KI Absicherung

Overview and Poster Abstracts

Making the safety of AI-based function modules for highly automated driving verifiable.

KI ABSICHERUNG
Safe AI for Automated Driving

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag

KI ABSICHERUNG
Safe AI for Automated Driving

KI FAMILIE

KI Absicherung is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

www.ki-absicherung-projekt.de   @KI_Familie   KI Familie

Welcome

Dear Readers,

We are pleased to share our insights with you into our three-year research in the field of safeguarding AI functions for automated driving. Initiated by the VDA Leitinitiative for autonomous and connected driving, KI Absicherung was the first project of the KI Familie. Three years ago we started to advance the use of AI in automated driving systems and to shed light on it from different perspectives. Our KI Absicherung project has achieved the goal of creating an exemplary stringent safety argumentation that will advance the release of AI function modules in the context of autonomous driving. All 24 project partners, have been working together towards this goal with four external technology providers over the past three years. Of course, our first big thank-you goes to the project consortium and to the subproject leads from BMW Group, Volkswagen AG and Robert Bosch GmbH. Our project partners, which include OEMs, suppliers, technology providers as well as research institutes and universities, have worked with relentless energy and commitment and thus contributed to the successful completion of our project this year. The wide array of expertise that they were able to contribute shows in the extensive and high-quality results of the research. In addition and on behalf of our consortium, we would like to thank our sponsor, the Federal Ministry for Economic Affairs and Climate Protection (BMWK) and in particular Mr. Ernst Stöckl-Pukall, who strongly supports the topic of autonomous driving. We would also like to thank our project officer TÜV Rheinland Consulting for supporting the project. Last but not least, we would also like to thank the subcontractors of external technology partners and scientific institutions, who were also very committed to the project. Particularly noteworthy is also the project office EICT, whose advice and organizational input was highly beneficiary to our work. We are confident that the results we have summarized on the following pages will be of great interest to readers from the industry and research alike. With our methodology for Safeguarding AI developed in the project, we want to contribute to Safe AI in autonomous driving. We hope that KI Absicherung will lay the foundation for a systematic approach to make the vision of safe AI in automated driving a reality.

Dr. Stephan Scholz
Project coordinator
Volkswagen AG

PD Dr. Michael Mock
Scientific coordinator and
consortium Co-Lead
Fraunhofer Institute for Intelligent
Analysis and Information Systems (IAIS)

Greeting from the Federal Ministry for Economic Affairs and Climate Action

With its National Artificial Intelligence Strategy, the German government has laid out a framework for promoting and ensuring the safe and responsible use of AI. The predominant role of AI in the mobility sector will significantly shape the future of this market. With the “KI Projektfamilie”, which includes a total of four projects, initiated as part of the VDA flagship initiative autonomous and connected driving the German automotive industry is addressing the central opportunities and challenges of using AI in highly automated driving systems.

The project “KI Absicherung” marked the starting point and has focused on the highly complex topic of safety over the past three years. The safety argumentation developed in the project is an important building block and brings us one step closer to the mobility of the future.

As the German Federal Ministry for Economic Affairs and Climate Protection, we are happy to support cross-institutional and interdisciplinary research projects such as “KI Absicherung”.

We are convinced that the major tasks that still lie ahead for the safe use of AI in the mobility market cannot be solved by individual companies, but by bringing together different expertise from the science and industry sectors.

Therefore, our thanks goes to the 24 partners who have worked together over the past three years to make the safety of AI-based functional modules in highly automated driving verifiable.

We are pleased that the project has come to a successful conclusion and that the funding from the Federal Ministry was able to play a decisive role in making the project a reality.

Ernst Stöckl-Pukall
Head of Division for
Digitalisation and Industry 4.0,
Federal Ministry of Economic
Affairs and Climate Action

Federal Ministry
for Economic Affairs
and Climate Action

Introduction to the KI Familie

Introduction

The application of AI is a key enabler for autonomous driving. In the KI Absicherung project, AI and safety experts from industry and academia develop a methodology for a safety argumentation that systematically identifies, makes measurable and mitigates weaknesses of AI functions. The aim is to achieve an industrial consensus for a methodical approach to assure AI functions for the use case of pedestrian detection.

The Project KI Absicherung

Assuring the safety of functions that make use of AI-based algorithms is crucial for the German automotive industry in international competition. In the KI Absicherung project [1], a consortium of OEMs, suppliers, technology providers and scientific institutions is developing an „industrial consensus“ on a methodology that can identify and systematically mitigate inherent weaknesses in AI functions. The methodology also includes a systematic approach for deriving a stringent evidence-based safety argumentation. KI Absicherung

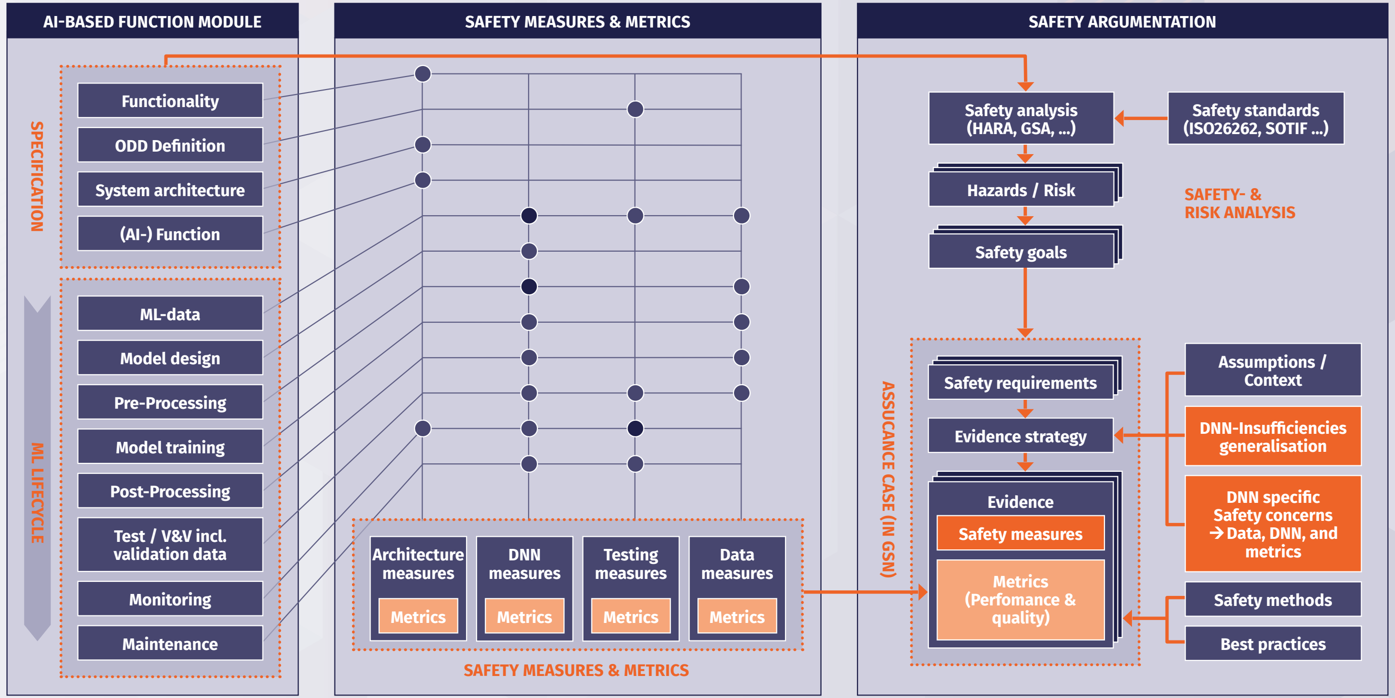


Figure 1: Approach to Safety Argumentation for AI-based Functions (© BMW Group | Robert Bosch GmbH | Fraunhofer IAIS | Volkswagen AG)

is part of the joint projects of the KI Familie. Figure 1 shows the specification and development steps of an AI function, as well as the methodology for building an evidence-based safety argumentation. This is based in particular on safety measures, metrics and tests used in the development and validation steps of the AI function.

The specification of the AI function is the elementary starting point, both for the construction of the safety argumentation and for the development of the function itself. In addition to the purely functional requirements, such as „recognition of persons on camera images“, the specification also includes the definition of the scope of use of the function, the so-called Operational Design Domain (ODD). When using Deep Neural Networks (DNN), the ODD specification also results in requirements for a systematic and representative selection of training and test data. The KI Absicherung project develops description languages and an ontology for the detailed specification of data and metadata. These are understandable for humans in order to be able to build up a comprehensible safety argumentation, as well as machine-readable in order to be able to carry out data analyses and test evaluations automatically.

DNNs can be understood as complex black box approximation functions that are optimized by training data. As such, they may have insufficiencies in the generalization capability, which in the unfavorable case can lead to a weakness of the software function. In order to systematically address these weaknesses, which can in particularly be the cause of functional insufficiencies, a list of „DNN-specific safety concerns“ has been developed by AI and safety experts in KI Absicherung. These must be taken into account during the safety analysis and appropriate mitigation measures must be identified.

By combining evidences in the safety argument it can be demonstrated that the DNN-specific safety concerns are adequately addressed.

The KI Familie

KI Absicherung is part of the KI Familie. The KI Familie represents a unique combination of projects that are of outstanding importance to Germany‘s industry and research landscape. Across the domains, all four projects and the interaction between them lay the foundation for the successful implementation of artificial intelligence for vehicle concepts and systems of the future.

In short: KI Absicherung aims to enable the safeguarded use of AI in vehicles; in KI Wissen, already existing knowledge is made available for AI. KI Delta Learning increases the learning competence of the networks and the KI Data Tooling project will provide a holistic database as well as various methods and tools for its efficient use in the context of training and validation of AI functions in vehicles.

Agenda

09:30 **Welcome and introduction to the event**
Dr. Stephan Scholz (Volkswagen AG)
PD Dr. Michael Mock (Fraunhofer IAIS)

09:40 **Welcome by the Federal Ministry for Economic Affairs and Climate Action**
Ernst Stöckl-Pukall

09:50 **Keynote: The future of mobility - where are we today?**
Wolfgang Müller-Pietralla (Volkswagen AG)

10:15 **KI Absicherung: Project overview, challenges and results**
Dr. Stephan Scholz (Volkswagen AG)
PD Dr. Michael Mock, Fraunhofer IAIS

10:30 **Coffee Break**

11:00 **KI Absicherung: The main results**
Fridolin Bauer (BMW Group), Frédérik Blank (Robert Bosch GmbH), Dr. Fabian Hüger (Volkswagen AG),
PD Dr. Michael Mock (Fraunhofer IAIS),
Andreas Rohatschek (Robert Bosch GmbH),
Dr. Thomas Stauner (BMW Group)

13:00 **Lunch break & interactive poster session**

14:00 **Poster session**

15:10 **Parallel presentations**
Parallel presentation 1:
Sensor Fusion for Robust Human Pose Estimation
Michael Fürst (DFKI)

Parallel presentation 2:
SegmentMeIfYouCan:
A Benchmark for Anomaly Segmentation & Entropy Maximization and Meta Classification for Out-of-Distribution Detection in Semantic Segmentation.
Svenja Uhlemeyer (Bergische Universität Wuppertal),
Robin Chan (Bergische Universität Wuppertal)
Parallel presentation 3:
Testing Deep Learning-based Visual Perception for Automated Driving
Dr. Christian Heinzemann (Robert Bosch GmbH)

15:35 **Outlook on standardization from the perspective of ISO**
Prof. Dr. Simon Burton (Fraunhofer IKS)

15:55 **Outlook on standardization from the perspective of DIN/BSI and Zertifizierte KI**
Christine Fuß (DIN), Daniel Lövenich (BSI),
Dr. Maximilian Poretschkin (Fraunhofer IAIS)

16:15 **Coffee Break**

16:45 **Podium: from the perspective of the KI Familie**
Hans-Jörg Vögel (BMW Group), Dr. Jörg Dietrich (Continental AG), Amin Hosseini (Mercedes-Benz)

17:30 **Wrap-up & Farewell**
Dr. Stephan Scholz (Volkswagen AG)
PD Dr. Michael Mock (Fraunhofer IAIS)

18:00 **End of the event**

Key Facts

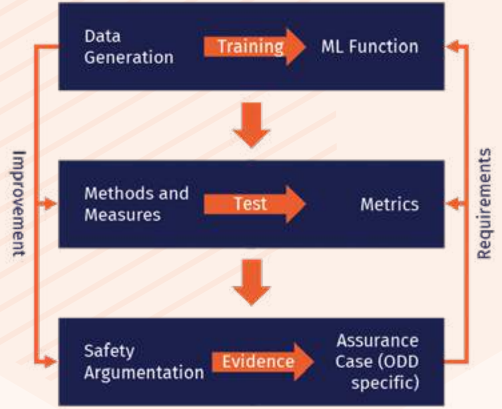
Goal & Challenge

KI Absicherung makes the safety of AI-based function modules for autonomous driving verifiable. An autonomous vehicle must be able to perceive its environment and react adequately to it. The solutions for such environment perception must be able to correctly interpret the movements of other road users and derive intentions for their continued behaviour. In highly automated vehicles, these tasks are increasingly performed by artificial Intelligence (AI). One of the greatest challenges in integrating these technologies in highly automated vehicles is to ensure the usual functional safety of previous systems. Existing and established safety processes cannot simply be transferred to machine learning methods. In the KI Absicherung project, a stringent and provable safety argumentation is being set up for the first time, with which AI-based function modules (AI-modules) can be secured and validated for autonomous driving.

Methodological approach

In the KI Absicherung project, methods and measures are developed that provide performance and safety smetrics. These methods, measures and metrics support the general safeguarding of an AI function in a car.
On the concrete use case of the AI-based perception of pedestrians, consensual approaches to the following focal points are developed:

- Selection and further development of AI algorithms for pedestrian detection with regards to their detection performance against the backdrop of their safeguarding capability.
- Development and combination of methods and measures to identify and reduce inherent insufficiencies of the AI modules.
- Stringent development of a safety argumentation and test methodology to prove the adequate mitigation of inherent insufficiencies of an AI module.
- In-process creation of synthetic training and test data sets for the analysis and evaluation of inherent insufficiencies of AI-based processes.



In order to move from the data-driven AI function to an assurance case, using the example of pedestrian detection to provide a stringent argument for demonstrating the safeguarding capability of this AI function, the following steps are taken (see illustration):

1. Process-accompanying generation of synthetic learning, test and validation data
2. Developing methods and measures to improve the AI function with respect to a wide range of metrics
3. Development and validation of test methods for these metrics
4. Stringent safety argumentation for an exemplary Operational Design Domain (ODD)

Unique features of the project

In the KI Absicherung project, experts from the fields of artificial Intelligence and machine learning, functional safety and synthetic sensor data generation are working together for the first time. In communication with standardisation committees and certification bodies, the findings gained in the project will be used to work towards building an industry consensus on a general AI test strategy.

Facts & Figures

Dr. Stephan Scholz

Volkswagen AG
Project coordinator

PD Dr. Michael Mock

Fraunhofer IAIS
Scientific coordinator and consortium co-lead

36 Months

Project runtime: (01/07/2019 - 30/06/2022)

€41 M

Project budget

€19.2 M

Funding budget

24 Project Partners

4 external technology partners

Overall Methodology

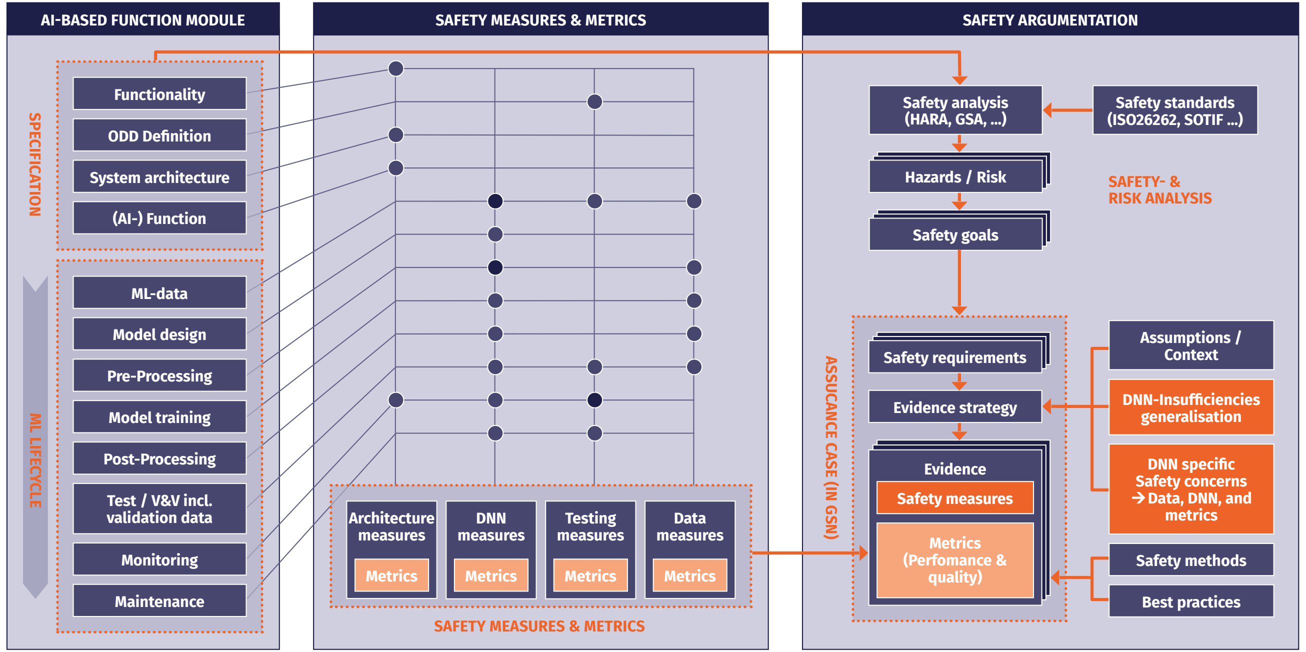
Assurance of AI-based functions requires a methodoloy that can identify and systematically mitigate inherent weaknesses in AI functions. The methodology also includes a systematic generation of Data for training and testing, definition of safety relevant metrics as bases of an approach for deriving a stringent evidence-based safety argumentation.

KI-Absicherung: Overall Approach

Stephan Scholz, Volkswagen AG, **Michael Mock,** Fraunhofer IAIS
Frédérik Blank, Robert Bosch GmbH, **Fabian Hüger,** Volkswagen AG
Andreas Rohatschek, Robert Bosch GmbH, **Thomas Stauner,** BMW Group

The application of AI is a key enabler for autonomous driving. In the KI Absicherung project, AI and safety experts from industry and academia develop a methodology for a safety argumentation that systematically identifies, makes measurable and mitigates weaknesses of AI functions. The aim is to achieve an industrial consensus for a methodical approach to assure AI functions for the use case of pedestrian detection.

AP5.1



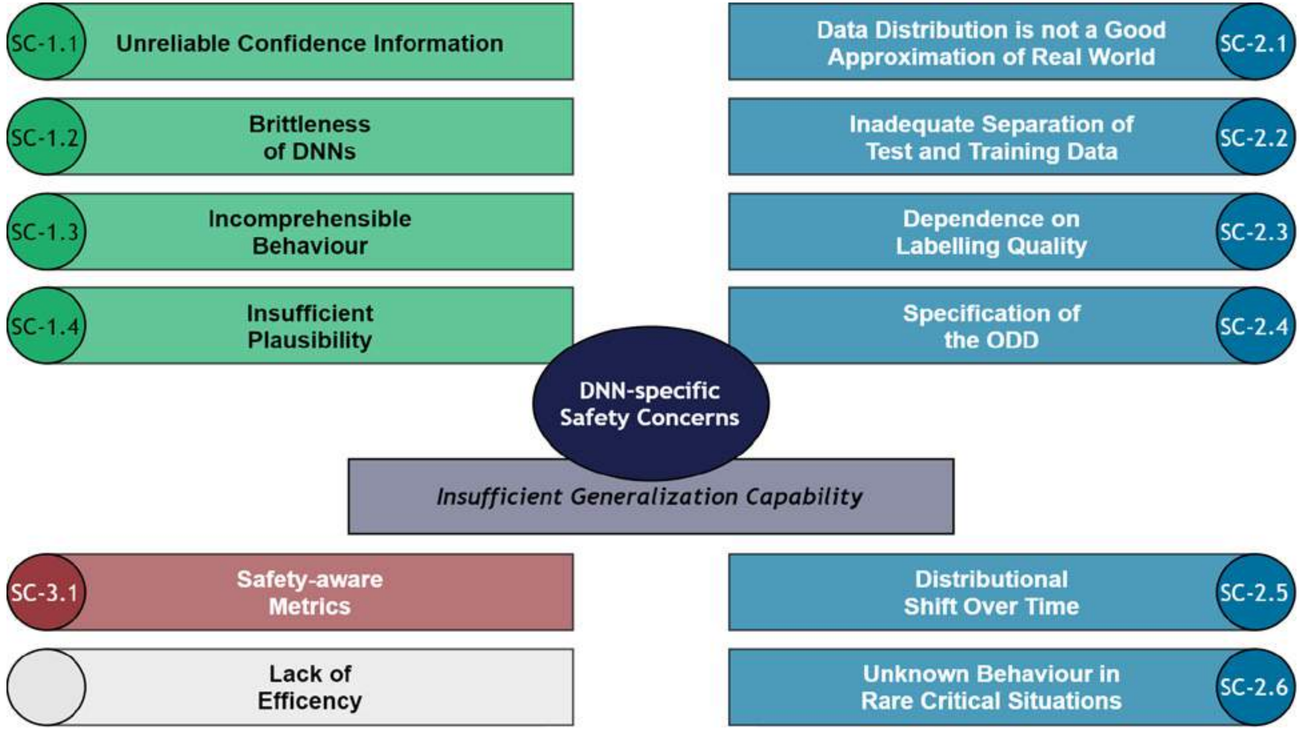
Overall approach for assurance of AI-based functions (© BMW Group | Robert Bosch GmbH | Fraunhofer IAIS | Volkswagen AG)

Towards Safety Metrics for Automated Driving

Christian Hellert, Continental AG, **Fabian Hüger**, Volkswagen AG
Lydia Gauerhof, Robert Bosch GmbH, **Timo Sämann**, Valeo Schalter und Sensoren GmbH
Dominik Brüggemann, BUW, **Christoph Thiem**, Opel Automobile GmbH

When deploying an AI function into a safety-critical system, safety metrics are required to quantify the remaining risk towards the defined safety goals. Thereby, two strategies can be followed to define appropriate safety metrics: From technology and from application perspective. We showed how to derive safety metrics from technology level, using specified DNN-specific safety concerns, and from functional requirements, deduced from system level.

AP3.2 | AP4.2 | AP4.3


SC-1.1
Unreliable Confidence Information
SC-1.2
Brittleness of DNNs
SC-1.3
Incomprehensible Behaviour
SC-1.4
Insufficient Plausibility
Data Distribution is not a Good Approximation of Real World
SC-2.1
Inadequate Separation of Test and Training Data
SC-2.2
Dependence on Labelling Quality
SC-2.3
Specification of the ODD
SC-2.4
DNN-specific Safety Concerns
Insufficient Generalization Capability
SC-3.1
Safety-aware Metrics
Lack of Efficency
Distributional Shift Over Time
SC-2.5
Unknown Behaviour in Rare Critical Situations
SC-2.6

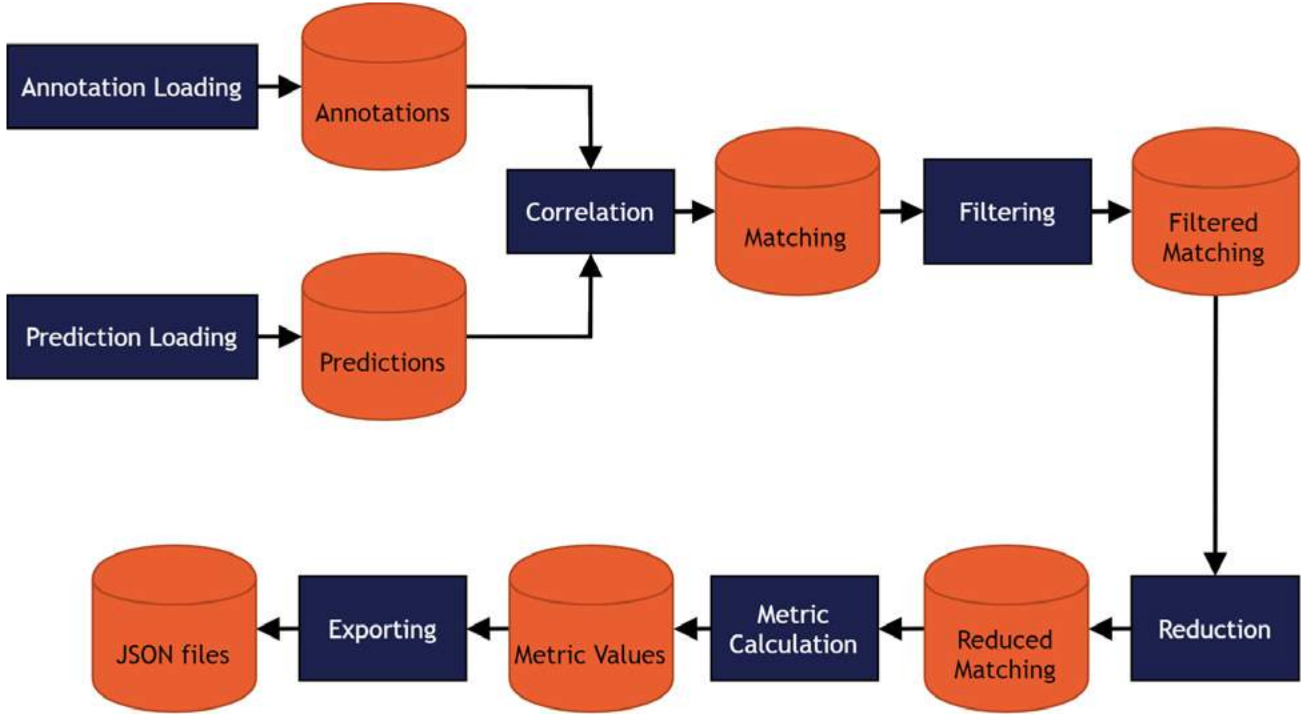
Overview of defined DNN-specific safety concerns (© BUW | BMW Group | Robert Bosch GmbH | Continental AG | Fraunhofer IAIS | Valeo Schalter und Sensoren GmbH | Volkswagen AG)

Metric Benchmarking Tool

Christian Hellert, Continental AG
Christian Brunner, **Tom Thielo**, **Jonas Schneider**, Elektronische Fahrwerksysteme GmbH
Dominik Brüggemann, BUW

The Metric Benchmarking Tool (MBT) is an application to perform standard benchmarks within the project KI-Absicherung. It is designed to enable and simplify evaluations regarding the effectiveness of mechanisms implemented to improve pedestrian recognition. The tool uses the KI Absicherung dataset with available enriched metadata and 2D bounding box predictions provided in the project specific output format to compute a user specified set of object detection metrics.

AP3.2 | **AP3.6**



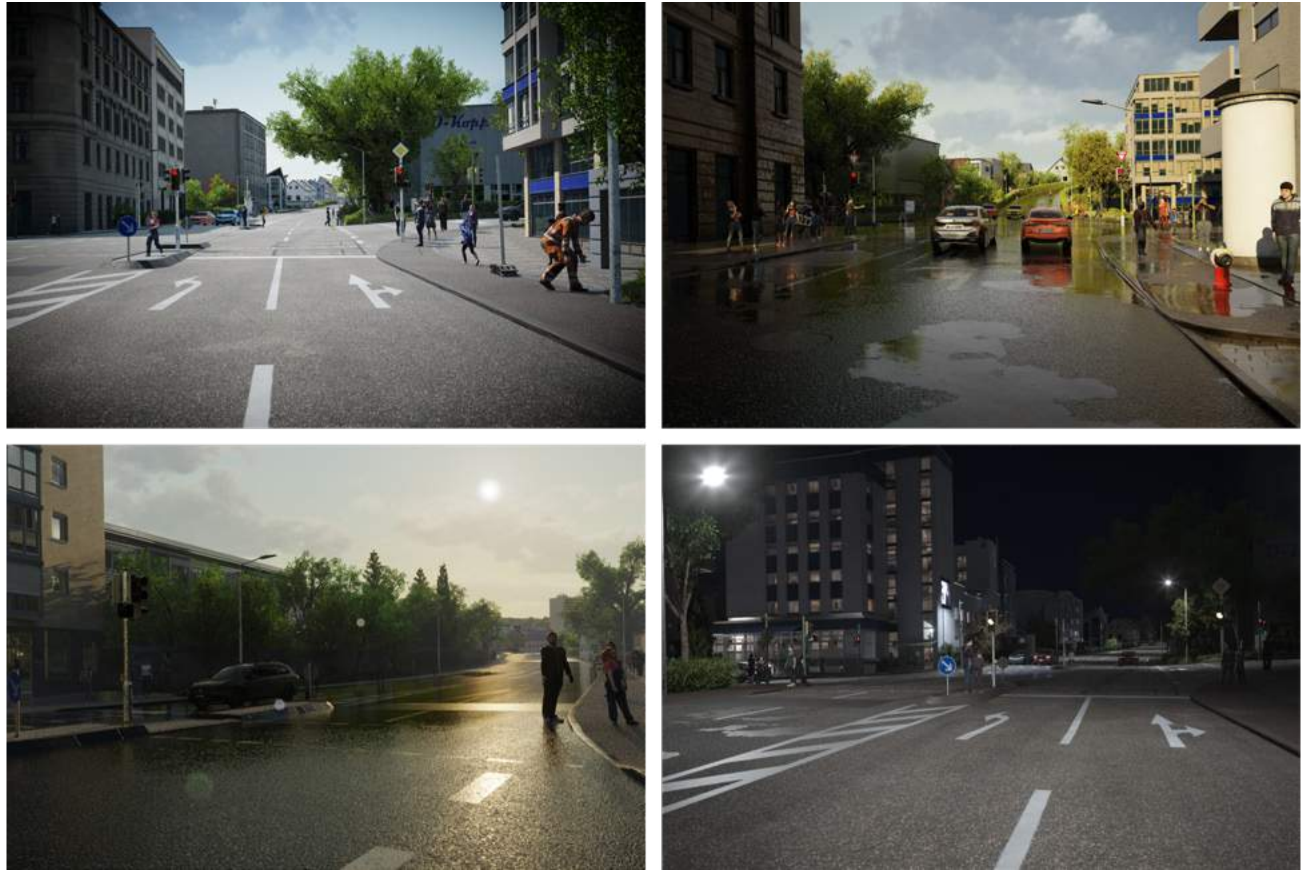
Workflow of the Metric Benchmarking Tool. Blue boxes represent processing modules and orange boxes represent data structures (© BUW | Continental AG | Elektronische Fahrwerksysteme GmbH)

Synthetic Data Generation based on a modern Game Engine

Markus Huber, **Christopher Hauck**, **Christian Zilliken**, Mackevision Medien Design GmbH

Modern game engines are perfectly suitable for synthetic data generation. With custom developed modules, the requirements of KI Absicherung data sets such as realism, a high degree of variance, varying light and weather conditions, and sensor effects are met with systematic data generation. The extensive amount of ground truth data and meta annotations allow a wider range of applications in testing and evaluation AI methods.

AP2.1 | AP2.5



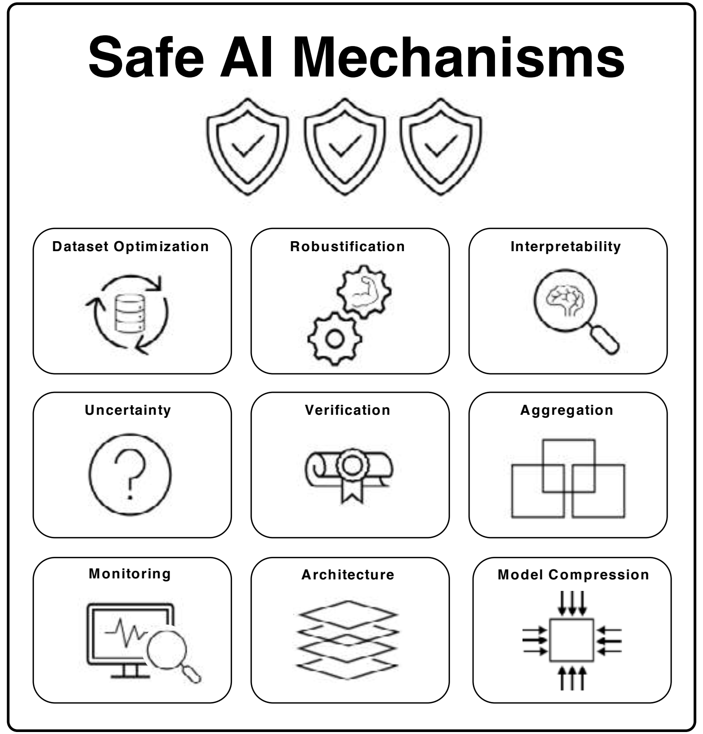
Synthetic data generation allows the adaption to different light and weather situations such as clear sky, wetness, or night scenarios (© Mackevision Medien Design GmbH)

Overview of Safe AI Mechanism Landscape and Taxonomy

Alexander Hirsch, Stephanie Abrecht, Robert Bosch GmbH
Gesina Schwalbe, Continental AG

The landscape of Safe AI Mechanisms, which are used to mitigate DNN-specific (or broader: AI-specific) Safety Concerns, is extremely diverse. With our work we are structuring this landscape by introducing a Safe AI Mechanism Taxonomy. Further we are providing a consistent and complete overview of the developed Safe AI Mechanisms in KI-Absicherung TP3 including a self evaluation using multiple classification criteria.

AP3.6



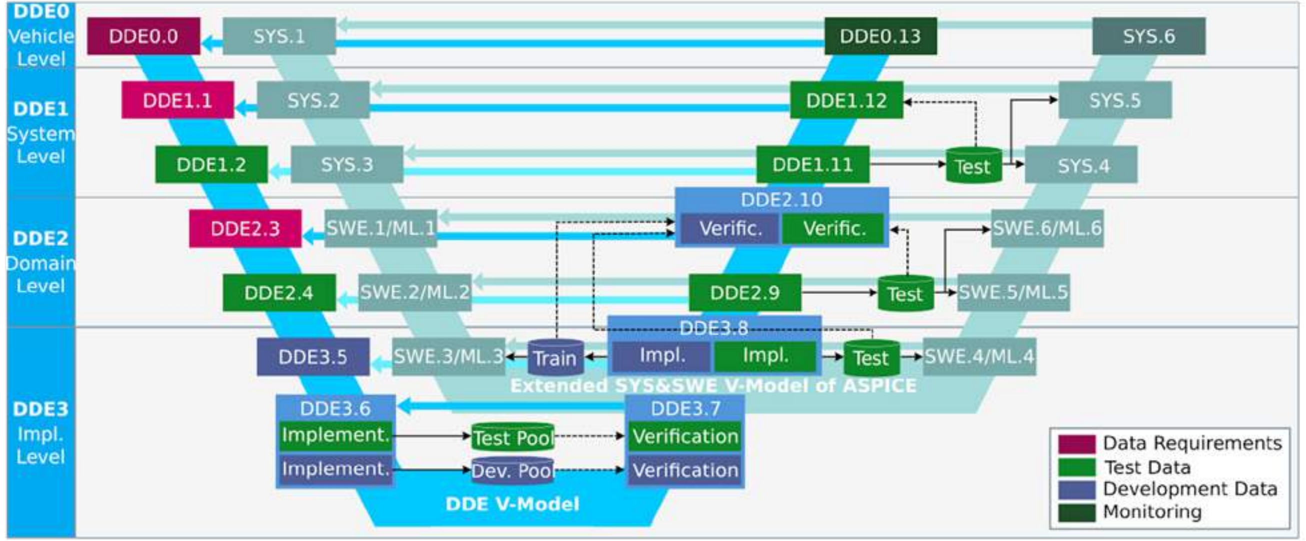
Overview of Safe AI Mechanism Taxonomy Domains (© Robert Bosch GmbH)

Data-Driven Engineering / ML Life Cycle | How to derive systematic data requirements?

Autoren: Andreas Albrecht, Thomas Geipel, Robert Bosch GmbH, **Henrik Putzer**, fortiss GmbH
Reviewer: Frédérik Blank, Robert Bosch GmbH, **Thomas Stauner**, BMW Group
Timo Dobberphul, Volkswagen AG, **Iwo Kurzidem**, Fraunhofer IKS

ML models learn their functional behavior implicitely from training data. If relevant information is missing the ML model will not learn it. So we need to collect well-structured and well-balanced data sets that comprehensively cover our problem. Due to the open context nature, we propose an iterative Data-Driven Engineering Process / ML-LifeCycle Model to systematically derive data requirements and data coverage.

A4.5, P1



Data-Driven Engineering Process & ML-LifeCycle Model that incorporates ML workflows and ML Test Strategy and maps to existing process standards (e.g. ASPICE) (© Robert Bosch GmbH)

Physically based synthetic data generation pipeline

Karl Leiss, BIT Technology Solutions, **Johannes Günther**, Intel Corporation
Anja Kleinke, Valeo Schalter und Sensoren GmbH, **Marzena Franeck**, Robert Bosch GmbH

Synthetic data is a scalable and flexible solution to systematically train & test AI based systems. A brand new glTF based pipline with exchangeable modules was developed to increase transferability of synthetic to real data. Apart from a 3D object and scenario management, automized scenario generation, physical sensor and material effects were incorporated. In the project this new pipeline was used to generate virtual scenarios along with labels and meta data.

AP2.1

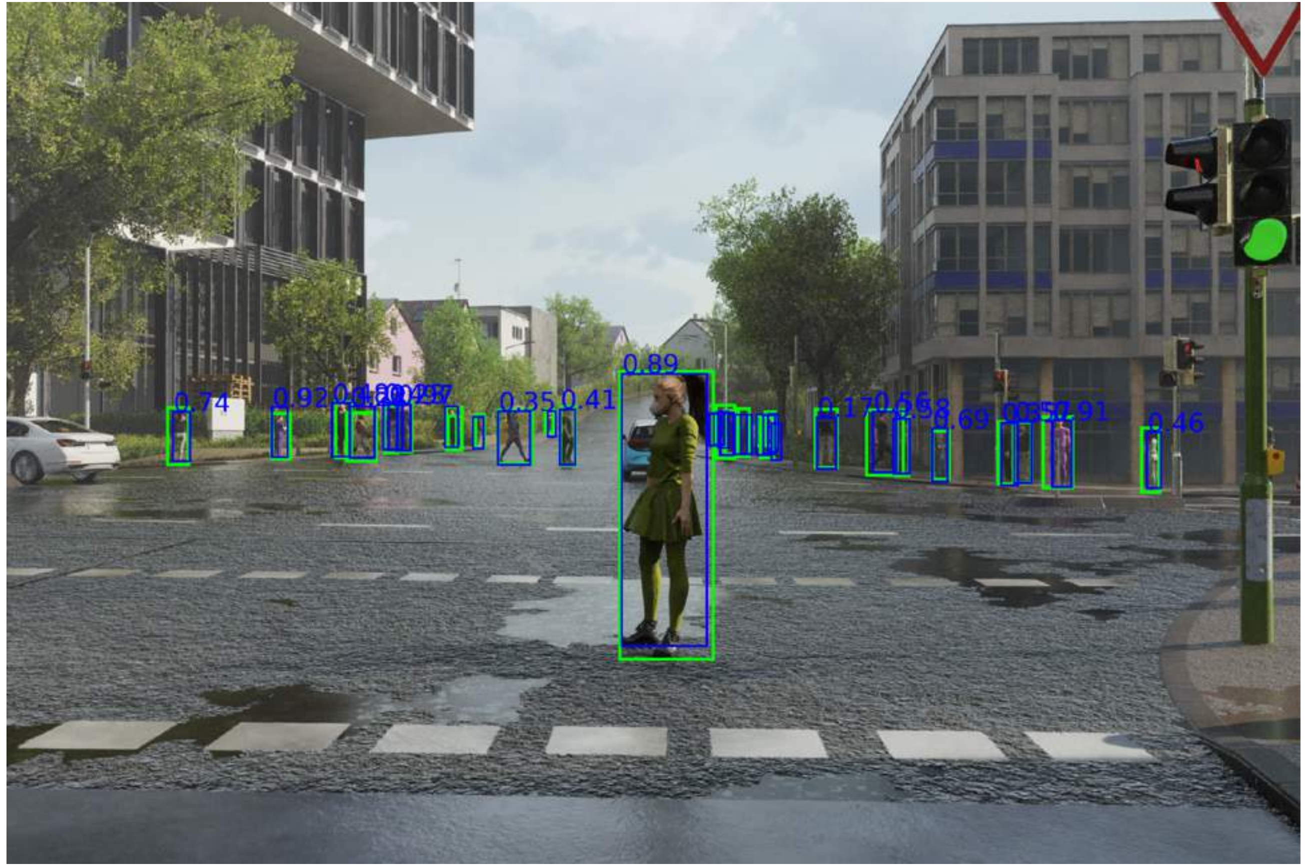
Raytracing generated synthetic image with procedural sun and physical sensor effects (© Robert Bosch GmbH | BIT Technology Solutions | Intel Corporation | Valeo Schalter und Sensoren GmbH)

Pedestrian detector development using the SSD and the KI Absicherung synthetic dataset (synPeDS)

Patrick Feifel, **Philipp Heidenreich**, Opel Automobile GmbH
Frédérik Blank, **Simon Heming**, Robert Bosch GmbH

The goal of this result is to provide a reference implementation for camera-only 2D bounding box pedestrian detection. To this end, the SSD has been selected as a traditional single-stage object detector using anchor boxes. To develop the SSD with the synPeDS dataset, we describe the necessary adaptions, including strategies to deal with many small and occluded pedestrians and the evaluation using safety-aware metrics.

AP1.3



Example SSD inference result of r4 (Mackevision Medien Design GmbH Seq84) (© Mackevision Medien Design GmbH | Opel Automobile GmbH)

Data structuring and analysis

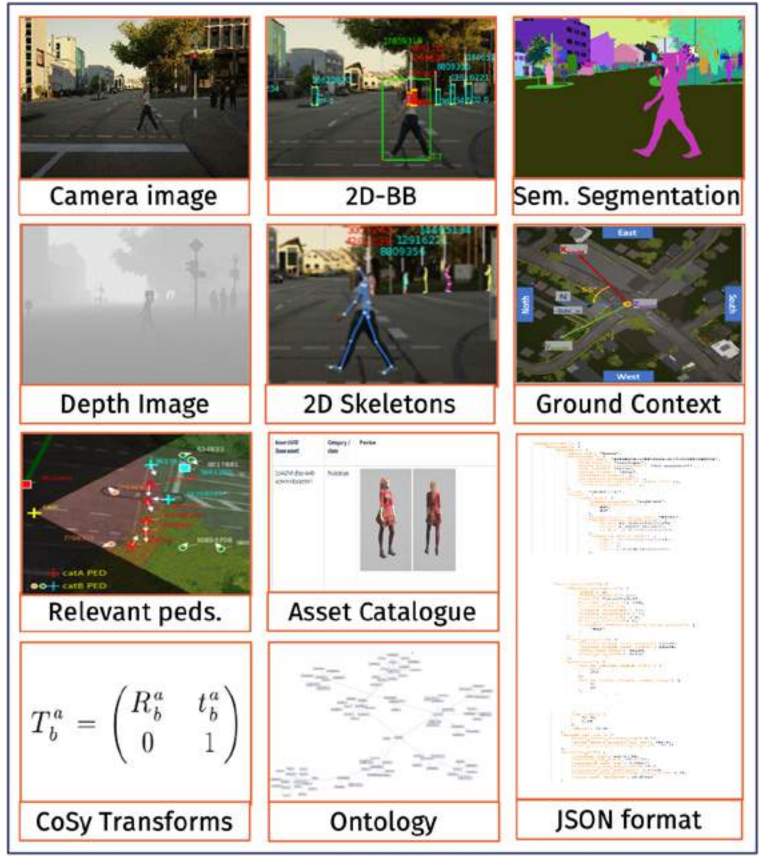
Approaches to systematically and where possible semantically describe and analyze the data input space are covered within this cluster. This includes the development of an ontology-based description language enabling concretizing the ODD and performing possible data coverage analyses as well as allows to deliver enriched metadata for performing in-depth data-analyses. Moreover, characteristics of the dataset generated in KI Absicherung and of the generation process itself are introduced.

Synthetic Dataset for Pedestrian Detection (synPeDS) - Overview

Bastian Knerr, QualityMinds GmbH
Thomas Stauner, BMW Group
Frédérik Blank, Robert Bosch GmbH
Michael Fürst, DFKI
Philipp Heidenreich, Opel Automobile GmbH

This synthetic dataset (video & LiDar) is aimed at being used for training, testing and assurance of ML-based pedestrian detection algorithms. It's vast amount of ground truth and metadata enables in-depth data, sensitivity and correlation analyses.

TP2 with TP1+P1, AP4.1



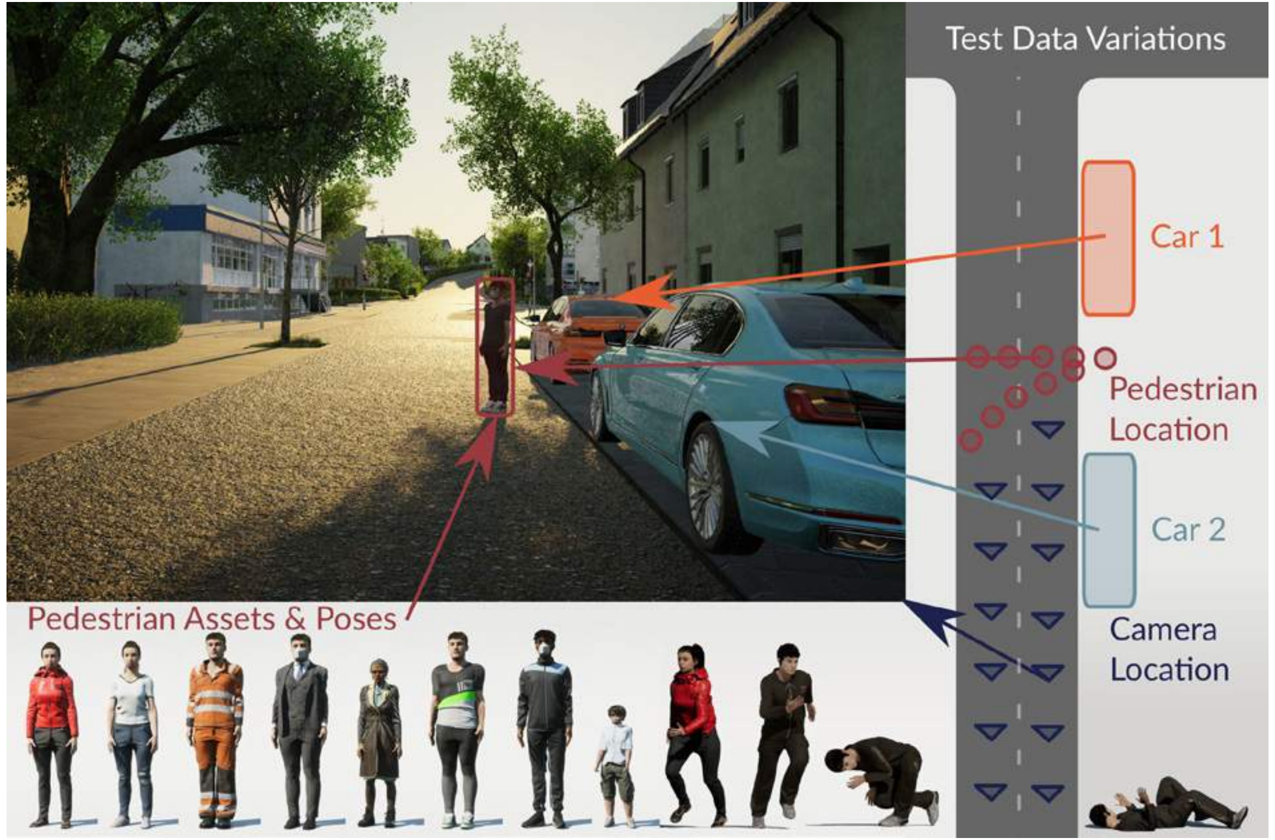
Groundtruth and meta-information for synthetic dataset (© Mackevision Medien Design GmbH | Robert Bosch GmbH | QualityMinds GmbH)

Using Ontologies in Automotive AI Applications

Christian Witt, Valeo Schalter und Sensoren GmbH
Martin Herrmann, **Christian Heinzemann**, **Frédérik Blank**, Robert Bosch GmbH
Frank Bonarens, Opel Automobile GmbH

Basis of a robust safety strategy for an automated driving function based on neural networks is a detailed description of its input domain. Ontologies fulfill the task to gather expert knowledge and model information to enable computer aided processing, while using a notion understandable for humans. We leveraged the KI-Absicherung ontology to define the operational design domain, to develop tools for structured data generation, to describe assets and metadata, and to analyze input domain coverage and DNN performance.

AP4.1



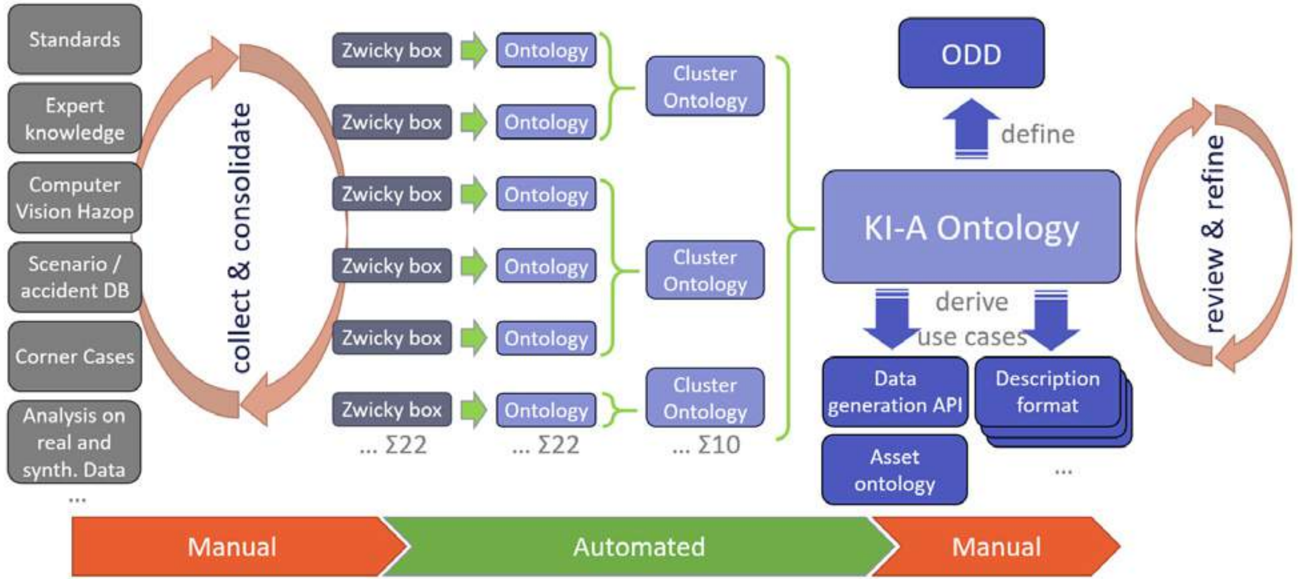
NCAP-like scenario “pedestrian crosses road between parked vehicles” generated by our tool for structured data generation and based on KI-Absicherung ontology (© Valeo Schalter und Sensoren GmbH | Mackevision Medien Design GmbH| Bosch)

Methodology of Creating an Ontology for Dataset Engineering

Christian Witt, Valeo Schalter und Sensoren GmbH
Martin Herrmann, **Christian Heinzemann**, **Frédérik Blank**, Robert Bosch GmbH

Basis of a robust safety strategy for an automated driving function based on neural networks is a detailed description of its input domain. Ontologies fulfill the task to gather expert knowledge and enable computer aided processing, while being understandable for humans. We developed and applied a methodology to create our ontology based on a domain analysis with a manifold of sources, followed by structuring, consolidation, refinement and review steps. The resulting KI-Absicherung ontology serves as a single point for meta data and semantic information.

AP4.1



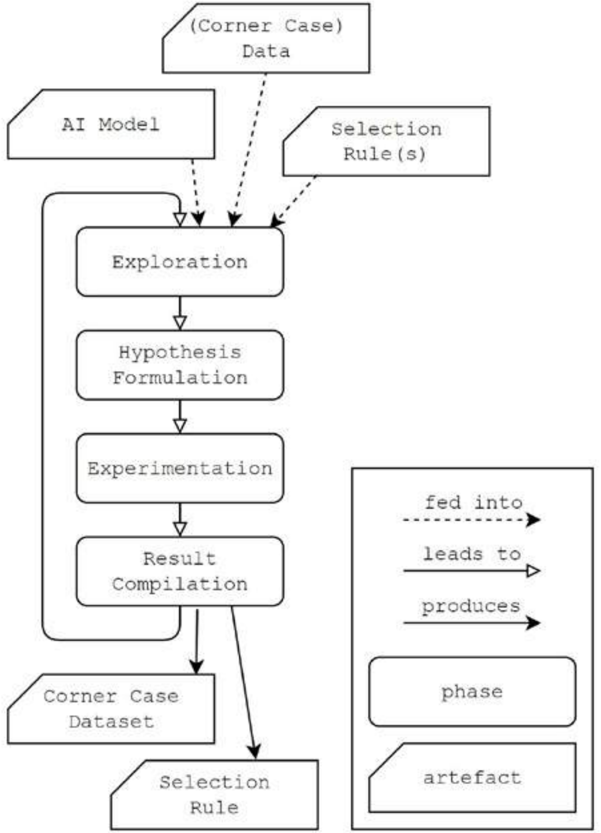
Methodology to develop the KI-Absicherung ontology with different manual and automated steps and applied use cases (© Robert Bosch GmbH)

Automated Corner Case Detection Pipeline

Namrata Gurung, Niels Heller, QualityMinds GmbH

A method for the identification and characterization of corner cases was developed. Applying this to the KI-A dataset, a total of eight selection rules were found, each based on a distinct performance inhibiting feature, which could be categorized under quantitative, perceptual, and situational inhibitors. Several data visualization tools were developed, including a tool that calculates the similarity of any given instance to the rest of the data.

AP2.2



The phase iteration model (© QualityMinds GmbH)

Enriched metadata in KI-Absicherung synthetic dataset (synPeDS)

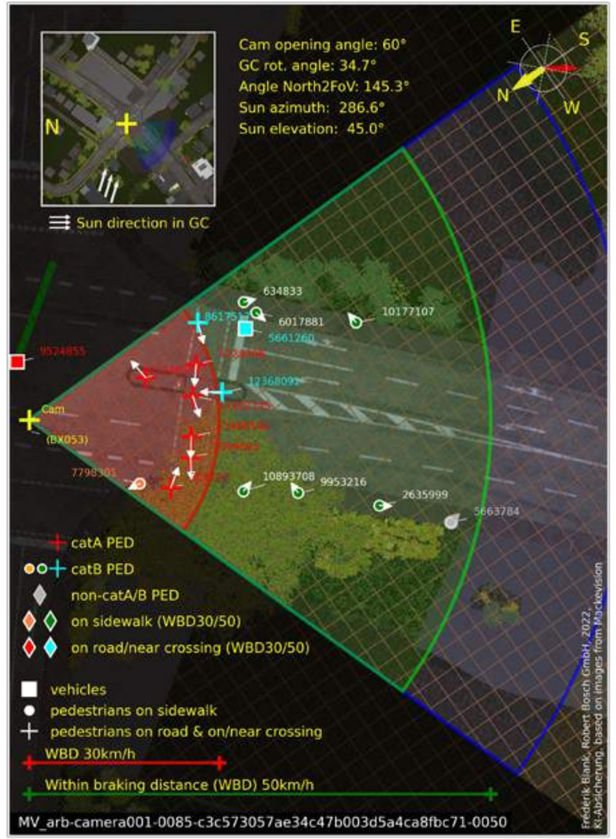
Frédérik Blank, Falko Matern, Robert Bosch GmbH
Philipp Heidenreich, Opel Automobile GmbH, **Michael Fürst**, DFKI
Markus Huber, Mackevision Medien Design GmbH, **Thomas Stauner**, BMW Group

Based on the collaborative work of AI, data and safety experts in KI Absicherung, the project‘s synthetic pedestrian dataset was highly enriched by adding more than 50 metadata variables valuable to:

- Evaluate DNN-performance based on **safety-related criteria** (safety-aware metrics)
- Search and cluster images by specific search criteria for **(statistical) image analysis**
- Define **specific training or test datasets**
- Link images and object / pedestrian instances to the **ontology**

This metadata was then used by several partners to develop their own project results. .

P1 | AP4.1



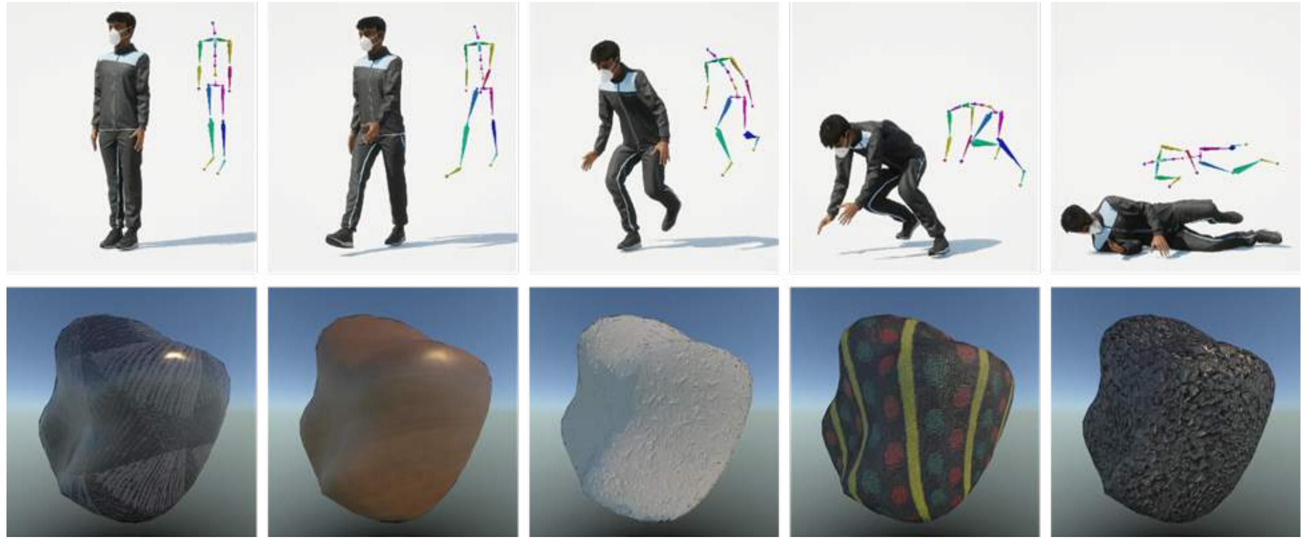
Bird-eye view with enriched metadata and safety-related evaluation zones and pedestrian classifications
(© Robert Bosch GmbH, Mackevision Medien Design GmbH)

Motion Capture & Material Measurements

Markus Bartnick, Markus Huber, Mackevision Medien Design GmbH
Johannes Günther, Intel Corporation

Synthetic data generation with a high degree of realism and accuracy requires the measurement of key scene elements such as pedestrian motion and material characteristics. The captured skeletal motion of several persons and their interaction with objects was transferred to 3D character models („retargeting“). Cloth and infrastructural materials were scanned by an X-Rite TAC7 scanner, processed and available in AxF and glTF file format.

AP2.5



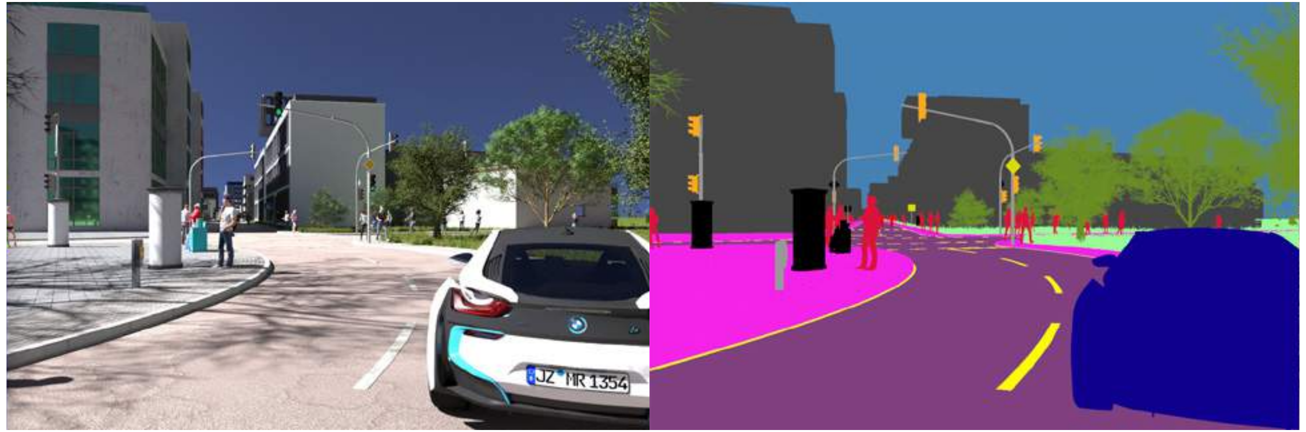
Top row: One of 90 captured pedestrian animations. Bottom row: Five examples of 80 measured material samples (© Mackevision Medien Design GmbH| Intel Corporation)

Lessons Learned on Synthetic Data Generation

Nicolas Gay, Maximiliano Cuevas, Ulrich Wurstbauer, Luxoft
Thomas Stauner, BMW Group
Oliver Grau, Korbinian Hagn, Intel Corporation
Falko Matern, Robert Bosch GmbH

The synthetic data generation process and the utilisation of such data for training and evaluation of AI models involves four well-defined stages: data specification, data production, data analysis and its usage in a real-world application. We summarise some of the lessons learned along the process.

AP2.4



Synthetic scenario and its corresponding semantic segmentation ground truth produced during the synthetic data generation process (© BIT Technology Solutions)

Method related Evidence Workstreams

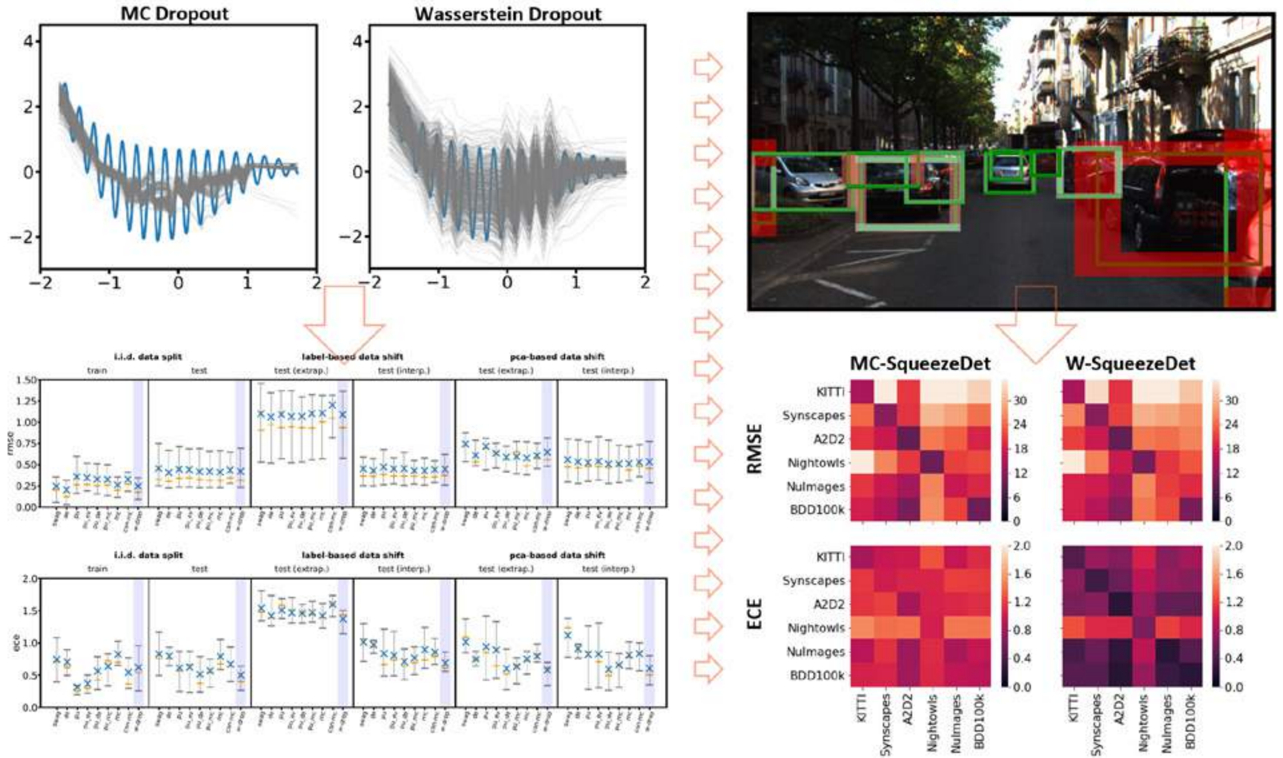
The method-related Evidence Workstreams cover the DNN-specific safety concerns about uncertainty, robustness, plausibility and explainability. In the following, the contributions show mechanisms to mitigate the safety concerns and describe related test approaches. Furthermore, aspects regarding the safety argumentation and determined evidence are described.

Non-Parametric Uncertainty Optimization for Bounding Box Regression

Joachim Sicking, Maximilian Pintz, Maram Akila, Fraunhofer IAIS

We propose Wasserstein dropout, an uncertainty estimator for regression tasks. It adjusts the widths of (dropout-based) sub-network distributions to match the local data uncertainty. Empirical analysis shows that it is on par with state-of-the-art methods and outperforms them in terms of consistency and robustness w.r.t. domain shift. Experiments indicate that such properties carry over to object detect when compared to (vanilla) MC Dropout estimation.

AP3.3



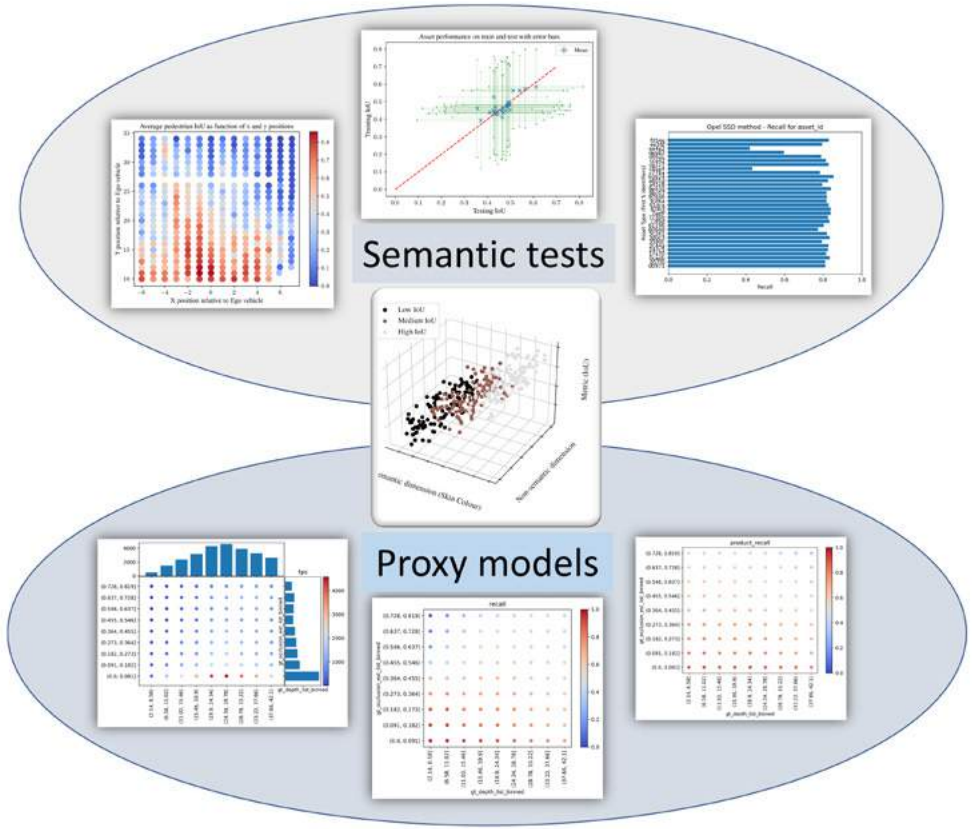
From toy experiments (upper left, showing widths of ensembles) over 1D regression (lower left) to object detection (r.h.s) including out-of-domain experiments (lower right), (© Fraunhofer IAIS | KIT)

Semantic Testing of DNNs with Proxy Models

Sujan Gannamaneni, Maram Akila, Fraunhofer IAIS

Our „Semantic Testing“ approach evaluates DNNs along semantic dimensions to uncover learnt weaknesses. In contrast to using aggregated metrics, our method enables more granular testing. In addition to identifying several weaknesses, we also evaluate whether the impact of semantic dimensions is independent or not. We perform this by the functional decomposition of observations using marginals of the data distributions, which acts as a simple proxy model.

AP4.4



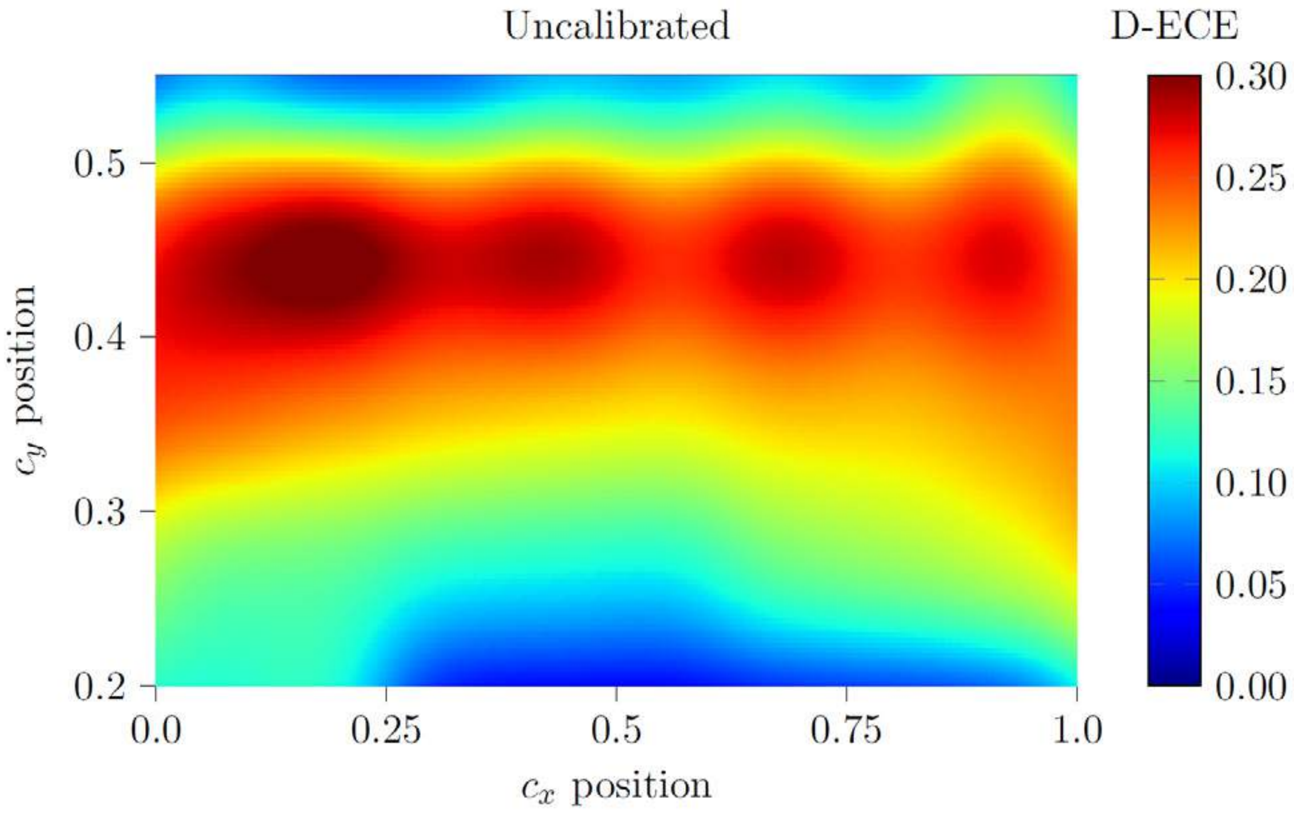
Semantic tests uncover weaknesses in DNNs, and Proxy models uncover independence of the impact of semantic dimensions. (© Fraunhofer IAIS)

Multivariate Confidence Calibration for Object Detection

Fabian Küppers, Hochschule Ruhr West
Anselm Haselhoff, Hochschule Ruhr West

For each detection, a neural network estimates its belief about the correctness. However, these estimates are known to be too overconfident, i.e., they are miscalibrated. We extend common calibration methods to include additional box information into calibration. These methods are trained and evaluated on the TP1 KI-A SSD predictions using the KI-A data sets. We found an underconfidence of the examined network which is successfully recalibrated by our methods.

AP3.5



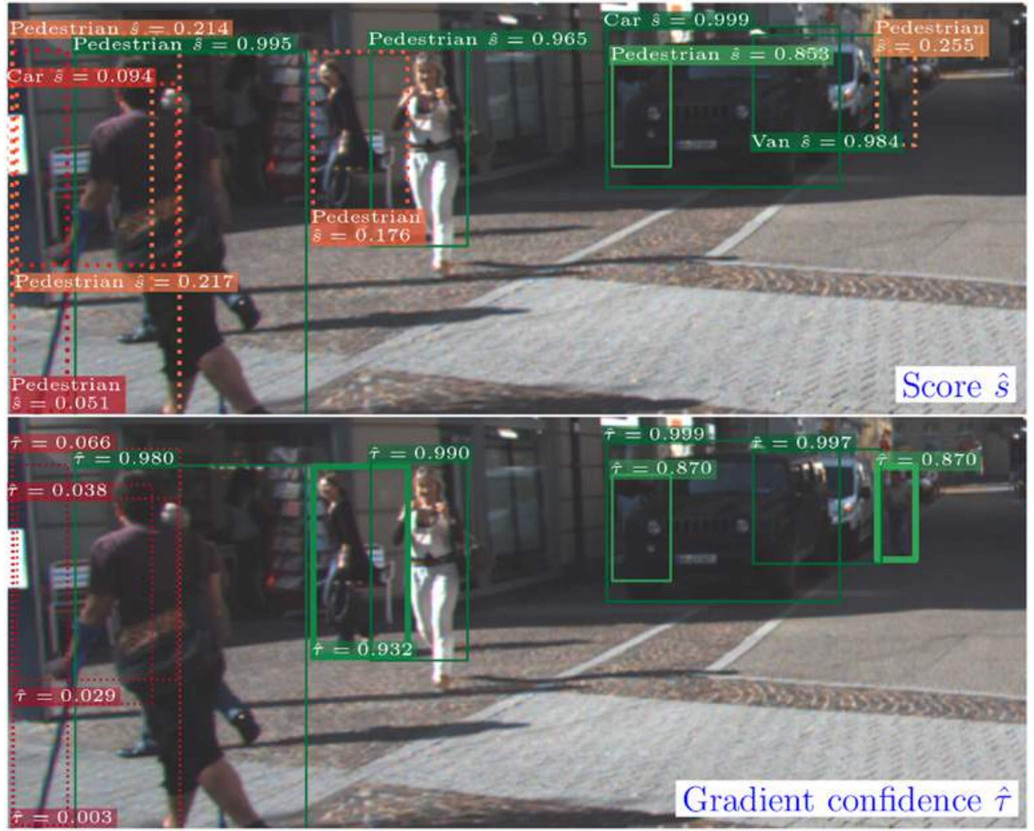
Position-dependent miscalibration of the TP1 KI-A SSD before calibration (© Hochschule Ruhr West)

Gradient-Based Uncertainty Estimation for Deep Object Detection

Tobias Riedlinger, Matthias Rottmann, Hanno Gottschalk, Bergische Universität Wuppertal

Common methods for quantifying prediction uncertainty tend to be based on sampling the network output. We introduce gradient-based instance-wise uncertainty measures for object detection refering to model parameters. We compare intrinsic network confidence to output- and gradient-based confidence estimates on the TP1 KI-A SSD architecture using TP2 data. A combination of output- and gradient-based uncertainty metrics yields the most accurate confidence estimation.

AP3.4



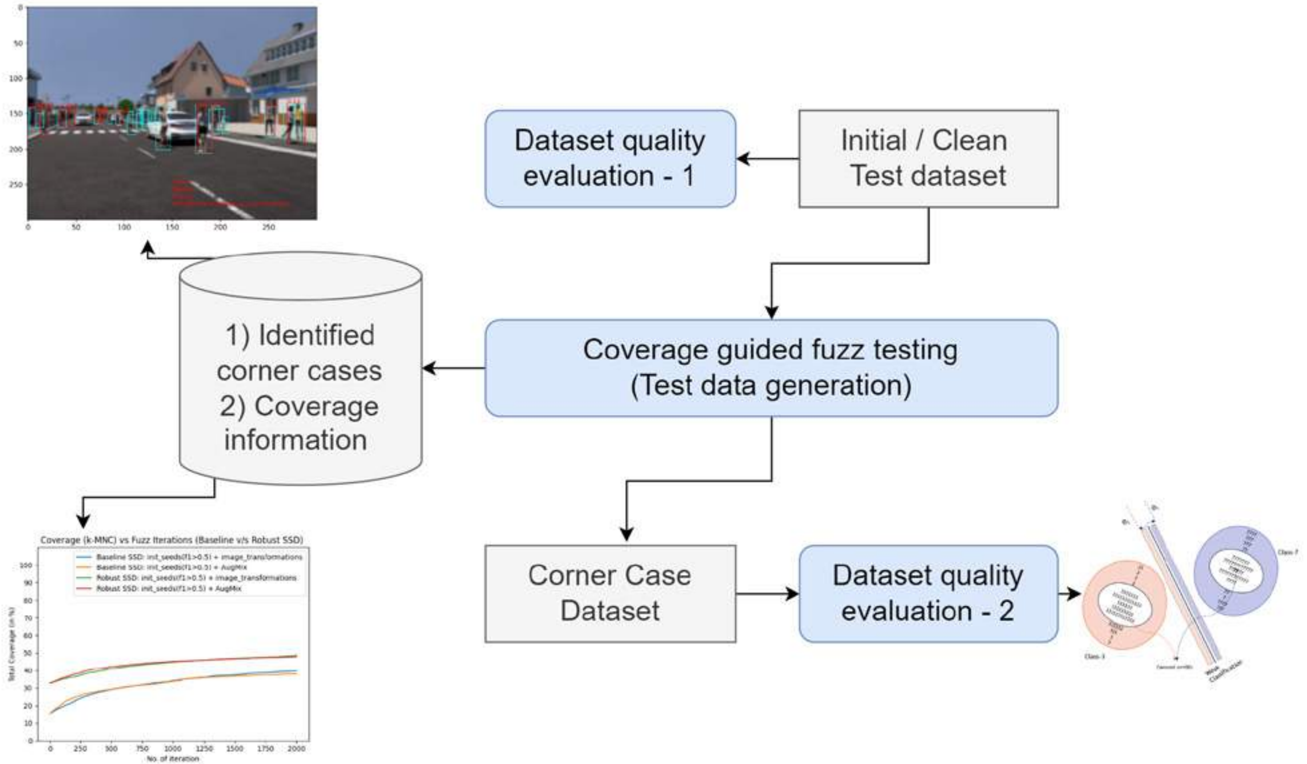
Comparison of network-intrinsic („Score“, top) and gradient-based confidence estimation (bottom) in a real-world street scene from the KITTI dataset (© BUW + KIT)

Coverage Guided Fuzz Testing Framework & Dataset Quality Metrics

Vivek Vekariya, Mojdeh Golagha, fortiss GmbH

The trustworthiness of the reported KPIs of DNN lies in the quality of its test dataset. We propose the test dataset quality metrics to infer various aspects including the latent space coverage of the DNN under test. We also use the coverage guided fuzz testing to sample the additional test data points. The performance of various DNNs can be compared using our testing framework. Also, the data points sampled using fuzzing help to maximize the latent space coverage.

AP4.4


Dataset quality evaluation - 1
Initial / Clean Test dataset
Coverage guided fuzz testing (Test data generation)
1) Identified corner cases
2) Coverage information
Corner Case Dataset
Dataset quality evaluation - 2
Coverage (k-MNC) vs Fuzz Iterations (Baseline v/s Robust SSD)
Baseline SSD: init_seeds(f1>0.5) + image_transformations
Baseline SSD: init_seeds(f1>0.5) + AugMix
Robust SSD: init_seeds(f1>0.5) + image_transformations
Robust SSD: init_seeds(f1>0.5) + AugMix
Total Coverage (in %)
0
10
20
30
40
50
60
70
80
90
100
0
250
500
750
1000
1250
1500
1750
2000
No. of iteration

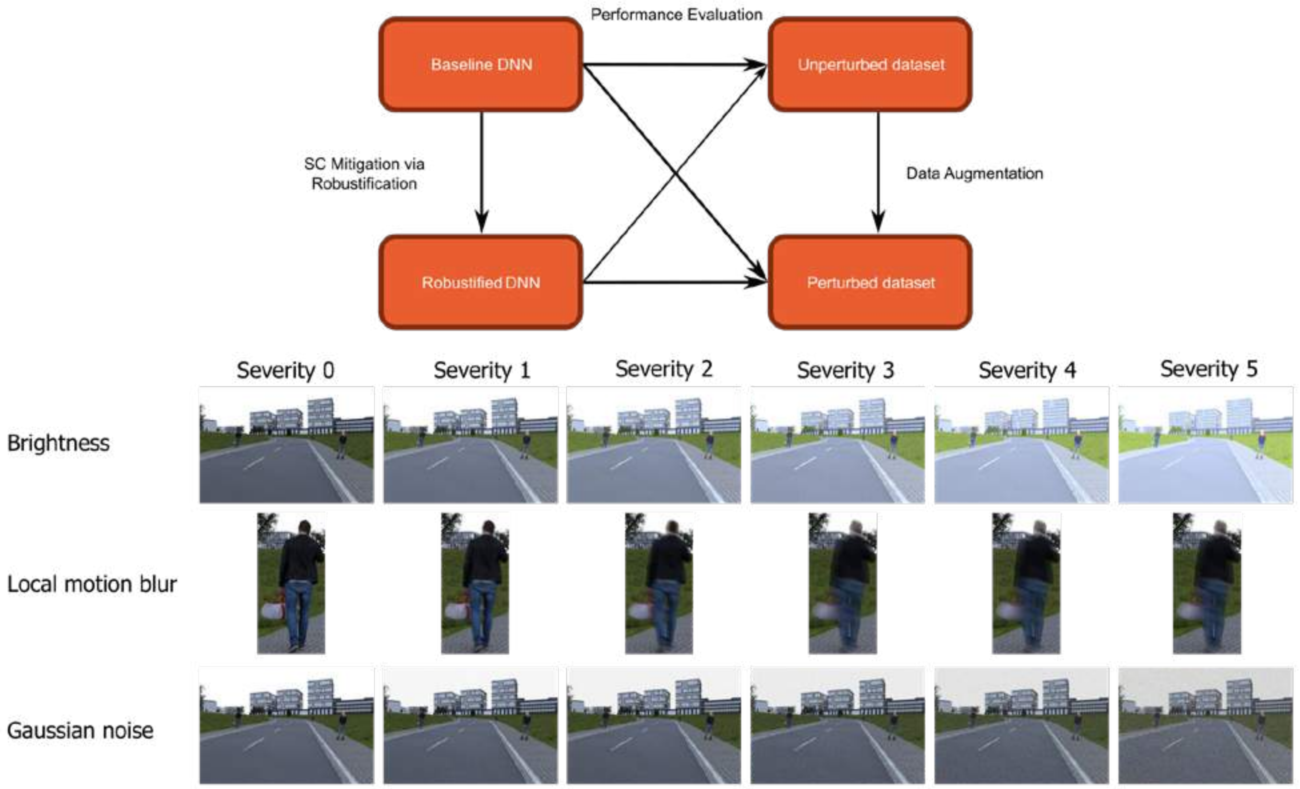
A framework to evaluate the dataset quality and maximize the latent space coverage using fuzz test generation (© Fortiss GmbH | BIT Technology Solutions)

Evidence Workstream: Analysis and Improvement of DNN Robustness

Thomas Schulik, ZF Friedrichshafen AG
Markus Bach, Valeo Schalter und Sensoren GmbH

The robustness of DNNs used for automotive perception systems is a crucial requirement for the deployment of such algorithms. Therefore, different mechanisms for robustness assessment and robustification are developed. Both of these tasks use data augmentation techniques that add natural perturbations or adversarial attacks to camera images. To obtain safety evidences, the performance in the form of the mAP metric is compared for different models and test datasets.

AP3.3 | AP3.5 | **AP4.3** | AP4.4 | AP4.5



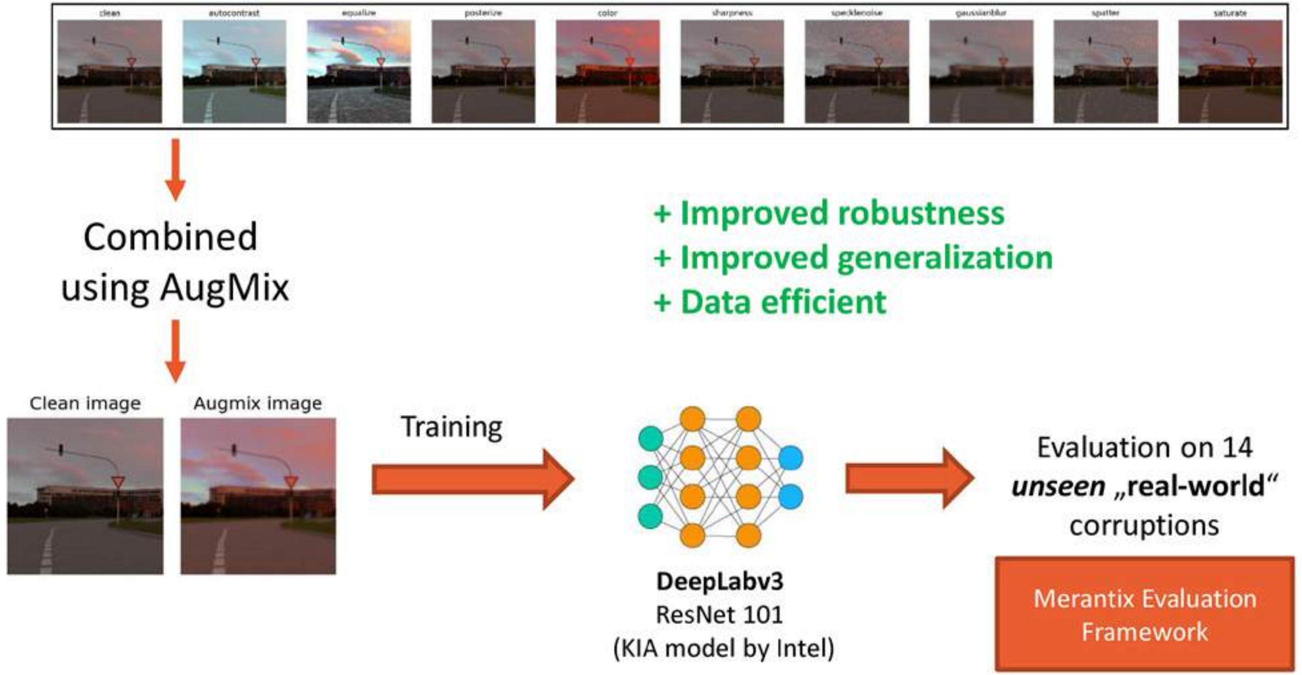
Main aspects of the Evidence Workstream (© ZF Friedrichshafen AG | BIT Technology Solutions)

AugMix: Improving Robustness via Data Augmentation

Nikhil Kapoor, Serin Varghese, CARIAD
Fabian Hüger, Volkswagen AG

Data augmentation is a powerful technique of achieving robustness and improved generalization on unseen data. AugMix is a state-of-the-art data augmentation technique that helps improve model robustness. It combines several augmentations that are sampled stochastically and layered together to produce high diversity of augmented images. The method helps improve model generalization on unseen data and cope with corner cases.

AP3.3



High-level overview of training and evaluation setup of AugMix (© Volkswagen AG | BIT Technology Solutions)

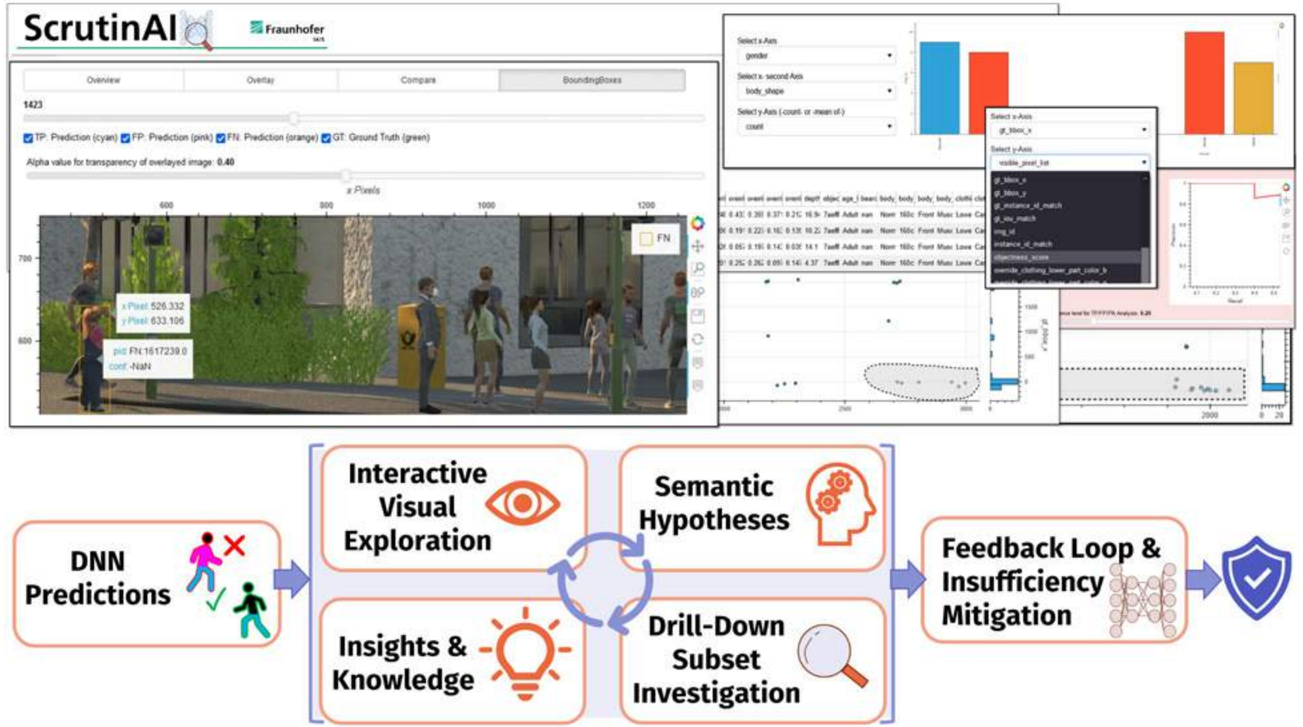
DNN specific Safety Concern: Brittleness of DNNs

Visual Exploration and Semantic Analysis of DNN Weaknesses with ScrutinAI

Elena Haedecke, Michael Mock, Fraunhofer IAIS

The interactive tool ScrutinAI is a visual analytics approach for the semantic analysis of DNN outputs. It supports analysts and/or auditors in utilizing their semantic knowledge to identify the causes of incorrect predictions by enabling a visual exploration of systematic DNN weaknesses. The method addresses the safety concern incomprehensible behavior. Insights gained support model improvement and foster a safety argumentation for AI applications.

AP3.6



☑ TP: Prediction (cyan) ☑ FP: Prediction (pink) ☑ FN: Prediction (orange) ☑ GT: Ground Truth (green)

Alpha value for transparency of overlayed image: 0.40

ScrutinAI supports the visual exploration of DNN predictions by various interlinked widgets (© Fraunhofer IAIS | Mackvision)

Evidence Work Stream: Incomprehensible Behavior and Insufficient Plausibility

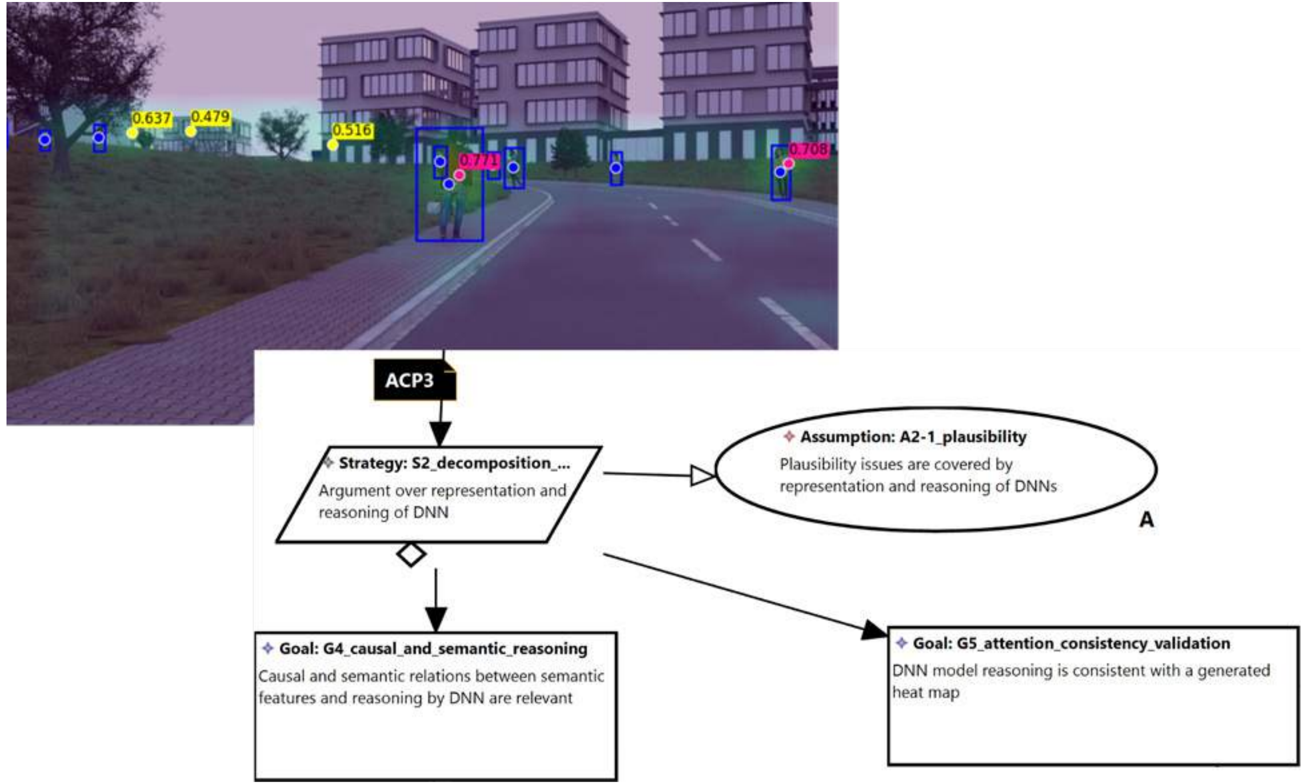
Martin Schels, Gesina Schwalbe, Continental AG
Esra Acar-Celik, Tianming Qiu, fortiss GmbH
Elena Haedecke, Michael Mock, Fraunhofer IAIS

In this poster, EWS-3, which revolves around the safety concerns „Incomprehensible Behavior and Insufficient Plausibility“ are introduced. We investigate 3 different TP3 methods that line up for this task:

- Concept embedding and hybrid learning (Continental AG)
- Visual analytics (Fraunhofer IAIS)
- Attention based heatmaps (fortiss)

We show corresponding GSN fragments and also how to bridge the gap between data driven development and our safety argumentation.

AP4.3


0.637
0.479
0.516
0.771
0.708
ACP3
Strategy: S2_decomposition_...
Argument over representation and reasoning of DNN
Assumption: A2-1_plausibility
Plausibility issues are covered by representation and reasoning of DNNs
A
Goal: G4_causal_and_semantic_reasoning
Causal and semantic relations between semantic features and reasoning by DNN are relevant
Goal: G5_attention_consistency_validation
DNN model reasoning is consistent with a generated heat map

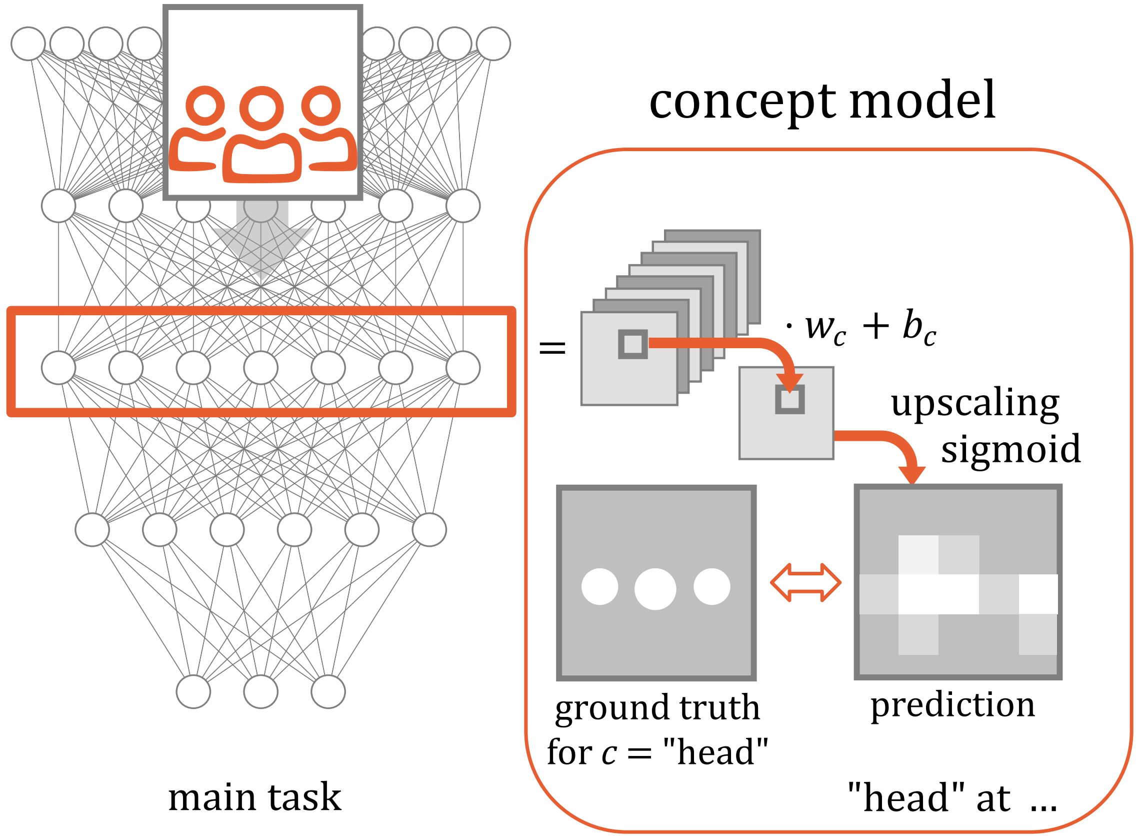
Contributions to EWS-3 (© BIT Technology Solutions | fortiss GmbH | Continental AG)

Using concept wce

Gesina Schwalbe, Martin Schels, Continental AG

Our approach optimizes prior work for associating a semantic concept (e.g. „arm“) with vectors (CAVs) in the DNN latent space. For a layer a small linear concept model (CM) is trained to predict presence of the concept from an activation map pixel in the layer output. The CM weights are the CAV. Several CM-based verification applications are investigated: Verify internal representation of concepts, inspect internal logic, and check compliance with fuzzy logic rules.

AP3.3 | AP3.4



Concept analysis approach (© Continental AG)

Data related Evidence Workstreams

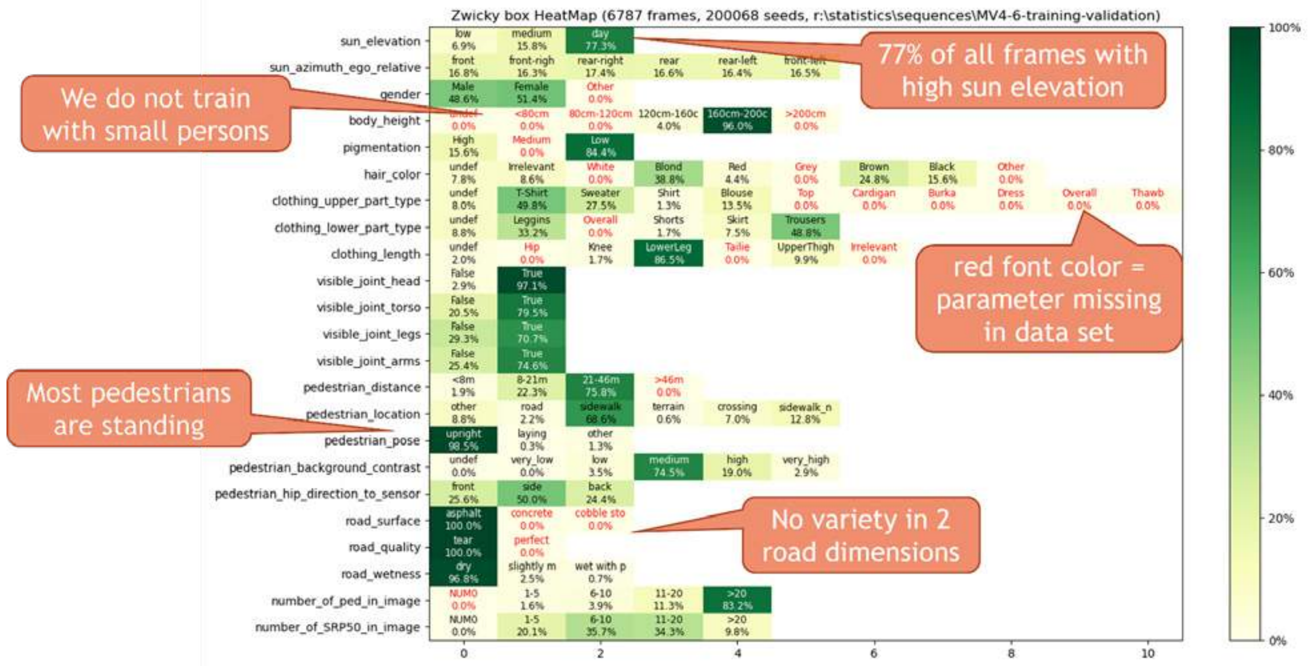
The data-related Evidence Workstreams cover the data related safety concern of inadequate data distribution as well as performance limiting factors. Therefore, the contributions presented in the following address coverage of the ODD and systematic analysis of the elements of the domain model on DNN performance.

Input Coverage Analysis using Domain Models and Combinatorial Testing

Christian Heinzemann, Martin Herrmann, Frédérik Blank, Lydia Gauerhof, Robert Bosch GmbH

Input coverage uses a domain model describing semantic features of input images of a DNN and argues coverage of a training or test dataset with respect to this domain model. Due to the usually high number of semantic features, a full exploration of a domain model is prohibitive. Therefore, we leverage combinatorial testing techniques for defining a weaker notion of coverage for dataset analysis that provides better scalability. We conducted an experiment indicating a relationship between low sample count of features and low DNN performance.

AP4.4. | AP4.1 | AP4.3 | P1



We do not train with small persons

77% of all frames with high sun elevation

red font color = parameter missing in data set

Most pedestrians are standing

No variety in 2 road dimensions

Input coverage for training data set in Mackevision Medien Design GmbH tranches #4, #5, and #6 on single parameters. Missing parameter values (marked by red font color) for n=1 result in even more missing value pairs for n=2 (© Robert Bosch GmbH)

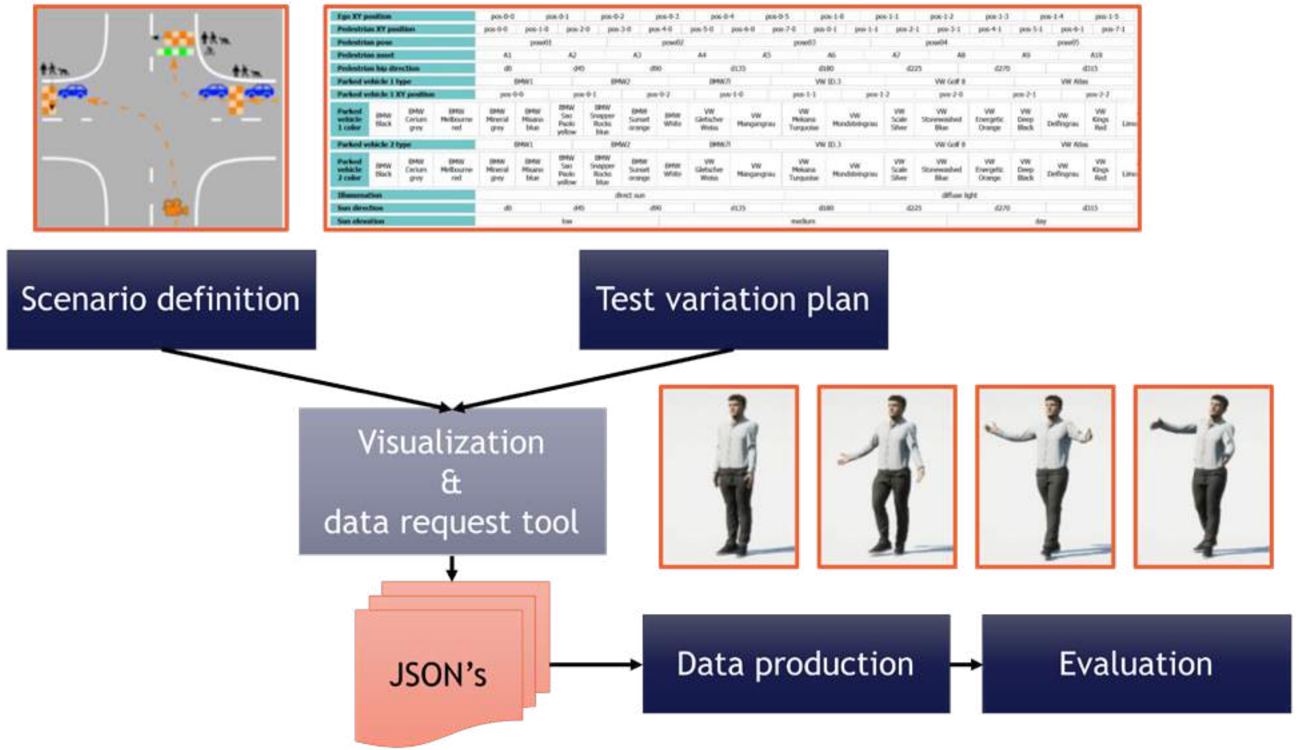
DNN specific Safety Concern: Data distribution is not a good approximation of the real world

ENCAP Standard oriented scenarios for DNN performance evaluation

Thomas Schulik, ZF Friedrichshafen AG
Michael Schuldes, FKA GmbH
Martin Herrmann, Frédérik Blank, Robert Bosch GmbH
Markus Huber, Mackevision Medien Design GmbH

The performance assessment within standardized scenarios like in the ENCAP specification is crucial and important step for the homologation of AD systems in future. For an efficient scenario definition and data production a semi-automated process with a machine-readable format was evolved. The evaluation has shown that the pedestrian pose, bounding box aspect ratio and contrast has a strong impact on the detection performance of the DNN (SSD) under test.

AP2.5 | AP4.1 | AP4.4 | P1



Parameter	Values
Ego XY position	pos-0-0 pos-0-1 pos-0-2 pos-0-3 pos-0-4 pos-0-5 pos-1-0 pos-1-1 pos-1-2 pos-1-3 pos-1-4 pos-1-5
Pedestrian XY position	pos-0-0 pos-1-0 pos-2-0 pos-3-0 pos-4-0 pos-5-0 pos-6-0 pos-7-0 pos-0-1 pos-1-1 pos-2-1 pos-3-1 pos-4-1 pos-5-1 pos-6-1 pos-7-1
Pedestrian pose	pose01 pose02 pose03 pose04 pose05
Pedestrian asset	A1 A2 A3 A4 A5 A6 A7 A8 A9 A10
Pedestrian hip direction	d0 d45 d90 d135 d180 d225 d270 d315
Parked vehicle 1 type	BMW1 BMW2 BMW7i VW ID.3 VW Golf 8 VW Atlas
Parked vehicle 1 XY position	pos-0-0 pos-0-1 pos-0-2 pos-1-0 pos-1-1 pos-1-2 pos-2-0 pos-2-1 pos-2-2
Parked vehicle 1 color	BMW Black BMW Cerium grey BMW Melbourne red BMW Mineral grey BMW Misano blue BMW Sao Paolo yellow BMW Snapper Rocks blue BMW Sunset orange BMW White VW Gletscher Weiss VW Mangangrau VW Makena Turquoise VW Mondsteingrau VW Scale Silver VW Stonewashed Blue VW Energetic Orange VW Deep Black VW Delfingrau VW Kings Red Lime
Parked vehicle 2 type	BMW1 BMW2 BMW7i VW ID.3 VW Golf 8 VW Atlas
Parked vehicle 2 color	BMW Black BMW Cerium grey BMW Melbourne red BMW Mineral grey BMW Misano blue BMW Sao Paolo yellow BMW Snapper Rocks blue BMW Sunset orange BMW White VW Gletscher Weiss VW Mangangrau VW Makena Turquoise VW Mondsteingrau VW Scale Silver VW Stonewashed Blue VW Energetic Orange VW Deep Black VW Delfingrau VW Kings Red Lime
Illumination	direct sun diffuse light
Sun direction	d0 d45 d90 d135 d180 d225 d270 d315
Sun elevation	low medium day

1. Overall process for test set defintion and parameter variation

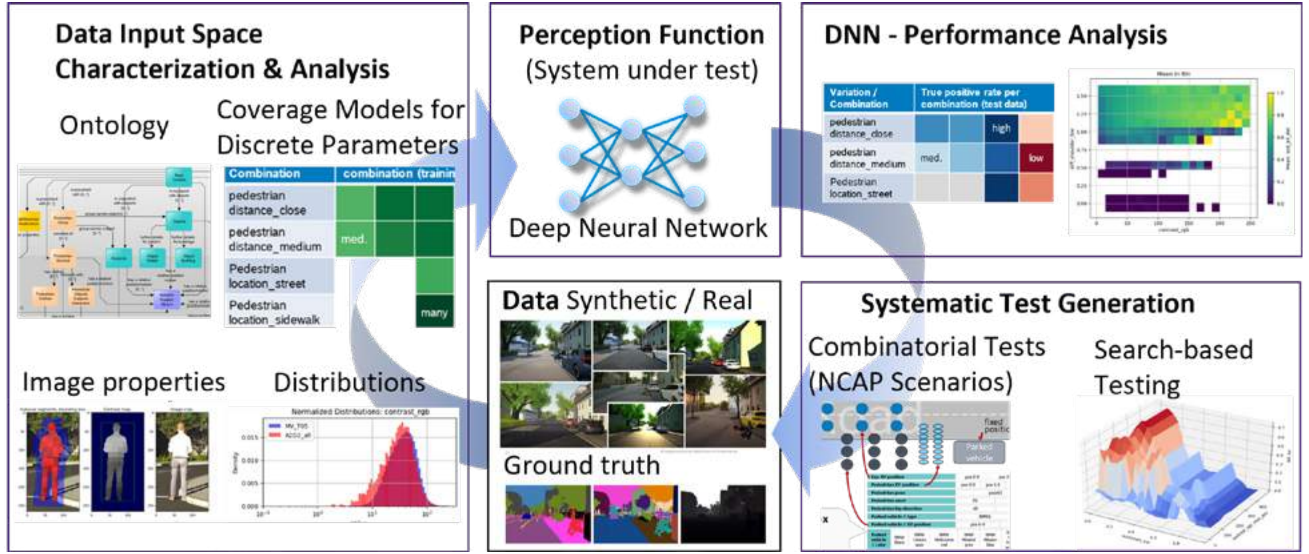
2. Efficient pose selection from pestrian animation and realization of machine readable format (© ZF Friedrichshafen AG | Robert Bosch GmbH | Mackevision Medien Design GmbH)

Applying Image Analysis, Combinatorial and Search-based Testing for DNN-Verification

Christoph Gladisch, Falko Matern, Frédérik Blank, Martin Herrmann, Simon Heming, Robert Bosch GmbH

Rigorous and systematic testing of AI requires new approaches focusing on data. Our approach is to extract information in form of ontology parameters (a.k.a. dimensions) from a labelled image dataset and from the DNNs predictions and to use a collection of black-box testing and analysis techniques. As experiments, we applied search-based testing, combinatorial testing, image analysis, distribution and correlation analysis, and structured image generation techniques.

AP4.4 | AP4.1 | AP2.4 | AP1.2 | P1



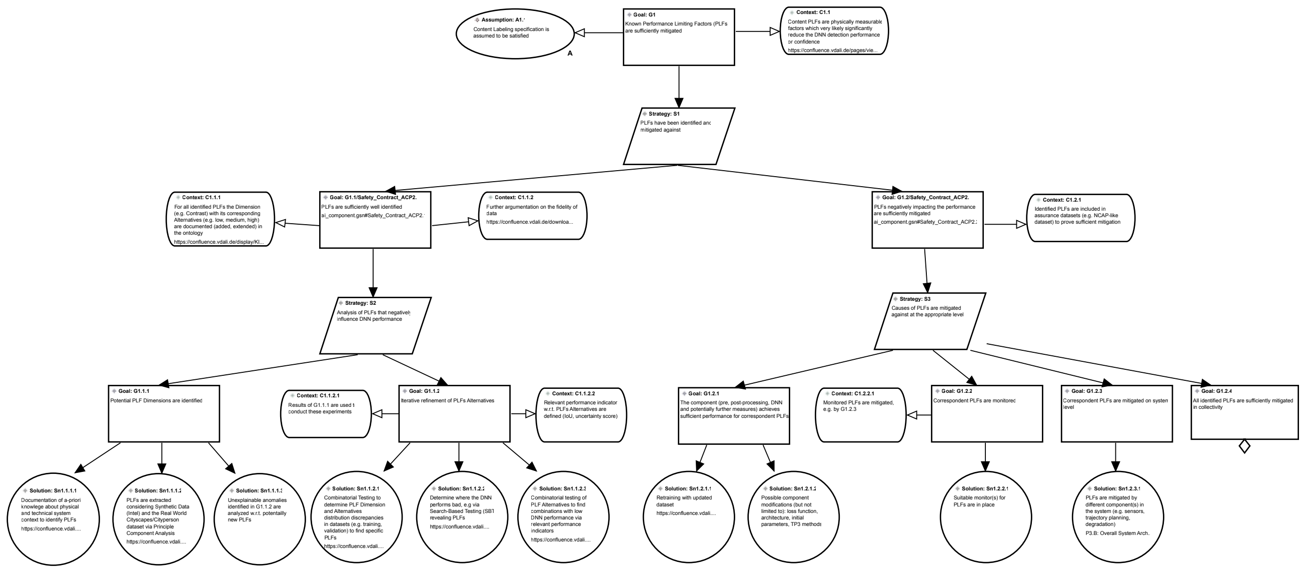
A schema of a verification loop to test a perception function consisting of image analysis and testing approaches like combinatorial and search-based testing (© Robert Bosch GmbH, Mackevision)

Performance Limiting Factors (PLFs) Data-related evidence Workstream

Frédérik Blank, Lydia Gauerhof, Christoph Gladisch, Falko Matern, Robert Bosch GmbH
Oliver Grau, Korbinian Hagn, Intel Corporation
Iwo Kurzidem, Fraunhofer IKS

Performance Limiting Factors are measurable factors, either of a direct physical or a model of an effect that leads to drops in perception performance. The argumentation provides evidence to identify and mitigate PLFs: Usage of a-priori knowledge about physical and technical system context and methods to identify PLFs. Mitigation includes: Retraining with updated dataset, Possible component modifications and by different component(s) on system-level.

AP1.2 | AP2.4 | **AP4.3** | **AP4.4** | P1


Assumption: A1.
Content Labeling specification is assumed to be satisfied
A
Goal: G1
Known Performance Limiting Factors (PLFs are sufficiently mitigated
Context: C1.1
Content PLFs are physically measurable factors which very likely significantly reduce the DNN detection performance or confidence
https://confluence.vdali.de/pages/vie...
Strategy: S1
PLFs have been identified anc mitigated against
Context: C1.1.1
For all identified PLFs the Dimension (e.g. Contrast) with its corresponding Alternatives (e.g. low, medium, high) are documented (added, extended) in the ontology
https://confluence.vdali.de/display/KI...
Goal: G1.1/Safety_Contract_ACP2.
PLFs are sufficiently well identified
ai_component.gsn#Safety_Contract_ACP2.
Context: C1.1.2
Further argumentation on the fidelity of data
https://confluence.vdali.de/downloa...
Goal: G1.2/Safety_Contract_ACP2.
PLFs negatively impacting the performance are sufficiently mitigated
ai_component.gsn#Safety_Contract_ACP2.
Context: C1.2.1
Identified PLFs are included in assurance datasets (e.g. NCAP-like dataset) to prove sufficient mitigation
Strategy: S2
Analysis of PLFs that negatively influence DNN performance
Strategy: S3
Causes of PLFs are mitigated against at the appropriate level
Goal: G1.1.1
Potential PLF Dimensions are identified
Context: C1.1.2.1
Results of G1.1.1 are used t conduct these experiments
Goal: G1.1.2
Iterative refinement of PLFs Alternatives
Context: C1.1.2.2
Relevant performance indicator w.r.t. PLFs Alternatives are defined (IoU, uncertainty score)
Goal: G1.2.1
The component (pre, post-processing, DNN and potentially further measures) achieves sufficient performance for correspondent PLFs
Context: C1.2.2.1
Monitored PLFs are mitigated, e.g. by G1.2.3
Goal: G1.2.2
Correspondent PLFs are monitored
Goal: G1.2.3
Correspondent PLFs are mitigated on system level
Goal: G1.2.4
All identified PLFs are sufficiently mitigated in collectivity
Solution: Sn1.1.1.1
Documentation of a-priori knowlege about physical and technical system context to identify PLFs
https://confluence.vdali....
Solution: Sn1.1.1.2
PLFs are extracted considering Synthetic Data (Intel) and the Real World Cityscapes/Cityperson dataset via Principle Component Analysis
https://confluence.vdali....
Solution: Sn1.1.1.3
Unexplainable anomalies identified in G1.1.2 are analyzed w.r.t. potentailly new PLFs
Solution: Sn1.1.2.1
Combinatorial Testing to determine PLF Dimension and Alternatives distribution discrepancies in datasets (e.g. training, validation) to find specific PLFs
https://confluence.vdali....
Solution: Sn1.1.2.2
Determine where the DNN performs bad, e.g via Search-Based Testing (SBT revealing PLFs
https://confluence.vdali....
Solution: Sn1.1.2.3
Combinatorial testing of PLF Alternatives to find combinations with low DNN performance via relevant performance indicators
https://confluence.vdali....
Solution: Sn1.2.1.1
Retraining with updated dataset
https://confluence.vdali....
Solution: Sn1.2.1.2
Possible component modifications (but not limited to): loss function, architecture, initial parameters, TP3 methods
Solution: Sn1.2.2.1
Suitable monitor(s) for PLFs are in place
Solution: Sn1.2.3.1
PLFs are mitigated by different component(s) in the system (e.g. sensors, trajectory planning, degradation)
P3.B: Overall System Arch..

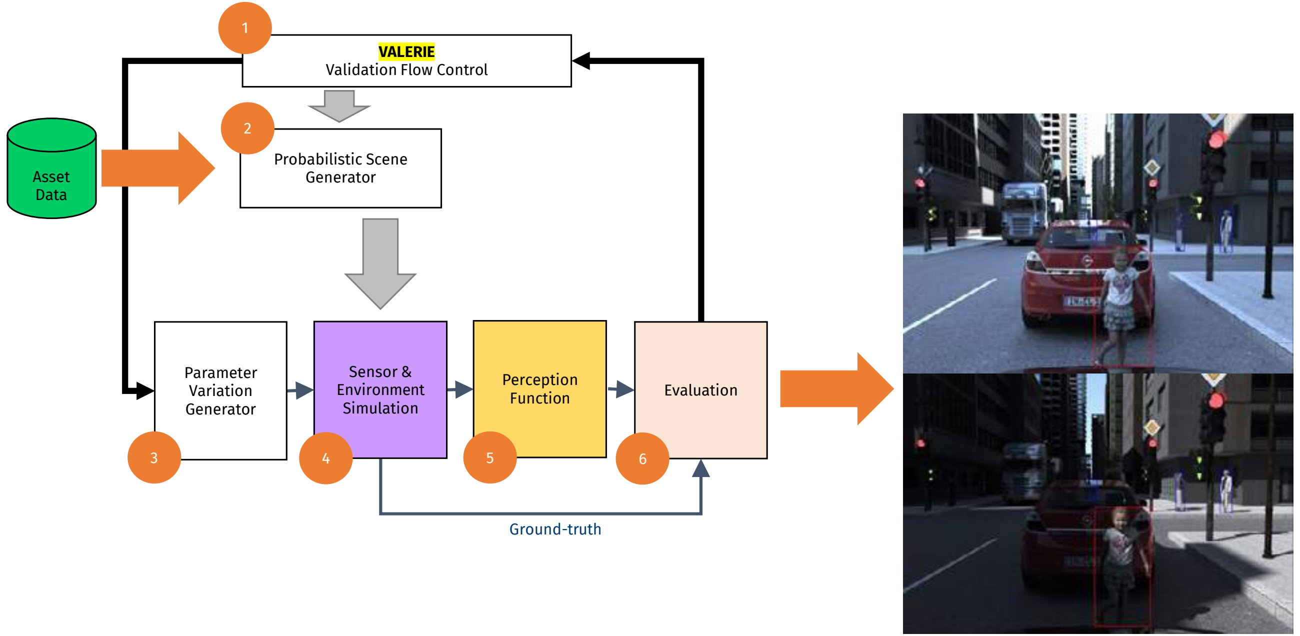
Schematic overview of the GSN safety argumentation for PLF mitigation (© Robert Bosch GmbH, FhG IKS)

Automated AI Validation using deep variational data synthesis

Korbinian Hagn, Oliver Grau, Intel Corporation

Automating AI validation through variational data synthesis enables detection of flaws in the DNN predictions. By generation and synthetization of highly parameterizable inner-city street scenes and a realistic sensor simulation, detection errors are spotted by the VALERIE flow control.

AP2.1 | AP2.4 | AP4.4


1
VALERIE
Validation Flow Control
2
Probabilistic Scene Generator
Asset Data
Parameter Variation Generator
3
Sensor & Environment Simulation
4
Perception Function
5
Evaluation
6
Ground-truth

VALERIE detects perception faults like missed pedestrian detections (red) in a scene through variations of the scene parameters, e.g., the sun position (© Intel Corporation)

Methods & Measures

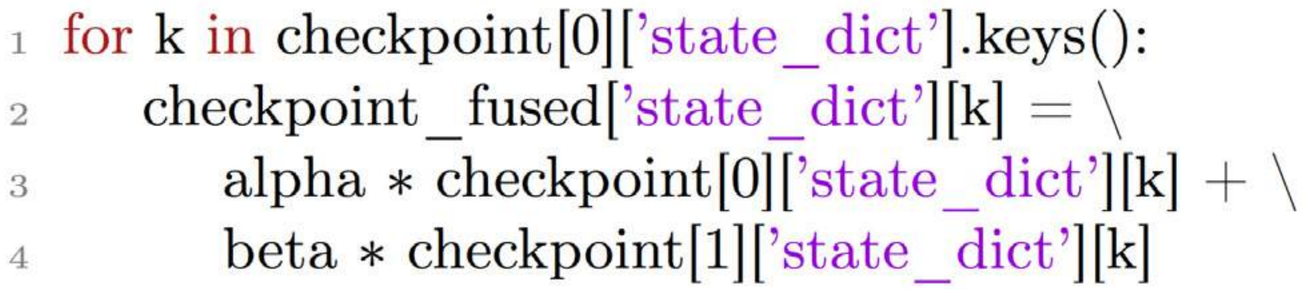
This section presents selected methods and measures from KI Absicherung that systematically determine and reduce inherent insufficiencies of AI functions. These mechanisms are examined and evaluated with respect to their significance in terms of safety and their effectiveness.

Improving Predictive Performance and Calibration by Weight Fusion in Semantic Segmentation

Timo Sämann, Valeo Schalter und Sensoren GmbH
Ahmed Hammam, Opel Automobile GmbH, **Andrei Bursuc**, Valeo.ai
Christoph Stiller, Karlsruhe Institute of Technology, **Horst-Michael Groß**, Ilmenau University of Technology

Contributions:
(i) Our weight fusion method improves predictive performance and calibration without impacting runtime cost. **(ii)** We introduce a new testing method that can measure the functional space between weights, called oracle testing. **(iii)** We show the superiority of our approach in a comparison with Stochastic Weight Averarging (SWA) and deep ensembles for in-distribution as well as for out-of-distribution data (ACDC).

AP3.6 | AP3.4



```
for k in checkpoint[0]['state_dict'].keys():
    checkpoint_fused['state_dict'][k] = \
        alpha * checkpoint[0]['state_dict'][k] + \
        beta * checkpoint[1]['state_dict'][k]
```

PyTorch code snippet of our weight fusion approach (© Valeo Schalter und Sensoren GmbH)

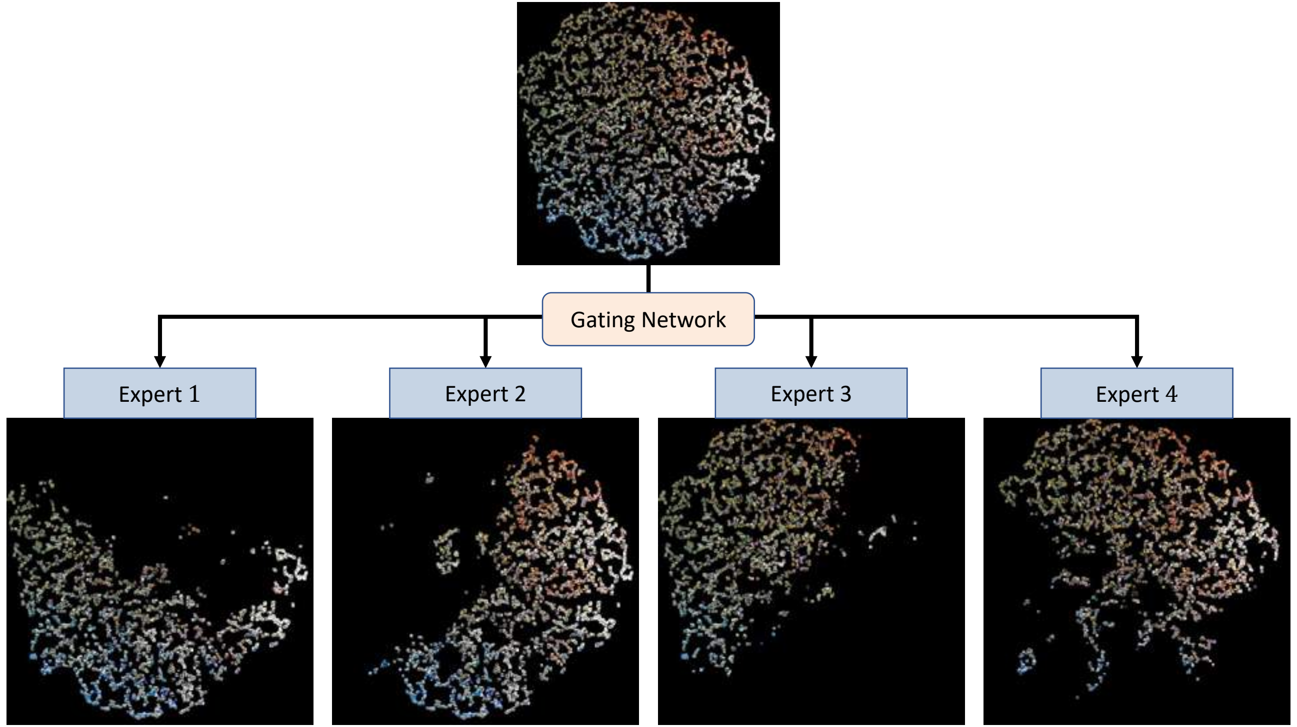
DNN specific Safety Concern: Lack of generalization

Insights into CNN Decision-Making via Embedded Sparse Mixture-of-Expert Layers

Svetlana Pavlitskaya, Christian Hubschneider, Michael Weber, J. Marius Zöllner,
FZI Forschungszentrum Informatik
Lukas Struppek, Karlsruher Institut für Technologie (KIT)

parsely-gated mixtures of experts (MoEs), embedded directly into the CNN layers, allow for end-to-end training without explicit dataset splits. We propose constraints to balance expert utilization during training and thus to control the trade-off between model performance and expert specialization. Embedded MoEs can provide additional insights into the decision-making process of CNNs, as experts can implicitly focus on individual sub-domains of the input space.

AP3.6



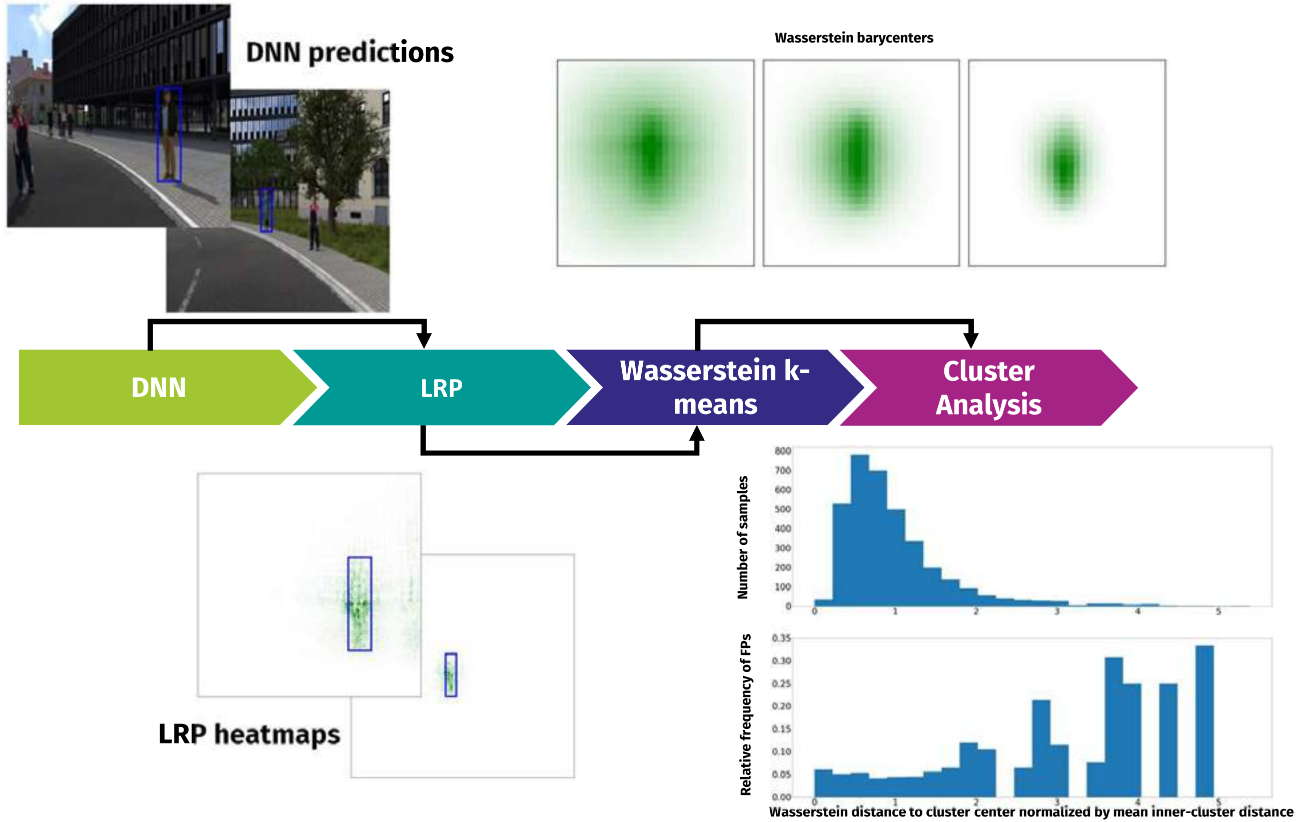
Assignment of input samples to specific experts in the MoE embedded into the last ResNet block of ResNet-18 for image classification on CIFAR-100 (© FZI Forschungszentrum Informatik)

Morphological aggregation of heatmaps with Wasserstein k-means

Gregor Richter, Alicia jimenez Herrera, Sabine Hug, umlaut
Dennis Herbik, ehemals umlaut

Saliency maps allow highlighting prediction relevant input features thus explaining individual predictions. We extend the saliency map Layer-wise Relevance Propagation (LRP) to object detection. Heatmaps generated via LRP are local, hence we consider aggregation by morphological clustering in order to reduce the complexity of the derived dataset of explanations. Yet, limitations of saliency map evaluation metrics make the benefit to a safety argumentation unclear.

AP3.1



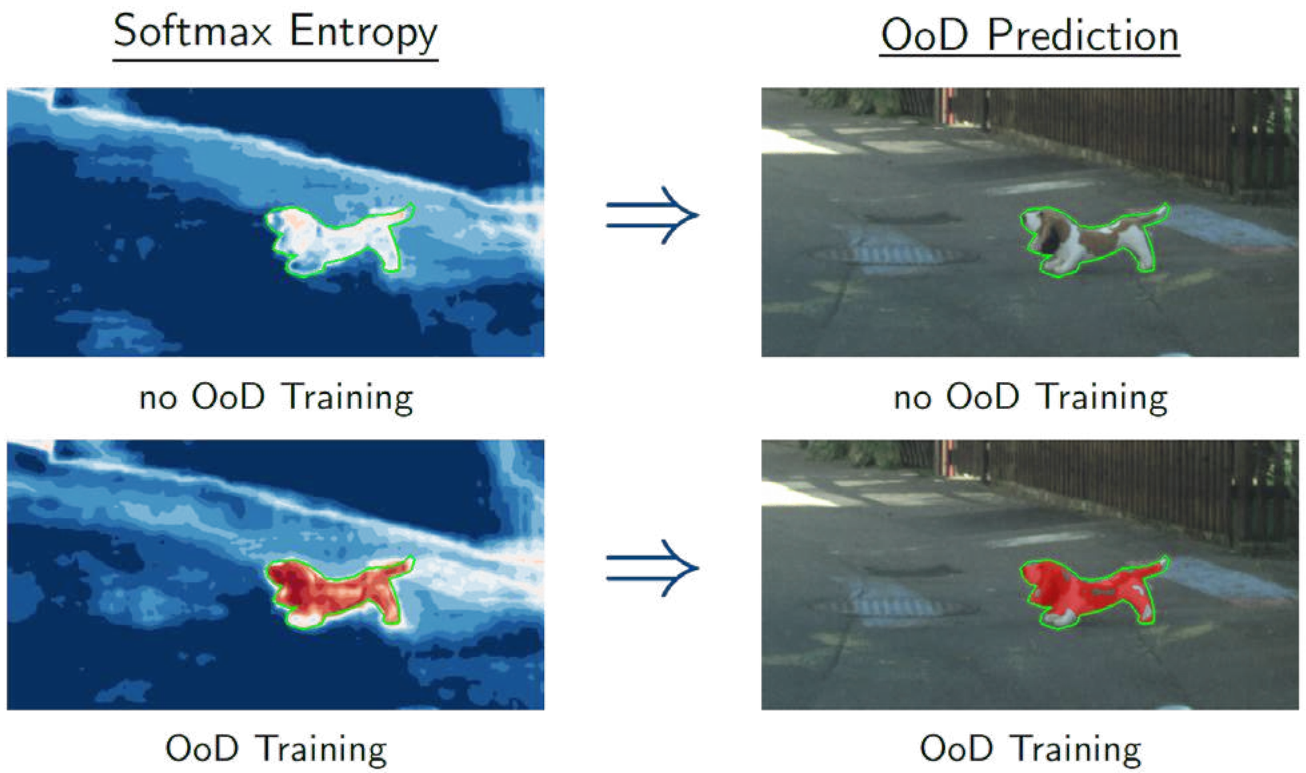
General workflow for aggregating heatmaps (© umlaut)

Out-of-Distribution Detection in Semantic Segmentation

Robin Chan, Svenja Uhlemeyer, Matthias Rottmann, Hanno Gottschalk, University of Wuppertal

Objects from unknown classes are also considered as „out-of-distribution“ (OoD) examples and their detection is extremely safety-relevant in many real-world applications, particularly in high-stakes applications like automated driving. In this work, we approach OoD detection in semantic segmentation. We present a method that achieves significant OoD detection improvements while sacrificing only marginally in original semantic segmentation performance.

AP3.3



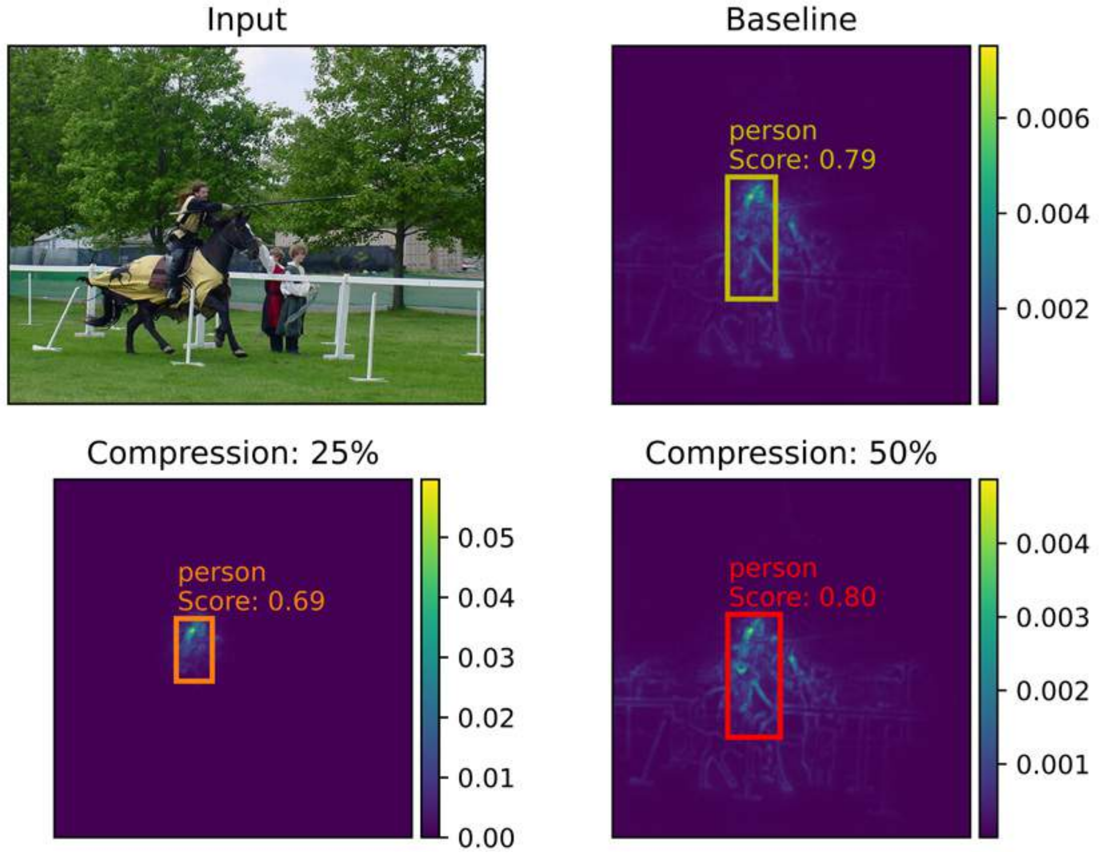
Comparison of OoD predictions with our OoD training approach (bottom row) and without (top row). The predictions are obtained by thresholding on the softmax entropy heatmaps (© BUW)

Analyzing the effect of pruning on the robustness of DNNs

Sven Mantowsky, Firas Mualla, ZF Group

In order for state of the art DNNs to meet the restrictions of embedded systems, they need to be compressed while not only keeping their accuracy but also maintain their safety requirements. To evaluate the robustness of models after compression in comparison to their baseline, we analyze two different metrics: Heatmap correlation and Expected Calibration Error.

AP3.3



Heatmap comparison between baseline and HRank-based compressed SSD models (compression rate of 25% and 50%) (© ZF Friedrichshafen AG | Continental AG)

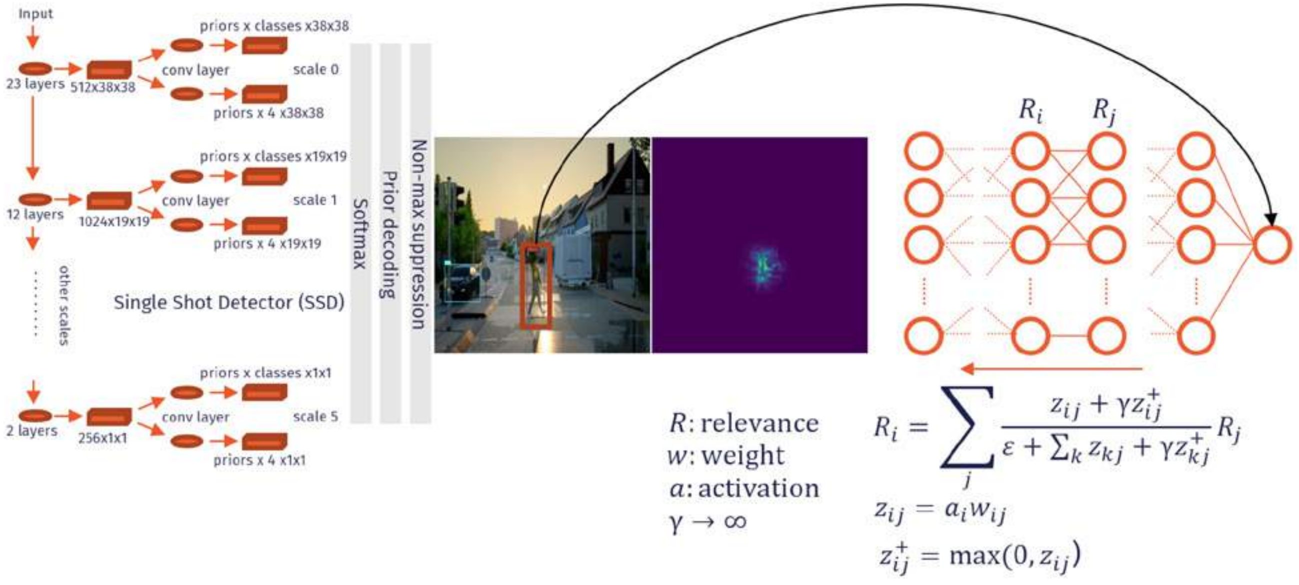
DNN specific Safety Concern: Unreliable confidence information & safety-aware metrics

Extension of Deep Taylor Decomposition to Object Detection

Firas Mualla, ZF Group

The Explainable-AI method Deep Taylor Decomposition (DTD) addresses the safety concern incomprehensible behavior. It delivers some insights, as to which pixels contribute to the model‘s decision. Compared to other heatmap methods, it also claims kind of theoretical soundness based on Taylor decomposition. We extended the method from classification to object detection, namely to the Single Shot Detector (SSD). As there is no corresponding ground truth, the evaluation was done based on an offline temporal stability analysis. In addition, we examined a possible application of the DTD method for the detection of adversarial attacks on SSD.

AP3.4



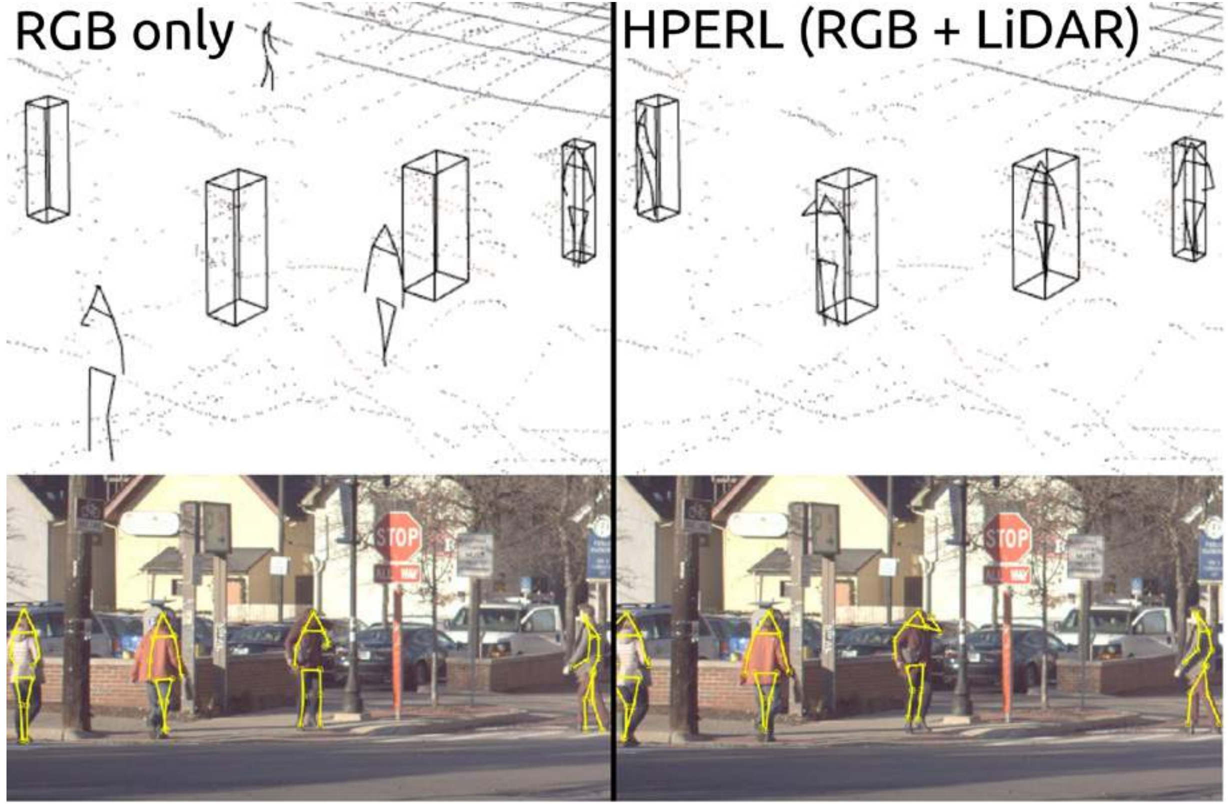
Extension of Deep Taylor Decomposition (DTD) to object detection. For relu-based networks such as SSD, DTD is equivalent to Layerwise Relevance Propagation (LRP) γ-rule when γ approaches infinity (© ZF Friedrichshafen AG | Mackevision Medien Design GmbH)

Sensor Fusion for Robust Pedestrian Detection and Human Pose Estimation

Michael Fürst, DFKI

Precise 3D localization of pedestrians is critical for AVs. However, current camera only approaches suffer from depth ambiguity. LiDAR only approaches have perfect depth perception, but lack resolution required for reliable long range pedestrian detection. By adequately fusing information from multiple types of sensors, the strengths are combined and weaknesses mitigated. Our work highlights different advantages and shortcomings of Camera, LiDAR and fusion approaches.

AP1.4 | **AP1.5**



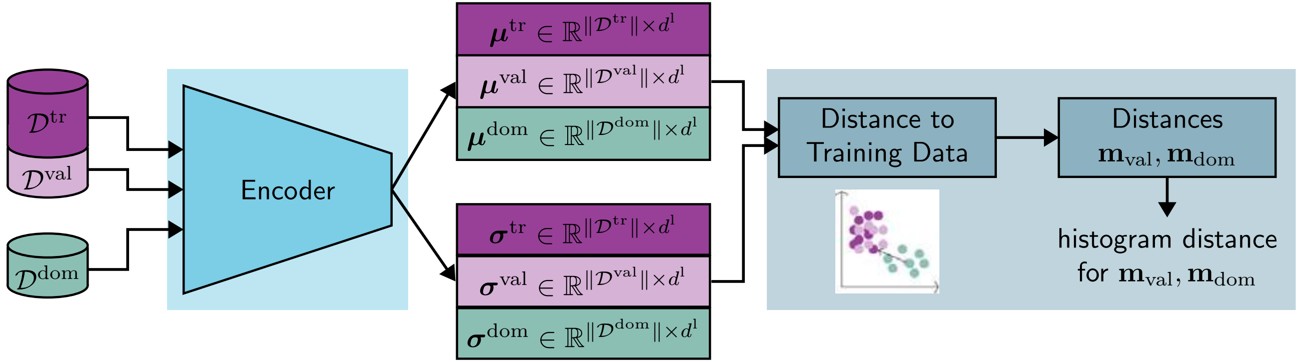
Depth ambiguity leads to the skeletons being far off the ground truth bounding box in RGB only (left). With fusion the error is significantly reduced (right) (© 2019 UM & Ford Center for Autonomous Vehicles (FCAV) | DFKI)

Analysis and Comparison of Datasets by Leveraging Data Distributions in Latent Spaces

Hanno Stage, Lennart Ries, Eric Sax, FZI Forschungszentrum Informatik

One insufficiency of DNNs is their ability to generalize from given training data. To mitigate this, methods for the detection of domain shifts between training and validation data are necessary. We showed that latent spaces of VAEs can be used find domain shifts between data sets during early development. We investigated numerous VAEs, distance metrics and exemplary domain shifts, where a Joint VAE with a probabilistic distance metric provided best results.

AP3.5


$\mathcal{D}^{\mathrm{tr}}$
$\mathcal{D}^{\mathrm{val}}$
$\mathcal{D}^{\mathrm{dom}}$
Encoder
$\boldsymbol{\mu}^{\mathrm{tr}} \in \mathbb{R}^{|\mathcal{D}^{\mathrm{tr}}| \times d^{\mathrm{l}}}$
$\boldsymbol{\mu}^{\mathrm{val}} \in \mathbb{R}^{|\mathcal{D}^{\mathrm{val}}| \times d^{\mathrm{l}}}$
$\boldsymbol{\mu}^{\mathrm{dom}} \in \mathbb{R}^{|\mathcal{D}^{\mathrm{dom}}| \times d^{\mathrm{l}}}$
$\boldsymbol{\sigma}^{\mathrm{tr}} \in \mathbb{R}^{|\mathcal{D}^{\mathrm{tr}}| \times d^{\mathrm{l}}}$
$\boldsymbol{\sigma}^{\mathrm{val}} \in \mathbb{R}^{|\mathcal{D}^{\mathrm{val}}| \times d^{\mathrm{l}}}$
$\boldsymbol{\sigma}^{\mathrm{dom}} \in \mathbb{R}^{|\mathcal{D}^{\mathrm{dom}}| \times d^{\mathrm{l}}}$
Distance to Training Data
Distances $\mathbf{m}_{\mathrm{val}}, \mathbf{m}_{\mathrm{dom}}$
histogram distance for $\mathbf{m}_{\mathrm{val}}, \mathbf{m}_{\mathrm{dom}}$

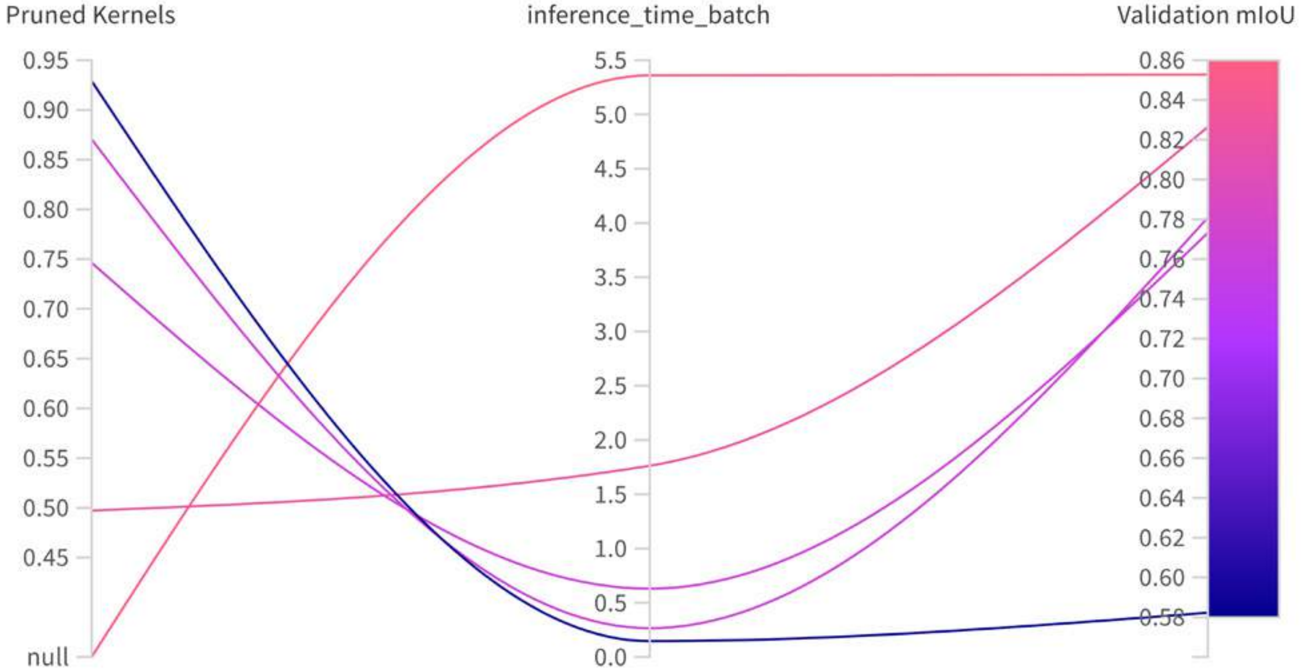
Overview of our approach. The encoder of a variational autoencoder (VAE) is used to transform data into a latent space and then compare the data to detect domain shifts (© FZI Forschungszentrum Informatik)

Self-compressing online pruning

Konstantin Ditschnuneit, Johannes Otterbach, Merantix Momentum GmbH

State-of-the-art semantic segmentation models are characterized by high parameter counts and slow inference times, making them unsuitable for deployment in resource-constrained environments - such as within autonomous driving vehicles. The proposed algorithm predictably prunes models to a given performance or minimum inference speed. Thus allowing the user to prune models up to the exact inference speed required for the given task and available hardware.

AP3.3



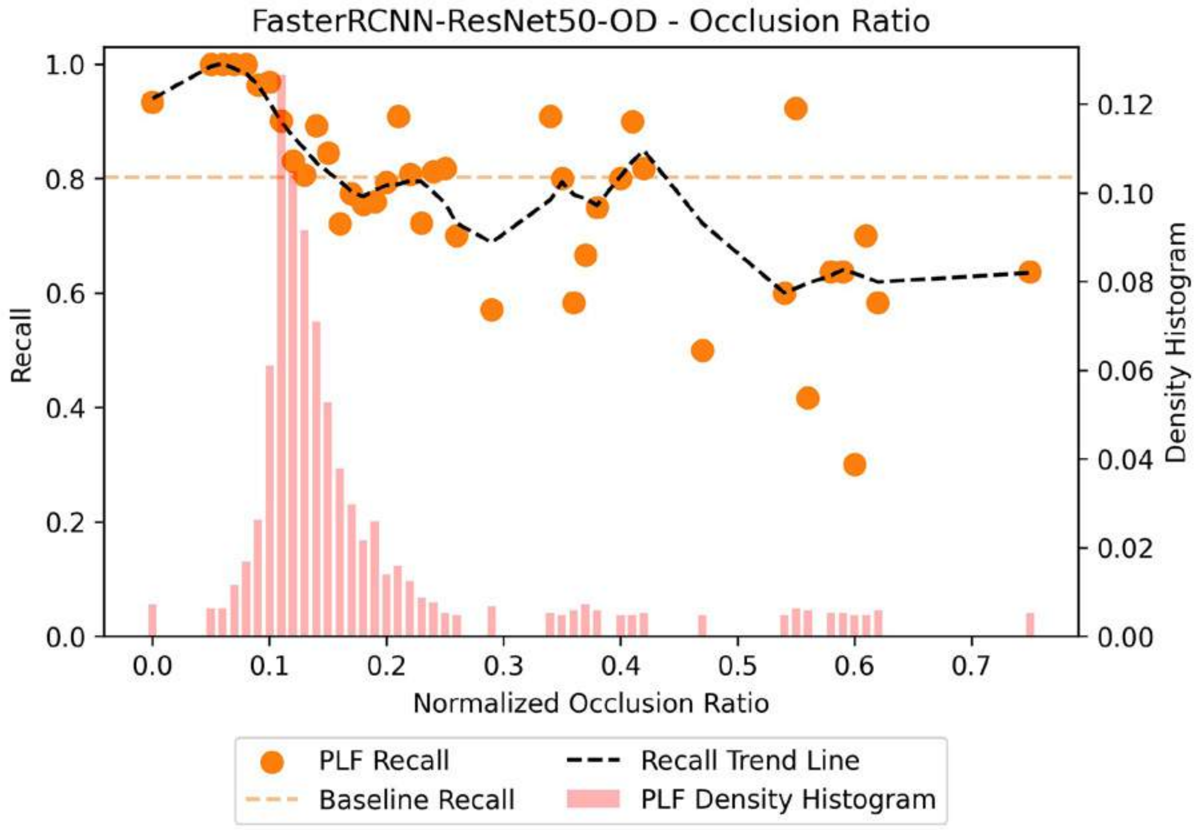
Correlation between pruned convolution kernels, inference time per batch and mean IoU on the validation split (© Merantix Momentum GmbH)

DNN Performance Limiting Factors Analysis

Yasin Bayzidi, Alen Smajic, Volkswagen AG

Deep Neural Networks might face challenging situations that hinder their detection performance. Recognizing such situations in a systematic way would help to recognize the factors that contribute to their mis-behaviour. To do so, 23 performance limiting factors are extracted from two data-sets and analyzed throughout six pedestrian detection models. The factors are categorized into two categories based on color and texture or geometrical properties.

AP2.2



The object occlusion rate correlation with the FasterRCNN Recall combined with the histogram of the occlusion rates throughout the CityPersons data-set (© Volkswagen AG)

Safety & Testing

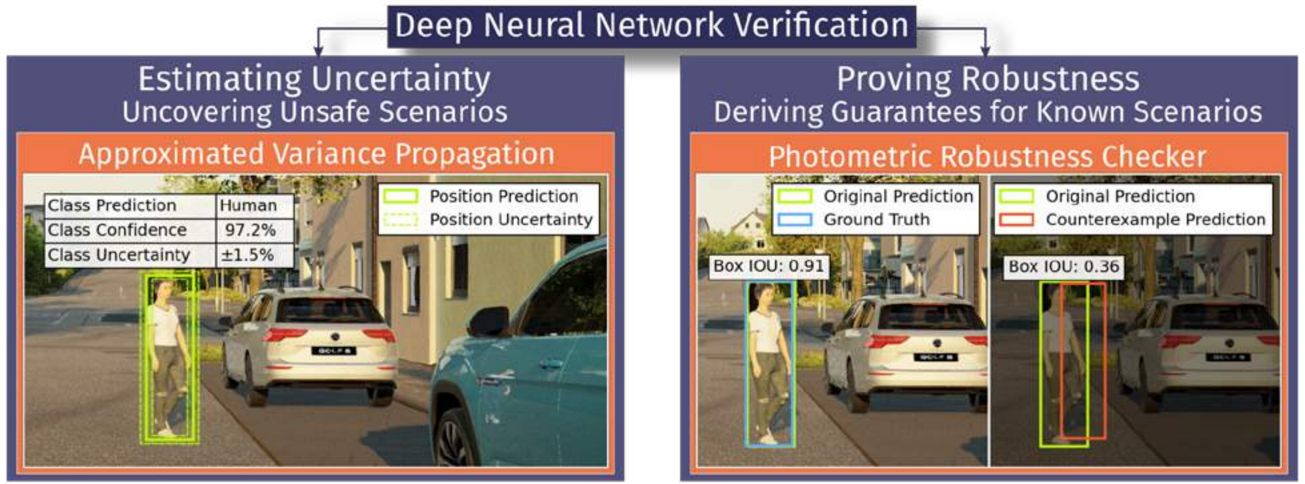
This section discusses contributions to an evidence-based safety argumentation in order to support a convincing safety assurance case for the ML-based pedestrian detection. Moreover, it showcases test and analysis methods that can be used to generate evidences and puts them into context.

Perspectives on Safety: Estimating and Proving

Tom Thielo, **Christian Brunner**, **Kai Fabi, Jonas Schneider**, Elektronische Fahrwerksysteme GmbH

To deploy a deep neural network in a safety critical application, it is crucial to verify its intended behavior. By applying approximated variance propagation for real-time uncertainty estimation, we uncover unknown unsafe scenarios during development and operation by live monitoring. Additionally, before releasing an AI model, we derive guarantees for known scenarios by proving the robustness of the models predictions, e.g., against photometric transformations.

AP3.4 | AP3.5


Deep Neural Network Verification
Estimating Uncertainty
Uncovering Unsafe Scenarios
Approximated Variance Propagation
Class Prediction | Human
Class Confidence | 97.2%
Class Uncertainty | ±1.5%
Position Prediction
Position Uncertainty
Proving Robustness
Deriving Guarantees for Known Scenarios
Photometric Robustness Checker
Original Prediction
Ground Truth
Box IOU: 0.91
Original Prediction
Counterexample Prediction
Box IOU: 0.36

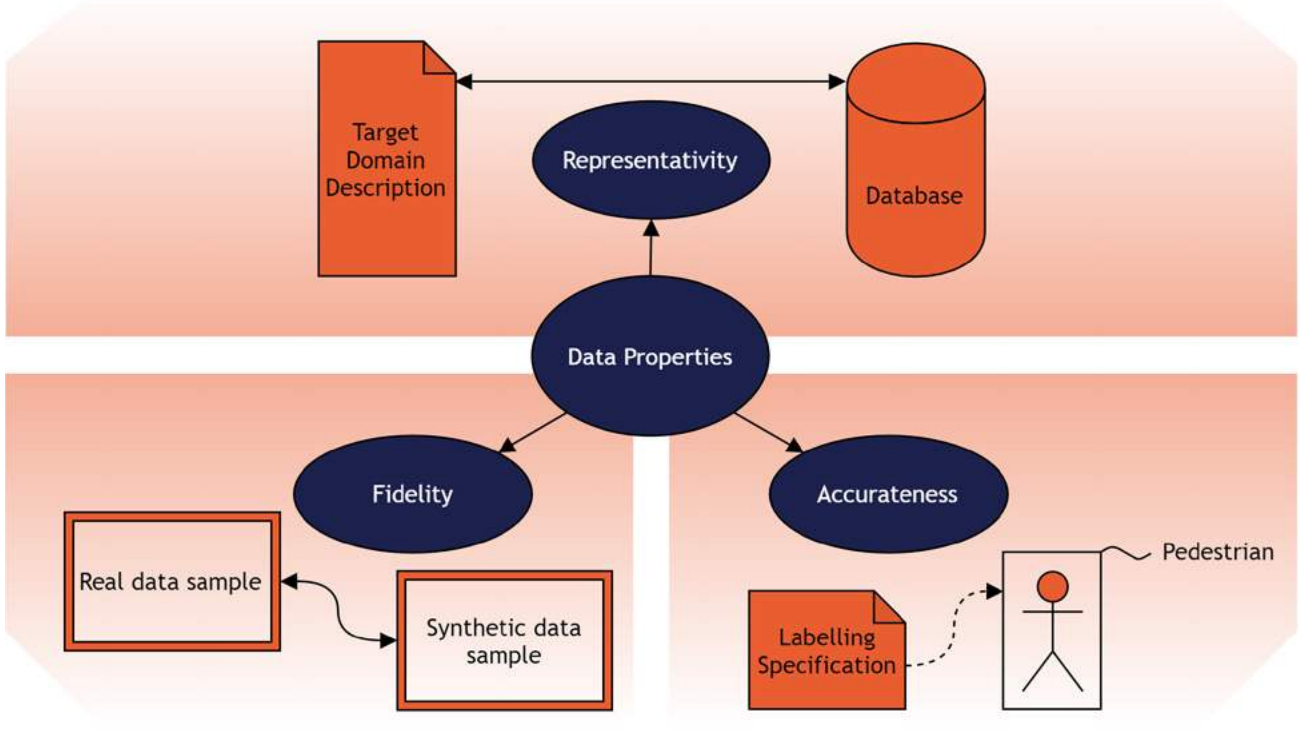
The neural network is certain in its prediction, but not robust under transformations. (© EFS GmbH | Mackevision Medien Design GmbH)

Safety Case Patterns for the Argumentation of a Sufficient Database

Markus Bach, Valeo Schalter und Sensoren GmbH
Christian Hellert, Continental AG
Lukas Bergmann, Volkswagen AG
Christian Pfister, Automotive Safety Technologies

The database has an essential influence on the development and evaluation of DNNs. Consequently, the safety argumentation must include evidences for a sufficient database for the intended functionality. In this context, three main data properties were identified: representativity, fidelity and accurateness. We have created a Safety Case Pattern for each property and instantiated it in the context of the KI Absicherung use case.

AP4.3



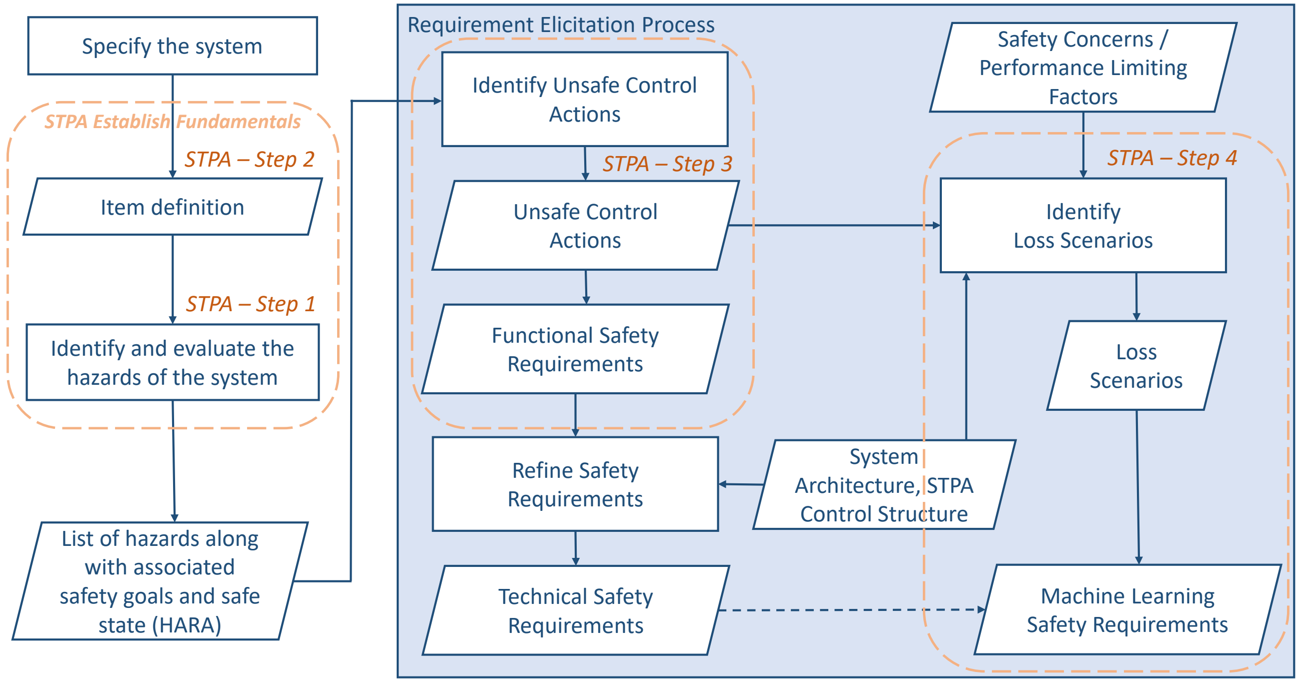
Overview of identified Data Properties – Representativity, Fidelity and Accurateness (© Continental AG)

Elicitation of Machine Learning Safety Requirements via STPA

Stefan Bläsius, Fridolin Bauer, BMW Group, **Esra Acar-Celik**, fortiss GmbH
Christian Pfister, Automotive Safety Technologies**, Martin Schels**, Continental AG
Markus Bach, Valeo Schalter und Sensoren GmbH, **Lukas Bergmann**, Volkswagen AG
Asim Abdulkhaleq, Shervin Raafatnia, Robert Bosch GmbH

We present our experience with applying System-Theoretic Process Analysis (STPA) to a Machine Learning (ML) based pedestrian collision avoidance system. STPA is integrated into the safety life cycle of functional safety (ISO 26262) complemented with Safety of the intended Functionality (ISO/PAS 21448) in order to elicit safety requirements. The requirements are derived using DNN-specific Safety Concerns and Performance Limiting Factors in ML specific Loss Scenarios.

AP4.2


Specify the system
STPA Establish Fundamentals
STPA – Step 2
Item definition
STPA – Step 1
Identify and evaluate the hazards of the system
List of hazards along with associated safety goals and safe state (HARA)
Requirement Elicitation Process
Identify Unsafe Control Actions
STPA – Step 3
Unsafe Control Actions
Functional Safety Requirements
Refine Safety Requirements
Technical Safety Requirements
System Architecture, STPA Control Structure
Safety Concerns / Performance Limiting Factors
STPA – Step 4
Identify Loss Scenarios
Loss Scenarios
Machine Learning Safety Requirements

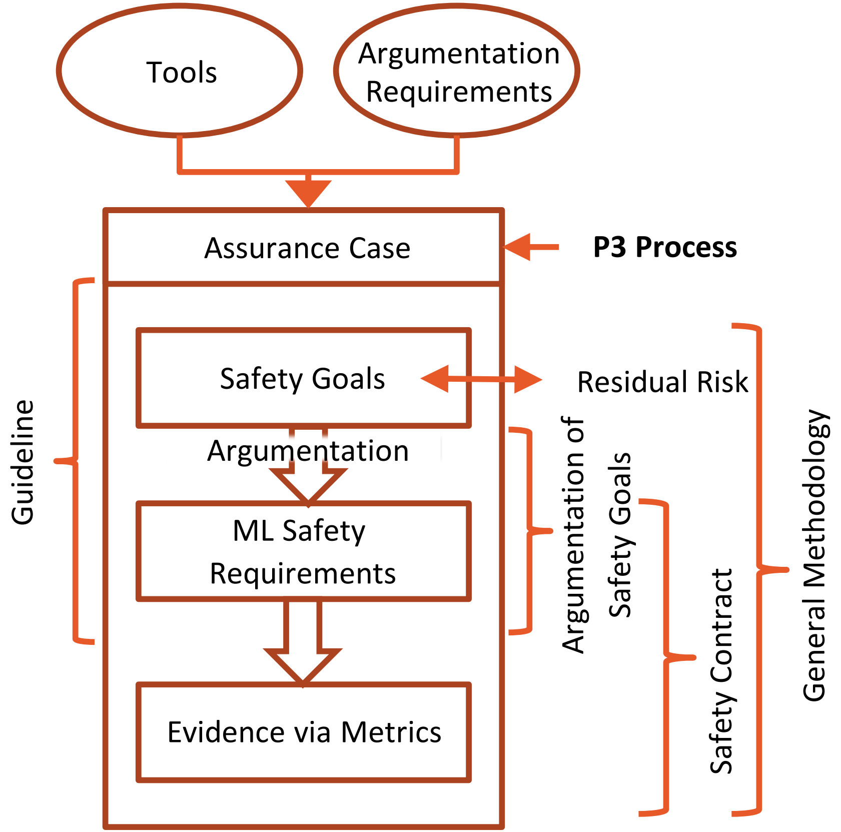
Workflow of the STPA based approach for the elicitation of Machine Learning Safety Requirements
(© BMW Group | fortiss GmbH | Automotive Safety Technologies | Continental AG | Valeo Schalter und Sensoren GmbH | Volkswagen AG | Robert Bosch GmbH)

Safety Argumentation Structure and Safety Requirements for the AI Function

Christian Pfister, Automotive Safety Technologies, **Martin Schels**, Continental AG
Esra Acar-Celik, fortiss GmbH, **Markus Bach**, Valeo Schalter und Sensoren GmbH, **Stefan Bläsius**, BMW Group
Iwo Kurzidem, Fraunhofer IKS, **Lukas Bergmann**, Volkswagen AG, **Lydia Gauerhof**, Robert Bosch GmbH

AP4.2 aims at arguing the safety of a system using a Deep Neural Network (DNN) for pedestrian detection down to the level of the AI function. Important blocks are the residual risk considerations, the elicitation of meaningful requirements along with suitable metrics and the definition of safety contracts. All these tasks are non-trivial when using machine learning and DNNs.

AP4.2



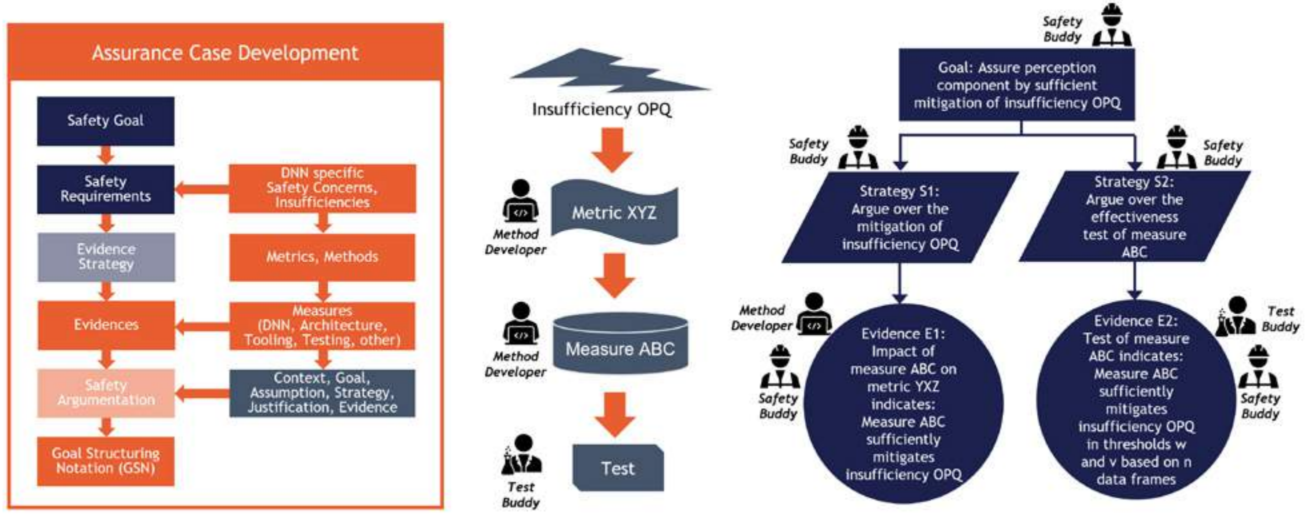
Overview of AP4.2 and contributions to the Assurance Case (© Continental AG | Robert Bosch GmbH | Automotive Safety Technologies)

Evidence-based Safety Argumentation: Approach and Organizational Setup

Andreas Rohatschek, Robert Bosch GmbH
Thomas Schulik, ZF Friedrichshafen AG
Christian Pfister, Automotive Safety Technologies GmbH

According to the principles of ISO 26262, ISO/DIS 21448, and ISO/TR 4084, the assurance case shall state in a convincing way: „The system is safe because…“. The central aspect of safety argumentation is to show that the mitigation of insufficiencies was successful. If the insufficiency is reduced to an acceptable level, this provides evidence to be used in the safety argumentation. This is supported by considering DNN-related safety concerns.

AP4.3


Assurance Case Development
Safety Goal
Safety Requirements
Evidence Strategy
Evidences
Safety Argumentation
Goal Structuring Notation (GSN)
DNN specific Safety Concerns, Insufficiencies
Metrics, Methods
Measures (DNN, Architecture, Tooling, Testing, other)
Context, Goal, Assumption, Strategy, Justification, Evidence
Insufficiency OPQ
Method Developer
Metric XYZ
Method Developer
Measure ABC
Test Buddy
Test
Safety Buddy
Goal: Assure perception component by sufficient mitigation of insufficiency OPQ
Safety Buddy
Strategy S1: Argue over the mitigation of insufficiency OPQ
Safety Buddy
Strategy S2: Argue over the effectiveness test of measure ABC
Method Developer
Safety Buddy
Evidence E1: Impact of measure ABC on metric YXZ indicates: Measure ABC sufficiently mitigates insufficiency OPQ
Evidence E2: Test of measure ABC indicates: Measure ABC sufficiently mitigates insufficiency OPQ in thresholds w and v based on n data frames
Test Buddy
Safety Buddy

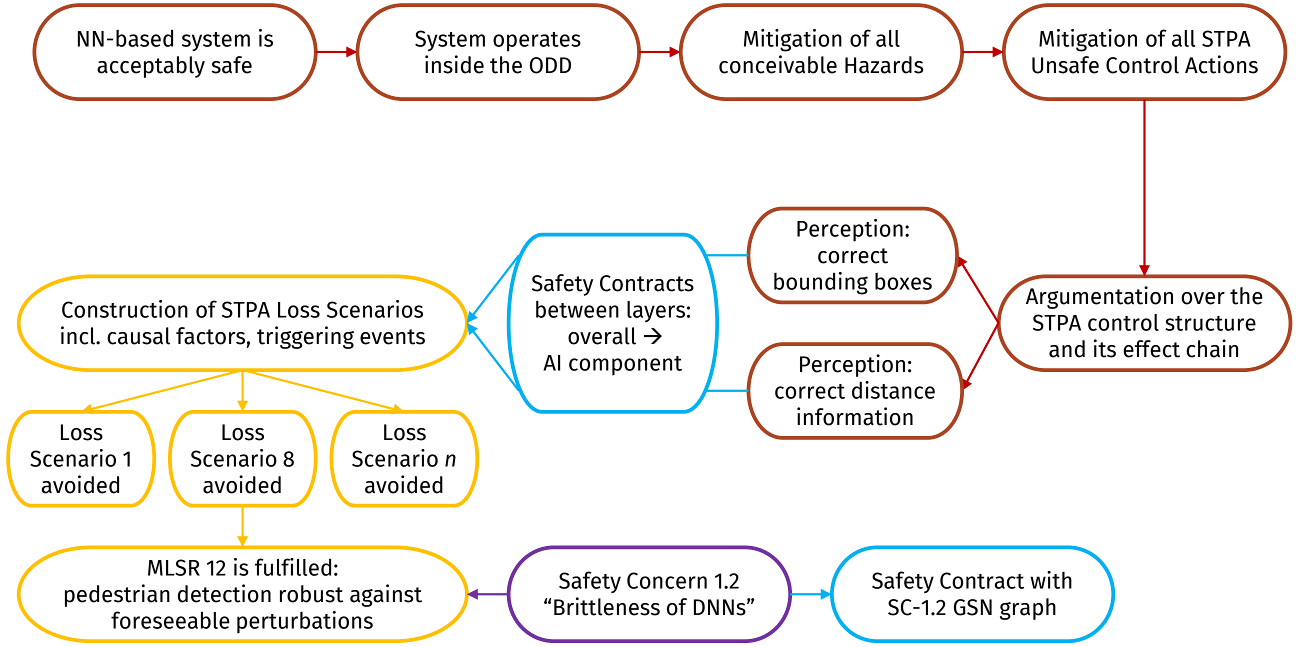
How to create evidences from methods and tests (© Robert Bosch GmbH)

Structure of the Overall Safety Argumentation

Christian Pfister, Automotive Safety Technologies
Esra Acar-Celik, fortiss GmbH
Andreas Rohatschek, Robert Bosch GmbH
Markus Bach, Valeo Schalter und Sensoren GmbH

The safety argumentation is structured in 2 layers: The overall GSN graph is the top-level part. We argue, inside the ODD, over the mitigation of Hazards and their corresponding Unsafe Control Actions down to the level of the AI component. Here, we argue over the avoidance of Loss Scenarios with corresponding causal factors (e.g. Safety Concerns, Performance Limiting Factors), supported by fulfilling the Machine Learning Safety Requirements.

AP4.2 | **AP4.3**


NN-based system is acceptably safe
System operates inside the ODD
Mitigation of all conceivable Hazards
Mitigation of all STPA Unsafe Control Actions
Argumentation over the STPA control structure and its effect chain
Perception: correct bounding boxes
Perception: correct distance information
Safety Contracts between layers: overall → AI component
Construction of STPA Loss Scenarios incl. causal factors, triggering events
Loss Scenario 1 avoided
Loss Scenario 8 avoided
Loss Scenario *n* avoided
MLSR 12 is fulfilled: pedestrian detection robust against foreseeable perturbations
Safety Concern 1.2 “Brittleness of DNNs”
Safety Contract with SC-1.2 GSN graph

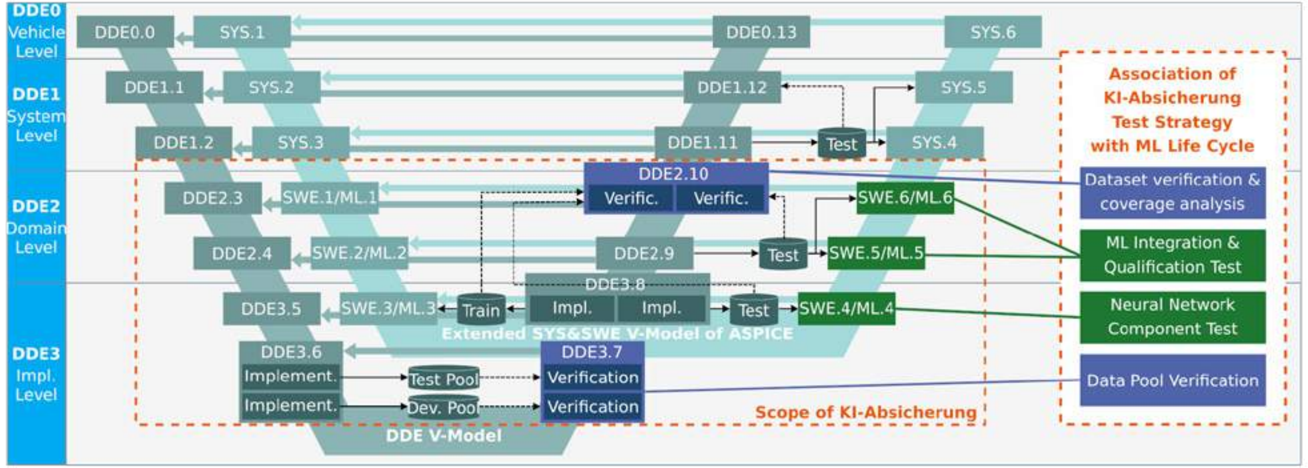
Overview of the Overall Safety Argumentation (© Automotive Safety Technologies | Valeo Schalter und Sensoren GmbH)

Proposal for a ML Test Strategy

Thomas Stauner, BMW Group
Andreas Albrecht, Robert Bosch GmbH

The ML test strategy of KI Absicherung consists of a set of recommendations of methods to be used for testing of DNN-based object detection functions. It is specified relative to the ML-LifeCycle and also addresses verification activities for the dataset.
The activities, their objectives and associated methods are described in the poster.
For most method classes several concrete methods were developed in the project.

AP4.5



Association of the test strategy with the ML-LifeCycle (© Robert Bosch GmbH)

Acknowledgments

Dear team members,

We are proud to say that we have succeeded in working very closely together and in a spirit of trust across all parties in the KI Absicherung funding project. We were fortunate to work with a management team that has shown deep commitment and professional competence in driving the project forward within the working teams. We would like to thank all the participants, the EICT team and in particular the persons mentioned here by name for all the good times and the shared enthusiasm that has always been in the foreground. We look back with joy and gratitude on the time with you and the many substantive discussions that made our success possible in the end - Thank you.

Dr. Stephan Scholz
Project coordinator

PD Dr. Michael Mock
Scientific coordinator and
consortium Co-Lead

TP / AP LEADS & CO-LEADS

TP1	**Fridolin Bauer** | Lead | **BMW Group**
AP1.1	Dr. Nicolas Gay | Lead | **Luxoft** M.Sc. Korbinian Hagn | Co-Lead | **Intel Corporation**
AP1.2	M.Sc. Simon Heming | Lead | **Robert Bosch GmbH** Philipp Heidenreich | Co-Lead | **Opel Automobile GmbH**
AP1.3	M.Sc. Korbinian Hagn | Lead | **Intel Corporation** M.Sc. Bharat Shinde | Co-Lead | **Valeo Schalter und Sensoren GmbH**
AP1.4	Martin Simon | Lead | **Valeo Schalter und Sensoren GmbH** Nikolas Brasch | Co-Lead | **TUM**
AP1.5	M.Sc. Michael Fürst | Lead | **DFKI** Andreas Blattmann| Co-Lead | **Universität Heidelberg**
TP2	**Dr. Thomas Stauner** | Lead | **BMW Group**
AP2.1	Dr. Johannes Guenther | Lead | **Intel Corporation**
AP2.2	M.Sc. Bastian Knerr | Lead | **QualityMinds GmbH**
AP2.3	Dr. Thomas Stauner | Lead | **BMW Group** M.Sc. Karl Amende | Co-Lead | **Valeo Schalter und Sensoren GmbH**
AP2.4	M.Eng. Timo Sämann | Lead | **Valeo Schalter und Sensoren GmbH** Ulrich Seger | Lead | **Robert Bosch GmbH**
AP2.5	Dr. Markus Huber | Lead | **Mackevision Medien Design GmbH**
TP3	**Dr. Fabian Hüger** | Lead | **Volkswagen AG**
AP3.1	PD Dr. Michael Mock | Lead | **Fraunhofer IAIS**

AP3.2	Dr. Christian Hellert | Lead | **Continental AG** Prof. Dr. Hanno Gottschalk | Co-Lead | **University of Wuppertal**
AP3.3	M.Sc. Jan David Schneider | Lead | **Volkswagen AG**
AP3.4	M.Eng. Timo Sämann | Lead | **Valeo Schalter und Sensoren GmbH** Jonas Schneider | Co-Lead | **EFS GmbH**
AP3.5	Jonas Schneider | Lead | **EFS GmbH** Sebastian Gerres | Co-Lead | **Merantix Momentum GmbH**
AP3.6	M.Sc. Alexander Hirsch | Lead | **Robert Bosch GmbH** M.Sc. Svetlana Pavlitskaya | Co-Lead | **FZI**
TP4	**Dipl.-Ing. Frédérik Blank** | Lead | **Robert Bosch GmbH** **Dipl.-Ing. Andreas Rohatschek** | Co-Lead | **Robert Bosch GmbH**
AP4.1	Christian Witt | Lead | **Valeo Schalter und Sensoren GmbH** Dipl.-Ing. Martin Herrmann | Co-Lead | **Robert Bosch GmbH**
AP4.2	M.Sc Christian Pfister | Lead | **ASTech GmbH** Dr. Martin Schels | Co-Lead | **Continental AG**
AP4.3	Dipl.-Ing. Andreas Rohatschek | Lead | **Robert Bosch GmbH** Thomas Schulik | Co-Lead | **ZF Friedrichshafen AG**
AP4.4	Thomas Schulik | Lead | **ZF Friedrichshafen AG** Dr. Maram Akila | Co-Lead | **Fraunhofer IAIS**
AP4.5	Dr. Thomas Stauner | Lead | **BMW Group** Dr. Andreas Albrecht | Co-Lead | **Robert Bosch GmbH**
TP5	**Dr. Stephan Scholz** | Lead | **Volkswagen AG** **PD Dr. Michael Mock** | Co-Lead | **Fraunhofer IAIS**
AP5.1	Dr. Stephan Scholz | Lead | **Volkswagen AG**
AP5.2	Dr. Stephan Scholz | Lead | **Volkswagen AG**
AP5.3	Dr. Stephan Scholz | Lead | **Volkswagen AG** Dipl.-Ing. Frédérik Blank | Co-Lead | **Robert Bosch GmbH**

PROCESS OWNER

P1	**Dipl.-Ing. Frédérik Blank** | Lead | **Robert Bosch GmbH** **Thomas Schulik** | Co-Lead | **ZF Friedrichshafen AG**
P2	**M.Eng. Timo Sämann** | Lead | **Valeo Schalter und Sensoren GmbH**
P3	**M.Sc Christian Pfister** | Lead | **ASTech GmbH**
P4	**Dr. Christian Hellert** | Lead | **Continental AG**
P5	**Dr. Oliver Grau** | Lead | **Intel Corporation**

EVIDENCE WORKSTREAM ORGANIZER

EWS1	**Prof. Dr. Hanno Gottschalk** | Lead | **University of Wuppertal**
EWS2	**Thomas Schulik** | Lead | **ZF Friedrichshafen AG**
EWS3	**Dr. Martin Schels** | Lead | **Continental AG**
EWS4	**Dr. Oliver Grau** | Lead | **Intel Corporation**
EWS8	**Dr. Niels Heller** | Lead | **QualityMinds GmbH**
EWS9	**Thomas Schulik** | Lead | **ZF Friedrichshafen AG**

Table of contents

Contact & further information

Project coordination

Dr. Stephan Scholz
Volkswagen AG I Autonomous Driving
Brieffach 011/1799/1
38436 Wolfsburg, Germany
ki-absicherung-konsortialfuehrung@eict.de

Scientific coordinator and consortium Co-Lead

PD Dr. Michael Mock
Fraunhofer Institute for Intelligent
Analysis and Information Systems (IAIS)
53754 Sankt Augustin, Germany
ki-absicherung-konsortialfuehrung@eict.de

Project management

European Center for Information and
Communication Technologies – EICT GmbH
EUREF Campus Haus 13
Torgauer Straße 12-15
10829 Berlin, Germany

Venue & event day

DRIVE. Volkswagen Group Forum | Berlin, Germany
June 23rd 2022 | 9:00am - 6:00pm

Poster download

www.ki-absicherung-projekt.de/poster
(many of the posters shown in the booklet are avaiable)

Project consortium

VOLKSWAGEN AKTIENGESELLSCHAFT · BMW GROUP · MINI · OPEL · BOSCH · Continental · Valeo · ZF
EFS · ASTech Automotive Safety Technologies · intel. · luxoft A DXC Technology Company · MACKEVISION Part of Accenture Interactive · MERANTIX·LABS · qualityminds · umlaut
DFKI Deutsches Forschungszentrum für Künstliche Intelligenz GmbH · DLR · Fraunhofer IAIS · Fraunhofer IKS · FZI · BERGISCHE UNIVERSITÄT WUPPERTAL · Technische Universität München TUM · UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

External technology partners

BIT TECHNOLOGY SOLUTIONS · neurocat · understand.ai · eict

Note on the legal form of the cooperation
The cooperation between the partners within the project has no independent legal personality. In fact a scientific exchange is conducted between the research centers, organizations and universities listed as cooperation partners. Thereby, no legal entity is established under company law, nor does this scientific cooperation constitute any form of association or similar entity. None of the cooperation partners is entitled to represent individuals, a defined group of cooperating partners nor the entirety of the cooperation partners towards third parties.