

Towards Virtual Testing of Perception in the Railway Domain

Michael Wild, Jan Steffen Becker, Anda Buinoschi-Tirpescu, Eike Möhlmann
German Aerospace Center, Oldenburg, Germany
Email: firstname.lastname@dlr.de

Abstract—To achieve higher automation levels in the railway domain (GoA-3/4), especially focusing on branch lines, an automated system must take over responsibilities which are currently allocated to the train driver. While existing automation technologies ensure operational safety at the macroscopic and technical levels, addressing perception at the operational level in open context remains a challenge. This paper describes key aspects of future railway perception components, and identifies challenges that they bring for verification and validation. It is shown how existing technologies may be combined into a holistic framework that finally enables scenario-driven simulative testing of perception components.

Index Terms—perception, testing, simulation, scenario, automation

I. INTRODUCTION

To achieve higher grades of automation (GoA3/4) in the railway domain, e.g. on branch lines, automation takes over responsibilities currently allocated to train drivers, dispatchers, or other personnel. Technologies like ATO over ETCS ensure safety at the macroscopic and technical level. Operational-level perception in open contexts remains a challenge [1].

The scope of the proposed concept is testing of perception components in the so called *autonomy pipeline*: *sense* → *perception* → *plan* → *prediction* → *control* → *actuation* [2].

Railway domain literature shows that perception components must support medium/long-term goals like maintaining knowledge of the railway system, line-specific features, protocols, train mechanics, as well as short-term goals like assessing track conditions, interpreting signals, anticipating events, and recognizing objects/obstacles within braking distance.

Existing work often tests isolated aspects, such as test case generation or synthetic data creation. This paper advances the field by proposing a concept for integrated, simulation-based evaluation of railway perception components. It identifies essential framework requirements and introduces a scenario-driven assessment method based on real-world sensor data.

The paper is structured as follows: Section II covers background and related work on validating highly automated systems, focusing on perception and railway applications. Section III outlines challenges and derives framework requirements. Section IV presents our assessment method including an architecture of our concept framework in Section V.

II. BACKGROUND INFORMATION AND RELATED WORK

This section provides background on validation and testing approaches for autonomous and highly automated systems

(II-A), with an emphasis on the railway domain. Particular focus is given to the perception component (II-B), covering requirements, completeness, sensor needs, and machine learning methods such as object detection, including explainable AI (xAI). Additional topics include maps (II-C) and comparison of world models (II-D).

A. Validation Approaches for Autonomous / Highly Automated Systems

Donà et al. [3] identify two high-level validation strategies for autonomous driving systems (ADS): integrated environment and subsystem-based solutions. Sifakis and Harel emphasize that simulation and testing are the only feasible validation approaches and propose an architecture for this [4]. Peleska et al. examine evaluation strategies for autonomous train controllers under ANSI/UL 4600, offering a reference architecture with perception-reliant components like *Obstacle Detection*, *Train Signal Classification*, and *Refined Positioning*, which feed the situational awareness in the planning *kernel* [2].

B. Validation and Testing of the Perception Component

The “black box” nature of machine learning in perception makes testing essential for safety. Outputs must be explainable to facilitate debugging and development. Ali et al. [5] survey explainability techniques at the data, model, and post-hoc levels. Hoss et al. [6] survey testing of automotive perception systems, focusing on test criteria, scenarios, and data.

1) *From real-world testing to simulation*: Real-world testing is a standard method for evidence generation. Efficiency can be improved by simulating elements of the railway system, including tracks, signals, weather, objects, sensors, and decision-making, while keeping the perception system intact. Hybrid methods, known as X-in-the-Loop (XiL), cover the spectrum between physical and simulated testing [3].

2) *Sensor and perception requirements in railways*: In GoA1/2, train drivers oversee operations. In GoA3/4, automated systems must e.g. detect and respond to track hazards. Sensor requirements for safe automated passenger train operation should meet or exceed human capabilities [7]. EN 50126 and related standards define a V-process for system development, starting with hazard analysis and safety requirement derivation. Perception systems must be trained and tested across relevant conditions, with traceability from requirements to test outcomes. For GoA3/4, functional requirements such as obstacle, smoke, animal, or anomaly detection are specified

along with performance constraints [8]. Interfaces to processing modules and the physical environment are also defined.

GoA4 systems also require perception for passenger transfer and vehicle health supervision [2]. Remote train operations (RTO) need an independent, lower-fidelity perception system. These topics, as well as security issues, e.g. adversary attacks on the perception component, are out of scope in this work.

3) *Operational Domain (OD) and Operational Design Domain (ODD)*: Safety validation requires a clear definition of the operational domain (OD) and its subset, the operational design domain (ODD) [9]. In closed systems like metros, the OD can be modeled comprehensively. In open contexts, full OD coverage is unfeasible. Systems must detect ODD exits and transition to safe states (halt, transfer control, ...) [10].

4) *Sensors and sensor models*: Schlager et al. [11] review state-of-the-art sensor models. Linhoff et al. [12] propose a cause-effect based modeling approach. Safe perception requires redundant, independent sensor technologies. Sensor data may be interpreted separately or fused. Gao et al. explore challenges and propose a virtual testing framework for early (raw data) or late (processed data) multi-sensor fusion [13].

5) *Object detection and relevant objects*: Grossmann et al. [14] introduce a framework for generating railway-specific scenes for object detection. D'Amico et al. [15] develop a simulation tool for LiDAR and camera data. Talwar et al. [16] evaluate neural networks trained on synthetic data, noting realism and domain match issues. De Gordo et al. [17] assess detection accuracy on synthetic images. Discher et al. [18] identify use cases for train- and infrastructure-side detection. Sifakis and Harel [4] highlight varying information relevance across decision horizons. Rakow [19] offers a formal model of relevance. Ristić-Durrant et al. review vision-based distance estimation for train-obstacle detection [20].

6) *Explainability*: Limited transparency of neural networks complicates interpretation. De Graaff et al. [21] improve explainability in time series classification for axle counter data. Roßbach et al. [22] enhance trust in train signal classification via runtime monitoring with verified vision algorithms.

C. Localization and maps

Perception updates a world model, including position, track boundaries, obstacles, and signals. GNSS is unreliable in forests, urban canyons, and tunnels. Since balises aren't ubiquitous, Simultaneous localization and mapping (SLAM) can be used. Cadena et al. [23] stress task-driven perception in SLAM, producing maps aligned with mission needs. Local dynamic maps (LDMs) enrich static maps with dynamic objects and semantic content, enabling real-time updates and information sharing [24].

D. World model comparison and inference

To fulfill a function like *Detect road vehicle that could endanger the train* [8], the world model must include beliefs about intentions. Nozari et al. [25] apply active inference [26] to model agent responses. Friston et al. view the world as a random dynamical system, where minimizing variational free energy aligns internal models with sensory inputs [27].

III. CHALLENGES FOR TESTING

While real-world testing of autonomous train systems is possible, we focus on simulation due to lower cost, greater safety, and the ability to systematically test rare or dangerous edge cases [1]. This section lists challenges and derives requirements (Req) for a simulation framework focused on testing and validating the perception component.

A. Realistic Environments

Talwar et al. show that the simulation precisely needs to mimic the characteristics of the ODD/OD (behavior of the environment, feature density, 3D models) [16]. Using real and synthetic data from German and US maps, they demonstrate that object detection performance degrades significantly when test and training data differ in localization (e.g., lane width)—a factor more critical than visual realism. This underscores the need to accurately simulate the system's ODD/OD (cf. [6]).

Req 1. *Test scenarios must completely cover the OD/ODD of the perception system, with respect to traffic participants, environmental conditions, and route-specific elements.*

Train drivers must assess situations, anticipate object trajectories, and infer intentions, which requires physical modeling consistent with their belief formation. Perception systems also exploit expected physical behavior in sensor fusion and filtering, often using Kalman filters that iteratively refine state estimates based on system dynamics. This extends to multi-sensor setups [28].

Req 2. *A simulation framework must implement realistic physics.*

B. Sensor Simulation

To maximize accuracy and robustness, perception systems integrate multiple sensors. Sengupta [29] identifies typical setups, while Gao et al. [30], [13] emphasize the need for consistent synthetic data across sensors for validation. Simulators must use a shared ground truth if generating data for different sensors. For instance, GNSS, radar, and accelerometers may be combined for localization. E.g. in the European Rail Traffic Management System (ERTMS), localization depends on sparse but precise balises integrated with other sensors [31].

Req 3. *A simulation framework must be able to create consistent data emulating different sensors, including railway-specific sensors such as balise antennas.*

Perception systems tailored to specific routes may rely on prior ODD knowledge such as high-resolution maps, which are essential for techniques like SLAM [23].

Req 4. *A simulation framework must be able to create sensor data that is consistent with real-world map data.*

Existing simulators typically cover only isolated aspects of perception system testing, e.g., ego-train physics or vendor-specific sensor models. Validating complex systems like collaborative train-infrastructure for shared world models and safe operation [18] requires integrating diverse simulators.

Req 5. *A simulation framework shall enable the co-simulation of simulation models from different vendors (with a standardized interface).*

C. Comparison of World Models

Sensor data carries inherent uncertainty, systematic and random, often amplified by environmental conditions. This uncertainty propagates through the perception system to the world model, making it non-deterministic [12]. Phillion et al. introduce the *Planning Kullback–Leibler divergence* to account for such uncertainties [32].

Req 6. *A validation framework must support modeling, propagation, and evaluation of uncertainty in sensor data and perception outputs.*

Most existing validation frameworks target specific stages of perception. While object detection metrics like Intersection over Union (IoU) [33] exist, they must be adapted to compare uncertain world models against ground truth (cf. [6]).

Req 7. *A validation framework must provide metrics to compare uncertain, perception-derived world models with ground truth data.*

Active perception [34] enables systems to enhance sensing through actions like repositioning or activating sensors (e.g., LiDAR, active lighting).

Req 8. *A validation framework must support active sensing and allow the agent to enhance perception through interaction.*

In deep artificial neural net (DNN) based classifiers, softmax outputs are often misinterpreted as probabilities. Calibration, necessary for interpreting these confidences as true probabilities, is essential [6], as DNNs tend to be overconfident [35].

Req 9. *A validation framework must support testing classifier calibration, ensuring output confidence reflects true classification probability.*

IV. CONCEPT OF A TEST METHODOLOGY

This concept aims to bridge the gap between real-world and virtual testing. Assume a novel highly automated (GoA3/4) train to be the system under test (SUT), which shall be certified for a specific (branch) line, which is its ODD. A line is assumed to be an end-to-end sequence of operational situations (e.g., open track, level crossings, stations, shunting) in which the SUT has to perform specific tasks. Existing work from the automotive domain identified that test cases for these tasks need to be generated in a systematic manner in order to achieve high coverage and confidence. Therefore, test scenarios are used on different abstraction levels [36], [37].

Abstract scenarios focus on the relevant OD/ODD concepts and their relations, allowing for potentially infinite many different evolutions. In contrast, concrete scenarios fix a concrete setting and evolution, allowing for reproducible simulations.

Wild et al. [38] presented a scenario-based simulation framework for testing of highly automated railway systems, enabling creation and closed loop play-out of test-cases from

such operational situations. These are instances of abstract scenarios modeled as Traffic Sequence Charts (TSC).

From abstract test cases, concrete test cases can be sampled by means of constraint solving [39], [40]. For TSCs, this is described in detail by Becker [40]. Borchers et al. [39] evaluate an exemplary solver setup and different sampling strategies. This technique is applied for railway scenarios [1], [38] and goes beyond sampling from pre-defined behavior patterns, exploring a larger part of the scenario space which is not characterized by a predefined parameter set. In [14], [15], [17] ontologies are used to sample test cases.

For satisfying Req. 1, we propose to use abstract scenarios as they allow fine-grained control and integrate with scenario-driven development methodology [36], [37], [41].

A test scenario – both abstract and concrete – consists of static and dynamic aspects. In order to satisfy Req. 4, the actual route for which the system shall be certified is used for instantiating the static parts of the environment in the simulation. This also supports Req. 1, as the static parts are complete wrt. the ODD match its individual characteristics. Static aspects include the track layout, static objects like trees, parked cars or trains, time of day, etc. Assume a sensor stream of the route with different sensor-technologies exists, e.g. camera, LiDAR, and radar. Objects, their position, orientation, shape and texture can be extracted from this dataset and used as the static part of the scenario. Rahman et al. demonstrate a robot equipped with the needed sensors and technology to create a precise model of a railway line [42]. Ariyachandra and Brilakis use LiDAR to build a digital model of railway overhead wires [43]. A common approach in the automotive domain (cf. [44]) is to generate 3D models from existing map data using generic models of infrastructure and environment objects (buildings, trees, ...). However, incomplete map data leads to unsatisfactory results. For railway applications it is more promising to accurately record and model the environment because the targeted railway line is usually fixed and accessible at development time.

Since it is unlikely, that critical situations are encountered in this dataset, e.g. a road user blocking an unprotected level-crossing, our approach is to sample the dynamic aspects of the test scenario from abstract test cases using the techniques already present in the literature [40], [39], [38]. Dynamic aspects of the scenario include the trajectory of road users and other objects, as well as non-ego trains. Using the findings from a preceding hazard analysis, they are governed in such a way that the SUT is guaranteed to encounter the (potentially critical) situations for which its behavior needs to be evaluated. Moreover, in order to achieve a higher coverage of critical scenarios, additional static / dynamic objects can be augmented into test-cases, if considered critical to the SUT.

The proposed process is shown in Fig. 1 and briefly described below.

- 1) **Record the route:** A traditional train equipped with a sensor suite with a train driver on board drives the route and a dataset with sensor outputs is created.

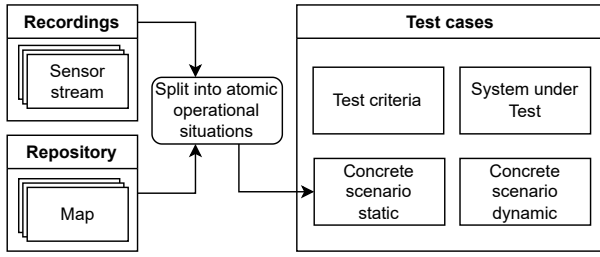


Fig. 1: Flowchart of the test creation framework

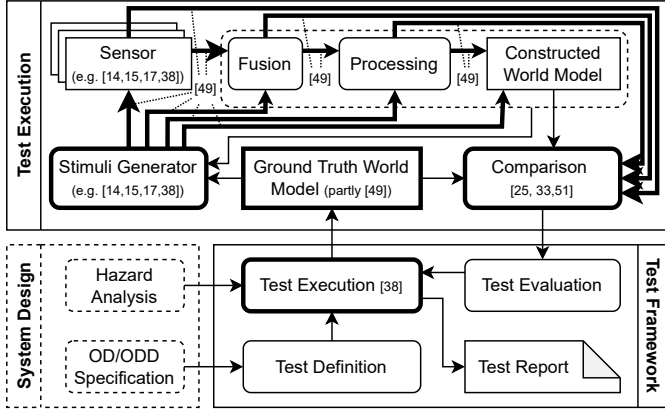


Fig. 2: Functional architecture of the testing framework. Bold elements have existing solutions given as literature references.

- 2) **Extract static parts:** The path of the train could be estimated from the dataset. If available, it can rather be extracted from existing recordings, e.g. from open street maps. Based on the path, sensor data is aggregated to a semantic map with object labels and 3D mesh models of the objects.
- 3) **Split the track into operational situations.** This may be automated – e.g., map-based mapping of relevant abstract scenarios on matching parts of the route – or expert-driven.
- 4) **Create test-cases for the occurring operational situations:** For this purpose Wild et al. describe a process and provided a framework to derive concrete test-cases from abstract scenarios [1] [38].
- 5) **Aggregate the results:** The tests are collected and a verdict for the SUT on the whole line is reached. Weak spots can be identified.

V. PROPOSED TEST FRAMEWORK ARCHITECTURE

Based on the findings from the previous sections, we propose the functional architecture shown in Fig. 2. Test cases, consisting of a concrete scenario and test criterion, are defined based on system design, including *hazard analysis* and the target *OD/ODD*. Putze and Böde [45] show how hazards can systematically be identified in a scenario and keyword-driven way. Following their approach, one could identify, e.g., a *hazard collision with a road user* and specify a test scenario with a blocked level crossing as done by Wild et al. [38]. By

linking identified hazards to test scenarios (c.f. [46]), an argument about test coverage can be made. Tests are executed via simulation on the designated route. The simulation maintains a *ground truth world model*, dynamically updated, containing all relevant information (e.g., object positions, ego train state, environment conditions) required for testing the perception system. This model drives a synthetic *stimuli generator*, which can generate physical/virtual sensor stimuli or simulate sensor outputs directly (single or fused). The stimuli may be purely synthetic, or previously recorded and augmented with simulated objects. The perception component under test derives a *constructed world model* from this data. To support active sensing (Req. 8), a feedback loop recalibrates virtual sensors during runtime. The perceived and ground truth models are compared to evaluate test criteria and generate reports.

Numerous simulators for railway applications produce synthetic sensor data [14], [15], [17], [38] and can take the role of the *stimuli generator* in our proposed test framework. Many of them are based on Unreal Engine, enabling realistic camera imagery. Automotive-focused tools also exist, as surveyed by Rosique et al. [47], though most remain standalone and limited to a subset of sensors (e.g., RGB, point cloud). For enabling use of these simulators in the validation framework, we propose to develop standardized interfaces based on ASAM OSI. The ASAM OSI standard [48], [49] provides an architecture and data formats for building modular multi-sensor simulations. OSI proposes to connect an environment simulation, vehicle agents, and sensor models in a co-simulation. The modular architecture allows to plug in simulators for specific sensors and combine them with customized environment behavior and a SUT. Wrapping modules as Functional Mockup Units, ASAM OSI also standardizes the technical interfaces and packaging. Using OSI, the existing sensor simulations could be modularized and combined with a scenario-based environment and physics simulator (such as described, e.g., by Wild et al. [38], or dedicated standalone train physics engines such as benchmarked by Wu et al. [50]). This would then satisfy Requirements 2, 3, and 5.

In OSI, ground truth (communicated from the environment simulation) and perceived world models (communicated from sensors/perception components) share the same data model. For perceived world models, OSI extends this data model with uncertainty information (Req. 6). Therefore, we see OSI being beneficial in providing a basis for the evaluation of test results (Req. 7). In accordance to classical metrics for evaluating object recognition in 2D images [33], Hartstern et al. define some distance metrics that operate directly on the world model [51]. Besides many different approaches for calibrating object classifiers [52], only few works consider its verification, for example, Kumar et al. [53]. From our point of view, it is an open research question how to integrate these with comparison models for non-probabilistic world models, yielding a test criteria for complex perception systems.

OSI has been designed for automotive applications and therefore only automotive ODs can be represented in the world model. We see several limitations with the current design.

For example, object classes for railway-specific infrastructure are missing. Besides a description of the rail network (e.g., rails, sleepers, turnouts, switches) this would include at least balises and signaling¹. Furthermore, OSI does not consider the exchange of mission data² – such as station stops, train schedules, or train compositions – or other sensor types such as odometry, GNSS, or sound sensors. Therefore, in order to enable it for the railway domain, the OSI data model needs to be extended. Although OSI’s co-simulation architecture is predestined for the integration of white box models, these require sensor-specific and detailed environment models, which cannot be expected to be provided out of the box [51].

VI. CONCLUSION AND FUTURE WORK

In this paper we described key challenges for the simulative testing of perception in the railway domain for GoA3/4 train operation. We compiled a list of requirements for a simulation framework capable of enabling this. Although current literature on the topic was considered and the list was carefully derived, it does not claim to be complete, but can serve as a starting point when deciding what tools to use to get a step closer in the big task which is enabling highly automatic railway operation in open context. Table I summarizes the identified requirements, the bibliographic resources that lead to their identification, and existing solutions that address most of the requirements. To our knowledge, no complete solution exists that addresses all of the requirements. A modular approach is needed that allows to integrate the existing solutions into a holistic framework.

TABLE I: Overview of the requirements

Requirement	Sources	Solutions
R1 Cover OD/ODD	[16]	[1], [39]
R2 Realistic Physics	[51]	[38], [50]
R3 Consistent data	[30], [13]	[49]
R4 Map data	[23]	[42], [44], [43]
R5 Sensor co-simulation	[51], [18]	[49]
R6 Handle uncertainty	[12], [32], [6]	[49]
R7 World model comparison	[6]	[51], [33]
R8 Active perception	[34]	[25]
R9 Classifier calibration	[35]	[53]

As a first contribution towards a complete testing framework, we provide an approach for creating realistic test cases, which is based on recording real sensor streams from the line for which the GoA3/4 system shall be certified. After splitting the collected data into atomic operational situations, information of the static components of a scenario are extracted. A test-case for the behavior of the SUT in the atomic operational scenario is modeled as an abstract scenario and a concrete scenario derived from it. The test evaluation of the perception hinges on the comparison between the ground truth

¹generic three-state traffic lights present in OSI cannot sufficiently describe railway signaling.

²The OSI ground truth can optionally contain map data, but the requirements for, e.g., ATO go far beyond that.

world model and the constructed world model. We hinted at some methods that might be applicable, but the details depend on the actual implementation / modeling choices.

Furthermore, we identified that the Open Simulation Interface (OSI) [49] known from automotive industry provides a viable approach for integrating the needed technology bricks.

Roadmap and Future Work: Although many building blocks are already available, realizing the proposed framework requires further research including the following steps:

- A detailed gap analysis of ASAM OSI towards the railway domain, and development of the needed extensions
- Modularization of existing sensor simulations and packaging them for use in an OSI-based co-simulation
- Integration of the extended OSI specification into existing simulators and test execution frameworks
- Further development of test generation methods emphasizing achievement of test coverage; development of methods for test aggregation

An evaluation of the approach further needs specification of a test system (based on existing research [18]) – including performance requirements –, and recording and modeling of a test track for evaluation.

ACKNOWLEDGEMENTS

This work has received funding from the European Union’s Horizon Europe research and innovation programme within the projects ‘MOTIONAL’ (under grant agreement No: 101101973) and ‘Pods4Rail’ (under grant agreement No: 101121853).

REFERENCES

- [1] M. Wild, J. S. Becker, G. Ehmen, and E. Möhlmann, “Towards Scenario-Based Certification of Highly Automated Railway Systems,” in *Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification*, ser. LNCS. Springer, 2023.
- [2] J. Peleska, A. E. Haxthausen, and T. Lecomte, “Standardisation Considerations for Autonomous Train Control,” in *Leveraging Applications of Formal Methods, Verification and Validation. Practice*, ser. LNCS. Springer, 2022.
- [3] R. Donà and B. Ciuffo, “Virtual Testing of Automated Driving Systems. A Survey on Validation Methods,” *IEEE Access*, vol. 10, 2022.
- [4] J. Sifakis and D. Harel, “Trustworthy Autonomous System Development,” *ACM Trans. Embed. Comput. Syst.*, vol. 22, no. 3, Apr. 2023.
- [5] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, and F. Herrera, “Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence,” *Information Fusion*, vol. 99, 2023.
- [6] M. Hoss, M. Scholtes, and L. Eckstein, “A review of testing object-based environment perception for safe automated driving,” *Automotive Innovation*, vol. 5, no. 3, 2022.
- [7] R. Tagiew, D. Leinhos, H. von der Haar, C. Klotz, D. Sprute, J. Ziehn, A. Schmelter, S. Witte, and P. Klasek, “Onboard sensor systems for automatic train operation,” in *Dependable Computing – EDCC 2022 Workshops*, ser. CCIS. Springer, 2022, vol. 1656.
- [8] S. Besure, “D5.1 WP5 GoA3/4 Specification,” Shift2Rail X2Rail-4, Tech. Rep., Dec. 2023.
- [9] P. Koopman and F. Fratrick, “How Many Operational Design Domains, Objects, and Events?” in *SafeAI@AAAI*, 2019.
- [10] G. Weiss, M. Zeller, H. Schoenhaar, C. Drabek, and A. Kreutz, “Approach for Argumenting Safety on Basis of an Operational Design Domain,” in *Proceedings of the IEEE/ACM 3rd Int. Conf. on AI Engineering - Software Engineering for AI*. ACM, Apr. 2024.

- [11] B. Schlager *et al.*, “State-of-the-Art Sensor Models for Virtual Testing of Advanced Driver Assistance Systems/Autonomous Driving Functions,” *SAE International Journal of Connected and Automated Vehicles*, vol. 3, no. 3, Oct. 2020.
- [12] C. Linnhoff, P. Rosenberger, S. Schmidt, L. Elster, R. Stark, and H. Winner, “Towards Serious Perception Sensor Simulation for Safety Validation of Automated Driving - A Collaborative Method to Specify Sensor Models,” in *2021 IEEE Int. Intelligent Transportation Systems Conf. (ITSC)*, Sep. 2021.
- [13] X. Gao, Z. Wang, Y. Feng, L. Ma, Z. Chen, and B. Xu, “Multitest: Physical-aware object insertion for testing multi-sensor fusion perception systems,” in *Proceedings of the IEEE/ACM 46th Int. Conf. on Software Engineering*. ACM, 2024.
- [14] J. Grossmann *et al.*, “Test and training data generation for object recognition in the railway domain,” in *Software Engineering and Formal Methods. SEFM 2022 Collocated Workshops*, ser. LNCS. Springer, 2023, vol. 13765.
- [15] G. D’Amico, M. Marinoni, F. Nesti, G. Rossolini, G. Buttazzo, S. Sabina, and G. Lauro, “Trainsim: A railway simulation framework for lidar and camera dataset generation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, 2023.
- [16] D. Talwar, S. Guruswamy, N. Ravipati, and M. Eirinaki, “Evaluating validity of synthetic data in perception tasks for autonomous vehicles,” in *2020 IEEE Int. Conf. On Artificial Intelligence Testing (AITest)*. IEEE, 2020.
- [17] J. A. I. de Gordo, S. García, L. P. V. de La Iglesia, I. Urbieto, N. Aranjuelo, M. Nieto, and D. O. de Eribe, “Scenario-based validation of automated train systems using a 3d virtual railway environment,” in *2023 IEEE 26th Int. Conf. on Intelligent Transportation Systems (ITSC)*. IEEE, 2023.
- [18] S. Discher, T. Herrmann, A. Schulz, M. R. Rehme, B. Heibutzki, M. Meinig, R. Otto, I. Schiering, A. Gabel, R. Schmidt, and P. Grenz, “Sensorbasierte Technologien im Bahnsystem: Markt- und Technologieanalyse,” Deutsches Zentrum für Schienenverkehrsforschung beim Eisenbahn-Bundesamt, Tech. Rep., 2024.
- [19] A. Rakow, “A Notion of Relevance for Safety-Critical Autonomous Systems,” in *Engineering Safe and Trustworthy Cyber Physical Systems, Essays Dedicated to Werner Damm on the Occasion of His 71st Birthday, Proceedings, ESTCPS*. Springer, 2024, accepted for publication.
- [20] D. Ristić-Durrant, M. Franke, and K. Michels, “A review of vision-based on-board obstacle detection and distance estimation in railways,” *Sensors-basel.*, vol. 21, no. 10, 2021.
- [21] T. de Graaff, M. Wild, T. Werner, E. Möhlmann, S. Seibt, and B. Ebrecht, “Increasing Explainability in Time Series Classification by Functional Decomposition,” in *Explainable Artificial Intelligence*. Springer, 2024.
- [22] J. Roßbach and M. Leuschel, “Certified Control for Train Sign Classification,” *Electronic Proceedings in Theoretical Computer Science*, vol. 395, Nov. 2023.
- [23] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, Dec. 2016.
- [24] H. Shimada, A. Yamaguchi, H. Takada, and K. Sato, “Implementation and Evaluation of Local Dynamic Map in Safety Driving Systems,” *Journal of Transportation Technologies*, vol. 05, no. 02, Mar. 2015.
- [25] S. Nozari, A. Krayani, P. Marin, L. Marcenaro, D. M. Gomez, and C. Regazzoni, “Modeling Autonomous Vehicle Responses to Novel Observations Using Hierarchical Cognitive Representations Inspired Active Inference,” *Computers*, vol. 13, no. 7, Jul. 2024.
- [26] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo, “Active Inference: A Process Theory,” *Neural Comput.*, vol. 29, no. 1, Jan. 2017.
- [27] K. Friston, L. D. Costa, N. Sajid, C. Heins, K. Ueltzhöffer, G. A. Pavliotis, and T. Parr, “The free energy principle made simpler but not too simple,” *Phys. Rep.*, vol. 1024, Jun. 2023.
- [28] C. Yang, J. Zheng, X. Ren, W. Yang, H. Shi, and L. Shi, “Multi-sensor kalman filtering with intermittent measurements,” *IEEE Trans. Autom. Control*, vol. 63, no. 3, 2018.
- [29] M. Sengupta, “Choice of sensor fusion framework for train positioning system,” *Computers in Railways XVII: Railway Engineering Design and Operation*, vol. 199, 2020.
- [30] X. Gao, Z. Wang, Y. Feng, L. Ma, Z. Chen, and B. Xu, “Benchmarking robustness of ai-enabled multi-sensor fusion systems: Challenges and opportunities,” in *Proceedings of the 31st ACM Joint European Software Engineering Conf. and Symposium on the Foundations of Software Engineering*. ACM, 2023.
- [31] M. Marinoni, P. Amato, C. D. Franco, S. Sabina, and G. Buttazzo, “A novel approach for reducing train localization errors by inertial measurements,” *IEEE Access*, vol. 11, 2023.
- [32] J. Pillion, A. Kar, and S. Fidler, “Learning to Evaluate Perception Models Using Planner-Centric Metrics,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020.
- [33] R. Padilla, S. L. Netto, and E. A. B. Da Silva, “A survey on performance metrics for object-detection algorithms,” in *2020 Int. Conf. on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2020.
- [34] R. Bajcsy, Y. Aloimonos, and J. K. Tsotsos, “Revisiting active perception,” *Autonomous Robots*, vol. 42, 2018.
- [35] L. Neumann, A. Zisserman, and A. Vedaldi, “Relaxed Softmax: Efficient Confidence Auto-Calibration for Safe Pedestrian Detection.”
- [36] T. Menzel, G. Bagschik, and M. Maurer, “Scenarios for development, test and validation of automated vehicles,” in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018.
- [37] C. Neurohr, L. Westhofen, T. Henning, T. de Graaff, E. Möhlmann, and E. Bode, “Fundamental considerations around scenario-based testing for automated driving,” in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020.
- [38] M. Wild, J. S. Becker, C. Schneiders, and E. Möhlmann, “A scenario-based simulation framework for testing of highly automated railway systems,” in *Proceedings of the 11th International Conference on Vehicle Technology and Intelligent Transport Systems - VEHITS*, 2025.
- [39] P. Borchers *et al.*, “TSC2CARLA: An abstract scenario-based verification toolchain for automated driving systems,” *Sci. Comput. Program.*, vol. 242, 2025.
- [40] J. S. Becker, “Safe linear encoding of vehicle dynamics for the instantiation of abstract scenarios,” in *Formal Methods for Industrial Critical Systems*, ser. LNCS, vol. 14952. Springer, 2024.
- [41] C. Neurohr, L. Westhofen, M. Butz, M. H. Bollmann, U. Eberle, and R. Galbas, “Criticality analysis for the verification and validation of automated vehicles,” *IEEE Access*, vol. 9, 2021.
- [42] M. Rahman, H. Liu, M. Masri, I. Durazo-Cardenas, and A. Starr, “A railway track reconstruction method using robotic vision on a mobile manipulator: A proposed strategy,” *Computers in Industry*, vol. 148, 2023.
- [43] M. Ariyachandra and I. Brilakis, “Digital twinning of railway overhead line equipment from airborne lidar data,” 2020.
- [44] T. Azfar, J. Weidner, A. Raheem, R. Ke, and R. L. Cheu, “Efficient procedure of building university campus models for digital twin simulation,” *IEEE Journal of Radio Frequency Identification*, vol. 6, 2022.
- [45] L. Putze and E. Böde, “Systematic identification and analysis of hazards for automated systems,” *INSIGHT*, vol. 28, no. 1, pp. 69–74, 2022.
- [46] R. Wallner, B. A. Gran, T. A. Pedersen, T. A. Johansen, and M. A. Lundteigen, “Identifying Test Scenarios for Simulated Safety Demonstration using STPA and CAST,” in *33rd European Safety and Reliability Conference*, 2023, pp. 2929–2936.
- [47] F. Rosique, P. J. Navarro, C. Fernández, and A. Padilla, “A Systematic Review of Perception System and Simulators for Autonomous Vehicles Research,” *Sensors-basel.*, vol. 19, no. 3, Jan. 2019.
- [48] C. Linnhoff, P. Rosenberger, M. F. Holder, N. Cianciaruso, and H. Winner, “Highly Parameterizable and Generic Perception Sensor Model Architecture,” in *Automatisiertes Fahren 2020*. Springer, 2021.
- [49] ASAM e. V., “ASAM OSI (Open Simulation Interface),” Jul. 2024, accessed on 2025-03-24. [Online]. Available: <https://opensimulationinterface.github.io/osi-antora-generator/asamosi/latest/specification/appendix/releases/v03.07.html>
- [50] Q. Wu *et al.*, “International benchmarking of longitudinal train dynamics simulators: results,” *Vehicle System Dynamics*, vol. 56, no. 3, 2018.
- [51] M. Hartstern, V. Rack, M. Kaboli, and W. Stork, “Simulation-based evaluation of automotive sensor setups for environmental perception in early development stages,” in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020.
- [52] T. Silva Filho, H. Song, M. Perello-Nieto, R. Santos-Rodriguez, M. Kull, and P. Flach, “Classifier calibration: a survey on how to assess and improve predicted class probabilities,” *Machine Learning*, vol. 112, no. 9, 2023.
- [53] A. Kumar, P. S. Liang, and T. Ma, “Verified uncertainty calibration,” *Advances in neural information processing systems*, vol. 32, 2019.