Towards Safe AI for Automated Driving

Fabian Hüger, Volkswagen & CARIAD

CSCS 2021 (online), November 30, 2021

Agenda

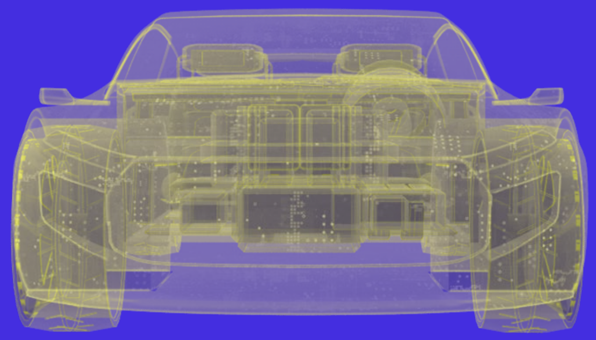
The car needs to be rethought

Connectivity

Making cars and mobility part of our customers' digital life.

Software driven

Rethinking the car from a software perspective, turning it into an intelligent companion.



Electrification

Making mobility more sustainable.

Autonomous driving

Making cars safer and more comfortable for everyone.

CARIAD is here to make automotive mobility safer, more sustainable, and more comfortable.

Comfort

From enjoying the ride to enjoying digital life in your car – everything will become easier, more convenient, and more fun to use.

Safety

Automated and assisted driving will be much safer than any human at the steering wheel.

Sustainability

Continuous software updates keep our cars fresh for many years. Our smart navigation features save kilometers and resources, while reducing congestion.

Our software platform delivers it all.

One software platform. Lots of benefits.

Updatability

Constant and efficient updatability enables attractive vehicles and the best, always fresh customer experiences.

Speed

The seamless software platform and intelligent data analysis speed up development and time to market.

Scalability

The digital platform suits any car model – from entry-level to top-end. Applications can easily be customized.

Simplicity

One unified platform reduces complexitiy – and less hardware reduces costs and weight.

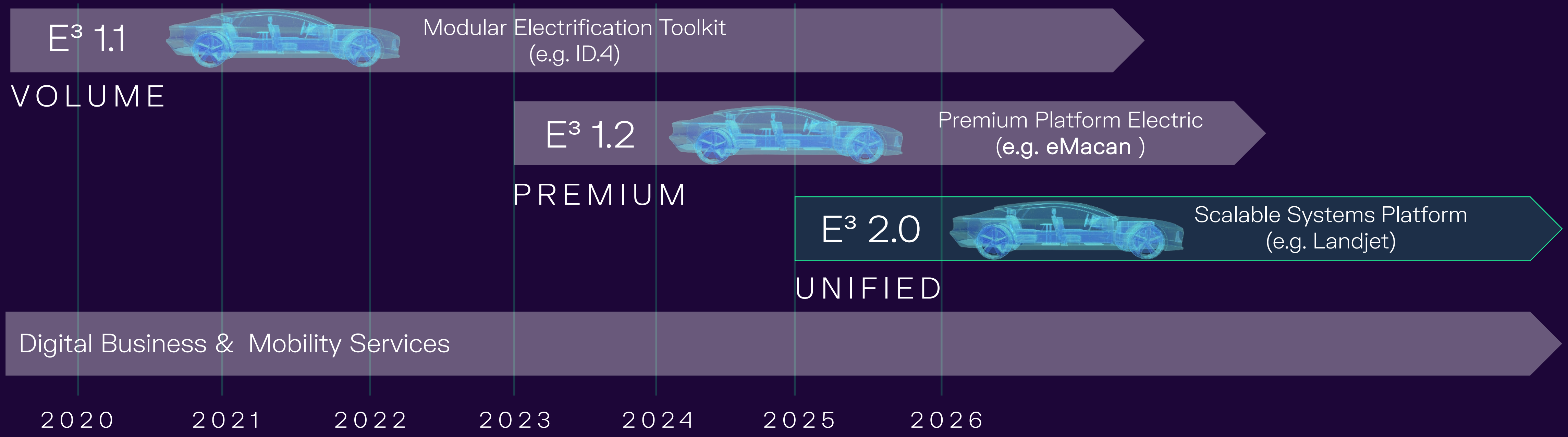
Customer orientation

Data-oriented development helps us to learn from and react to customers' needs and desires.

New revenue streams

Car brands can generate new digital business models– from after sales to monetizing data or third-party apps.

Our platforms E^3 1.1 and E^3 1.2 are technological front runners, while E^3 2.0 will be the one platform in the Group starting 2025.



By 2030 up to 40 million vehicles running on CARIAD software platforms

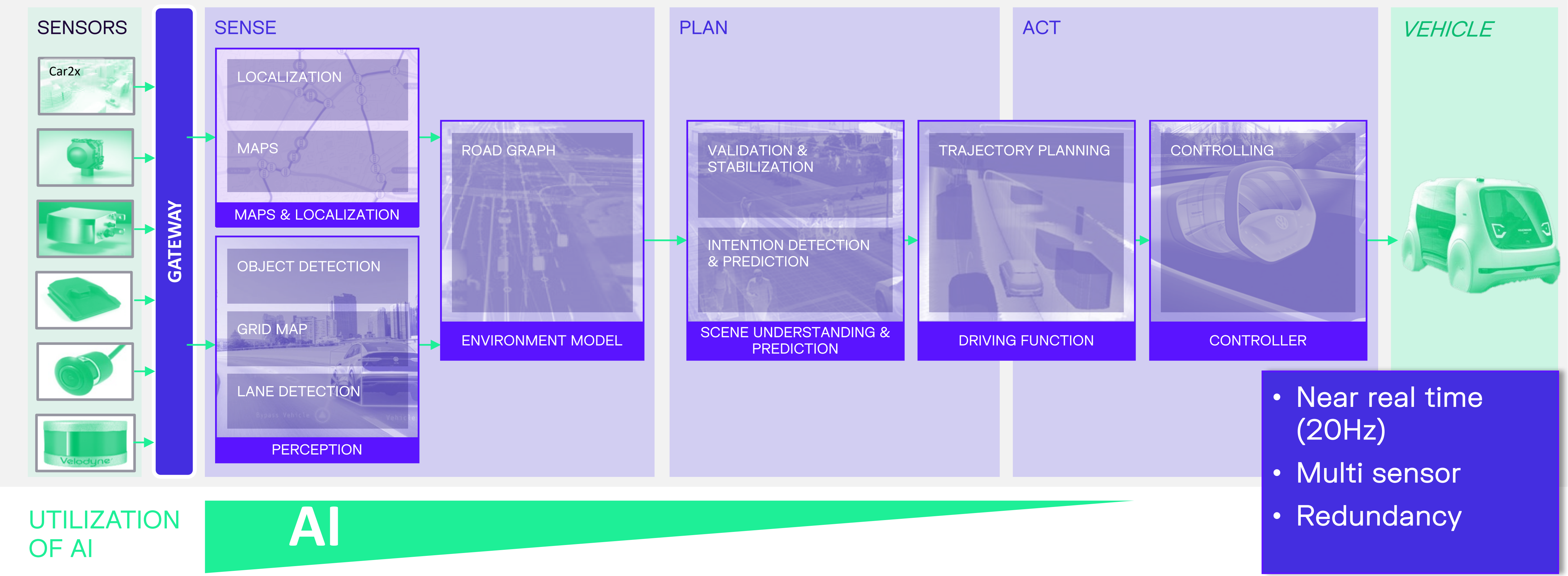
E^3 - End-to-End Electronic Architecture

Agenda

1. Introduction – CARIAD
2. DNNs and Safety in Automated Driving
3. KI-Absicherung Project & Approach
4. Consequences

Automated Driving and AI

Processing chain of autonomous driving & the use of AI along



- Near real time (20Hz)
- Multi sensor
- Redundancy

Arguing Safety in Automated Driving Systems
AI goes safety critical

CENTRAL CHALLENGE

SAFETY

(FuSa + SOTIF)

Central Challenge in bringing highly automated driving on the road.

Argument on safe functioning needed to allow for acceptance & road permission

COMPLEXITY DRIVERS

Mere driving will not suffice to plausibilize safety – particularly challenging with respect to software updates over time. "Black-Box" approach seems impracticable

Handling complexity of the driving environment – open world, unknown unknowns, etc.

Need for continual safety monitoring & assurance – continuous monitoring

Agenda

Acknowledgement: The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Energy within the project "Methoden und Maßnahmen zur Absicherung von KI basierten Wahrnehmungsfunktionen für das automatisierte Fahren (KI-Absicherung)". The authors would like to thank the consortium for the successful cooperation.

KI ABSICHERUNG

Safe AI for Automated Driving

KI-Absicherung Project & Approach

The results, opinions and conclusions expressed in this publication are not necessarily those of Volkswagen Aktiengesellschaft.

Gefördert durch:

Bundesministerium für Wirtschaft und Energie

aufgrund eines Beschlusses des Deutschen Bundestages

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


Making the safety of AI-based function modules for highly automated driving verifiable
KI ABSICHERUNG
Safe AI for Automated Driving
Pedestrian detection

Challenge

AI Land

Promising new technology with unimagined possibilities

Established safety processes cannot be applied

KI ABSICHERUNG

Safe AI for Automated Driving

Safety Land

Safe, trustworthy driving function

Industry consensus (Safe AI): Methodology for joint safety argumentation

Our Team: Experts from AI, Safety and Virtual Reality

OEMs

- VOLKSWAGEN AKTIENGESELLSCHAFT
- BMW GROUP

Tiers

- BOSCH Invented for life
- Continental
- Valeo
- ZF
- EFS

Technology Provider

- ASTech Automotive Safety Technologies
- intel
- luxoft A DXC Technology Company
- MACKEVISION
- MERANTIX
- qualityminds

Research Institutes

- DFKI Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
- DLR
- Fraunhofer IAIS
- Fraunhofer IKS
- FZI

Universitys

- Bergische Universität Wuppertal
- Technische Universität München
- Universität Heidelberg

External Partners

- BIT Technology Solutions
- neurocat
- understand.ai
- ejct

Consortium Lead: Volkswagen AG

Co - Lead: Fraunhofer IAIS

Budget: 41 Mil. €

BMWi Fundinng: 19.2 Mil. €

24 Partners

Duration: 36 month

01.07.2019 - 20.06.2022

Gefördert durch:

Bundesministerium für Wirtschaft und Energie

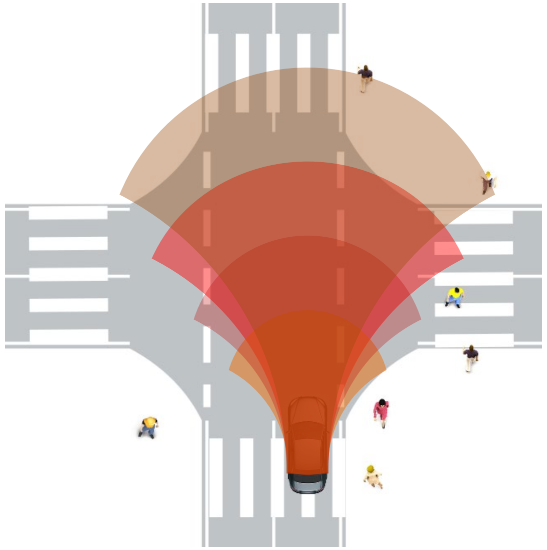
aufgrund eines Beschlusses des Deutschen Bundestages

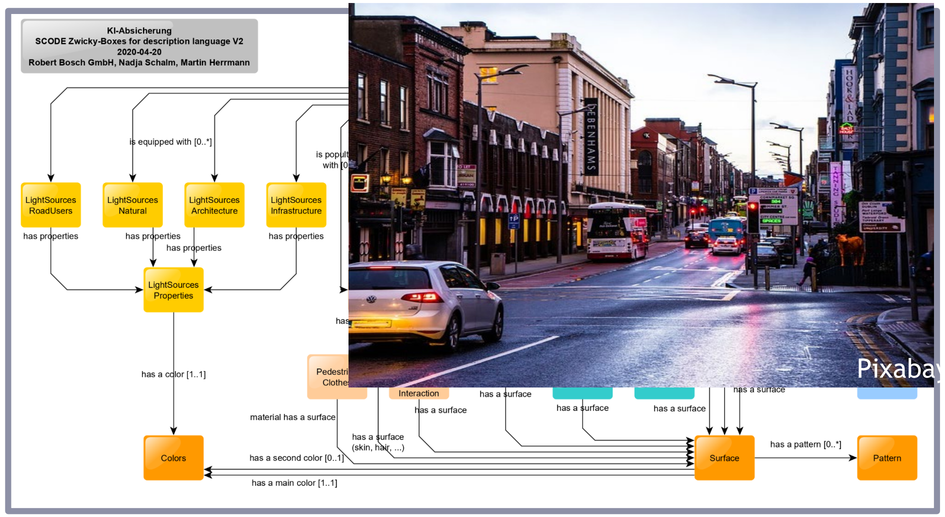
Our Approach: Specification

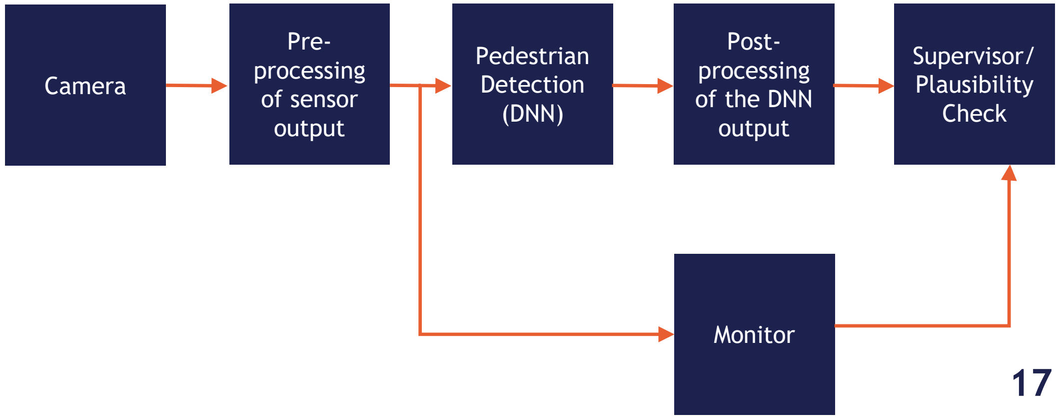
Function

ODD

System Architecture





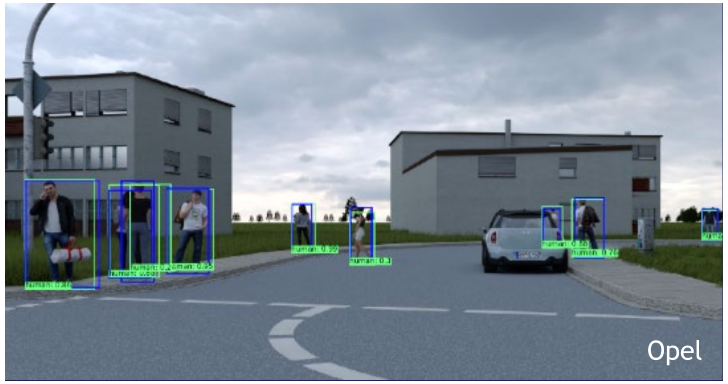


Our Approach: AI Function Pedestrian detection

- Function
- ODD
- System Architecture
- **AI Function**

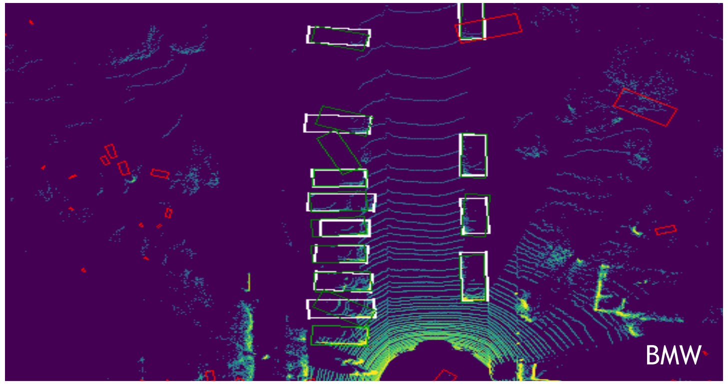
Semantic Segmentation

2D Bounding Box Detection



Instance Segmentation

3D Bounding Box Detection

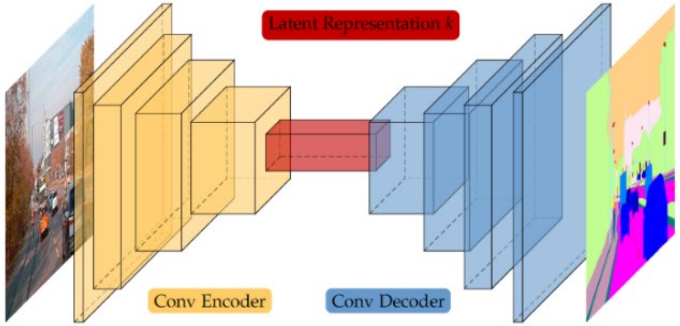


3D Pose estimation

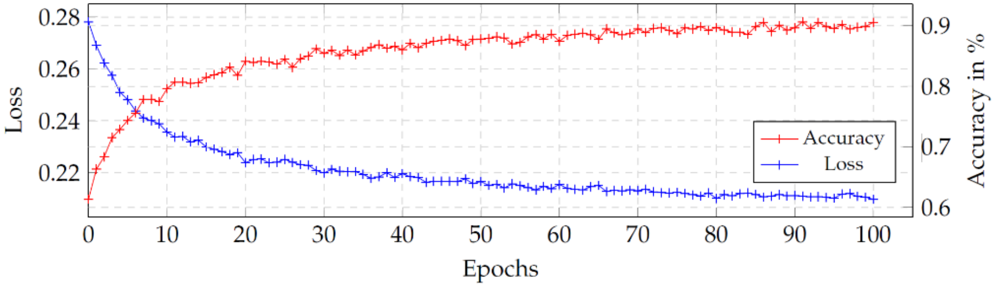
Our Approach: Synthetic Data and ML-Lifecycle

ML LIFECYCLE

- ML-Data
- Model design
- Pre-Processing
- Training
- Post-Processing
- Test/ V&V incl. Data
- Monitoring
- Maintenance



Volkswagen AG



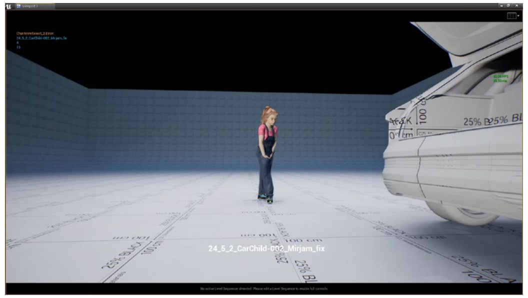
Volkswagen AG

Our Approach: ML-Lifecycle-Validation data

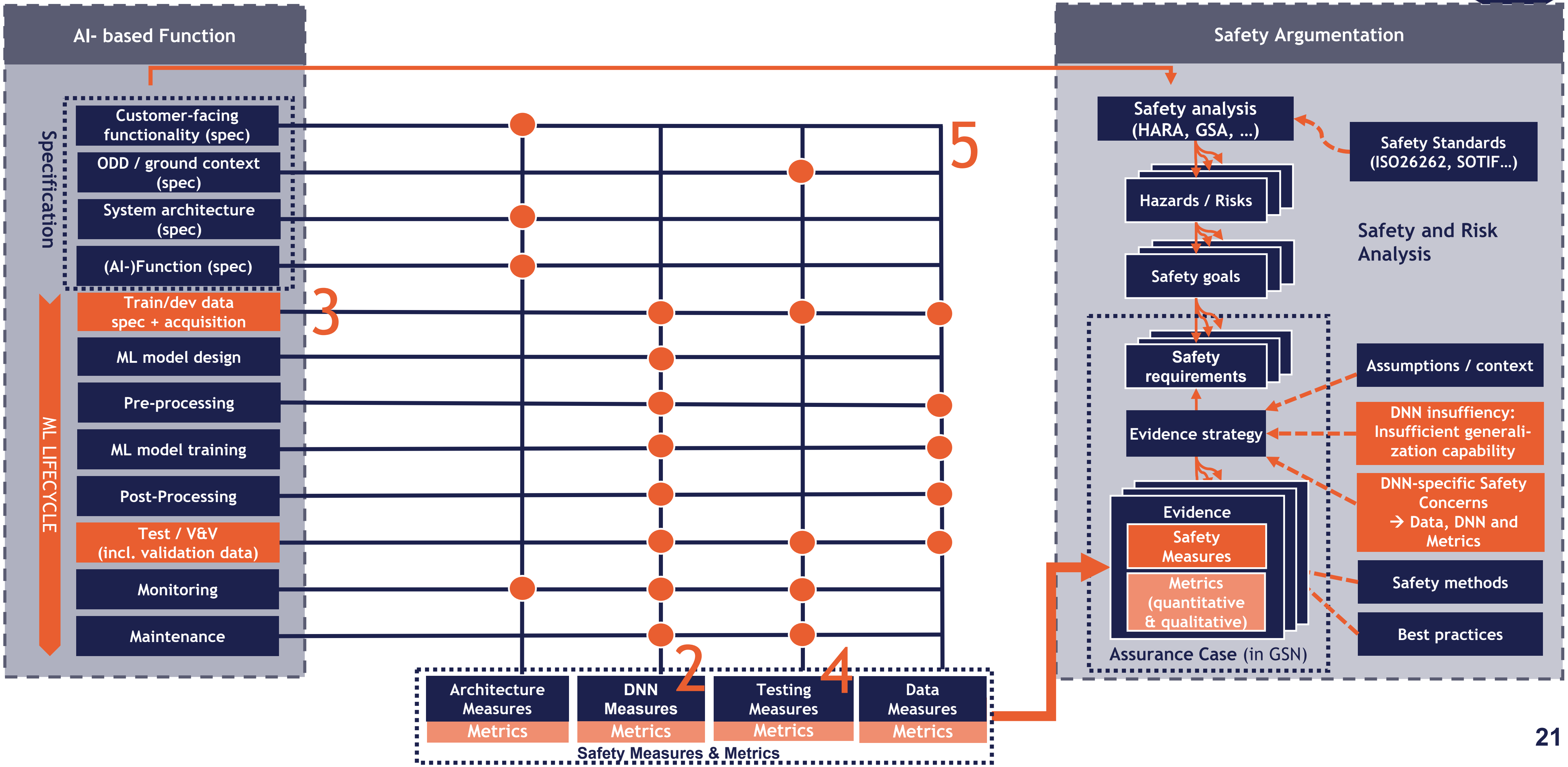
ML LIFECYCLE

- ML-Data
- Model design
- Pre-Processing
- Training
- Post-Processing
- **Test/ V&V incl. Daten**
- Monitoring
- Maintenance

Continuous process for identification, specification and generation of synthetic data



Our Approach: Big picture



1

DNN-specific safety concerns

Our Approach: DNN-specific Safety Concerns (1/2)

We define **DNN-specific Safety Concerns (SCs)** as underlying issues of DNN-based perception which may negatively affect the safety of a system.

e.g., fog, snow, camera issues

Triggering Event

Erroneous Output

- DNN-characteristics-related concerns
- Data-related concerns
- Metric-related concerns

DNN-specific Safety Concerns

Insufficient Generalization Capability

Functional Insufficiency according to ISO/PAS 21448

FI-1 INSUFFICIENT GENERALIZATION CAPABILITY
Wrong outputs by an AI-based function that was trained on a limited database. Erroneous input to output mapping or wrong approximation.

SC-1.1 UNRELIABLE CONFIDENCE INFORMATION
DNNs tend to be overconfident in their predictions under certain conditions or in general outputting unreliable confidence information.

SC-1.2 BRITTLENESS OF DNNS
Non-robustness against common perturbations such as noise or certain weather conditions as well as targeted perturbations known as adversarial examples

SC-1.2.1 LACK OF TEMPORAL STABILITY
Detection results rapidly changing in time whereas little change occurs in the ground truth

SC-1.3 INCOMPREHENSIBLE BEHAVIOUR
Inability to explain exactly how DNNs come to a decision.

SC-1.4 INSUFFICIENT PLAUSIBILITY
AI based functions usually lack basic plausibility checks, which are intended to identify detections of the perception function that violate physical laws.

SC-2.1 DATA DISTRIBUTION IS NOT A GOOD APPROXIMATION OF REAL WORLD
The distribution of data used in the development should be a valid approximation of the ODD in the real world.

SC-2.2 INADEQUATE SEPARATION OF TEST AND TRAINING DATA
Test data might be correlated to training data which might induce overfitting on test data.

SC-2.3 DEPENDENCE ON LABELLING QUALITY
Labelling quality can directly affect the resulting model performance. Moreover, due to missing labelling quality, evaluation results might be misleading.

SC-2.3.1 MISSING LABEL DETAILS OR META-LABELS
Missing meta-labels or label details possibly leads to improper data selection or insufficient training objectives.

SC-2.4 SPECIFICATION OF THE ODD
An incomplete or incorrect ODD specification leads to incomplete data records for training and testing.

SC-2.5 DISTRIBUTIONAL SHIFT OVER TIME
A DNN is trained and tested at a certain point in time. Changes will occur naturally and therefore can potentially harm the performance of DNNs.

SC-2.6 UNKNOWN BEHAVIOUR IN RARE CRITICAL SITUATIONS
The long tail problem describes the fact that there exists an enormous amount of possibly safety-critical street scenes that have a low occurrence probability.

SC-3.1 SAFETY-AWARE METRICS
Some state-of-the-art metrics only evaluate the average performance of DNNs. Safety-aware metrics are required to sophistically evaluate the performance of DNNs.

Based on:

O.Willers, S. Sudholt, S. Raafatnia, S. Abrecht: Safety Concerns and Mitigation Approaches Regarding the Use of Deep Learning in Safety-Critical Perception Tasks

T. Sämann, P.Schlicht, F. Hüger: Strategy to Increase the Safety of a DNN-based Perception for HAD Systems

G. Schwalbe, B. Knie, T. Sämann, T. Dobberphul, L. Gauerhof, S., V. Rocco: **Structuring the Safety Argumentation for Deep Neural Network Based Perception in Automotive Applications**

Functional Insufficiencies

DNN-characteristics-related concerns

Data-related concerns

Metric-related concerns

DNN-specific Safety Concerns

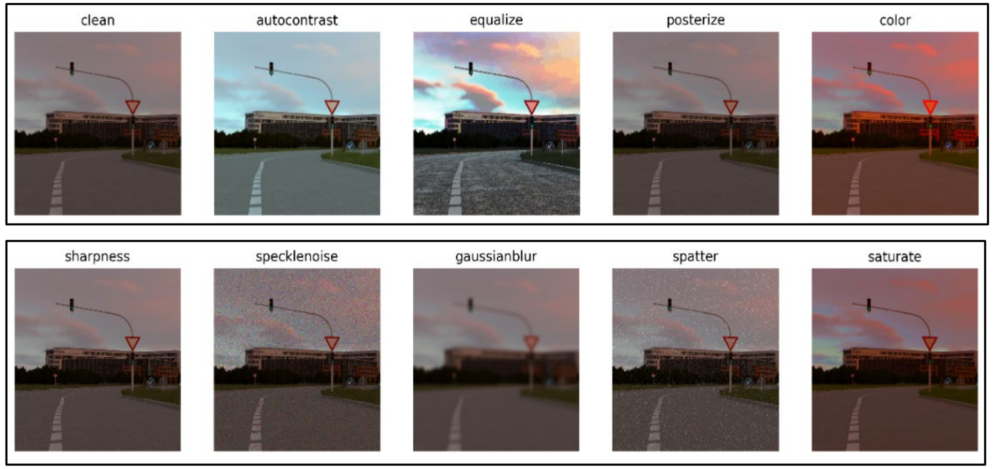
2

Exemplary Measures

Our Approach: Identify, Measure and & Counteract „DNN-specific Safety Concerns”

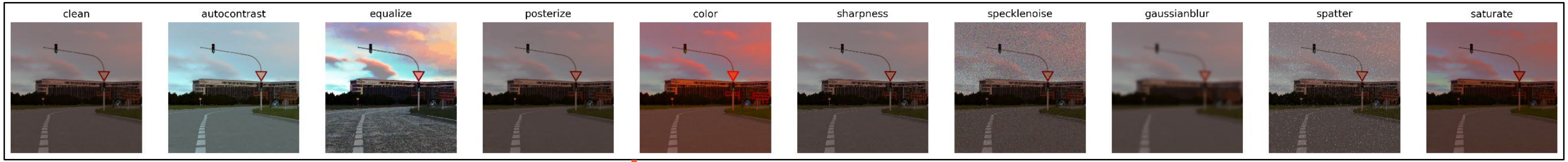
Adressed Safety Concern: Brittleness of DNNs

- Adressing **“Brittleness of DNNs” (Example)**
  - **Requirement:** Robustness = Performance even under reasonable perturbations (gained from ODD definition, data analysis and sensor specs)
  - **Metric:** Performance under corruption
  - **Methods (e.g.)**
    - Augmentation Training (**AugMix**)
    - From a Fourier-Domain Perspective on Adversarial Examples to a **Wiener Filter** Defense for Semantic Segmentation
  - **Evidence**: Effectiveness of measure via metric

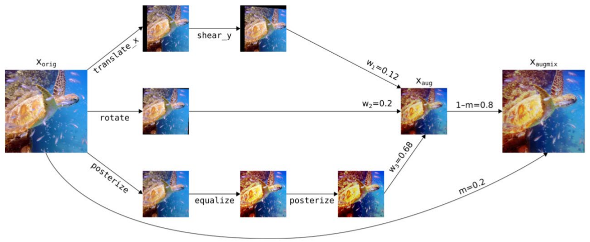


Our Approach: Identify, Measure and & Counteract „DNN-specific Safety Concerns” via AugMix

Adressed Safety Concern: Brittleness of DNNs Corruption Robustness



Combined using AugMix



AUGMIX: A SIMPLE DATA PROCESSING METHOD TO IMPROVE ROBUSTNESS AND UNCERTAINTY

Dan Hendrycks*
DeepMind
hendrycks@berkeley.edu

Norman Mu*
Google
normanmu@google.com

Ekin D. Cubuk
Google
cubuk@google.com

Barret Zoph
Google
barretzoph@google.com

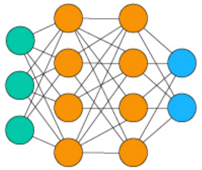
Justin Gilmer
Google
gilmer@google.com

Balaji Lakshminarayanan†
DeepMind
balajiln@google.com

+ Improved robustness
+ Improved generalization
+ Data efficient augmentation strategy

Clean image Augmix image

Training

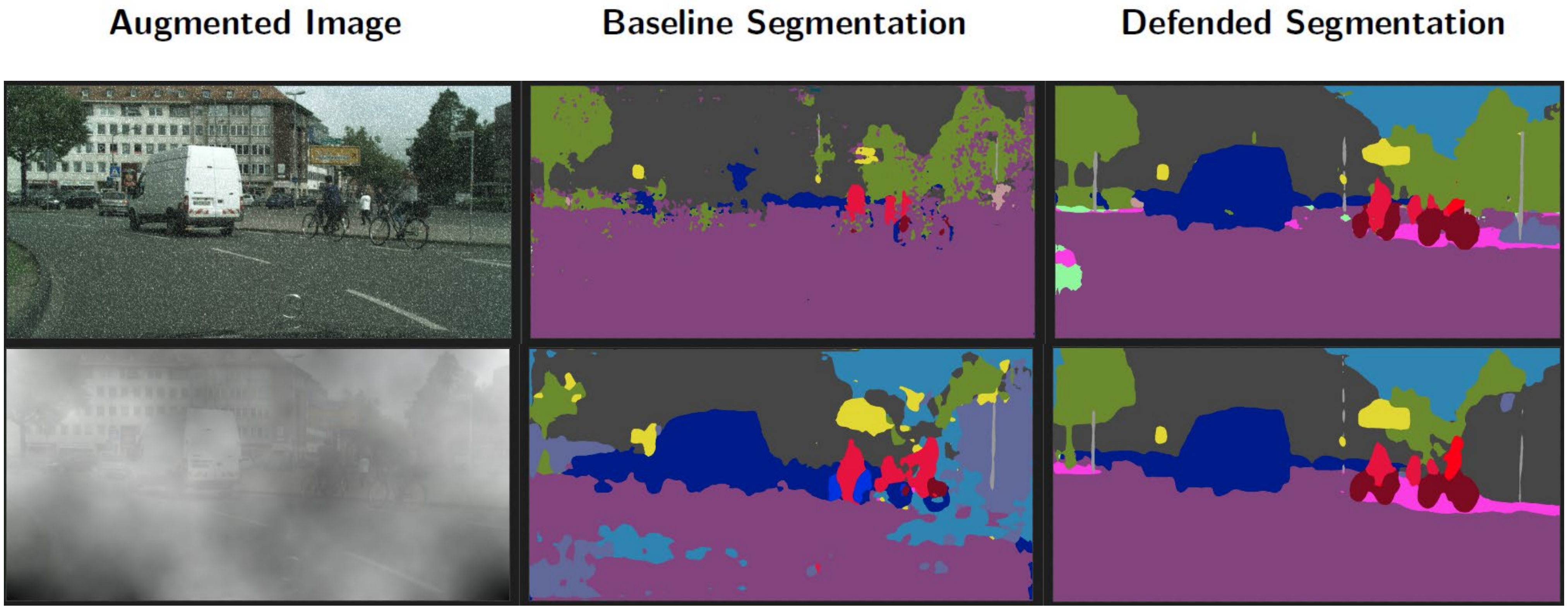


Evaluation on 14 *unseen* „real-world“ corruptions

AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty, D. Hendrycks et al, https://arxiv.org/abs/1912.02781

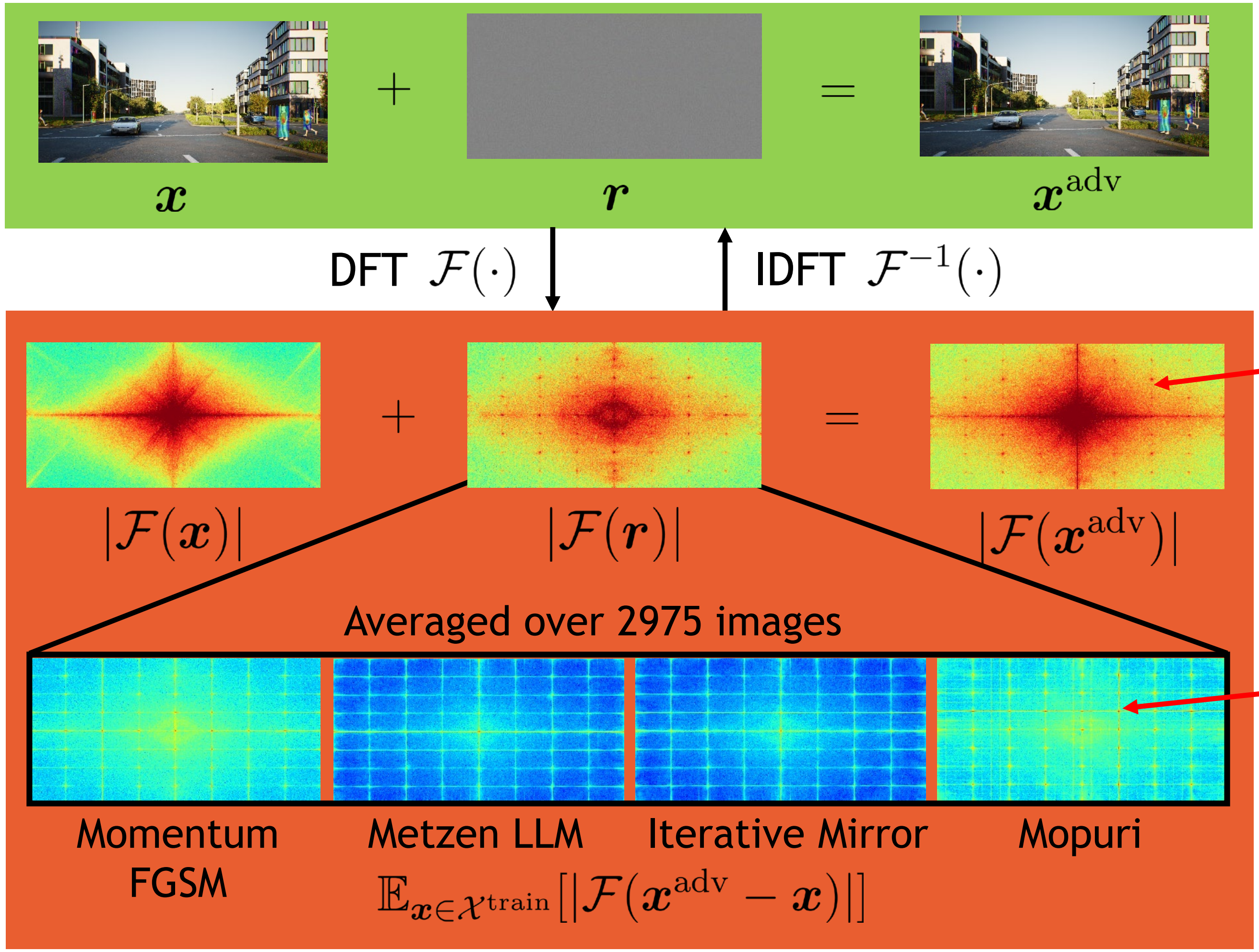
Our Approach: Identify, Measure and & Counteract „DNN-specific Safety Concerns" via AugMix

Adressed Safety Concern:
Brittleness of DNNs
Corruption Robustness



Our Approach: Identify, Measure and & Counteract „DNN-specific Safety Concerns"

Adressed Safety Concern:
Brittleness of DNNs
Adversarial Attacks



Adversarial examples are imperceptible in the spatial domain

Strong visible artifacts in the frequency domain

These artifacts are image-type and attack-type independent

Spatial domain

Frequency domain

From a Fourier-Domain Perspective on Adversarial Examples to a Wiener Filter Defense for Semantic Segmentation, N. Kapoor et al. *https://arxiv.org/abs/2012.01558*

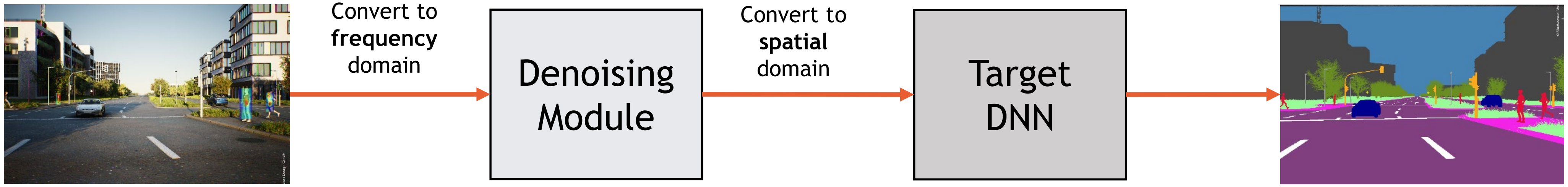
Our Approach: Identify, Measure and & Counteract „DNN-specific Safety Concerns" via Wiener Filters

Adressed Safety Concern:
Brittleness of DNNs
Adversarial Attacks

Wiener Filters (WF) as an online denoising module

Steps:

1. Convert input image to DFT domain.
2. Apply pre-computed WF as a multiplicative filter.
3. Convert to spatial domain using IDFT.
4. Feed image to target DNN.



Attacked or clean image | Input Pre-processor | Perception DNN | Robust Segmentation

From a Fourier-Domain Perspective on Adversarial Examples to a Wiener Filter Defense for Semantic Segmentation, N. Kapoor et al. *https://arxiv.org/abs/2012.01558*

Our Approach: Explore Mechanisms!

- Heatmap-based Attention Consistency Validation
- Mixture of Experts
- Domain Randomization in Optimized Dataset Selection
- MC Dropout
- Uncertainties For Anomaly Detection
- Hybrid Learning using Concept Enforcement
- Active Learning
- Adverserial Training
- Hybrid and robustness-focussed Compression
- …

Approx 80 Mechanisms are developed and evaluated

Inspect, Understand, Overcome: A Survey of Practical Methods for AI Safety

Sebastian Houben[1], Stephanie Abrecht[2], Maram Akila[1], Andreas Bär[15], Felix Brockherde[10], Patrick Feifel[8], Tim Fingscheidt[15], Sujan Sai Gannamaneni[1], Seyed Eghbal Ghobadi[8], Ahmed Hammam[8], Anselm Haselhoff[9], Felix Hauser[11], Christian Heinzemann[2], Marco Hoffmann[16], Nikhil Kapoor[7], Falk Kappel[13], Marvin Klingner[15], Jan Kronenberger[9], Fabian Küppers[9], Jonas Löhdefink[15], Michael Mlynarski[16], Michael Mock[1], Firas Mualla[13], Svetlana Pavlitskaya[14], Maximilian Poretschkin[1], Alexander Pohl[16], Varun Ravi-Kumar[4], Julia Rosenzweig[1], Matthias Rottmann[5], Stefan Rüping[1], Timo Sämann[4], Jan David Schneider[7], Elena Schulz[1], Gesina Schwalbe[3], Joachim Sicking[1], Toshika Srivastava[12], Serin Varghese[7], Michael Weber[14], Sebastian Wirkert[6], Tim Wirtz[1], and Matthias Woehrle[2]

[1] *Fraunhofer Institute for Intelligent Analysis and Information Systems*
[2] *Robert Bosch GmbH*
[3] *Continental AG*
[4] *Valeo S.A.*
[5] *University of Wuppertal*
[6] *Bayerische Motorenwerke AG*
[7] *Volkswagen AG*
[8] *Opel Automobile GmbH*
[9] *Hochschule Ruhr West*
[10] *umlaut AG*
[11] *Karlsruhe Institute of Technology*
[12] *Audi AG*
[13] *ZF Friedrichshafen AG*
[14] *FZI Research Center for Information Technology*
[15] *Technische Universität Braunschweig*
[16] *QualityMinds GmbH*

Survey: available at https://www.ki-absicherung-projekt.de/

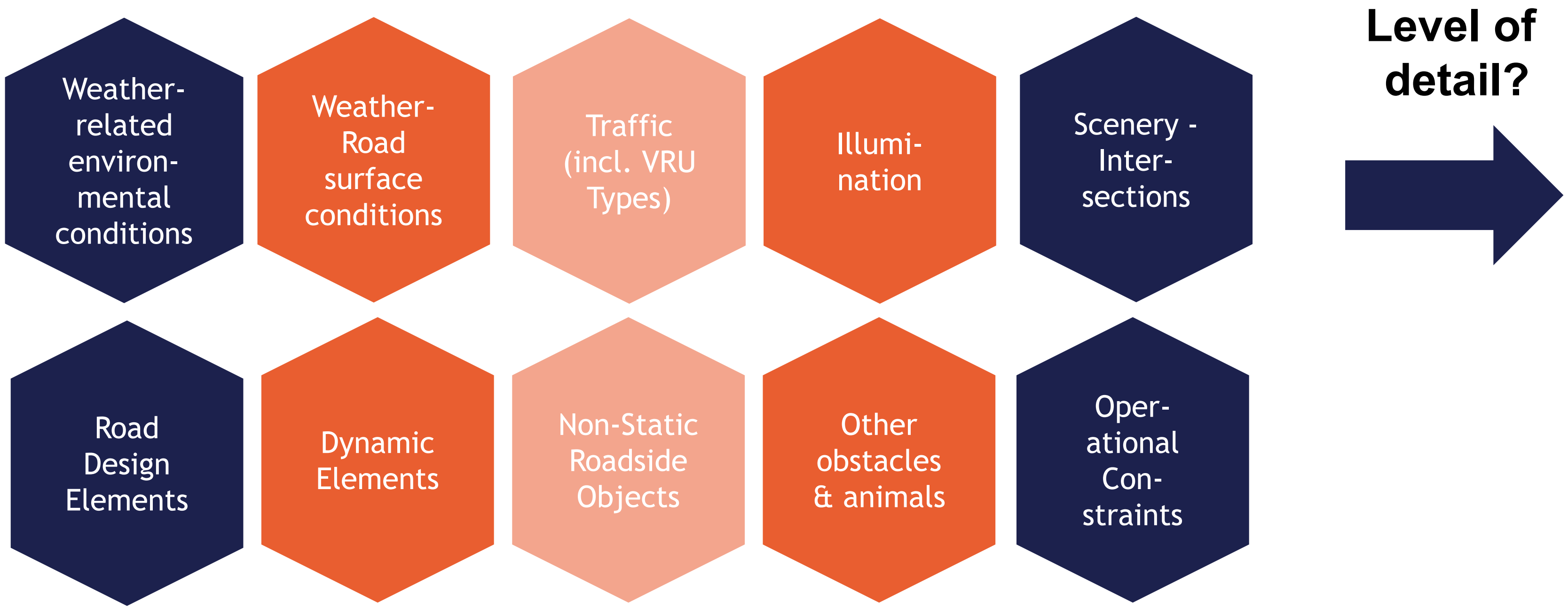
3

Our systematic approach for data

Source: Robert Bosch GmbH, Frederik Blank

Operational design domain (ODD)

- An ODD describes / specifies operating conditions under which a given driving automation system or feature is specifically designed to function […]
  - Taxonomy and Definitions for Terms Related to Driving Automation Systems (examples)



Description language

Target:
DNN-(detection) capability to cover ODD

Source: Robert Bosch GmbH, Frederik Blank

A description language & data input space modeling is needed to…

Complexity of language

Be able to describe / specify operating conditions (and edges of ODD*) as of PAS 1883:2020 and others

Systematically capture important knowledge and describe the (expected) key input space dimensions and their possible variations having an influence on the functional performance of a DNN-based function (→ Zwicky Boxes & Ontology)

Perform training and assurance data coverage estimations for data driven AI-based systems

Describe Corner cases / rare critical situations to be considered in training / test data sets

For synthetic perception data production & meta-data: describe data dimensions that should be variated & incrementally generate new data by analyzing coverage and generating missing combinations

DNN-specific Safety Concerns (examples)

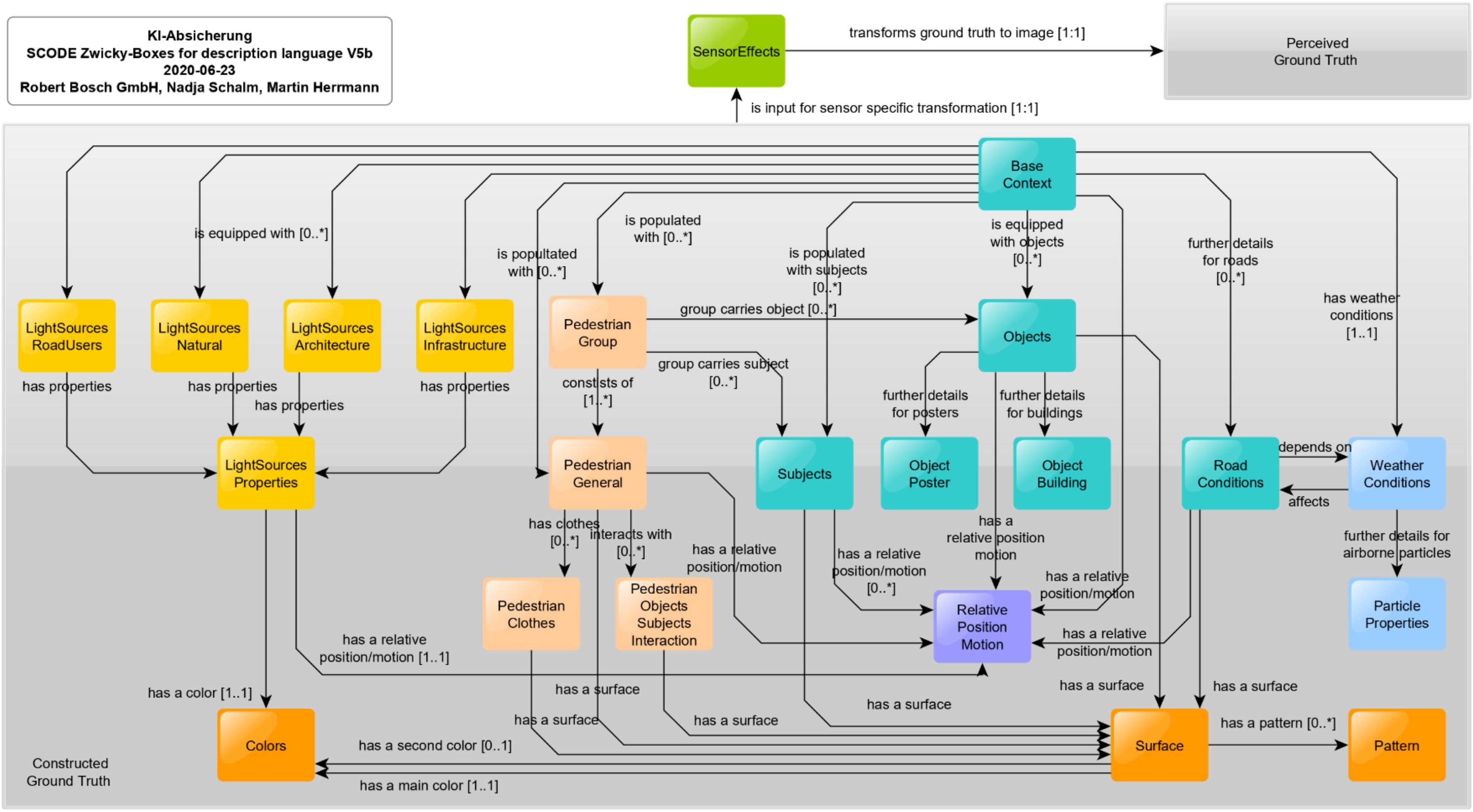
Data distribution is not a good approximation to target domain

Unknown behavior in rare safety-critical situations

Based on: O. Willers, S. Sudholt, S. Raafatnia, S. Abrecht: Safety Concerns and Mitigation Approaches Regarding the Use of Deep Learning in Safety-Critical Perception Tasks

Source: Robert Bosch GmbH, Frederik Blank

High Level view of Ontology / Domain model derived from SCODE Zwicky-Boxes



Total

- ~250 dimensions
- ~1000 alternatives
- Several Sub-domains

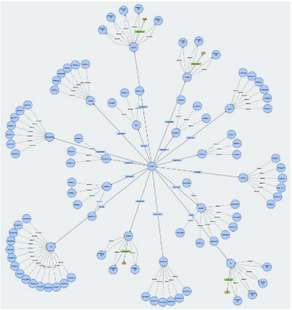
Approach

- Review of public data sources / existing standards
- Brainstorming with experts
- Expert interviews
- Iterative refinement
- Needs to be challenged / extended by identified corner cases

Source: Robert Bosch GmbH, Frederik Blank

Data representations of the data input space aligned to ontology

Ontology Graph (Relations)



Visualization of KI Absicherung pedestrian sub ontology

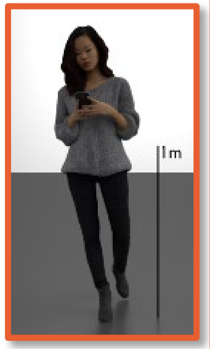
Representations of variations

Daytime	morning	day	evening	night	
Haze/Fog	no	yes			
Street condition	dry	wet	icy	snow	broken
Sky	cloudy	no	clear		
Rain	no	yes			
Reflection on road	no	yes			
Shadow on road	no	yes			
VRU type	adult	child			
VRU pose	pedestrian	jogger	cyclist		
VRU contrast to BG	low	high			

Zwicky Box – Discretized variations of important dimensions (Bosch)

Source: Robert Bosch GmbH, Frederik Blank

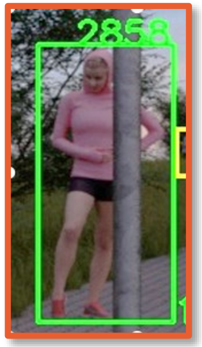
Asset & Object descriptions for data analytics



Pedestrian:Age "adult"
Pedestrian:BodyHeight "160cm-200cm"
Pedestrian:BodyShape "thin"
Pedestrian:BodyType "hourglass"
Pedestrian:FaceShape "oval"
Pedestrian:Gender "female"
Pedestrian:HairColor "black"
Pedestrian:HairLength "long"
Pedestrian:HairStyle "other"
Pedestrian:Pigmentation "medium"
Pedestrian:Pose "walking"
Pedestrian:SkinModification "no"
Pedestrian:SpecialHandicap "no"

Source: BIT-TS

Object GT Annotations for DNN-Training & Testing



Height = 55 px
Width = 10 px
Occlusion_level: 80%
Occluded_body_part: arm
Occluder: lamp
Within_breaking_distance_30kph: true

Source: BIT-TS

Systematic Combination of variations

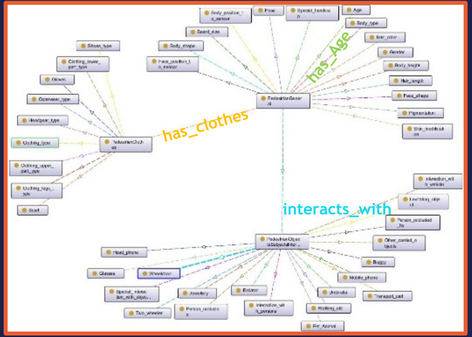
Dimension	Person1	Person2	Person3	...
Age	Child	Teenager	Adult	
Gender	Male	Female	Male	
Body height	80-120 cm	120-160 cm	160-200 cm	
Pose	Running	Lying	Walking	
Pedestrian Location	Middle of street	Left side walk	Right side walk	
...	...	...	...	

Systematically identify and describe the (known / expected) key input space dimensions and their possible variations & combinations having an influence on the functional performance of a DNN-based function

*) GT = Ground Truth

Structured Incremental dataset generation to boost data coverage (Vision)

Domain model / data input space



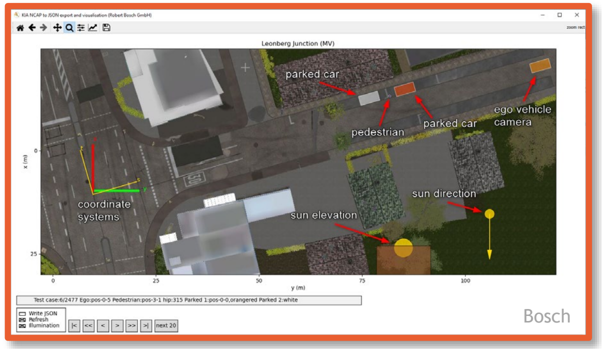
Known Performance Limitations

(expert know-how)

- Person occlusions
- Low contrast: similar color to background
- Uncommon person locations
- Uncommon poses

Possible Variations based on Ontology

Data Request tool



JSON

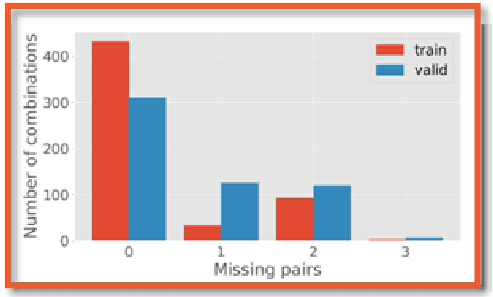
Synth. Data production

Annotations & Meta data

Incremental dataset generation

Missing Combinations

Constrained test-space



Optimized Combinatorial Testing (example)

Data Coverage Analysis

Dimension	Person1	Person2	Person3
Age	Child	Teenager	Adult
Gender	Male	Female	Male
Body height	80-120cm	120-160cm	160-200 cm
Pose	Running	Lying	Walking
Pedestrian Location	Middle of street	Left side walk	Right side walk
...	...	...	...

Test method result as input to Assurance Case (to be combined with other data related evidences)

Source: Robert Bosch GmbH, Frederik Blank

4

NCAP inspired test data production process

Source: ZF, Thomas Schulik

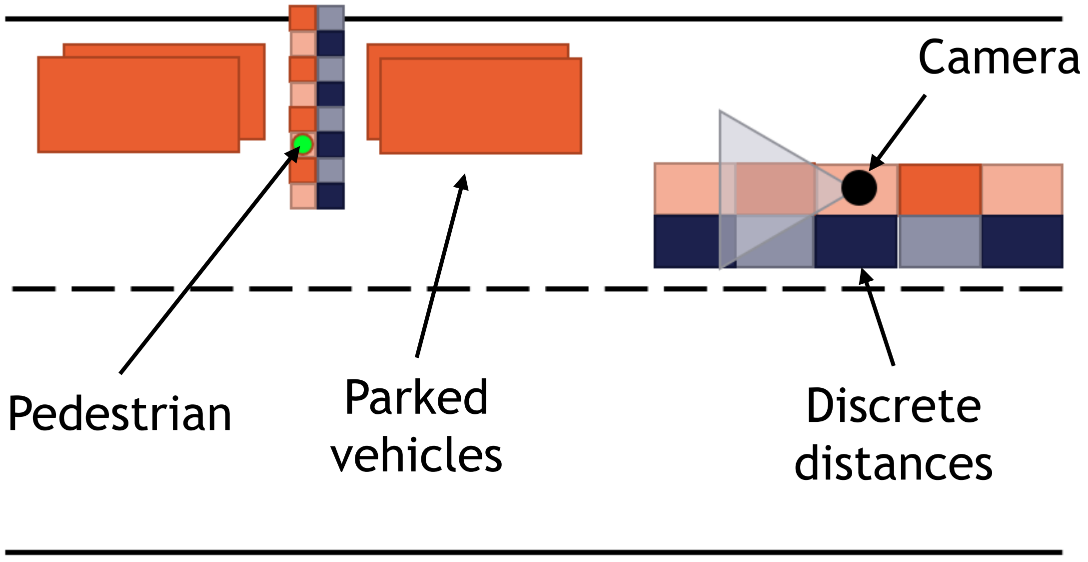
ML-Lifecycle-Validation data

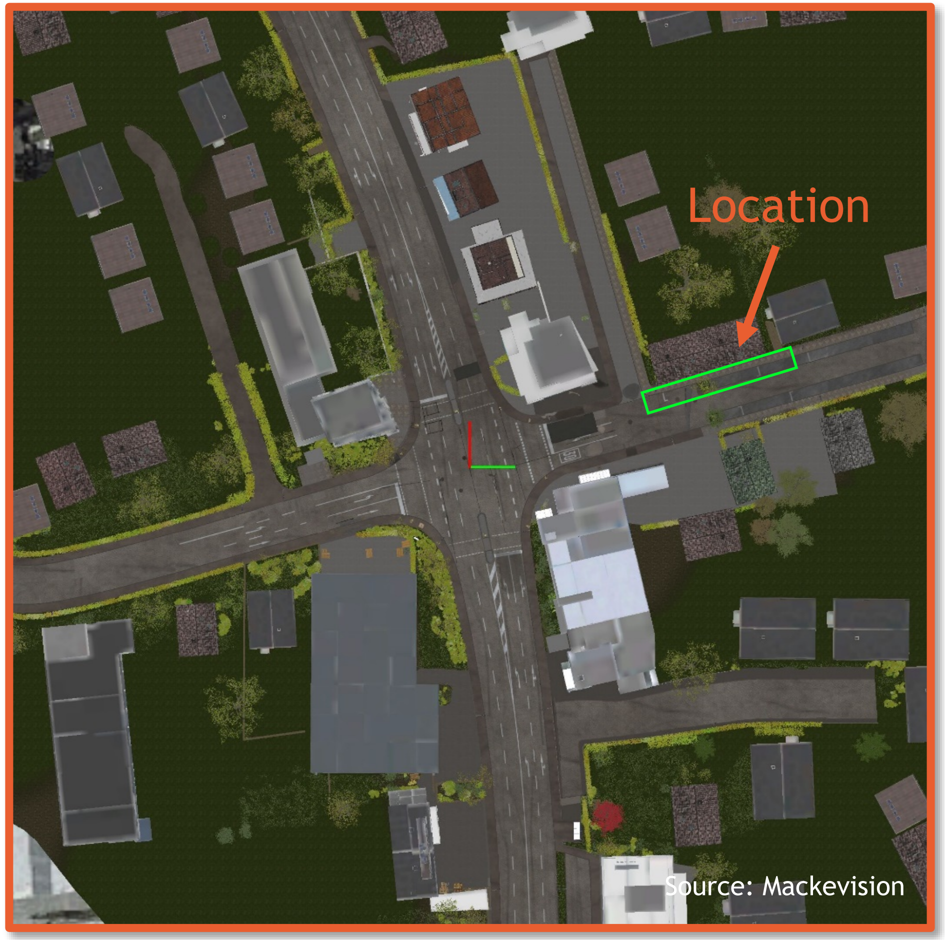
ML LIFECYCLE

- ML-Data
- Model design
- Pre-Processing
- Training
- Post-Processing
- Test/ V&V incl. Daten
- Monitoring
- Maintenance

Definition of base scenario and location on base context

Story
A pedestrian is approaching the ego vehicle between two parking cars under different environment conditions





Source: ZF, Thomas Schulik

Discretization of dimensions in “Zwicky Boxes”

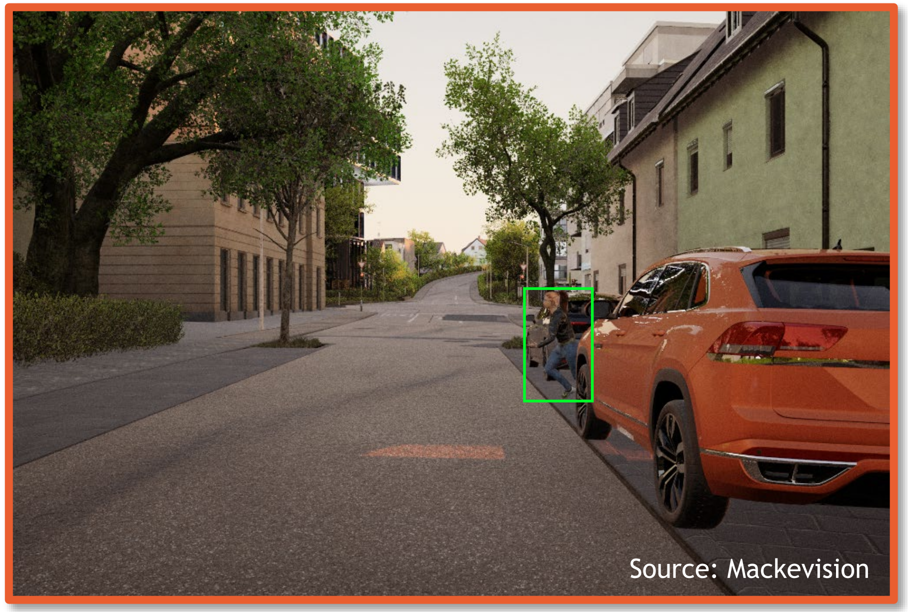
Ego XY position	pos-0-0	pos-0-1	pos-0-2	pos-0-3	pos-0-4	pos-0-5	pos-1-0	pos-1-1	pos-1-2	pos-1-3	pos-1-4	pos-1-5							
Pedestrian XY position	pos-0-0	pos-1-0	pos-2-0	pos-3-0	pos-4-0	pos-5-0	pos-6-0	pos-7-0	pos-0-1	pos-1-1	pos-2-1	pos-3-1	pos-4-1	pos-5-1	pos-6-1	pos-7-1			
Pedestrian pose	pose01	pose02	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	BMW7l	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 Mekana Turquoise	VW Mondsteingrau	VW Scale Silver	VW Stonewashed Blue	VW Energetic Orange	VW Deep Black	VW Delfingrau	VW Kings Red
Parked vehicle 2 type	BMW1	BMW2	BMW7l	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 Mekana Turquoise	VW Mondsteingrau	VW Scale Silver	VW Stonewashed Blue	VW Energetic Orange	VW Deep Black	VW Delfingrau	VW Kings Red
Illumenation	direct sun	diffuse light																	
Sun direction	d0	d45	d90	d135	d180	d225	d270	d315											
Sun elevation	low	medium	day																
Road surface	A	B	C	D															

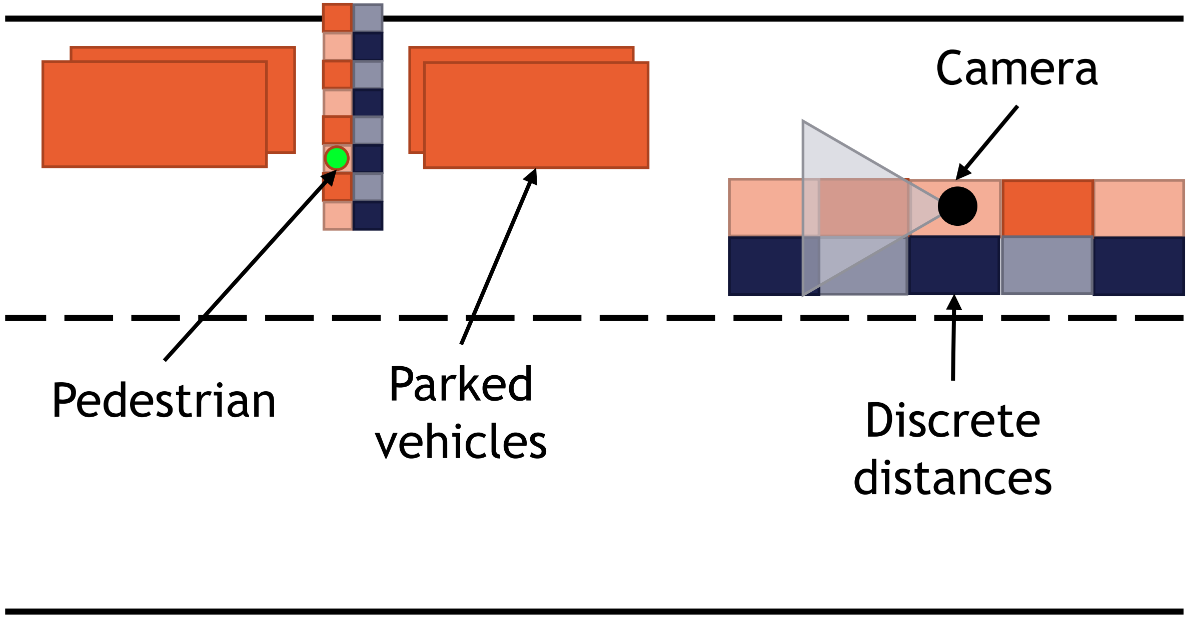
Source: Robert Bosch GmbH

- **Discretization:** The most critical dimensions are identified and discretized
- **Test coverage:** With pairwise testing it’s possible to achieve a high error coverage in traditional software testing

Source: ZF, Thomas Schulik

Data production - Example data snapshot 1



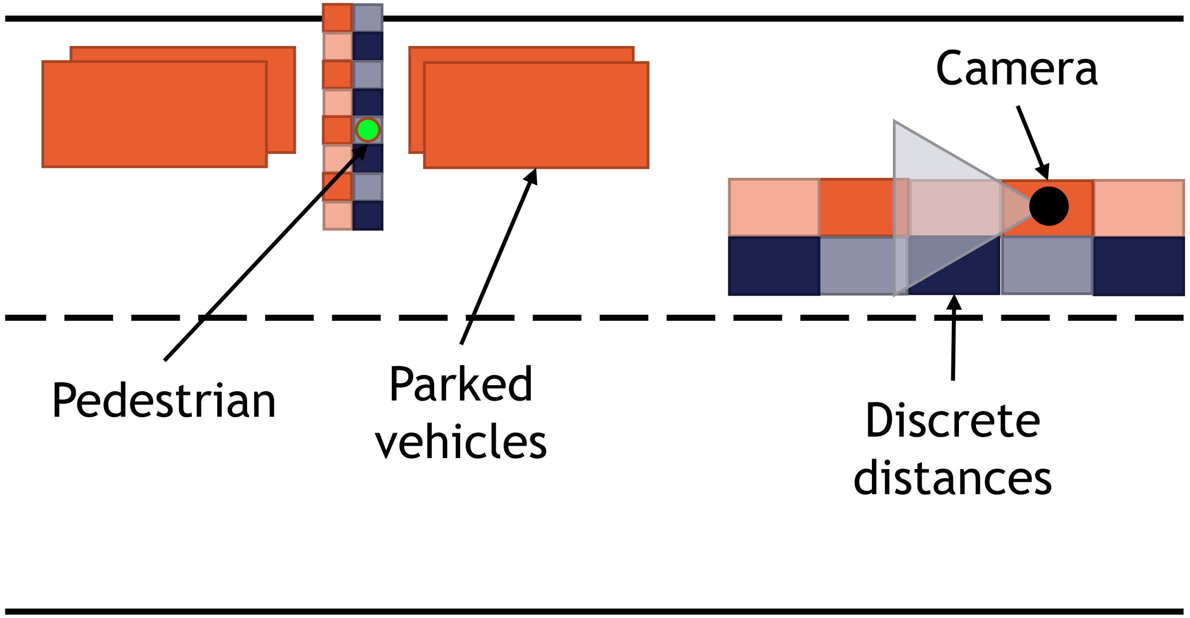


- **Safety critical:** Pedestrian has a running pose towards the camera
- The perception function shall be able to detect the pedestrian early enough without any image perturbations

- Those images are well suited as a reference for the analysis of brittleness in DNN's

Source: ZF, Thomas Schulik

Data production - Example data snapshot 1



- **Safety critical:** The legs are extended to the driving lane
- **Uncommon pose:** Pedestrian lays between two vehicles and is difficult to see

- In which combinations is the object detector **not** capable to perceive the pedestrian?

Source: ZF, Thomas Schulik

Examples for data post processing

original

fog

frost

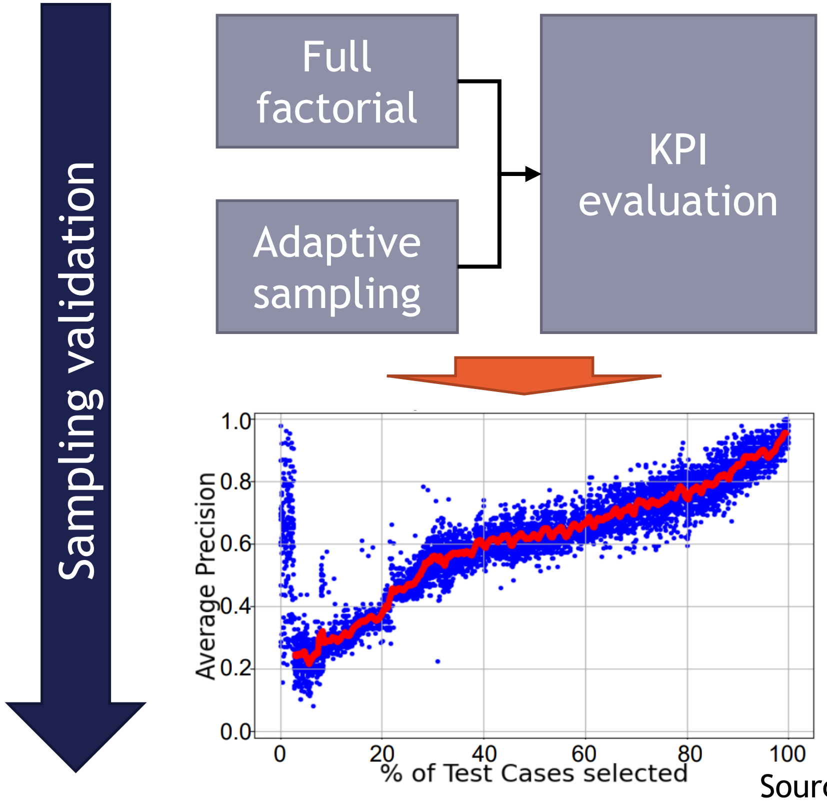
brightness

contrast

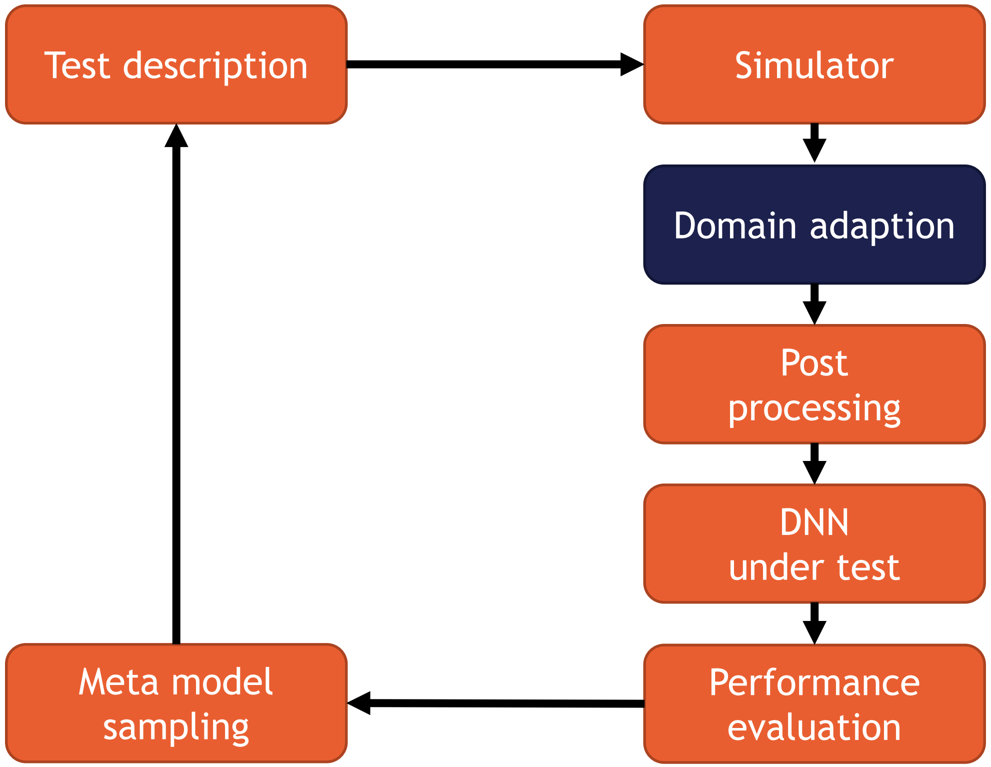
Motion blur

Source: ZF, Thomas Schulik

Test space exploration optimization



Source: ZF



The most performance critical test cases are identified early in the test exploration
"Adaptive test case selection for DNN-based perception functions"
Paper release: https://ieeexplore.ieee.org/document/9582499

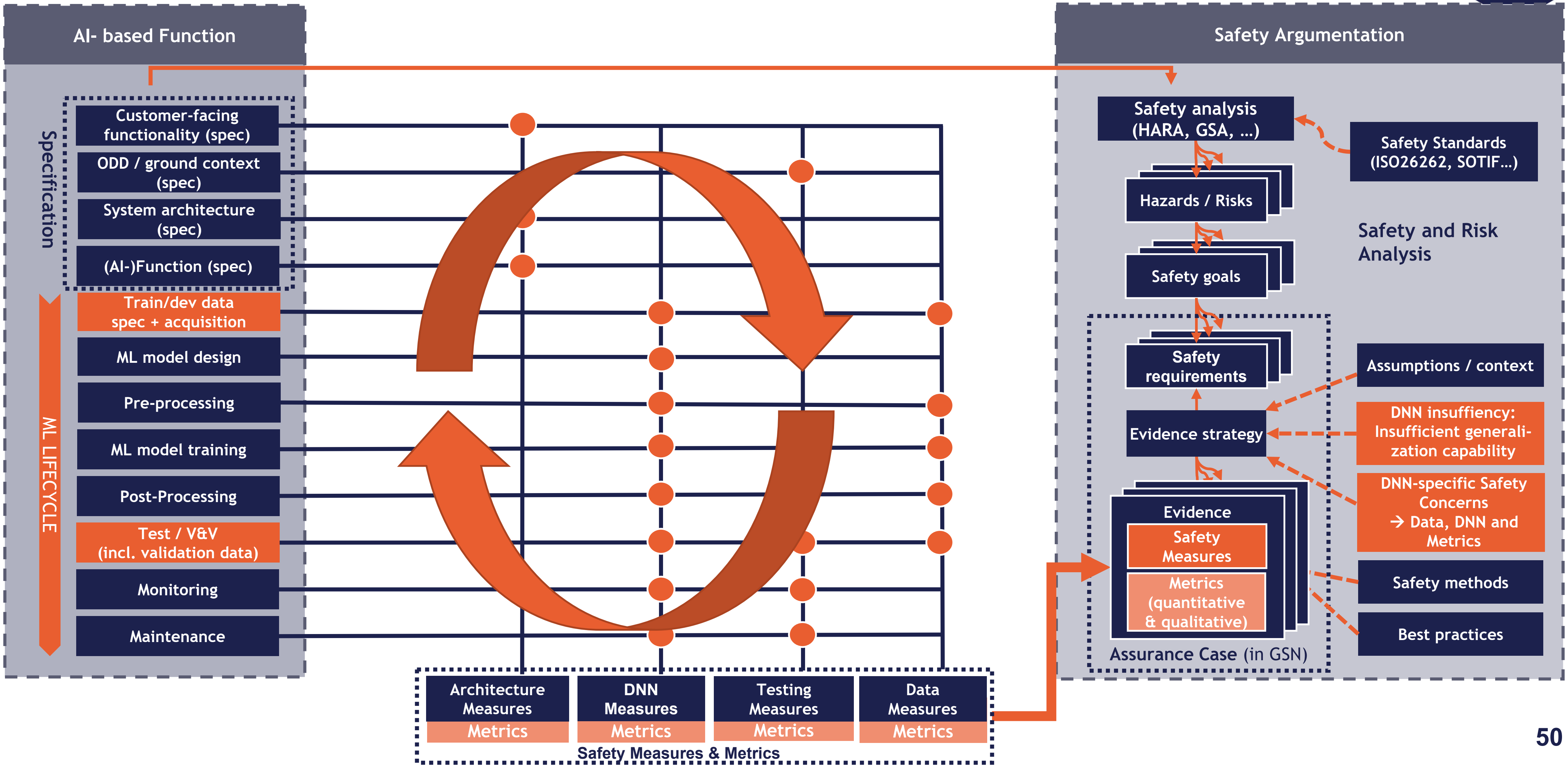
Source: ZF, Thomas Schulik

Adaptive test case selection for DNN-based perception functions, Bernhard, J.; Schulik, T.; Schutera, M.; Sax, E., 2021 IEEE International Symposium on Systems Engineering (ISSE)

5

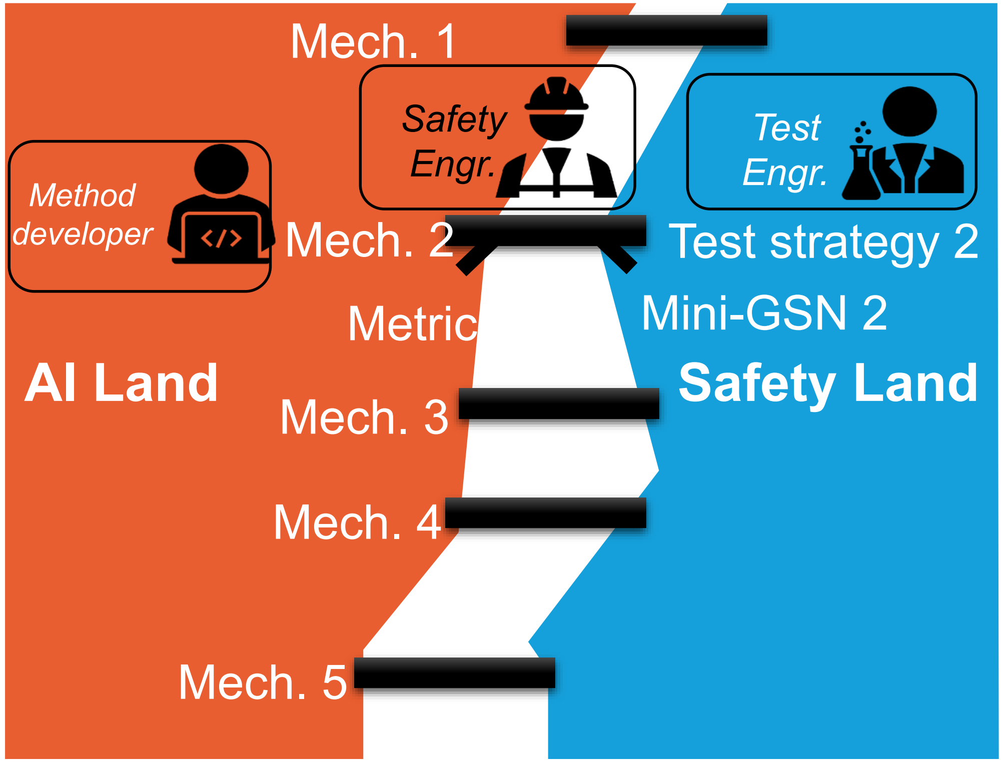
How do we work

Our Approach: Summary



Our Approach: Evidence Workstreams

Empowering experts from safety engineering and ML to produce measures and evidences



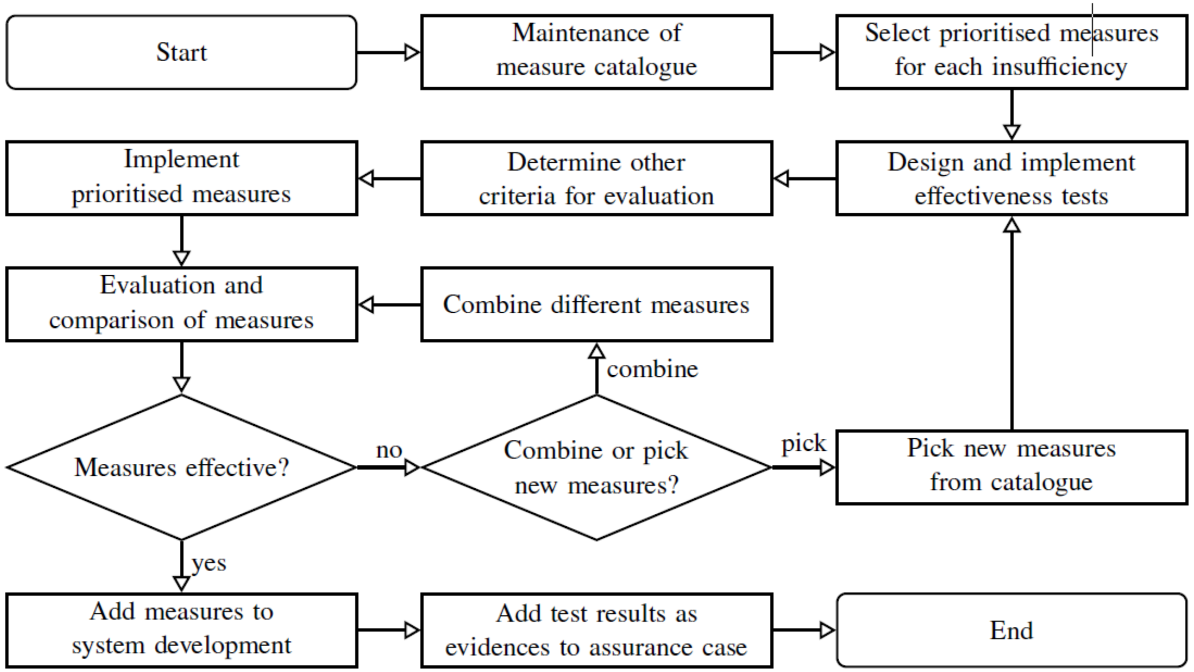


Fig. 3. Flowchart of the evidence workstream process

S. Burton et al (2022): Safety Assurance of Machine Learning for Perception Functions, to be published

Agenda

Summary

Findings & Consequences

- Safe AI is a central challenge for highly automated driving
- KI-Absicherung provides an approach for Safe AI
- Approach may serve as template for the industry and beyond (see ISO PAS 8800)
- Deep integration of AI-specifics into development PMT is necessary (continuous assurance of AI)

Contact:

Fabian Hüger

Artificial Intelligence Safety
@Volkswagen CARIAD

Contact: **fabian.hueger@volkswagen.de**

https://scholar.google.de/citations?user=lSPOi1UAAAAJ

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

Thank you!

QUESTIONS?