In-Service Monitoring and Assessment (ISMA) of Autonomous Driving Vehicles with AI based Algorithms

Bangarevva Patil and Thomas Corell
Autonomous Mobility, Systems & Software Business
Continental Autonomous Mobility Germany GmbH
Frankfurt am Main, Germany
email:{firstname}.{lastname}@continental-corporation.com

Rudra Narayan Hota
AI & Data Engineering, Software Central Technologies
Continental Automotive Technology GmbH
Frankfurt am Main, Germany
e-mail: rudra.hota@continental-corporation.com

Abstract— **Autonomous Driving technology is anticipated to be a key aspect for achieving a higher level of road safety and efficient mobility. A major challenge for this automation is to verify and validate safety aspects of Highly Autonomous Vehicles appropriately. Due to high system complexity and costs, it requires an exponential increase of test efforts for real world testing. Use of In-service Monitoring and Assessment (ISMA) system can efficiently reduce the verification and validation effort with respect to both cost and time. ISMA is an approach to ensure safety of an Autonomous Driving vehicle during the entire product life cycle by continuously intelligently monitoring and evaluating Autonomous Driving functionality during operation. In this paper, we propose an in-service monitoring and assessment concept with a set of exemplary implementations of Artificial Intelligence based algorithms. This concept aims to identify leading indicators for safety critical scenarios and addresses the assumptions made during the development. The research results demonstrate that the proposed approach can safely and efficiently monitor Autonomous Driving functions in both offline and online mode, and helps to discover critical and unknown unsafe scenarios even before an accident occurs.**

Keywords-verification and validation; in-service monitoring and assessment; silent testing; artificial intelligence functions.

I. INTRODUCTION

The development of highly automated driving (HAD) system is making rapid progress, but there is still no satisfying answer on how to prove that autonomous vehicles are safe. The safety validation of HAD system still remains a major challenge. State-of-the-art research on safety validation of highly automated functions (from SAE L3) [1] has found that autonomous vehicles would have to be driven hundreds of millions of miles to demonstrate their reliability in terms of fatalities and injuries compared to human drivers [2]. With such a large number of testmiles , validation of highly automated functions is economically and methodically not feasible [3].

Automated driving (AD) of Society of Automotive Engineers (SAE) L3 levels and above poses much higher requirements for the development and validation of safe systems. We need to improve the way HAD systems are developed and tested. For this, existing design and testing processes need to be extended to processes covering the full product life cycle of a HAD systems. These extended processes enable usage of data collected in the field for continuous improvement of HAD functions [4]. In service monitoring and assessment makes it possible to collect evidence from the field operation to demonstrate that the AD vehicle is safe and remains safe throughout its lifecycle.

In this research work, we introduce a concept called in-service monitoring and assessment, which is a real-time monitoring approach for validating the safe operation of AD systems. In some literature, this approach is also called Silent Testing. Here we describe how Artificial Intelligence (AI) based algorithms can be used to identify leading indicators for a possible accident. For this purpose, we implement prototypes of multiple AI-based event algorithms for pedestrian detection in urban intersection scenarios and evaluate their effectiveness.

The following sections are organized as follows: section II formally describes the problem at hand, section III discussed related work, after-which section IV describes our proposed solutions known as in-service monitoring and assessment, section V describes our example triggers, and finally section VI and VII present our performance evaluation and conclusions.

II. PROBLEM DESCRIPTION

Autonomous driving requires highly complex systems and is expected to be used in an unstructured real-world operational design domains (open contexts) with high levels of uncertainty. In this research, we will focus on how the safety of an automated driving system can be validated against two major challenges.

1. There will be situations or phenomena in the environment that we either cannot predict or are unaware that they may influence the behavior of the vehicle. In other words, there will be “unknown unknowns” events.
   This makes currently considered AI models prone to errors caused by events that are underrepresented in training data. That means, event classes that are safety critical for an AD application are limited, hence it is difficult to accurately evaluate models in such situations.
2. An unstructured real-world operational design domain spawns infinitely many possible scenarios, in which the intended behavior is based on implicit expectations, which are difficult to express formally [5]. For this reason, developing complex systems for open contexts

 ISBN: 978-1-68558-101-5

essentially deals with simplified representations. The validation challenge is closely related to the many necessary simplifications applied during development because every simplification is certainly based on explicit and most often even implicit assumptions. If these assumptions are not justified (even temporarily), the simplified representation is invalid. Therefore, hypotheses need to be formulated and their legitimacy needs to be demonstrated, e.g., through real-world monitoring. In this work, we will be referring to these hypotheses as "trigger functions" of an in-service monitoring and assessment system.

III. RELATED WORK

The impementation of new test strategies for the verification and validation of automated vehicles is required to provide sufficient test coverage while ensuring feasibility and maximum efficiency. State of the Art approaches are expained below.

A. Scenario based testing

In automated systems, research has shown that scenario-based testing can be used to provide an evidence of safety in a manner that is cost-efficient and time-efficient.

In the PEGASUS project [6] a holistic method for scenario-based safety assessment of HAD functions has been developed, using highway chauffeur as an exemplary test object, (i.e., a SAE L3 conditional automation system). The main idea of this approach is to identify relevant scenarios and then generalize them to generate more test scenarios. Scenario based testing exposes the Object Under Test (OUT) to a (pre)defined scenario and the OUT reaction is assessed. In this work, scenarios are derived by combining a data driven and a knowledge-based approach [7]. Scenario-based testing is a promising approach which enables the reduction of the effort required to test a HAD through the identification of relevant scenarios. Nevertheless, the determination of relevant scenarios, the definition of an appropriate parameter space within a scenario and the combination of parameters are great challenges [8].

An AD system will be exposed to a variety of scenarios throughout its deployment lifecycle. As part of the development and assessment process, it is therefore necessary to test against these scenarios. However, it would not be possible to perform such a large number of scenarios in the real world. The use of virtual test environments (i.e., simulations) to perform these tests is therefore essential.

B. VAFOO Approach

A new approach introduced by Wachenfeld and Winner [9] is the Virtual Assessment of Automation in Field Operation (VAAFO), which extracts relevant cases from a huge number of kilometers driven in the random nature of the real world. In the VAAFO approach, instead of testing automated driving in an unsafe environment, functionality is tested virtually in real traffic while the vehicle is driven by a human driver. The decisions of the driver contain additional information about the environment and are used as further input to compare it with the information from the perception sensors. This comparison is used for the event-based trigger.

The method uses differences between the trajectories of the OUT and the test vehicle as a trigger for event-based data recording. The recorded data is evaluated offline after the test drive. Hereby, the environment model of the OUT is corrected retrospectively to create a ground-truth environment model. As the simulation is open loop, the behavior of other traffic participants cannot be influenced by the OUT and therefore only short time frames can be simulated when the behavior of the OUT is different to the test vehicle [10].

C. Shadow mode testing

The shadow mode approach is proposed by Tesla [11]. The idea is very similar to the VAAFO approach. In shadow mode testing, a vehicle is being driven by a human, receiving data from the sensors but not taking control of the car in any way. Rather, it makes decisions about how to drive based on the sensors, and those decisions can be compared to the decisions of a human driver. Both the recorded data and the comparison are used, amongst others, to discover unthought of edge and corner cases, and to evaluate and demonstrate the safety of autonomous functionalities. Based on this technique, shortcomings in the system could be identified, and the collected data can in turn be used to improve their camera-based machine learning significantly. This approach is also used for a fleet of vehicles with human driver behavior as reference system.

IV. IN-SERVICE MONITORING AND ASSESSMENT AS A NEW TEST METHOD

To ensure the safe behavior of automated vehicles, the challenges of the open traffic context must be considered throughout the whole product life cycle. In-Service Monitoring and Assessment will collect evidence from the field operation to demonstrate that the ADS continues to be safe when operated on the road and thus supporting the safety argumentation [12] by addressing the following:

- the dynamic nature of road transportation,
- the assumptions made during development,
- the residual risk of unknown unsafe events.

In-Service Monitoring and Assessment can be used as part of Safety Management System (SMS) [13], DevOps and Learning-driven product life cycle processes for highly automated vehicles [14].

A. In-Service Monitoring and Assessment Framework

An ISMA system provides a framework that allows the performance of a driving function to be evaluated during operation. Figure 1 illustrates the framework using the following steps: a) data acquisition-provides all the data required to evaluate the functionality b) allows the recording & storage of data c) allows the transfer of stored data to a cloud for data management & analysis d) provides an interface that allows an operator to configure the system.

 ISBN: 978-1-68558-101-5

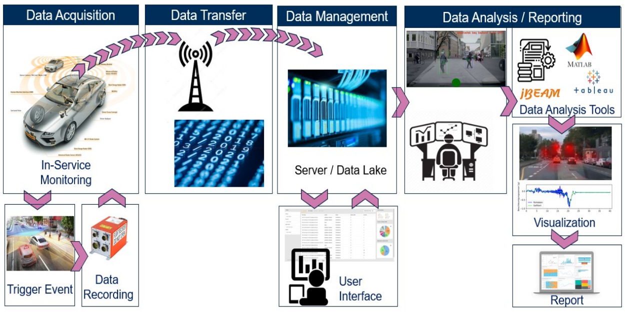

Figure 1. ISMA Dataflow Framwork

B. Monitoring and Data Collection Approaches

One possibility to reduce the amount of data and the analysis effort is real-time data monitoring during operation. Basically, the two different methods that can be used to trigger data collection are rule-based approach (e.g., a threshold exceeding case analysis) and data driven approach [15]. Data collection depends upon multiple different factors and their possible alternative choices as shown in the below Table 1. For many data collection pipelines presented in literature, the dependent factors are theirs inputs, along with the types of said inputs, the applications being considered, trigger function definitions, and reference systems. And for each of these factors, there are different alternatives already applied in various applications. Some of them are mentioned in the different rows of the Table 1. As monitoring and collecting data mainly depends upon the reference system, this is the most important factor of all.

TABLE I. DEPENDEMT FACTORS AND ALTERNATE CHOICES FOR MONIOTORING AND DATA COLLECTION.

Inputs	Application	Trigger Function	Reference System
Real world or Simulation Scenarios	Cut-in, Lange Change	Object state & Position	Reference Map
Urban or Highway Scenarios	Object state uncertainty	Driving behavior error	Virtual Scenario
Offline or Online Streams	Driving path and Trajectories	Trajectory deviation	Drivers Input
Single or Fleet of Vehicles	Anomalies & Corner Cases	Corner Cases	Complementary Sensors

1) *Rule based approach:* One possible method for triggering data collection is when the measurements or derived attributes exceeds certain threshold values. Thresholding events are equivalent to leading indicators as they are known in the field of safety assessment.

The advantage with rule-based approach is that it is simple to define and develop. In many cases, they are based on semantics of the scene. Safety critical cases can be triggered with simple rules on derived feature for e.g., setting threshold on the speed of pedestrian crossing the road. It is also possible to use model parameters and model confidence to set the triggers. The challenges with rule-based approaches are in both the selection of relevant subset of rules and in finding appropriate triggers for AD functions. Below are some of the rule-based examples.

a) Lane mark Recognition [16]: In this rule-based approach the estimated lane marks, gathered from sensorics inputs, are compared to equivalent marks from online map data and GPS localization, as shown in Figure 2. Both left and right lane marks are used for this comparison.The trigger is then defined using the deviation between the lane marking recognition and the lane markings retrieved from map data. Selected scenes where the recognition algorithm does not perform well are saved for analysis.

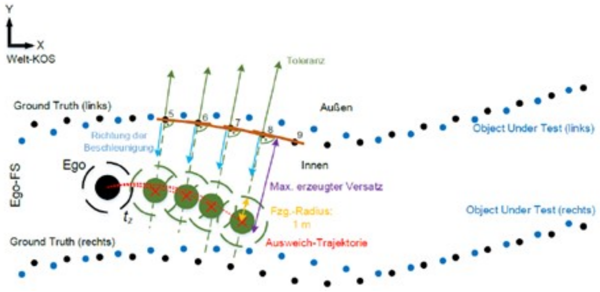

Figure 2. Ground truth (black dots) and estimated lane marks (blue dots) with ego vehi cle position (black and green dots)

b) Trajectory prediction: We used prediction error with pedestrian trajectories. Some of the example cases are shown in Figure 3. A high prediction error with respect to future positions causes a trigger to be generated and can be used to filter out special and atypical cases.

Figure 3. Cases showing mismatch between pedestrian actual future position (green track) and predictions (ged track).

2) *Data driven approach:* This method can be useful for detecting anomalies in a specific scenario by first establishing the profile of a “nominal” vehicle behavior and then mathematically identifying outliers. This method does not require a problem to be known in advance in order to detect the outliers (unlike an analysis of exceedances), but there are also certain limitations. The problems detected are not clearly contextualized and once detected, extensive expert analysis is required to understand patterns of causality. Data driven methods, on the other hand, are easier to implement but lack contextualization. Below is an example of the data driven approach.

a) Success and failure case estimation [17]: This is a data driven approach designed to predict failures as early as possible by using sensor data from up to ten seconds before each disengagement. In this work different sensor measurements are taken as input for categorization of success or failure cases. They categorize the event into success and failures cases which is very similar to classifying it as normal and anomaly.

V. SUITABLE ISMA TRIGGERS DEVELOPED WITH AI FUNKTIONS

In our examples, we mainly focus on developing our triggers using mostly rule-based approaches along with a few data driven approaches. We will now briefly describe the development of our various example triggers.

A. *Pedestrian detection:*

The pipeline for pedestrian detection and trigger definition is shown in Figure 4. Here the trigger is defined to check presence of pedestrian within the defined Region Of Interest (ROI).

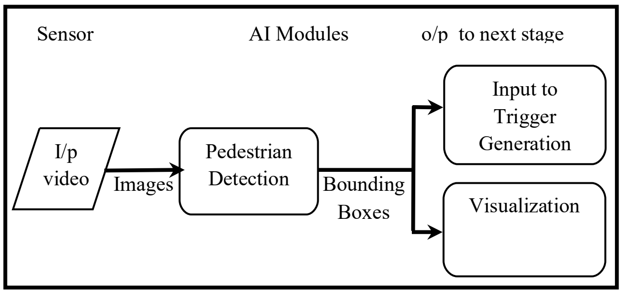

Figure 4. Pipeline for the pedestrian detection on video stream and trigger generation

Some of the active trigger examples with pedestrian detection are shown below in Figure 5, where the pedestrians are withing the defined region of interest, hence satisfying defined triggered condition.

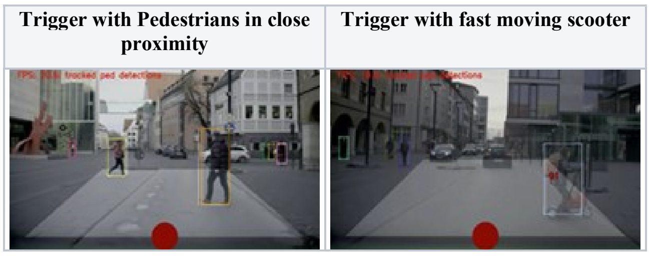

Figure 5. Trigger examples with pedestrian detections

B. *Pedestrian trajectory prediction:*

The algorithmic pipeline for trajectory prediction, which takes both video stream and vehicle speed as inputs to predict future paths, is shown in Figure 6. The inputs to the trajectory prediction are object detections and multiple object trackings. The past positions of tracked objects are used in the algorithm to predicts future paths.

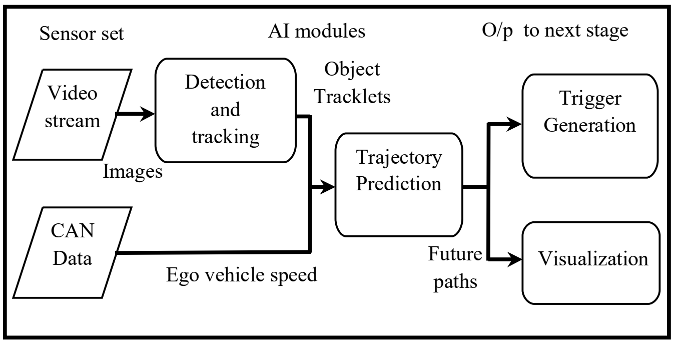

Figure 6. Pipeline for the pedestrian trajectory prediction on video stream and trigger generation.

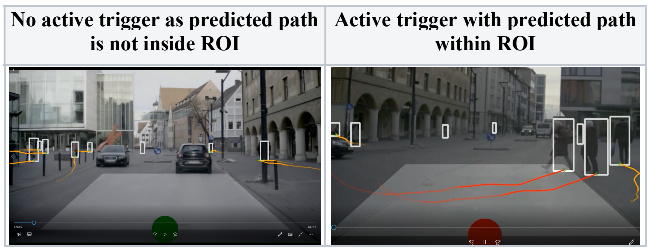

Figure 7. Trigger examples with trajectory prediction

The trigger is then activated using the definition as the predicted position within, and crossing, the center of the defined ROI. Some of the examples are shown in Figure 7.

C. *Event Detection:*

This AI function works on a sequence of input images and estimates the category of events going to happen (pipeline as shown in Figure 8) [18]. The trigger is defined to flag all of these anomaly cases that are considered as rare and unique driving situations, which can further be used for model improvement through continuous updates.

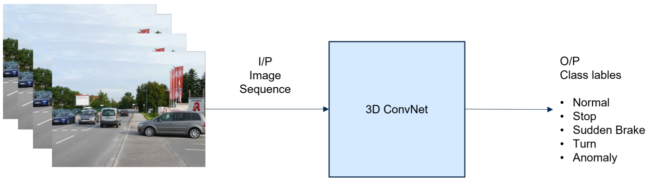

Figure 8. Pipeline for the Event detection on video stream.

Some of the examples, belonging to known event classes, are normal driving, sudden breaking and turning. It also can also predict anomaly events such as: camera falling, heavy jittering due to driving on uneven roads and scenes with low visibility. Examples of event classification and anomaly events are shown in Figure 9.

Figure 9. Examples of Anomaly event classification.

D. *Event Discovery:*

Here we applied an approach called Generalized Event Discovery (GED), which is an extension of "Generalized Object Discovery (GOD)" [19]. This is a fairly new approach to estimate the number of classes in the unlabeled data. This approach can categorize repeatedly occurring new objects or events, by attribute clustering from the unlabeled set. An illustration of this pipeline can be found in Figure 10.

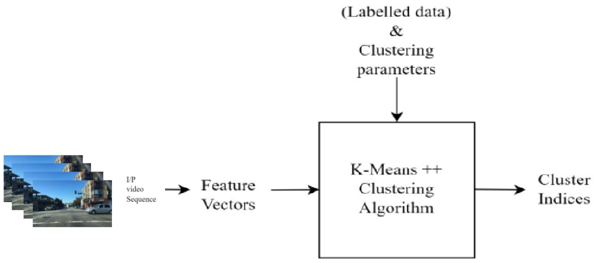

Figure 10. Pipeline for the Event discovery on video stream.

In case of anomaly detection or unknow class of event detection, approaches are focused on the boundary elements with respect to the distribution of the learned data. Furthermore, in GED approach, similar unique events are grouped together, and can later be labelled and used for continuous updates and learning. Figure 11 shows two example situations discovered in the driving data: on the left, a near accident case, and on the right a sudden breaking case is depicted. These kinds of samples can be used as either triggers or filters.

Figure 11. Example of safety critical event cases.

VI. PERFORMANCE EVALUATION AND ANALYSIS

We evaluate our approaches using prepared sub-sequences from BDD100k video data. The total number of sub-sequences are around 14000. The overall classification performance of different event categorization is 79.85%.

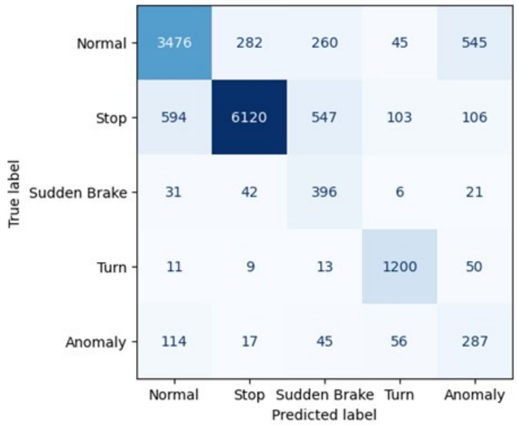

Figure 12. Confusion Matrix for Event detection Performance.

The above confusion matrix plot shows the overall correct classification of event categories. The off-diagonal elements in the table show the different classifications from the true class labels. The top right corner element shows that up to 10% data is categorized as anomaly in comparison to the normal labeled data. These samples can be selected using a defined trigger, some of them can be safety critical in nature. With further analysis we can select out safety critical cases from these and use them for other systems such as self-adaptive systems or for DevOps process for continuous learning.

VII. CONCLUSION

Safety assessment and validation for HAD systems is still very challenging considering the system complexity and cost of deployment. In this paper, we presented an In-Service Monitoring and Assessment approach as a new method for safety validation of automated driving functions under real world conditions. This article covers relevant literature on verification and validation, different approaches for monitoring of HAD Systems during operation and state of the art development of triggers. Along with rule-based approaches, we also covered various data driven monitoring approaches to identify rare, unusual, and critical driving situations.

Future work on this topic includes the exploration of appropriate sets of triggers, defining suitable metrices for their evaluation, and selection of context specific trigger subests. Use of safety critical data for continuous learning and improvement is also a topic that needs further study.

 ISBN: 978-1-68558-101-5

ACKNOWLEDGMENT

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project "VVM - Verification and Validation Methods for Automated Vehicles Level 4 and 5". The authors would like to thank the consortium for the successful cooperation.

REFERENCES

[1] SAE International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," 2018.

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

[3] W. Wachenfeld, and H, Winner, "The new role of road testing for the safety validation of automated vehicles. Automated driving: Safer and more efficient future driving", pp.419-435, 2017.

[4] N. Marko, E. Möhlmann, D. Ničković, J. Niehaus, P. Priller,and M. Rooker, "Challenges of engineering safe and secure highly automated vehicles", arXiv preprint arXiv:2103.03544, 2021.

[5] J. E: Stellet,T. Brade, A. Poddey, S. Jesenski, and W. Branz, "Formalisation and algorithmic approach to the automated driving validation problem", IEEE Intelligent Vehicles Symposium (IV) (pp. 45-51). IEEE, June 2019.

[6] German Aerospace Center: PEGASUS-Project https://www.pegasusprojekt.de/en/home, 2019, Access 22.03.2021.

[7] S. Riedmaier, T. Ponn, T., D. Ludwig, B. Schick, B. and F. Diermeyer, "Survey on scenario-based safety assessment of automated vehicles". IEEE access, 8, pp.87456-87477, 2020..

[8] P. Junietz, W. Wachenfeld, W., K. Klonecki,and H. Winner,"Evaluation of different approaches to address safety validation of automated driving", 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 491-496). IEEE, November, 2018.

[9] W. Wachenfeld, and H. Winner, H., "Virtual assessment of automation in field operation a new runtime validation method", In Workshop Fahrerassistenzsysteme (Vol. 161), September, 2015.

[10] C. Wang. "Silent Testing for Safety Validation of Automated Driving in Field Operation". In Technische Universität Darmstadt. Dissertation - (Retrieval date: 01.06.2022), 2021.

[11] Tesla: What is Shadow Mode Tesla Autonomy, https://www.youtube.com/watch?v=SAceTxSelTI, 2019.

[12] VMAD-SG3-20-02 Concept and agenda for the VMAD SG3 technical workshop on ADS in-service monitoring and reporting. 17th March 2022.

[13] FAA Order 8000.369C - Safety Management System 2020-06-24, PDF.

[14] M. Haiber, et al. „Learning Driven Product Lifecycle for Automated Driving Systems ", Systems Engineering day. 12.11.2021. https://www.tdse.org/, PDF.

[15] A. Scarinci, "Monitoring safety during airline operations: A systems approach", Doctoral dissertation, Massachusetts Institute of Technology, 2017.

[16] Wadim Tribelhorn "Conceptual design and prototypical implementation of a "Silent Testing" method for automated lane marking detection evaluation for automated vehicles" Master's thesis,TU Damstadt, Nr. 703/18,p. 68, 2018.

[17] C. B. Kuhn, M. Hofbauer, M., G. Petrovic,and E. Steinbach, "Introspective black box failure prediction for autonomous driving". IEEE Intelligent Vehicles Symposium (IV) (pp. 1907-1913). October, 2020.

[18] G. Yu, S. Wang, Z. Cai, X. Liu, C. Xu, C. and C. Wu, "Deep anomaly discovery from unlabeled videos via normality advantage and self-paced refinement", In Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (pp. 13987-13998), 2022.

[19] S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Generalized category discovery". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7492-7501), 2022.

[20] C. Wang,K. Storms, and H. Winner, H., "Online safety assessment of automated vehicles using silent testing", IEEE Transactions on Intelligent Transportation Systems, 23(8), pp.13069-13083, 2021.

[21] A. Koenig, K. Witzlsperger, F. Leutwiler, and S. Hohmann, "Overview of HAD validation and passive HAD as a concept for validating highly automated cars",Automatisierungstechnik, 66(2), pp.132-145. 2018.

[22] J. A. Bolte, A. Bar, D. Lipinski, and T. Fingscheidt, "Towards corner case detection for autonomous driving," In 2019 IEEE Intelligent vehicles symposium (IV), pp. 438-445, June, 2019.