, in: ArXiv, Vol. 1708.06374, 2017

Simons, D. J.: Current Approaches to Change Blindness (2000)
Simons, D. J.: Current Approaches to Change Blindness, in: Visual Cognition, Vol. 7, pp. 1–15, 2000

Simons, D. J. et al.: Failure to detect changes to people during a real-world interaction (1998)
Simons, D. J.; Levin, D. T.: Failure to detect changes to people during a real-world interaction, in: Psychonomic Bulletin & Review, Vol. 5, pp. 644–649, 1998

Singh, B. et al.: R-FCN-3000 at 30fps: Decoupling Detection and Classification (2018)
Singh, B.; Li, H.; Sharma, A.; Davis, L. S.: R-FCN-3000 at 30fps: Decoupling Detection and Classification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1081–1090, 2018

Snowden, R. J. et al.: Motion Perception in the Ageing Visual System (2006)
Snowden, R. J.; Kavanagh, E.: Motion Perception in the Ageing Visual System: Minimum Motion, Motion Coherence, and Speed Discrimination Thresholds, in: Perception, Vol. 35, pp. 9–24, 2006

Song, B. et al.: A Stochastic Graph Evolution Framework for Robust Multi-target Tracking (2010)
Song, B.; Jeng, T.-Y.; Staudt, E.; Roy-Chowdhury, A. K.: A Stochastic Graph Evolution Framework for Robust Multi-target Tracking, in: Hutchison, David; Kanade, Takeo; Kittler, Josef; Kleinberg, Jon M.; Mattern, Friedemann; Mitchell, John C.; Naor, Moni; Nierstrasz, Oscar; Pandu Rangan, C.; Steffen, Bernhard; Sudan, Madhu; Terzopoulos, Demetri; Tygar, Doug; Vardi, Moshe Y.; Weikum, Gerhard; Daniilidis, Kostas; Maragos, Petros; Paragios, Nikos (Hrsg.): Computer Vision - ECCV 2010, Springer, 2010

Sontges, S. et al.: Worst-case Analysis of the Time-To-React Using Reachable Sets (2018)
Sontges, S.; Koschi, M.; Althoff, M.: Worst-case Analysis of the Time-To-React Using Reachable Sets, in: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 1891–1897, 2018

Spanfelner, B. et al.: Challenges in applying the ISO 26262 for driver assistance systems (2012)
Spanfelner, B.; Richter, Detlev; Ebel, S.; Wilhelm, U.; Patz, C.: Challenges in applying the ISO 26262 for driver assistance systems, in: (Hrsg.): 5. Tagung Fahrerassistenz, 2012

Srikanth, S. et al.: INFER: INtermediate representations for FuturE pRediction (2019)
Srikanth, Shashank; Ansari, Junaid Ahmed; R, Karnik Ram; Sharma, Sarthak; J, Krishna Murthy; K, Madhava Krishna: INFER: INtermediate representations for FuturE pRediction, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 942–949, 2019

Stapelberg, R. F.: Handbook of Reliability (2009)
Stapelberg, R. F.: Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design, Springer London, 2009

State of California Department of Motor Vehicles: Autonomous Vehicle Testing Permit Holders (2022)
State of California Department of Motor Vehicles: Autonomous Vehicle Testing Permit Holders, URL: https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-testing-permit-holders/, 2022

Statistisches Bundesamt: Verkehr: Verkehrsunfälle April (2022)
Statistisches Bundesamt: Verkehr: Verkehrsunfälle: April 2022, URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-monat-2080700221044.pdf?__blob=publicationFile, 2022

Storms, K. et al.: SURE-Val: Safe Urban Relevance Extension and Validation (2023)
Storms, K.; Mori, K.; Peters, S.: SURE-Val: Safe Urban Relevance Extension and Validation, in: (Hrsg.): 15. Workshop Fahrerassistenz und automatisiertes Fahren - FAS 2023, 2023

Storms, K.: Context Aware Data Reduction (2023)
Storms, Kai: Context Aware Data Reduction for Highly Automated Driving, Dissertation, Technische Universität Darmstadt, 2023

Strauss, M. et al.: Distance Estimation Error in a Roadway Setting (2009)
Strauss, M.; Carnahan, J.: Distance Estimation Error in a Roadway Setting, in: The Police Journal, Vol. 82, pp. 247–264, 2009

Sun, P. et al.: Scalability in Perception (2020)
Sun, P.; Kretzschmar, H.; Dotiwalla, X.; Chouard, A.; Patnaik, V.; Tsui, P.; Guo, J.; Zhou, Y.; Chai, Y.; Caine, B.; Vasudevan, V.; Han, W.; Ngiam, J.; Zhao, H.; Timofeev, A.; Ettinger, S.; Krivokon, M.; Gao, A.; Joshi, A.; Zhao, S.; Cheng, S.; Zhang, Y.; Shlens, J.; Chen, Z.; Anguelov, D.: Scalability in Perception for Autonomous Driving: Waymo Open Dataset, in: (Hrsg.): 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020

Takeuchi, T. et al.: Velocity discrimination in scotopic vision (2000)
Takeuchi, T.; Valois, K. K.: Velocity discrimination in scotopic vision, in: Vision Research, Vol. 40, pp. 2011–2024, 2000

Tang, Z. et al.: CityFlow (2019)
Tang, Zheng; Naphade, Milind; Liu, Ming-Yu; Yang, Xiaodong; Birchfield, Stan; Wang, Shuo; Kumar, Ratnesh; Anastasiu, David; Hwang, Jenq-Neng: CityFlow: A City-Scale Benchmark for Multi-Target Multi-Camera Vehicle Tracking and Re-Identification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8789–8798, 2019

Tesla: Upgrading Autopilot: Seeing the World in Radar (2016)
Tesla: Upgrading Autopilot: Seeing the World in Radar, URL: https://www.tesla.com/blog/upgrading-autopilot-seeing-world-radar, 2016

Thorn, E. et al.: A Framework for Automated Driving System Testable Cases and Scenarios (2018)
Thorn, E.; Kimmel, S.; Chaka, M.: A Framework for Automated Driving System Testable Cases and Scenarios, 2018

Thorpe, S. et al.: Speed of processing in the human visual system (1996)
Thorpe, S.; Fize, D.; Marlot, C.: Speed of processing in the human visual system, in: Nature, Vol. 381, pp. 520–522, 1996

Topan, S. et al.: Interaction-Dynamics-Aware Perception Zones (2022)
Topan, S.; Leung, K.; Chen, Y.; Tupekar, P.; Schmerling, E.; Nilsson, J.; Cox, M.; Pavone, M.: Interaction-Dynamics-Aware Perception Zones for Obstacle Detection Safety Evaluation, in: (Hrsg.): 2022 IEEE Intelligent Vehicles Symposium (IV), 2022

Treisman, A. et al.: Illusory conjunctions in the perception of objects (1982)
Treisman, A.; Schmidt, H.: Illusory conjunctions in the perception of objects, in: Cognitive Psychology, Vol. 14, pp. 107–141, 1982

Ullah, I. et al.: A brief survey of visual saliency detection (2020)
Ullah, Inam; Jian, Muwei; Hussain, Sumaira; Guo, Jie; Yu, Hui; Wang, Xing; Yin, Yilong: A brief survey of visual saliency detection, in: Multimedia Tools and Applications, Vol. 79, pp. 34605–34645, 2020

van Kempen, R. et al.: Combined Registration and Fusion (2023)
van Kempen, Raphael; Adrian Heidrich, Laurenz; Lampe, Bastian; Woopen, Timo; Eckstein, Lutz: Combined Registration and Fusion of Evidential Occupancy Grid Maps for Live Digital Twins of Traffic, in: 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1–6, 2023

van Rullen, R. et al.: The time course of visual processing (2001)
van Rullen, R.; Thorpe, S. J.: The time course of visual processing: from early perception to decision-making, in: Journal of cognitive neuroscience, Vol. 13, pp. 454–461, 2001

Vázquez, J. L. et al.: Deep Interactive Motion Prediction and Planning (2022)
Vázquez, José Luis; Liniger, Alexander; Schwarting, Wilko; Rus, Daniela; van Gool, Luc: Deep Interactive Motion Prediction and Planning: Playing Games with Motion Prediction Models, in: ArXiv, Vol. abs/2204.02392, 2022

Victor, T. et al.: Safety Performance of the Waymo System (2023)
Victor, Trent; Kusano, Kristofer; Gode, Tilia; Chen, Ruoshu; Schwal, Matthew: Safety Performance of the Waymo Rider-Only Automated Driving System at One Million Miles, URL: https://storage.googleapis.com/waymo-uploads/files/documents/safety/Safety%20Performance%20of%20Waymo%20RO%20at%201M%20miles.pdf, 2023

Volk, G. et al.: A Comprehensive Safety Metric (2020)
Volk, G.; Gamerdinger, J.; Betnuth, A. v.; Bringmann, O.: A Comprehensive Safety Metric to Evaluate Perception in Autonomous Systems, in: (Hrsg.): 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020

Vul, E. et al.: The structure of illusory conjunctions (2020)
Vul, E.; Rieth, C. A.; Lew, T. F.; Rich, A. N.: The structure of illusory conjunctions reveals hierarchical binding of multipart objects, in: Attention, perception & psychophysics, Vol. 82, pp. 550–563, 2020

Wachenfeld, W. et al.: Virtual Assessment of Automation in Field Operation (2015)
Wachenfeld, W.; Winner, H.: Virtual Assessment of Automation in Field Operation ? A New Runtime Validation Method, in: 10. Workshop Fahrerassistenzsysteme, 2015

Wachenfeld, W. et al.: The worst-time-to-collision metric (2016)
Wachenfeld, Walther; Junietz, Phillip; Wenzel, Raphael; Winner, Hermann: The worst-time-to-collision metric for situation identification, in: 2016 IEEE Intelligent Vehicles Symposium (IV), pp. 729–734, 2016

Wang, C. et al.: Overcoming Challenges of Validation (2019)
Wang, C.; Winner, H.: Overcoming Challenges of Validation Automated Driving and Identification of Critical Scenarios, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 2639–2644, 2019

Wang, T. et al.: FCOS3D (2021)
Wang, T.; Zhu, X.; Pang, J.; Lin, D.: FCOS3D: Fully Convolutional One-Stage Monocular 3D Object Detection, in: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 913–922, 2021

Wang, Y. et al.: 1st Place Solutions for Waymo Open Dataset Challenges - 2D and 3D Tracking (2020)
Wang, Y.; Chen, S.; Huang, L.; Ge, R.; Hu, Y.; Ding, Z.; Liao, J.: 1st Place Solutions for Waymo Open Dataset Challenges - 2D and 3D Tracking, in: ArXiv, Vol. 2006.15506, 2020

Wang, Y. et al.: Train in Germany, Test in The USA (2020)
Wang, Y.; Chen, X.; You, Y.; Erran, L.; Hariharan, B.; Campbell, M.; Weinberger, K. Q.; Chao, W.-L.: Train in Germany, Test in The USA: Making 3D Object Detectors Generalize, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11710–11720, 2020

Waymo LLC: 3D Detection: Waymo Open Dataset: Challenge 1 (2019-2022)
Waymo LLC: 3D Detection: Waymo Open Dataset: Challenge 1, URL: https://waymo.com/open/challenges/2020/3d-detection/, 2019-2022

Waymo LLC: Waymo Safety Report (2020)
Waymo LLC: Waymo Safety Report, 2020

Weng, X. et al.: 3D Multi-Object Tracking (2020)
Weng, X.; Wang, J.; Held, D.; Kitani, K.: 3D Multi-Object Tracking: A Baseline and New Evaluation Metrics, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 10359–10366, 2020

Westhofen, L. et al.: Criticality Metrics for Automated Driving (2023)
Westhofen, L.; Neurohr, C.; Koopmann, T.; Butz, M.; Schütt, B.; Utesch, F.; Kramer, B.; Gutenkunst, C.; Böde, E.: Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art, in: Archives of Computational Methods in Engineering, Vol. 30, pp. 1–35, 2023

Willers, O. et al.: Safety Concerns and Mitigation Approaches (2020)
Willers, O.; Sudholt, S.; Raafatnia, S.; Abrecht, S.: Safety Concerns and Mitigation Approaches Regarding the Use of Deep Learning in Safety-Critical Perception Tasks, in: Casimiro, António (Hrsg.): Computer safety, reliability, and security, Springer, 2020

Wilson, B. et al.: Argoverse 2 (2021)
Wilson, Benjamin; Qi, William; Agarwal, Tanmay; Lambert, John; Singh, Jagjeet; Khandelwal, Siddhesh; Pan, Bowen; Kumar, Ratnesh; Hartnett, Andrew; Kaesemodel Pontes, Jhony; Ramanan, Deva; Carr, Peter; Hays, James: Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting, in: Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, Vol. 1, 2021

Wolf, M. et al.: Safety-Aware Metric for People Detection (2021)
Wolf, M.; Douat, L. R.; Erz, M.: Safety-Aware Metric for People Detection, in: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), pp. 2759–2765, 2021

Wolfe, J. M. et al.: Evidence for a Binding Problem in Human Vision (1999)
Wolfe, J. M.; Cave, K. R.: The Psychophysical Evidence for a Binding Problem in Human Vision, in: Neuron, Vol. 24, pp. 11–17, 1999

Wu, C. et al.: A hierarchical loss (2019)
Wu, Cinna; Tygert, Mark; LeCun, Yann: A hierarchical loss and its problems when classifying non-hierarchically, in: PloS one, Vol. 14, 2019

Wu, X. et al.: Recent Advances in Deep Learning for Object Detection (2019)
Wu, X.; Sahoo, D.; Hoi, S. C. H.: Recent Advances in Deep Learning for Object Detection, in: ArXiv, 2019

Wu, Y. et al.: Size-to-depth (2018)
Wu, Y.; Ying, S.; Zheng, L.: Size-to-depth: A New Perspective for Single Image Depth Estimation, in: ArXiv, 2018

Xu, R. et al.: V2V4Real (2023)
Xu, Runsheng; Xia, Xin; Li, Jinlong; Li, Hanzhao; Zhang, Shuo; Tu, Zhengzhong; Meng, Zonglin; Xiang, Hao; Dong, Xiaoyu; Song, Rui; Yu, Hongkai; Zhou, Bolei; Ma, Jiaqi: V2V4Real: A Real-World Large-Scale Dataset for Vehicle-to-Vehicle Cooperative Perception, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13712–13722, 2023

Xu, Z. et al.: Zero-shot Deep Reinforcement Learning Driving (2018)
Xu, Zhuo; Tang, Chen; Tomizuka, Masayoshi: Zero-shot Deep Reinforcement Learning Driving Policy Transfer for Autonomous Vehicles based on Robust Control, in: (Hrsg.): 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018

Yamaguchi, K. et al.: Who are you with and where are you going? (2011)
Yamaguchi, K.; Berg, A. C.; Ortiz, L. E.; Berg, T. L.: Who are you with and where are you going?, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1345–1352, 2011

Yan, J.-J. et al.: Visual processing of the impending collision (2011)
Yan, J.-J.; Lorv, B.; Li, H.; Sun, H.-J.: Visual processing of the impending collision of a looming object: time to collision revisited, in: Journal of Vision, Vol. 11, pp. 1–25, 2011

Yang, B. et al.: Auto4D: Learning to Label 4D Objects (2021)
Yang, B.; Bai, M.; Liang, M.; Zeng, W.; Urtasun, R.: Auto4D: Learning to Label 4D Objects from Sequential Point Clouds, in: ArXiv, Vol. 2101.06586, 2021

Yang, Z. et al.: IPOD: Intensive Point-based Object Detector for Point Cloud (2018)
Yang, Z.; Sun, Y.; Liu, S.; Shen, X.; Jia, J.: IPOD: Intensive Point-based Object Detector for Point Cloud, in: ArXiv, Vol. 1812.05276, 2018

Yin, T. et al.: Center-based 3D Object Detection and Tracking (2021)
Yin, T.; Zhou, X.; Krähenbühl, P.: Center-based 3D Object Detection and Tracking, in: (Hrsg.): 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021

You, R. et al.: VoNet (2016)
You, R.; Kwon, J.-W.: VoNet: vehicle orientation classification using convolutional neural network, in: Proceedings of the 2nd International Conference on Communication and Information Processing - ICCIP '16, pp. 195–199, 2016

Yu, D. et al.: Continuous Pedestrian Orientation Estimation (2019)
Yu, D.; Xiong, H.; Xu, Q.; Wang, J.; Li, K.: Continuous Pedestrian Orientation Estimation using Human Keypoints, in: 2019 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5, 2019

Yu, F. et al.: BDD100K: A Diverse Driving Dataset (2020)
Yu, Fisher; Chen, Haofeng; Wang, Xin; Xian, Wenqi; Chen, Yingying; Liu, Fangchen; Madhavan, Vashisht; Darrell, Trevor: BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2633–2642, 2020

Yu, H. et al.: DAIR-V2X (2022)
Yu, Haibao; Luo, Yizhen; Shu, Mao; Huo, Yiyi; Yang, Zebang; Shi, Yifeng; Guo, Zhenglong; Li, Hanyu; Hu, Xing; Yuan, Jirui; Nie, Zaiqing: DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21329–21338, 2022

Yu, K. et al.: Benchmarking the Robustness (2023)
Yu, Kaicheng; Tang, Tao; Xie, Hongwei; Lin, Zhiwei; Wu, Zhongwei; Xia, Zhongyu; Liang, Tingting; Sun, Haiyang; Deng, Jiong; Hao, Dayang; Wang, Yongtao; Liang, Xi; Wang, Bing: Benchmarking the Robustness of LiDAR-Camera Fusion for 3D Object Detection, in: (Hrsg.): 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023

Zalevski, A. et al.: Size Estimation with Night Vision Goggles (2001)
Zalevski, A.; Meehan, J. W.; Hughes, P. K.: Size Estimation with Night Vision Goggles: Air Operations Division Aeronautical and Maritime Research Laboratory, URL: https://apps.dtic.mil/sti/pdfs/ADA390742.pdf, 2001

Zendel, O. et al.: WildDash - Creating Hazard-Aware Benchmarks (2018)
Zendel, Oliver; Honauer, Katrin; Murschitz, Markus; Steininger, Daniel; Dominguez, Gustavo Fernandez: WildDash - Creating Hazard-Aware Benchmarks, in: Ferrari, Vittorio (Hrsg.): Computer vision - ECCV 2018, Springer, 2018

Zendel, O. et al.: Unifying Panoptic Segmentation for Autonomous Driving (2022)
Zendel, Oliver; Schörghuber, Matthias; Rainer, Bernhard; Murschitz, Markus; Beleznai, Csaba: Unifying Panoptic Segmentation for Autonomous Driving, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21319–21328, 2022

Zhang, H. et al.: Faraway-Frustum (2021)
Zhang, Haolin; Yang, Dongfang; Yurtsever, Ekim; Redmill, Keith A.; Özgüner, Ümit: Faraway-Frustum: Dealing with Lidar Sparsity for 3D Object Detection using Fusion, in: (Hrsg.): 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021

Zhang, Y. et al.: Key Issues of Salient Object Datasets (2020)
Zhang, Yi; Zhang, Lu; Hamidouche, Wassim; Déforges, Olivier: Key Issues for the Construction of Salient Object Datasets with Large-Scale Annotation, in: 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 117–122, 2020

Zhao, H. et al.: Suraksha: A Quantitative AV Safety Evaluation Framework (2021)
Zhao, H.; Hari, S. K. S.; Tsai, T.; Sullivan, M. B.; Keckler, S. W.; Zhao, J.: Suraksha: A Quantitative AV Safety Evaluation Framework to Analyze Safety Implications of Perception Design Choices, in: 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 35–38, 2021

Zhou, X. et al.: Probabilistic two-stage detection (2021)
Zhou, X.; Koltun, V.; Krähenbühl, P.: Probabilistic two-stage detection, in: ArXiv, Vol. 2103.07461, 2021

Zhu, B. et al.: Class-balanced Grouping and Sampling for Point Cloud 3D Object Detection (2019)
Zhu, Benjin; Jiang, Zhengkai; Zhou, Xiangxin; Li, Zeming; Yu, Gang: Class-balanced Grouping and Sampling for Point Cloud 3D Object Detection, in: ArXiv, Vol. abs/1908.09492, 2019

Zoox: Safety Innovation at Zoox (2018)
Zoox: Safety Innovation at Zoox: Setting the Bar for Safety in Autonomous Driving, URL: https://zoox.com/wp-content/uploads/zoox_safety_report_volume1_2018.pdf, 2018

Own Publications

Mori, Ken T., Liang, Xu; Elster, Lukas and Peters, Steven: The Inadequacy of Discrete Scenarios in Assessing Deep Neural Networks, in IEEE Access, vol. 10, pp. 118236-118242, 2022, doi: 10.1109/ACCESS.2022.3220904.

Mori, Ken; Brown, Trent; Peters, Steven: Towards Establishing Systematic Classification Requirements for Automated Driving, 2023 IEEE Intelligent Vehicles Symposium (IV), Anchorage, AK, USA, 2023, pp. 1-8, doi: 10.1109/IV55152.2023.10186542.

Mori, Ken; Storms, Kai; Peters, Steven: Conservative Estimation of Perception Relevance of Dynamic Objects for Safe Trajectories in Automotive Scenarios, in 2023 IEEE International Conference on Mobility, Operations, Services and Technologies (MOST), Detroit, MI, USA, 2023 pp. 83-95, doi: 10.1109/MOST57249.2023.00017.

Storms, Kai; **Mori, Ken**, Peters, Steven: SURE-Val: Safe Urban Relevance Extension and Validation, in 15. Workshop Fahrerassistenz und automatisiertes Fahren, Darmstadt, Germany, 2023.

Elster, Lukas; Rosenberger, Philipp; Holder, Martin; **Mori, Ken**; Staab, Jan; Peters, Steven: Introducing the Double Validation Metric for Radar Sensor Models, 2023, PREPRINT (Version 1) available at Research Square, doi: 10.21203/rs.3.rs-3088648/v1.

Mori, Ken and Peters, Steven: SHARD: Safety and Human Performance Analysis for Requirements in Detection, in IEEE Transactions on Intelligent Vehicles, 2023,
doi: 10.1109/TIV.2023.3320395.

Supervised Theses

Gröger, Simon: Analyse und Entwicklung einer Objektklassifizierung von Zusatzzeichen der StVO basierend auf Kameradaten eines automatisierten Fahrzeugs.
Bachelor-Thesis Nr. 1380/21

Wu, Di: Entwicklung einer Methodik zur Ableitung von Parameterbeschränkungen für kombinatorisches Testen.
Master-Thesis Nr. 816/21

Stahl, Patrick: Entwicklung einer Selektionsmethodik von sicherheitsrelevanten Bildbereichen aus Szenarien für das Hochautomatisierte Fahren
Master-Thesis Nr. 821/21

Bai, Juntao: Umfelderkennung für Automatisiertes Fahren basierend auf Punktdetektionen von Radar
Master-Thesis Nr. 823/21

Cai, Wei: Redundanz von Farbkanälen für Objekterkennung in Kamerabildern
Master-Thesis Nr. 834/21

Liang, Xu Development of a Methodology for Deriving Interpretable Causes for Perception Failures
Master-Thesis Nr. 847/22

Tsai, Cheng-Ting: Sim2Real Transfer for Lidar Perception
Master-Thesis Nr. 848/22

Hsu, Wei: Development of an Evaluation Methodology for Late Fusion Algorithms
Master-Thesis Nr. 854/22

Baroianu, Rares-Cristian Development of a Methodology for Targeted Fulfillment of Decomposed Perception Requirements
Master-Thesis Nr. 863/22

Tahvildari, Hatef: Al Based Semantic Segmentation In Topo Maps Created by Aerial Radar Recordings
Master-Thesis Nr. 864/22

Wang, Xince; LoCascio, Luigi; Rayudu, Aniketh; Virgo, Benjamin: Design of a Toolbox for Testing of Metrics for 3D Object Detection Evaluation
Advanced Design Project Nr. 171/23

Cui, Yi: Development of a Reinforcement-Learning-Based Virtual Driver for Model-Based Tests of Driving Dynamics
Master-Thesis Nr. 873/23

Chen, Zhaoguo: Conceptualization and Evaluation of a Probabilistic Free Space Estimation
Master-Thesis Nr. 876/23