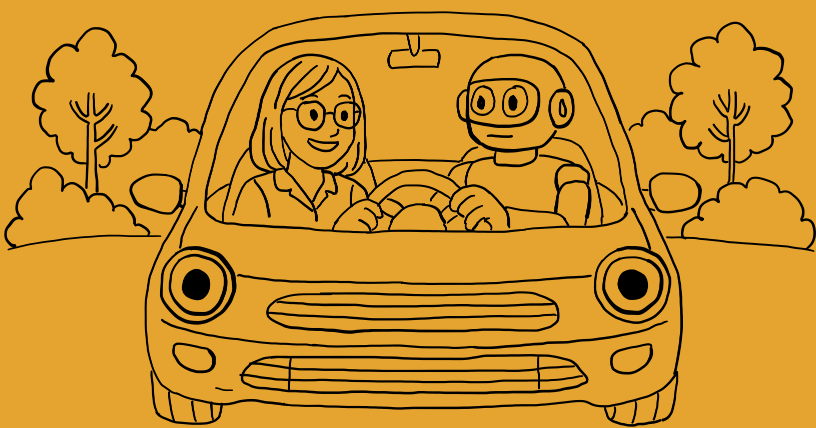
# Understanding the Interplay Between the Driver, the Vehicle, and the Environment for Adapting Driving Automation



a PhD dissertation by **Anaïs Halin**

supervised by Prof. Marc Van Droogenbroeck
& by Prof. Christel Devue

January 2026

*This dissertation has been submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Engineering Sciences and Technology.*

# Jury Members

**Prof. Marc Van Droogenbroeck** *Advisor*
Department of Electrical Engineering and Computer Science,
University of Liège, Belgium

**Prof. Christel Devue** *Co-Advisor*
Department of Psychology, University of Liège, Belgium

**Prof. Pierre Geurts** *President*
Department of Electrical Engineering and Computer Science,
University of Liège, Belgium

**Prof. Mario Cools**
Department of Architecture, Geology, Environment and Constructions (ArGEnCo), University of Liège, Belgium

**Dr. Clémentine François**
Ephyca, Belgium

**Prof. Frank O. Flemisch**
Institute of Industrial Engineering and Ergonomics (IAW), RWTH Aachen University, & Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE), Germany

**Prof. Christer Ahlström**
Department of Biomedical Engineering, Linköping University (LiU),
& Swedish National Road and Transport Research Institute (VTI),
Sweden

# Abstract

Since the invention of the automobile at the end of the 19th century, driving has continually evolved. From rudimentary vehicles consisting of little more than an engine, a seat, and wheels, today's cars have become technological marvels equipped with hundreds of sensors and intelligent algorithms. Consequently, driving has transformed into a complex activity involving multiple interacting entities: the human driver, the vehicle automation, and the driving environment. Despite major technological progress, how to best combine driving automation and driver monitoring systems to dynamically allocate driving tasks for safety and comfort purposes remains a key research challenge. Achieving such adaptive driving automation requires a deep understanding of the interplay between the driver, the vehicle, and the environment.

**Part I** describes the context of this thesis, tracing the evolution of the automobile from mechanical innovation to the integration of driving automation and driver monitoring. It also reviews the state of the art in driver monitoring, with a particular focus on mental workload and distraction.

**Part II** presents human studies conducted in a driving simulator to examine whether drivers' cognitive distraction and the complexity of the driving environment influence reliance on Adaptive Cruise Control (ACC) and whether such reliance affects driving performance. Furthermore, it investigates whether and how physiological and behavioral indicators reflect drivers' cognitive distraction under varying traffic conditions and ACC use. Specifically, three Electrodermal Activity (EDA)-based and three gaze-based indicators were analyzed.

**Part III** introduces engineering approaches for analyzing the driving environment. In particular, it presents a novel Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) setup in which computer vision models adapt on the fly to a dynamic environment divided into cells. To evaluate a method derived from this setup, a new multi-stream, large-scale synthetic semantic segmentation dataset, called DADE, was released.

In addition, a probabilistic approach to domain characterization is proposed, where domains are characterized as probability distributions. A method is presented for predicting the likelihood of different weather conditions from images captured by vehicle-mounted cameras.

**Part IV** proposes a closed-loop framework, called DEV, for risk-aware adaptive driving automation that captures the dynamic interplay between the driver, the environment, and the vehicle. The thesis concludes with insights and future perspectives stemming from this research, aimed at fostering safer and more adaptive human–automation cooperation.

# Acknowledgments

I would like to express my deepest gratitude to all the people who, in different ways, made this PhD possible and rewarding. Although this journey came with its challenges, it has been an enriching experience through which I learned a great deal, gained confidence, built many connections, and made many friends.

First, I would like to thank Jacques Verly, without whom I would never have embarked on a PhD in the first place. Thank you for introducing me to the research world and to the field of driver monitoring, which became the foundation of my work. Through the endless revisions of my first paper, you taught me how to write in a scientific and rigorous way, and these skills proved invaluable throughout this PhD.

I am profoundly grateful to Marc Van Droogenbroeck for giving me the opportunity to pursue this PhD, and for his trust and guidance. Thank you for understanding what motivated me and for helping shape this project in a way that allowed me to flourish. I am also thankful for the many conferences I was lucky to attend, which provided invaluable professional and personal experiences.

I warmly thank Christel Devue for accepting to be my co-supervisor and for her thoughtful advice and support. I particularly enjoyed the time we spent getting to know each other during ECVP.

Beyond their individual roles, I would like to jointly thank Jacques Verly, Marc Van Droogenbroeck, and Christel Devue for their guidance throughout this PhD journey. Their support, complementary perspectives, and advice helped me grow as a researcher, overcome challenges, and gradually gain the independence and confidence needed to complete this work.

I am grateful to the members of the jury for taking the time to read this dissertation and for their valuable feedback.

I would like to thank FEDER and ARIAC for financially supporting this PhD and making this research possible. I also thank AISIN, in particular Frédéric Burguet and Richard Virlouvet, for providing excellent conditions to conduct my simulator study. A special thank you to Richard for his help whenever the simulator decided to misbehave. I also thank all the participants who contributed to the data collection.

I am thankful to all my co-authors for the collaborations that made this work richer and more pleasant; to all the people who made conferences such amazing experiences; and to everyone involved in ARIAC (at ULiège and partner institutions) for the various events and workshops that combined productive workdays with relaxing and enjoyable moments.

More broadly, I am deeply grateful to all the people I was lucky to meet during the eight years I spent at Montefiore, for creating such a warm, stimulating, and friendly work environment.

I would like to thank Jean-Michel Redouté. It has been a real pleasure to work with you over the years on the Digital Electronics course.

I also thank Quentin and Clémentine, who were the first to welcome me and share their office with me, even though I quickly moved to join my long-time office mate and friend Anthony. I warmly thank all my past and present colleagues from VIULab/Telim: Adrien, Anthony, Benjamin, Jan, Jérôme, Marc, Maxim, Michaël, Philippe, Renaud, and Sébastien. Thank you to this incredible team for the friendly and collaborative atmosphere that made everyday work a pleasure.

Beyond my direct colleagues, there are many other people from Montefiore I would like to thank for greatly contributing to the warm, stimulating, and friendly environment I mentioned earlier. This includes everyone gathering at the Aquarium during lunch, the Rikiki card-game players (including Damien, Yann, Arnaud, Axelle, Maxime, Florent, and Ana), but also Marie, Laurine, Nicolas, Antoine, Pascal, Antonio, and all the others whom I cannot mention here, lest this acknowledgment become interminable.

Finally, I thank my parents for their unwavering emotional and practical support, and for believing in me even when I doubted myself. I of course also thank Loïc for his patience and understanding during busy and stressful periods, and for his support throughout this journey.

# Contents

# List of Acronyms

**EOR** Eyes-Off-Road. 54, 55, 57, 58, 61

**EVs** Electric Vehicles. 133

**GLMM** Generalized Linear Mixed Models. 84, 85, 91

**GNSS** Global Navigation Satellite System. 137, 141

**GRS** General Safety Regulation. 11

**HMI** Human-Machine Interface. 182, 185, 187

**HR** Heart Rate. 37, 43, 45, 49–52, 62

**IMU** Inertial Measurement Unit. 5, 129

**IoT** Internet of Things. 126, 129, 130, 148

**LMM** Linear Mixed Models. 84, 88, 90, 110, 113

**MOR** Mind-Off-Road. 61

**NDRT** Non-Driving Related Task. 33, 54, 55, 76

**NHTSA** National Highway Traffic Safety Administration. 56, 60, 102

**NPE** Neural Posterior Estimation. 157, 160, 161, 166, 167

**NSF** Neural Spline Flow. 157, 161

**ODD** Operational Design Domain. 7, 12, 33, 153, 155–158, 168, 170–172, 177, 181, 184, 192

**PDF** Probability Density Function. 157, 161, 164

**REML** Restricted Maximum Likelihood. 88, 90, 110, 113

**SAE** Society of Automotive Engineers. 6, 7, 9, 15, 26, 31, 33, 66, 67, 73–77, 85, 120, 153, 177, 179, 181, 184, 185, 190

**SBI** Simulation-Based Inference. 157, 161

**SC** Skin Conductance. 102, 103

**SCL** Skin Conductance Level. 17, 98, 101–103, 105, 109–112, 117, 119, 121

**SCR** Skin Conductance Response. 17, 98, 101–103, 105, 109, 111, 112, 117, 119, 121

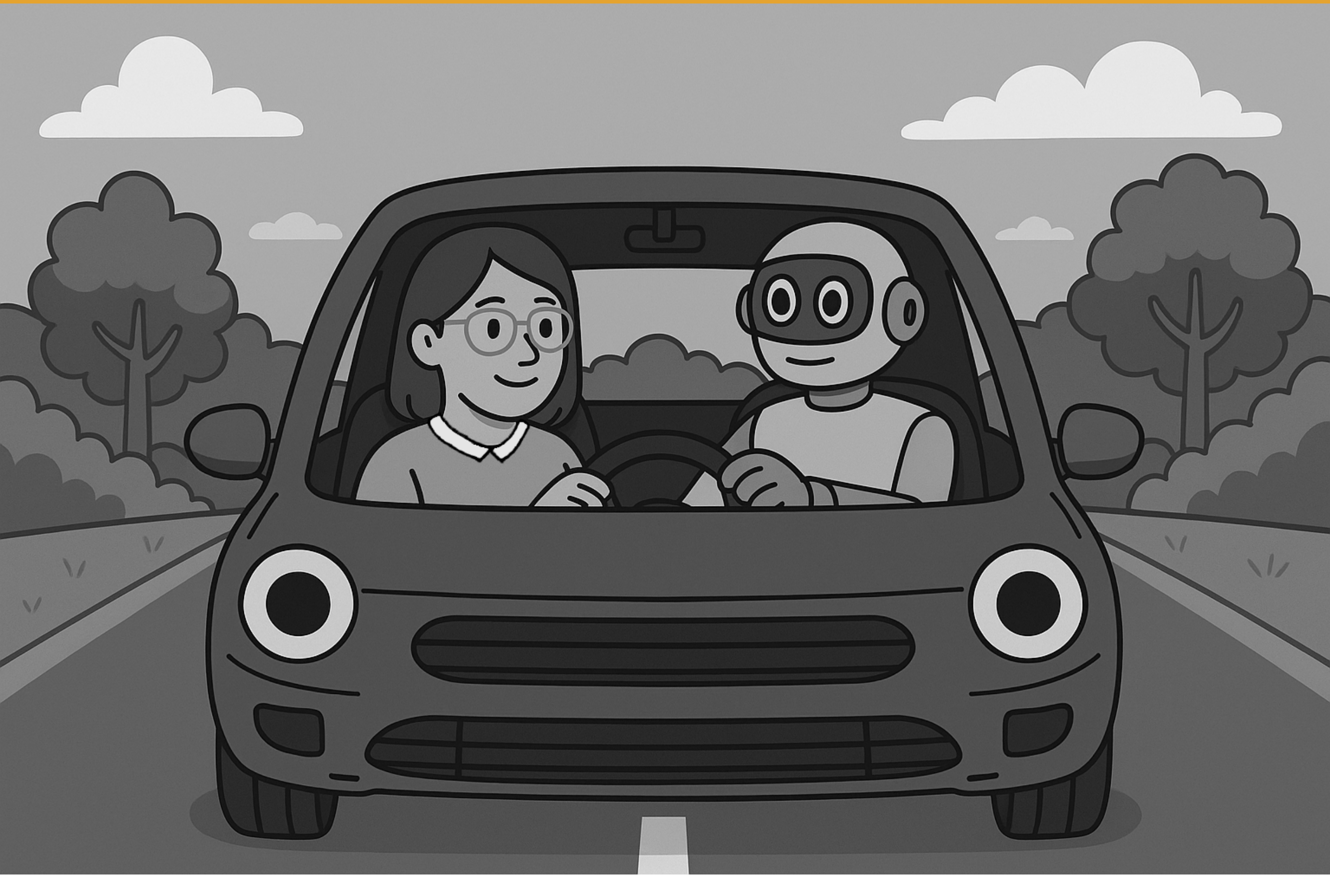
**SDLP** Standard Deviation of Lane Position. 48, 52, 63, 83, 90, 91, 93, 94, 96

**TOR** Take-Over Request. 33, 75, 186

**TTA** Test-Time Adaptation. 131, 132, 134

# Part I

## How Involved Should the Driver Be in Driving Automation?



# 1 Introduction

This thesis addresses the interplay between the driver, the vehicle, and the environment in the era of driving automation. In this introduction, Section 1.1 first presents the context, outlining the evolution of the automotive domain and driving automation—specifically, how vehicles have gained capabilities and progressively taken over parts of the driving task—as well as the evolution of driver monitoring, which ensures driver engagement alongside automation. Section 1.2 then describes the challenges that emerged within this context and defines the main objectives of the thesis. Section 1.3 provides an overview of the thesis' structure and summarizes the main contributions. Finally, Sections 1.4 and 1.5 list the publications on which the content of this thesis is based and additional works published during the PhD journey, respectively.

## 1.1 Context

### 1.1.1 From Mechanical Innovation to Driving Automation

The automotive timeline begins in the late 19th century with the invention of the automobile. The first modern car and the first car produced in series appeared in 1886, when Carl Benz developed the "Benz Patent Motor Car", a gasoline-powered three-wheeled automobile (see Fig. 1.1), and manufactured several identical copies.

**Figure 1.1: Benz Patent Motor Car.** This car build by Carl Benz in 1886 is widely considered the first automobile. It was a practical, marketable automobile for everyday use. It was a two-seater vehicle powered by a gasoline engine. The compact high-speed single-cylinder four-stroke engine was installed horizontally at the rear. The car had a tubular steel frame and three wire-spoked wheels. (Source: `https://group.mercedes-benz.com`)

Early developments in automotive engineering focused primarily on mechanical reliability and usability. A major milestone was the introduction of the *electric starter* by Charles Kettering in 1912 [157], which replaced the hazardous hand-crank mechanism. This advancement not only enhanced convenience but also significantly improved safety by reducing manual strain and risk of backfire. As vehicles became more common, designers started focusing on aerodynamics. Streamlining became a design principle in the 1930s, with manufacturers such as Chrysler and Studebaker adopting smoother body lines to reduce drag. In the same decade, the widespread adoption of all-steel bodies improved both safety and durability.

Although rudimentary speed regulators existed since the early 1900s, the first modern *cruise control* system, capable of maintaining a constant speed without driver input, was commercialized in 1958 on the Chrysler Imperial. Invented by Ralph Teetor [318], it used a mechanical speed sensor linked to the throttle and a dashboard dial to set the desired speed. This innovation can be regarded as the first driver assistance system, *i.e.*, a technology designed to assist the driver in operating the vehicle more safely or comfortably.

In 1959, Nils Bohlin, a Volvo engineer, patented the three-point *seat belt* [30], a simple yet transformative safety device that would save countless lives. Around the same time, Mercedes-Benz introduced the *crumple zone*, typically located in the front part of the vehicle (engine compartment), designed to absorb crash energy and reduce passenger injury during collisions. *Airbags*, conceptualized in the 1950s, underwent experimental testing by Ford and General Motors during the 1970s and appeared in production vehicles in the 1980s. These advances marked a paradigm shift toward occupant protection.

In 1971, Chrysler introduced the “Sure Brake” system on the Chrysler Imperial, the first computer-controlled *Anti-lock Brake System (ABS)* acting on all four wheels. This technology prevented wheel lock-up during hard braking, improving control and stability. Other manufacturers, such as Ford and Toyota, quickly followed and ABS became a standard safety feature in the following decades. Subsequent developments included *traction control* system, introduced in the late 1980s by Toyota, BMW, and Mercedes-Benz to prevent wheel slip during acceleration. In 1995, the introduction of *electronic stability control* by Mercedes-Benz, Toyota, and General Motors further enhanced stability by detecting skids and selectively braking individual wheels to help drivers maintain directional stability.

The 1990s marked the emergence of Advanced Driver Assistance System (ADAS), *i.e.*, systems that employ advanced sensors, such as cameras, radars, or LiDARs[1], together with software algorithms that enable environmental perception and decision-making capabilities to assist the driver more actively. While earlier driver assistance systems, such as cruise control, ABS, or electronic stability control, relied on vehicle-internal inputs (*e.g.*, data from onboard Inertial Measurement Unit (IMU)), ADAS leverage information about the surrounding environment [90]. In 1992, Mitsubishi Motors became the first manufacturer to offer a *LiDAR-based distance warning* system in a production vehicle, which alerted drivers to closing gaps with vehicles ahead. The same model also featured the world’s first *lane departure warning* system,

[1] A *camera* captures images of the scene; a *radar* (radio detection and ranging) uses radio waves to measure the distance and relative speed of surrounding objects; and a LiDAR (light detection and ranging) uses laser pulses to obtain 3D distance measurements.

which used a camera to monitor road markings and triggered an alert if the vehicle drifted out of its lane. In 1995, Mitsubishi expanded these concepts with the "Preview Distance Control" system, which used LiDAR and a camera to automatically modulate throttle and downshift to maintain a safe following distance. Although it could not apply brakes, this was the first *Adaptive Cruise Control (ACC)* system in a production vehicle. Four years later, in 1999, Mercedes-Benz introduced "Distronic", the first radar-based ACC capable of actively braking to maintain distance. This milestone marked the beginning of what the Society of Automotive Engineers (SAE) would later define as Level 1 driving automation (driver assistance) in its J3016 taxonomy, introduced in 2014 and revised in 2021 [283]. According to this taxonomy, conventional cruise control is not classified as Level 1, since it does not respond to external events and thus does not provide sustained vehicle control. Similarly, systems such as ABS and electronic stability control provide only momentary interventions in longitudinal and/or lateral vehicle motion control but do not perform any part of the dynamic driving task (DDT) on a sustained basis. A detailed description of the SAE levels of driving automation is provided in Section 2.2 in the following chapter.

The 2000s marked a period of rapid proliferation and diversification of ADAS technologies. *Forward collision warning*, introduced by Mercedes-Benz in 2000, monitored relative speed and distance to warn drivers of potential collisions. *Blind-spot monitoring*, first deployed by Volvo in 2003, used side-mounted cameras to detect vehicles in adjacent lanes and alert the driver. *Intelligent parking assistance*, launched by Toyota in 2003, enabled semi-automatic parallel parking with minimal driver input. *Automatic emergency braking*, introduced by Honda in 2003, automatically applied brakes when detecting imminent collisions, significantly reducing accident severity. Cicchino [48] compared the rates of police-reported crashes in the US from 2010 to 2014 between passenger vehicle models with automatic emergency braking and the same models without; and showed that automatic emergency braking reduced front-to-rear crash rates by $43\%$ and front-to-rear injury crash rates by $45\%$. In 2004, the *lane keeping assistance* system released by Toyota became the first production system to actively steer the vehicle back into its lane when unintentional departure was detected. Finally, in 2009, *traffic sign recognition* systems entered production in Opel vehicles, using forward-facing cameras to identify and display speed limits and other road signs to the driver.

By the 2010s, advancements in cameras, radars, LiDARs, and on-board computing power enabled more integrated and intelligent ADAS functionalities. Manufacturers began combining longitudinal (ACC) with lateral support (*lane centering assistance*), giving rise to Level 2 partial driving automation. Notably, Mercedes-Benz introduced “Steering Assist” alongside “Distronic” in 2013, while Audi and BMW offered *traffic jam assistance* systems capable of simultaneously controlling steering and speed in slow-moving traffic. A major milestone came in October 2015, when Tesla deployed Autopilot via a software update. This Level 2 system used cameras, radar, and ultrasonic sensors to automatically steer within a lane and maintain distance from other vehicles, and even change lanes with the *lane change assistance* feature under driver supervision. Tesla’s approach showcased the potential of software-driven automation and pushed the entire industry toward more advanced ADAS capabilities.

The late 2010s also saw significant progress in driver monitoring and conditional automation. In 2018, General Motors introduced “Super Cruise”, the first *hands-free highway driving system*. Using high-definition maps and a driver-facing camera, it enabled hands-off driving on pre-mapped highways while ensuring driver alertness. In 2022, Mercedes-Benz launched “Drive Pilot”, the world’s first certified Level 3 system for conditional automation, available to customers under certain conditions on German highways up to 60 km/h (*i.e.*, in heavy traffic) and extended to 95 km/h for flowing traffic in 2025. Unlike Level 2 systems, which require the driver to continuously supervise the environment and remain responsible for vehicle control, a Level 3 system allows the vehicle to perform the entire dynamic driving task under specific conditions. However, the driver must remain ready to intervene upon request.

As of 2025, no highly automated passenger car (*i.e.*, Level 4, as defined by SAE) is available yet for private ownership, although Waymo operates Level 4 driverless taxis in several U.S. cities within defined geofenced areas. At this level, the vehicle is capable of performing the entire dynamic driving task within its Operational Design Domain (ODD) without human supervision, but can only operate autonomously in restricted environments (*e.g.*, specific city zones or highways). The path to Level 5, *i.e.*, full automation under all conditions, remains a long-term objective. Meanwhile, ADAS features such as automatic emergency braking, lane keeping assistance, and ACC have become standard even in mass-market vehicles, substantially improving safety and reducing driver workload. Figure 1.2 synthesizes the major milestones of driver assistance systems in production vehicles available to private owners.

### 1.1.2 From Driver Assistance to Driver Monitoring

As vehicles progressively assumed greater control of the driving task, maintaining appropriate driver engagement became a critical safety concern. Until full automation becomes available—if ever—even the most advanced systems cannot guarantee safety if the driver becomes inattentive, fatigued, or unable to take over when required.

Consequently, research and industry efforts began focusing not only on how vehicles "perceive" [2] and act within their environment but also on how they monitor and interpret the human driver's state. This led to the development of Driver Monitoring System (DMS)—technologies designed to assess the driver's state and issue alerts or trigger countermeasures to prevent accidents. These systems fall into two main categories: passive (indirect) and active (direct). Passive DMS has traditionally relied on vehicle-internal parameters such as steering wheel sensors or driving trajectory to infer the driver's state. Active systems, by contrast, are far more sophisticated. They use in-vehicle cameras and sensors to directly monitor the driver's physiological and behavioral state, analyzing parameters such as eyelid closure, gaze direction, blinking rate, or yawning.

Toyota's luxury brand Lexus launched the world's first active camera-based DMS in 2006. The initial version monitored the driver's gaze direction, and if the driver looked away from the road when a forward obstacle was detected, the system issued a warning. This system focused on *distraction monitoring*, alerting the driver to refocus on the road ahead.

In 2007, Volvo introduced "Driver Alert Control", the world's first dedicated *driver drowsiness detection system*. Rather than observing the driver's face, it used a windshield-mounted camera and sensors to mon-

[2] By vehicles that *perceive* their environment, we mean that they use onboard sensors and algorithms to gather and interpret information about their surroundings, similarly to how humans use their senses.

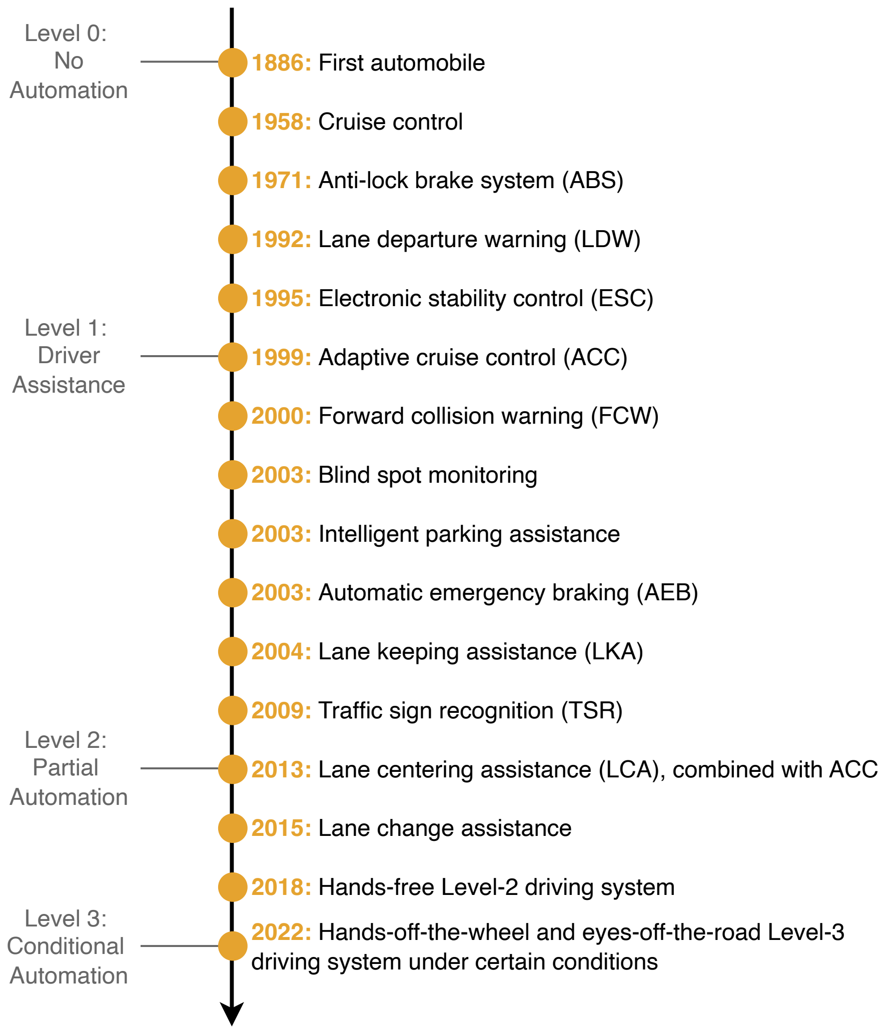



**Figure 1.2: Timeline of major (advanced) driver assistance systems milestones in production vehicles.** The timeline presents key driver assistance systems and Advanced Driver Assistance System (ADAS) introduced in production passenger vehicles available to private owners, along with the first appearance of corresponding Society of Automotive Engineers (SAE) Levels of driving automation.

itor the vehicle's path within the lane. The system evaluated whether the car was being driven erratically (*e.g.*, weaving or drifting due to fatigue or inattention). When such behavior was detected, it emitted a sound signal and displayed a coffee cup icon, suggesting the driver take a break. This system was considered passive, as it detected fatigue from driving behavior rather than by directly observing the driver.

In 2009, Mercedes-Benz introduced "Attention Assist", a driver monitoring feature that analyzed steering inputs and other signals (including road conditions, crosswinds, and driver's interaction with vehicle controls) to detect signs of fatigue. If the driver's steering corrections became inconsistent or erratic, the system sounded a warning and displayed a coffee cup icon on the dashboard, encouraging the driver to rest.

Significant advances in DMS were achieved nearly a decade later, enabled by progress in both hardware and software. The development of infrared, near-infrared, and low-light-sensitive cameras made it increasingly feasible to install in-cabin sensors capable of reliably tracking the driver's eyes and faces under variable lighting conditions. At the same time, advances in computer vision and machine learning improved the robustness of face detection, gaze estimation, and blink recognition.

In 2017, the PSA Group (France) launched Europe's first active camera-based DMS in a production vehicle, anticipating the forthcoming EU safety regulations mandating fatigue monitoring. The "DS Driver Attention Monitoring" system used an infrared camera mounted above the steering wheel to continuously track the driver's eyes (for blink rate), face, and head movements, as well as the vehicle's trajectory. When signs of drowsiness (*e.g.*, heavy eyelids) or distraction (*e.g.*, eyes off the road) were detected, the system alerted the driver.

As "Super Cruise" enabled hands-free driving in 2018, ensuring that the driver remained attentive and able to take over became essential. The system therefore incorporated a driver-facing infrared camera, developed by Seeing Machines, to monitor the driver's head pose and eye gaze and to verify that the driver was monitoring the road whenever semi-autonomous mode (Level 2) was active. This marked a global milestone, integrating driver distraction monitoring for driving automation purposes.

The years 2019-2020 saw the widespread adoption of active DMS across the automotive industry. By the late 2010s, driver monitoring had become a core safety feature. BMW first added an internal camera-based DMS in the 2019 X5 SUV to verify that the driver's eyes remained open and directed toward the road, and Mercedes-Benz transitioned to camera-based attention monitoring with the 2020–2021 S-Class, integrating an infrared driver-facing camera into its system. Many other manufacturers, such as Mazda, Nissan, Ford, Hyundai, and others, introduced similar driver-facing camera systems between 2018 and 2021. This wave of adoption represented a turning point: active driver monitoring moved into the mainstream, driven by the demands of Level 2/3 of driving automation and impending safety regulations.

In 2022, a major milestone was reached as DMS became mandatory in new vehicles in Europe. The European Union's General Safety Regulation (GRS) 2019/2144 [73], effective July 2022, requires all new passenger vehicle models to include a "Driver Drowsiness and Attention Warning (DDAW)" system, with the obligation extending to all new cars sold from July 2024 onward. Furthermore, an "Advanced Driver Distraction Warning (ADDW)" system becomes compulsory for new models launched from July 2024 and for all new vehicles from July 2026. Under the GRS, DDAW is defined as a system that "assesses the driver's alertness through vehicle systems analysis and warns the driver if needed". The law does not prescribe a specific technology. Manufacturers may use passive systems (relying, *e.g.*, on steering patterns or lane positioning) or active systems (with camera-based eye-tracking) to evaluate drowsiness. In contrast, ADDW is defined as "a system that helps the driver to continue to pay attention to the traffic situation and that warns the driver when he or she is distracted". Unlike DDAW's focus on fatigue, ADDW specifically monitors the driver's gaze direction. EU Delegated Regulation 2023/2590 [74] requires real-time gaze tracking. For instance, the system must detect if the driver's eyes are diverted downward (toward a phone, lap, or console) beyond a prescribed duration—approximately 6 seconds at urban speeds (20–50 km/h) or 3.5 seconds at higher speeds ($>$ 50 km/h). Warnings must combine a visible alert with audible and/or haptic signals that escalate until the driver refocuses attention on the road. The regulation however permits drivers to manually deactivate the DDAW system warnings at each drive. This regulatory step cemented driver monitoring as a standard safety feature, aiming to reduce accidents caused by drowsy or distracted driving.

Europe is thus making driver-monitoring cameras as integral to modern vehicles as seatbelts or electronic stability control once were.

Each of these milestones represents a major step in the evolution of DMS. From early passive methods based on vehicle trajectory or steering behavior to modern active camera-based systems, the technology has evolved from luxury flagships to widespread adoption. Today, production vehicles routinely integrate DMS as a core component of advanced safety and driving automation systems. These advances have been made with the goal of improving road safety by helping to ensure that drivers remain alert and attentive [73]. However, although these systems are widely recognized to effectively detect driver fatigue and distraction [80, 221], there is still very limited evidence that they actually have a real impact on road safety [23, 135]. Some studies even question the effectiveness of the warnings issued following the detection of fatigue or distraction [81, 167]. Further research is thus needed to use DMSs in a way that truly improve road safety.

## 1.2 Challenges and Objectives

Looking at the automotive timeline, it is clear that vehicle design is no longer limited to engineering aspects such as engine performance or aerodynamic efficiency. Increasingly, it relies on intelligent algorithms capable of analyzing the driving environment for advanced driver assistance systems that can support—or even partially take over—the driving task. In parallel, in-vehicle monitoring systems, in particular DMS, are designed to assess the driver's state in real time, aiming to mitigate risks associated with lapses in driving abilities due to, *e.g.*, fatigue or distraction. These systems monitor not only the driver while manually driving, but also when driving with automation systems to assess, *e.g.*, readiness to take back control if required. As a result, driving has evolved into a complex activity involving multiple interacting entities: the human driver, the vehicle's automation, and the surrounding or driving environment.

Despite significant technological progress, many challenges persist regarding the safe and efficient integration of driving automation. A central difficulty lies in the dynamic allocation of control between the driver and the vehicle. Most current driving automation systems can operate only within a limited ODD [283], defined by specific environmental, infrastructural, and contextual conditions (*e.g.*, road type, weather, visibility, or traffic). Therefore, automation features cannot be activated under all circumstances, and their activation or deactivation remains largely under the driver's responsibility. Under these conditions, both overtrust (or over-reliance) and undertrust in automation are problematic. Drivers whose trust exceeds the system's actual reliability may fail to supervise the automation properly or may simply use it inappropriately (misuse) [180, 252]. Conversely, undertrust can lead drivers to disregard available assistance (disuse) [160, 249], thereby negating potential safety benefits. Related issues include mode confusion, in which drivers are uncertain about the system's current operational mode or its capabilities [356]. Moreover, human performance in this context is also strongly influenced by task demand, which is shaped by both the level of driving automation and the driving environment. Humans perform optimally within a moderate range of workload [53, 368]. Excessive task demands—such as in sudden takeover situations or dense urban traffic—can lead to overload, stress, and performance degradation [319], while low demands—such as during prolonged monitoring of automation or monotonous highway driving—can result in underload, boredom, fatigue, or reduced vigilance [77, 168]. When drivers supervise automation for extended periods, they tend to lose situation awareness and may react more slowly when intervention is required. Ensuring that the driver remains appropriately involved when interacting with driving automation is therefore essential for safety. The interplay between the driver, the vehicle's automation, and the driving environment constitutes a dynamic and interdependent system that requires careful management.

Based on the current context and identified challenges, this thesis aims to advance our understanding of how driver monitoring can contribute to reducing risk across the different levels of driving automation and to define the degree to which the driver should remain involved in automated driving. The goal is to promote a human-centered approach to driving automation, in which automation serves the driver rather than the driver adapting to automation.

A promising way to achieve this objective is through the development of *adaptive automation* [36], capable of adjusting dynamically not only to the driver state but also to the driving situation. Adaptive automation dynamically allocates control between the driver and the vehicle according to contextual factors such as the driver state or the complexity of the driving environment. Such adaptability, however, depends on our ability to understand, analyze, and characterize both the driver and the environment, as well as their interactions with the vehicle's automation. This perspective therefore calls for a deeper understanding of the interplay between the driver, the vehicle, and the environment.

With respect to the driver, it is essential not only to monitor their states but also to understand how they interact with automation under varying conditions. This understanding is key to designing adaptive automation that responds to drivers' needs and fosters both trust and acceptance. These aspects are explored in Part II of this thesis.

Regarding the driving environment, the objective is to analyze its content (for instance, the number of vehicles, pedestrians, or obstacles) and to characterize its properties (such as weather or visibility conditions) to assess its complexity. This characterization is essential, as environmental conditions influence both the range of driving automation features that can be safely used and the driver state. These aspects are investigated in Part III.

Finally, Part IV introduces a closed-loop framework that jointly considers the driver, the environment, and the vehicle to dynamically adapt the level of automation. This framework aims to reduce risk and enhance safety by integrating three key components: the driver state or level of involvement, the complexity of the environment, and the available driving automation features.

The present part, Part I, introduces the thesis before providing a comprehensive review of the state of the art in driver monitoring, establishing the foundation for the subsequent parts of this work.

## 1.3 Detailed Thesis Outline, Contributions, and Key Findings

This thesis is structured into four main parts, as synthesized in Fig. 1.3. First, Part I provides the general introduction in Chapter 1 and reviews the state of the art in driver monitoring in Chapter 2. Then, Part II focuses on human studies to analyze driver behavior and includes Chapters 3 and 4. Part III presents engineering approaches to analyze the driving environment and includes Chapters 5 and 6. Finally, Part IV introduces an integrated framework for adaptive driving automation in

Chapter 7, and Chapter 8 concludes the thesis. The remainder of this section summarizes the content of each chapter (excluding the general introduction and conclusion) and highlights their main contributions.

**Chapter 2** reviews the state of the art in driver monitoring, focusing on how to characterize the driver state. Since driver monitoring is increasingly intertwined with driving automation, the chapter first clarifies its role across the six SAE levels of driving automation. Then, it surveys and synthesizes existing approaches to provide a unique, structured, polychotomous view of the many characterization techniques, focusing on two key dimensions of driver state: mental workload and distraction. The polychotomous view of driver monitoring is presented through a pair of interlocked tables that relate these states to their indicators (*e.g.*, the eye-blink rate) and the sensors that can access each of these indicators (*e.g.*, a camera). The tables factor in, not only the effects linked directly to the driver, but also those linked to the (driven) vehicle and the (driving) environment. These tables serve as a resource for researchers, equipment providers, and vehicle manufacturers to identify available options for developing advanced DMS and to highlight areas for further research and innovation.

**Contributions**

1. A structured overview of the role of driver monitoring across the six SAE levels of driving automation.
2. The introduction and formal definition of the notion of *indicator*.
3. Two interlocked tables connecting driver states, their indicators, and the sensors providing the corresponding data, accompanied by detailed explanations.

**Chapter 3** presents a simulator study that adopts a human-centered approach to investigate how drivers interact with ACC. Specifically, this chapter examines whether drivers' cognitive state and the complexity of the driving environment influence reliance on ACC, and whether such reliance affects driving performance. In this study, participants operated a simulated vehicle equipped with ACC across six predefined driving

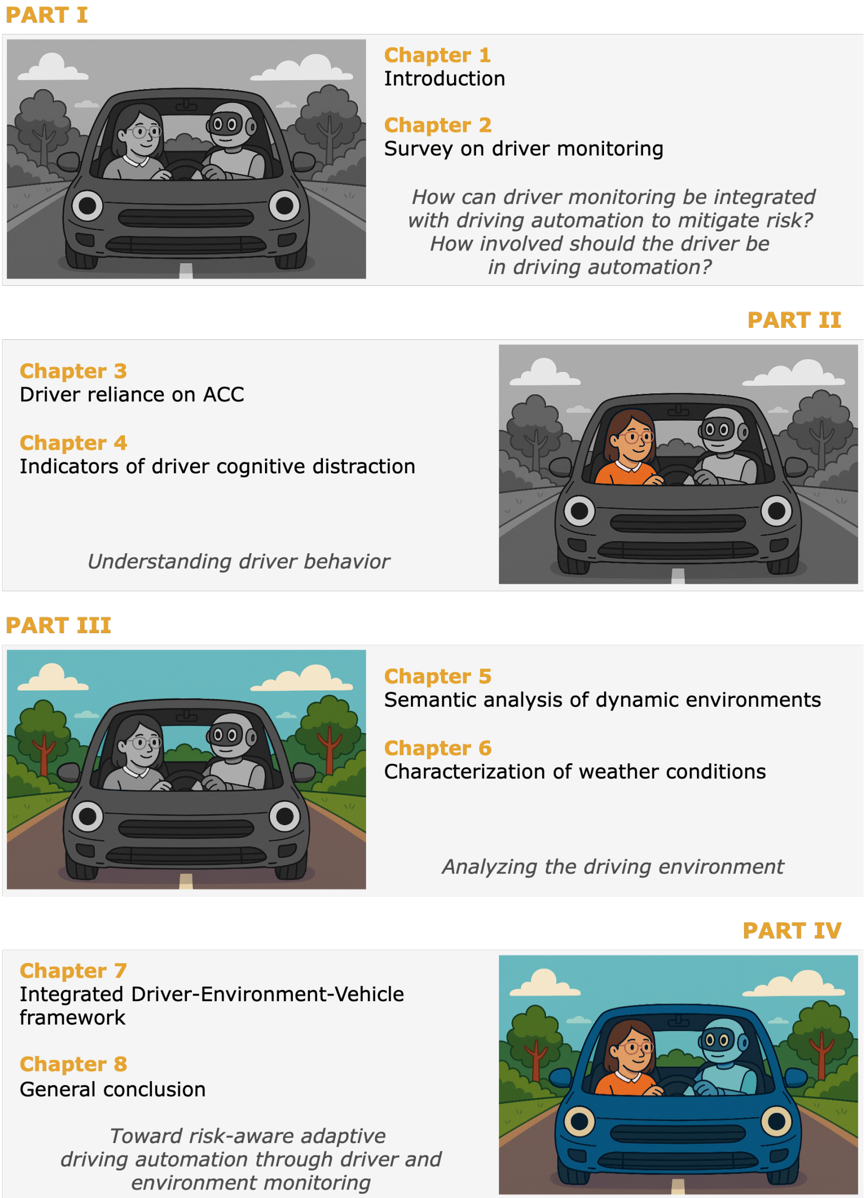



**Figure 1.3: Overview of the thesis structure.** We first introduce the research context and review the state of the art in driver monitoring (Part I). We then focus on human studies examining driver behavior (Part II) and on engineering approaches for analyzing the driving environment (Part III). Finally, we propose an integrated framework for adaptive driving automation and conclude the thesis (Part IV).

scenarios that varied in traffic conditions, while either performing a cognitively demanding task or not. They were instructed to respect speed limits and could freely activate or deactivate ACC.

> **Contributions and Key Findings**
>
> (1) The complexity of the driving environment significantly influences ACC use, whereas cognitive load does not.
>
> (2) ACC use improves lateral control and affects compliance with speed limits but does not significantly influence lane-changing behavior.
>
> (3) Overall, this chapter contributes to understanding how driver cognitive distraction and traffic conditions shape driver reliance on ACC and the resulting driving performance.

**Chapter 4** investigates whether and how Electrodermal Activity (EDA) and gaze parameters reflect driver cognitive distraction under varying traffic conditions and ACC use, through a driving simulator study (the same as in Chapter 3). Participants completed six driving scenarios that combined two levels of cognitive distraction (presence or absence of a mental calculation task) and three levels of driving environment complexity (different traffic conditions). Throughout the experiment, they were free to activate or deactivate the ACC, resulting in two levels of ACC use. We analyze (1) three EDA-based indicators of cognitive distraction: mean skin conductance level (SCL), mean amplitude of skin conductance responses (SCR amplitude), and rate of skin conductance responses (SCR rate); as well as (2) three gaze-based indicators: percent road center, horizontal gaze dispersion, and vertical gaze dispersion.

> **Contributions and Key Findings**
>
> (1) All three EDA-based indicators are significantly influenced by cognitive distraction and ACC use, whereas environment complexity significantly influences SCL and SCR amplitude, but not SCR rate.

2. All three gaze-based indicators are significantly influenced by ACC use, whereas environment complexity only significantly influences vertical gaze dispersion, but not percent road center and horizontal gaze dispersion, and cognitive distraction significantly influences percent road center and vertical gaze dispersion, but not horizontal gaze dispersion.

3. Surprisingly, cognitive distraction reduces road center gaze and increases vertical dispersion. However, complementary analyses reveal that these effects primarily occur between mental calculations periods, whereas active mental calculations phases are characterized by a temporary increase in gaze concentration.

4. Overall, this chapter provides an analysis of physiological and behavioral indicators of driver state, examining how they are jointly influenced by driver cognitive distraction, environment complexity, and ACC use.

**Chapter** 5 presents a new multi-stream large-scale synthetic semantic segmentation dataset, called *DADE*, motivated by a novel Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) setup in which models adapt on the fly to a dynamic environment divided into cells. *DADE* was used to evaluate a real-time adaptive student-teacher method that leverages the multiple streams available in each cell to quickly adapt to changing data distributions. Cells are defined based on location and weather conditions. This multi-stream approach yields better scene analysis performance than a single-stream baseline.

**Contributions**

1. A new Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) setup in which models adapt on the fly to a dynamic environment divided into cells.

2. A novel real-time adaptive student-teacher method to aggregate knowledge across different agents evolving in the same cell.

3. A new synthetic dataset, called *DADE*, for the semantic segmentation task on board autonomous vehicles.

**Chapter** 6 introduces a probabilistic approach to domain characterization, in which domains are characterized as probability distributions. The chapter presents a method for predicting the likelihood of different weather conditions from images captured by vehicle-mounted cameras, by estimating distributions of physical parameters using normalizing flows. Experiments conducted in the context of autonomous vehicles, focusing on predicting the distribution of weather parameters, demonstrate how domains can be characterized both by physical parameters (absolute characterization) and by arbitrarily predefined domains (relative characterization). Finally, the chapter evaluates whether a system can safely operate within a target domain by comparing it to multiple source domains where safety has already been established.

**Contributions**

1. A novel probabilistic methodology to characterize domains in the case of autonomous vehicles driving in various weather conditions.
2. Demonstration that simulation-based inference (normalizing flows) can effectively estimate weather parameter distributions, with comparison of different backbones for features extraction.
3. A method for characterizing a new target domain as a mixture of source domains.

**Chapter** 7 presents the *DEV* framework, a closed-loop framework for risk-aware adaptive driving automation that captures the dynamic interplay between the driver, the environment, and the vehicle. The framework promotes continuous adjustment of the operational level of automation based on a risk management strategy. The real-time risk assessment supports smoother transitions and effective cooperation between the driver and the automation system. Furthermore, we introduce a nomenclature of indexes corresponding to each core component—namely driver involvement, environment complexity, and vehicle engagement—and discuss how their interaction influences driving risk. The *DEV* framework provides a comprehensive perspective to align multidisciplinary research efforts and guide the development of dynamic, risk-aware driving automation systems.

**Contributions**

1. A closed-Loop framework, called *DEV*, for risk-aware adaptive automation of driving integrating the driver, the environment, and the vehicle.

2. A nomenclature of indexes corresponding to each core component of the framework: driver involvement, environment complexity, and vehicle engagement.

## 1.4 Publications

Apart from the introduction and conclusion, the scientific content of this manuscript is predominantly derived from original research contributions published in peer-reviewed venues. The manuscript is organized around the following publications, listed in chronological order:

[115] *Survey and Synthesis of State of the Art in Driver Monitoring*, **Anaïs Halin**, Jacques G. Verly, and Marc Van Droogenbroeck. Sensors, 2021.
→ Chapter 2.

[92] *Multi-Stream Cellular Test-Time Adaptation of Real-Time Models Evolving in Dynamic Environments*, Benoît Gérin*, **Anaïs Halin***, Anthony Cioppa*, Maxim Henry, Bernard Ghanem, Benoît Macq, Christophe De Vleeschouwer, and Marc Van Droogenbroeck. In IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Workshop on Autonomous Driving, 2024.
→ Chapter 5.

* Equal contributions.

[112] *Physically Interpretable Probabilistic Domain Characterization*, **Anaïs Halin***, Sébastien Piérard*, Renaud Vandeghen, Benoît Gérin, Maxime Zanella, Martin Colot, Jan Held, Anthony Cioppa, Emmanuel Jean, Gianluca Bontempi, Saïd Mahmoudi, Benoît Macq, and Marc Van Droogenbroeck. In Asian Conference on Computer Vision Workshops (ACCV Workshops), Workshop on AI-based All-Weather Surveillance System (AWSS), 2025.
→ Chapter 6.

[114] *Effects of Cognitive Distraction and Driving Environment Complexity on Adaptive Cruise Control Use and Its Impact on Driving Performance: A Simulator Study*, **Anaïs Halin**, Marc Van Droogenbroeck, and Christel Devue. In International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI), 2025.
→ Chapter 3.

[113] *Are Electrodermal Activity-Based Indicators of Driver Cognitive Distraction Robust to Varying Traffic Conditions and Adaptive Cruise Control Use?*, **Anaïs Halin**, Marc Van Droogenbroeck, and Christel Devue. In Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI), 2025.
→ Chapter 4.

[108] *Gaze-Based Indicators of Driver Cognitive Distraction: Effects of Different Traffic Conditions and Adaptive Cruise Control Use*, **Anaïs Halin**, Adrien Deliège, Christel Devue, and Marc Van Droogenbroeck. In Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI), 2025.
→ Chapter 4.

[109] *DEV: A Driver-Environment-Vehicle Closed-Loop Framework for Risk-Aware Adaptive Automation of Driving*, **Anaïs Halin**, Christel Devue, and Marc Van Droogenbroeck. In Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI), 2025.
→ Chapter 7.

At the beginning of each chapter, a statement indicates whether its content is based, in whole or in part, on a publication and provides the corresponding reference. In addition, my specific contributions as an author are detailed when I am not the sole first author of the publication.

## 1.5 Additional Publications

During my PhD journey, I had the opportunity to engage in valuable collaborations beyond the direct scope of this thesis. These collaborations led to the following co-authored publications:

[134] *Online Distillation with Continual Learning for Cyclic Domain Shifts*, Joachim Houyon, Anthony Cioppa, Yasir Ghunaim, Motasem Alfarra, **Anaïs Halin**, Maxim Henry, Bernard Ghanem, and Marc Van Droogenbroeck. In IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Workshop on Continual Learning (CLVision), 2023.

[256] *Mixture Domain Adaptation to Improve Semantic Segmentation in Real-World Surveillance*, Sébastien Piérard, Anthony Cioppa, **Anaïs Halin**, Renaud Vandeghen, Maxime Zanella, Benoît Macq, Saïd Mahmoudi, and Marc Van Droogenbroeck. In IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Workshop on Real-World Surveillance: Applications and Challenges, 2023.

[259] *Foundations of the Theory of Performance-Based Ranking*, Sébastien Piérard, **Anaïs Halin**, Anthony Cioppa, Adrien Deliège, and Marc Van Droogenbroeck. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.

[258] *The Tile: A 2D Map of Ranking Scores for Two-Class Classification*, Sébastien Piérard, **Anaïs Halin**, Anthony Cioppa, Adrien Deliège, and Marc Van Droogenbroeck. arXiv preprint, 2024.

[111] *A Hitchhiker's Guide to Understanding Performances of Two-Class Classifiers*, **Anaïs Halin***, Sébastien Piérard*, Anthony Cioppa, and Marc Van Droogenbroeck. arXiv preprint, 2024.

[257] *A Methodology to Evaluate Strategies Predicting Rankings on Unseen Domains*, Sébastien Piérard, Adrien Deliège, **Anaïs Halin**,

and Marc Van Droogenbroeck. In IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Workshop on Big Surveillance Data Analysis and Processing (BIG-Surv), 2025.

# 2

# Background and Related Work

**Contents of this chapter**

CONTEXT. Monitoring the state of the driver has become a safety necessity—both to ensure that the driver remains fit to drive and to verify that he or she correctly supervises driving automation or is ready to resume control when required. However, states such as mental workload or distraction are not directly measurable. They must instead be inferred from observable indicators captured through various sensors. The present chapter first discusses the evolving role of driver monitoring across the six SAE levels of driving automation. It then provides a review of driver monitoring, focusing on two key driver states: mental workload and distraction, including its manual, visual, auditory, and cognitive forms. The chapter identifies the indicators that can be used to characterize these states and the sensors allowing access to the values of these indicators. These elements are synthesized into two interlocked tables that map the relationships between driver states, their corresponding indicators, and the sensors used to capture them. One of the conclusions drawn in this work is that a Driver Monitoring System (DMS) should ideally monitor not only the driver (D), but also the environment (E), and the vehicle (V). This insight shaped the structure of the two tables in this chapter and influenced the broader direction of my thesis: Part II focuses on the driver (analyzing both interaction with automation and indicators of driver state), Part III addresses the environment (analyzing its content and characterizing it), and Part IV presents an integrated driver-environment-vehicle (*DEV*) framework.

**Contributions**

1. A structured overview of the role of driver monitoring across the six SAE levels of driving automation.
2. The introduction and formal definition of the notion of *indicator*.
3. Two interlocked tables connecting driver states, their indicators, and the sensors providing the corresponding data, ac-

companied by detailed explanations.

Related publication. This chapter builds upon Halin et al. [115]. The original publication discussed five driver states: drowsiness, mental workload, distraction, emotions, and driving under the influence. In the present chapter, the scope is restricted to mental workload and distraction, as these are the states further investigated in this thesis (in Chapters 3 and 4). Moreover, the survey and synthesis of the state of the art in driver monitoring have been updated, and recent references published after the original publication have been integrated.

## 2.1 Introduction

A report published in 2018 [304] provides the results of an analysis performed on data about the events and related factors that led to crashes of small road vehicles from 2005 to 2007 across the USA. It indicates that the critical reasons for these crashes are likely attributable to the driver (in $94\%$ of the cases), the vehicle ($2\%$), the environment ($2\%$), and unknown causes ($2\%$). An overwhelming proportion of these crashes is thus due to human error. It is widely recognized that most of them could be avoided by constantly monitoring the driver [12, 357], and by taking proper, timely actions when necessary.

Monitoring the driver is thus critically important, and this applies to all vehicles except for those that are fully autonomous, that is, where the driver does not control the vehicle under any circumstances. Given that the average driver will not own a fully-autonomous vehicle for decades to come, driver monitoring will remain critically important during all this time.

This chapter focuses on the topic of driver monitoring, which is usefully viewed as consisting of two successive steps. In the first, one characterizes the driver, or more precisely the state of the driver, and, in the second, one decides what safety actions to take based on this characterization. For example, in the monitoring of drowsiness, the first step might compute the level of drowsiness, whereas the second might check whether this level is at, or will soon reach, a critical level. More generally, the decision process should ideally fuse the various characterization parameters available and predict the future state of the driver based on them. This chapter focuses almost exclusively on the characterization of the state of the driver, that is, on the first step in driver monitoring, which is also the one that is almost exclusively considered in the literature.

By “state of the driver” or “driver state”, we mean, in a loose way, the state or situation that the driver is in from various perspectives, in particular physical, physiological, psychological, and behavioral. To deal with this driver state in a manageable, modular way, we consider a specific number of distinct facets (such as distraction) of this driver state, which we call “driver (sub)states”. In the sequel, “state” thus refers either to the global state of the driver or to one of its facets, or substates. The core of the chapter focuses on the characterization of (sub)states, using indicators (of this state) and sensors (to access the values of these indicators in real time and in real driving conditions). In the example of the (sub)state of distraction, an indicator thereof is the gaze, and it can be accessed using a camera.

Driver monitoring is important, whether the vehicle is equipped with some form of driving automation (except for full automation) or not. In future vehicles, driving automation and driver monitoring will need to increasingly interact, and they will need to be designed and implemented in a synergistic way. While the chapter focuses on driver monitoring (and, more precisely, on its characterization part), it considers and describes, at a high-level, how driver monitoring and driving automation interact at the various, standard levels of driving automation.

The chapter comprises two main phases: (1) it reports on a systematic survey of the state of the art of driver monitoring conducted early 2021 (see Section 2.3); (2) it provides a synthesis of the many characterization techniques of driver monitoring for the (sub)states of mental workload and distraction (see Section 2.5 to Section 2.7). This synthesis leads to an innovative, structured, polychotomous view of the recent developments in the characterization part of driver monitoring. In a nutshell, this view is provided by two interlocked tables that involve the main driver (sub)states, the indicators of these states, and the sensors allowing access to the values of these indicators. The polychotomy presented should prove useful to researchers, equipment providers, and vehicle manufacturers for organizing their approach concerning the characterization and monitoring of the state of the driver.

Section 2.2 describes the standard levels of driving automation, and the role played by driver monitoring for each. Section 2.3 indicates the strategy for, and the results of, our survey of the literature on driver monitoring. Section 2.4 describes the rationale and strategy for expressing the characterization of the driver state as much as possible in terms of the triad of the (sub)states, indicators, and sensors. Section 2.5 provides our innovative, structured, polychotomous view of the characterization part of driver monitoring. Sections 2.6 and 2.7 describe, respectively, two out of the five driver (sub)states that the survey revealed as being the most important. Section 2.8 summarizes, and Section 2.9 concludes the chapter.

**Table 2.1: SAE Levels and role of key actors.** This table shows the role played by each of the four key actors, *i.e.*, driver, driver support features, automated driving features, and driver monitoring, at each of the six SAE Levels of driving automation (from 0 to 5).

| SAE Levels / Actors | 0 No Driving Automation | 1 Driver Assistance | 2 Partial Driving Automation | 3 Conditional Driving Automation | 4 High Driving Automation | 5 Full Driving Automation |
|---|---|---|---|---|---|---|
| **Driver** | Driving and supervising DS features | | | Driving when AD features request it | Driving (if desired) when AD features reach their limits | / |
| **Driver Support Features** | Warning and temporary support | Lateral or longitudinal support | Lateral and longitudinal support | / | / | / |
| **Automated Driving Features** | / | / | / | Driving when AD features permit it | | Driving |
| **Driver Monitoring (DM)** | Monitoring | Monitoring with relevant indicators | | Monitoring fallback-ready driver | Monitoring when driver in control | / |

## 2.2 Driving Automation and Driver Monitoring

In autonomous vehicles—also called self-driving or fully-automated vehicles—driver monitoring plays a critical role as long as the automation allows the driver to have some control over the vehicle. This section describes the interaction between driver monitoring and driving automation in the context of the six levels of driving automation defined by the Society of Automotive Engineers (SAE) [283], ranging from zero (no automation) to five (full automation).

Table 2.1, inspired by the SAE J3016 Levels of Driving Automation Graphic, describes the role of each of the three key actors in the driving task, namely the driver, the driver support features, and the automated driving features, at each of the six SAE levels. We also integrated into this table a fourth actor, that is, driver monitoring, as its role is crucial at all levels except the highest, to ensure that the state of the driver allows him/her[1] to perform the driving task safely, when applicable.

We now discuss some terminology. In Section 2.1, we introduced the term "driving automation" (as a convenient, companion term for driver monitoring) and, in the previous paragraph, the SAE-suggested term "automated driving". While these two terms seem to further add to a

[1] Throughout, we use the inclusive pronoun "he/she" and adjective "his/her" to refer to the driver.

jumble of terms and abbreviations, they both appear in the literature through their corresponding systems, that is, the "Driving Automation System (DAS)" and "Automated Driving System (ADS)". An ADS is a system consisting of the automated driving features, and a DAS is a system that includes, among other things, both driver support features and automated driving features. One could also view the driver support features as constituting a system, but this is not needed here.

In future vehicles with progressively increasing degrees of automation, the development of DASs and, in particular, of ADSs should go hand-in-hand with the development of Driver Monitoring System (DMS). The next four paragraphs complement the information in Table 2.1.

At Levels 0 to 2, the driver is responsible for the driving task and he/she may be aided by a variable number of driver support features such as automatic emergency braking, adaptive cruise control, and lane centering. At Level 1, the driver support features execute the subtask of controlling either the lateral motion or the longitudinal motion of the vehicle (but not both), expecting the driver to perform the rest of the driving task. At Level 2, the driver support features execute the subtasks of controlling both the lateral motion and the longitudinal motion, expecting the driver to complete the object-and-event-detection-and-response (OEDR) subtask and to supervise these features. At Levels 0 to 2, a DMS should thus be used continuously. At Levels 1 and 2, for monitoring the state of the driver, a vehicle-related indicator of driving performance should be either avoided or used only when compatible with the driver support features that are engaged. The speed cannot, for instance, be used as an indicator of the driver state when an adaptive cruise control is regulating this speed. As more and more driver support features are introduced in vehicles, vehicle-related indicators of driving performance become less and less relevant for monitoring the state of the driver, whereas driver-related parameters (both physiological and behavioral) remain reliable indicators.

At any of Levels 3 to 5, and when the corresponding automated driving features are engaged, the driver is no longer in charge of the driving task and does not need to supervise them. Additionally, at Level 3, and at any time, the driver must, however, be fallback-ready, namely, ready to take over the control of his/her vehicle when the automated driving features request it (*i.e.*, ask for it). A DMS should, therefore, be capable of 1 assessing whether the current state of the driver allows him/her to take over the control of his/her vehicle if requested now or in the near future, and of 2 monitoring his/her state as long as he/she is in control. El Khatib et al. [70] discuss the potential need for a DMS even when the vehicle is in control and does not require the driver to supervise the driving or to monitor the driving environment. Whenever the driver has the option of, for example, engaging in some entertainment activity, he/she must be prepared to regain control in due course. Therefore, at Level 3, even though the driver is allowed to perform a secondary task, a DMS is still necessary to ensure that the driver is ready to take control at any time. During automated driving, drivers readily engage in Non-Driving Related Task (NDRT) [234]. Such engagement can help maintain adequate arousal levels and prevent mental underload [222]. However, performing an NDRT can impair takeover performance because attention to the roadway environment is reduced and drivers need to switch task and regain situation awareness before resuming control [350]. Although findings regarding the potential benefits of NDRT engagement are sometimes contradictory [60], Johns et al. [148] suggest that maintaining a moderate level of mental workload while the vehicle is operated by a DAS may enhance performance during a transfer of control from automated to manual. Such an optimal workload could be achieved, for instance, either by engaging the driver in a suitable secondary task (*e.g.*, one that supports, rather than degrades, situation awareness) or, conversely, by keeping him/her involved in the driving task at an appropriate level.

At Level 4, the automated driving features can only drive the vehicle under limited conditions, but they will not require the driver to respond within some specified time delay to a Take-Over Request (TOR). The Operational Design Domain (ODD) specifies the conditions under which the DAS is specifically designed to operate, including, but not limited to, 1 environmental, geographical, and time-of-day restrictions, and/or 2 the requisite presence or absence of certain traffic or roadway characteristics. Still at Level 4, the automated driving features are capable of automatically 1 performing a fallback of the driving task and 2 reaching a minimal-risk condition (*e.g.*, parking the car) if the driver neither intervenes nor takes over the driving task within the delay. If the driver decides to respond to the TOR, one can assume that the DMS would check that his/her state allows for this, even though the SAE J3016 does not say so explicitly.

At Level 5, the driving is fully automated under all possible conditions, and no DMS is required as the driver is never in control, and becomes, in effect, a passenger of the vehicle.

## 2.3 Survey of Literature on Driver Monitoring

This section presents a survey of the literature on driver monitoring and DMSs. The following subsections successively describe (1) our strategy for building an initial set of references, (2) some conclusions drawn from these references, and (3) trends observable in the summary tables or in some particular references. The analysis performed in this section guides the developments in subsequent sections.

The tables, along with comments on their design and content, are provided in Chapter A of the appendix.

### 2.3.1 Strategy for Building an Initial Set of References

To build an initial set of relevant references, we used an approach inspired from Gutiérrez et al. [105]. The block (or flow) diagram of Fig. 2.1 describes it.

Our search focused on surveys, reviews, and similar studies about driver monitoring and DMSs. We independently performed two searches during February 2021. The first focused on publications from IEEE, ScienceDirect, and Sensors, and the second on publications from ResearchGate; these four databases appeared well-suited for providing a useful set of initial references. We used the search engine specific to each database and a boolean query equivalent to *("survey" OR "review") AND ("driver" OR "driving") AND ("detection" OR "detecting" OR "behavior" OR "state" OR "monitoring")*. We limited the search to publications in English, and did not place any constraint on the dates of publication. The two searches yielded $124$ and $30$ items, respectively. After removing $16$ duplicates, we obtained a set of $138$ references. We manually screened these, and only kept the ones satisfying the two criteria of (1) being in scientific journals or conference proceedings, and (2) providing a survey, review, or similar study of one or more aspects of the domain of interest. This screening led to $56$ references. A summary describing, for each reference, the states, the indicators, and the sensors considered therein can be found in three separate tables

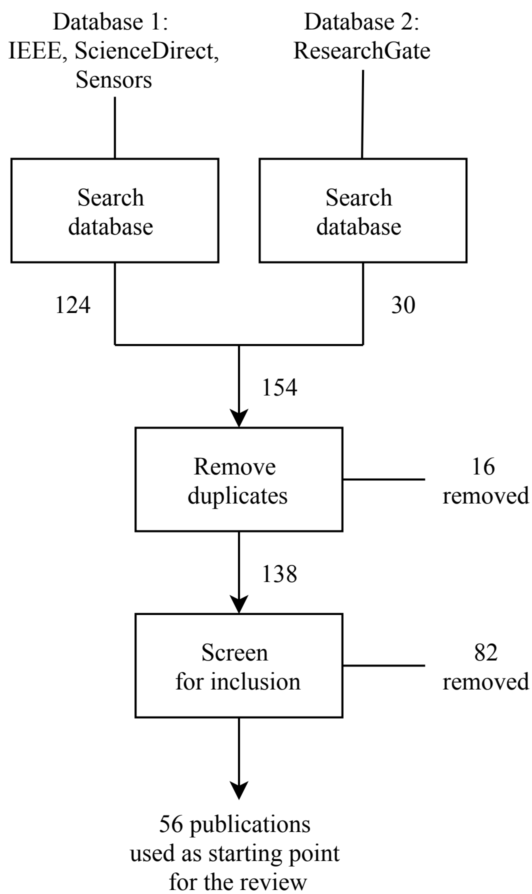



**Figure 2.1: Strategy fo building the initial set of references.** The flow diagram illustrates the strategy used for our survey of the literature on driver monitoring and Driver Monitoring System (DMS), and shows the number of publications at each stage of the process.

in Chapter A, along with explanations about the resulting structure of these tables. The main conclusions drawn from the preliminary analysis of the initial set of references and the observed trends are described in the following sections.

### 2.3.2 Conclusions from the Preliminary Analysis

The preliminary analysis of the $56$ initial references led to the following high-level conclusions:

1. To characterize the (global) state of a driver, one should consider the five main substates of *drowsiness*, *mental workload*, *distraction*, *emotions*, and *under the influence*.
2. A wide variety of parameters, which we call "indicators", are used to characterize each of these substates, and some indicators are applicable to more than one substate.
3. Ideally, a DMS should monitor not only the driver, but also the (driven) vehicle and the (driving) environment.
4. A value for each indicator is obtained by processing data (mainly signals and images) obtained from sensors "observing" the driver, the vehicle, and the environment.
5. A DMS generally involves one or more types and/or instances of each of the following: substate, indicator, and sensor.

These conclusions guided the structuring and writing of the bulk of the paper [115] on which this chapter is based. However, in contrast to [115], the present chapter specifically focuses on mental workload and distraction.

When the context is clear, we use "state" for the global state and each of the five substates. The phrase "state i" and the plural "states" imply that one is talking about one substate and several substates, respectively.

### 2.3.3 Observed Trends

The three tables summarizing the initial set of references (Table A.1, Table A.2, and Table A.3 in Chapter A of the appendix) reveal the following trends.

Drowsiness is the most covered state (with $44$ references among the total of $56$), distraction is the second most covered (with $20$ references), and more than one (sub)state is considered in only $19$ references.

Indicators are widely used in most references, in various numbers and combinations. Subjective indicators are not frequent, which is to be expected given the constraints of real-time operation. While several authors, such as Dong et al. [65] and Sahayadhas et al. [284], emphasize the importance of the environment and of its various characteristics (*e.g.*, road type, weather conditions, and traffic density), few references—and, specifically, only $6$—take them into account.

While the three columns for the sensors seem well filled (in Table A.3), several references either neglect to mention the sensor(s) they use, or cover them in an incomplete way. Some references give a list of indicators, but do not say which sensor(s) to use to get access to them. References simply saying that, for example, drowsiness can be measured via a camera or an eye tracker do not help the reader. Indeed, these devices can be head- or dashboard-mounted, and they can provide access to various indicators such as blink dynamics, percentage of closure (PERCLOS), and gaze parameters.

Many systems are tested in real conditions, perhaps after initial development and validation in a simulator. Many papers do not, however, systematically document the test conditions for each method that they describe.

Other trends are not directly observable looking at the three tables, but can be identified in some individual references.

Experts agree that there does not exist any globally-accepted definition for each of the first four states (that is, drowsiness, mental workload, distraction, and emotions). There is thus a need to define, as precisely as possible, what the states are.

In the more recent references, one sees a trend, growing with time, in the use of mobile devices, and in particular of smartphones [14, 42, 45, 70, 151, 152, 210, 218, 231, 323]. A smartphone is relatively low-cost, and one can easily link it to a DMS. This DMS can then use the data provided by the smartphone's many sensors, such as its inertial devices, microphones, cameras, and navigation system(s). A smartphone can also receive data from wearable sensors (*e.g.*, from a smartwatch), which can provide information such as Heart Rate (HR), skin temperature, and Electrodermal Activity (EDA). A smartphone can also be used for its processing unit.

## 2.4 Driver-State Characterization through States, Indicators, and Sensors

Our survey of the field of driver monitoring and DMSs led us to the idea of synthesizing this field in terms of the three key components of states, indicators, and sensors.

Figure 2.2 shows a system block diagram that uses the terminology introduced above, that is, sensors, indicators (and values thereof), and states (and levels thereof). The block diagram is drawn for a single, generic state, and one must specialize it for each of the five states of interest (or others).

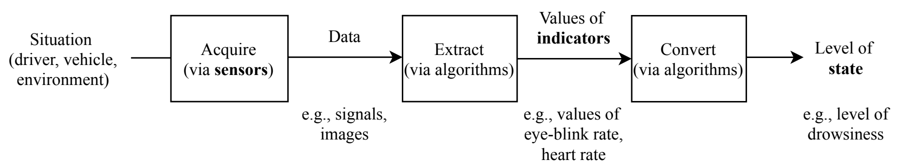



**Figure 2.2: Block diagram for driver-state characterization.** The figure shows, for the context of driver monitoring, the system block diagram applicable to the characterization of a generic (sub)state. The input is the situation of interest and the output is the level of the state. The operation of each of the three subsystems is described in the text.

The block diagram is self-explanatory. The input is the situation of interest (with the driver, vehicle, and environment). One or more sensors acquire data, typically signals and images. Algorithms extract the values of the indicators that are deemed relevant for the state of interest. Other algorithms convert these values into a level of the state. The three successive subsystems are labeled with the operation they perform, that is, acquire, extract, and convert. The input and output of each subsystem should ideally be viewed as being functions of time.

If several states are used simultaneously, the value of a given indicator can be used to compute the level of any state that this indicator relates to.

### 2.4.1 States

Our survey convinced us that the (global) state of a driver should be characterized along at least the five dimensions—called here states—of drowsiness, mental workload, distraction, emotions, and under the influence.

One goal of a DMS is to determine the levels of one or more of these states in real time, nearly continuously, and, preferably, in a non-invasive way. We use “level” in a very general sense. The level can take several forms, such as a numerical value or a label. The numerical value can be on a continuous scale or on a discrete scale. A label can be the most likely (output) class of a classifier together with its probability, likelihood, or equivalent. A level can be binary, *e.g.*, 0 and 1, or “alert” and “drowsy”. The levels of one or more of the five states can then be used to issue alerts or take safety actions; this is, however, not the object of this chapter.

It is challenging to quantify the four states of drowsiness, mental workload, distraction, and emotions in that they are not defined in a precise way and cannot be measured directly, by contrast with, say, physical quantities such as voltage and power. The fifth state (under influence) can be defined precisely, at least in the case of alcohol, but the measurement of its level requires asking the driver to blow in a breathalyzer and/or to submit to a blood test, both of which can be performed neither in real time nor non-invasively. In short, for all practical purposes, one cannot directly measure or obtain the level of any of the five states in any simple way. This is the reason for having recourse to “indicators” of each of these states.

### 2.4.2 Indicators

While one may have an intuitive idea of what an indicator is, it is useful to define, as precisely as possible, what it is. In a nutshell, an indicator must be well defined, and there must be a clear procedure for computing its values (at a succession of time instants) based on input data provided by one or more sensors.

For the purpose of this chapter and of the paper [115], a “quantity” or “item” is called an indicator for a given (sub)state if it satisfies all the following conditions:

1. it has a precise definition based on science (*e.g.*, physics, mechanics, chemistry, biology, physiology);
2. it can be measured, or characterized in some way, with real-time constraint when necessary, based upon data obtained from relevant sensors available in the application of interest;
3. it must take values (such as numbers or labels) within a pre-specified domain, and these values must preferably correspond to physical units (such as seconds or Hertz);
4. it is not a unique and full descriptor of the state;
5. it is recognized, in the literature, as being linked, in some meaningful way, to the state or trend thereof;
6. it is possibly useful with respect to one or more related, or unrelated, states;
7. it is reproducible, meaning that its value is always the same for fixed data.

For example, the eye-blink rate (*i.e.*, the blink rate of the left or right pair of eyelids) is scientifically recognized as being indicative of drowsiness. This parameter obeys all conditions above, and is thus an indicator of drowsiness.

Similarly to the level of a state, we talk about the value of an indicator. We use both “value” and “level” simply as a way to implicitly communicate whether one is talking about an indicator or a state. Ultimately, a set of values of the indicators of a state must be converted into a level of this state. The conversion may require the use of an advanced, validated algorithm.

Indicators are generally imperfect. In most cases, an indicator cannot be guaranteed to be fully correlated with a related state. Due to the presence of complex interrelationships between each (sub)state and its indicators, it is important to use as many indicators as possible to promote a valid and reliable interpretation of the (sub)state of the driver and, ultimately, of the (global) state of the driver.

The values of indicators are obtained through algorithms applied to data collected via sensors.

## 2.5 Synthesis of Driver-State Characterization via Two Interlocked Tables

The previous section shows the key role played by the triad of states, indicators, and sensors (also emphasized in Fig. 2.2) in driver-state characterization, which is the first of two key steps in driver monitoring. The present section describes our approach to synthesizing, in terms of this triad, the techniques for driver-state characterization found in the literature.

Our approach aims at answering, in a simple, visual way, the two following questions: 1 For a given state, what indicator(s) can one use? 2 For a given indicator, what sensor(s) can one use? We achieve this goal by naturally providing two tables (or matrices) of "states vs indicators" and "sensors vs indicators". These two tables can be viewed as being two-dimensional (2D) views of a 3D table (or array) of "states vs indicators vs sensors", as illustrated in Fig. 2.3, where the positions shown for the three dimensions and for the "dihedral" they subtend make the tables on the right appear in numerical order from top to bottom. The figure shows visually that the tables share the "Indicators" dimension, and are thereby interlocked. It gives a simplified representation of each of the tables that are progressively filled in Sections 2.6 and 2.7, that is, Tables 2.2 and 2.3.

In Fig. 2.3, the simplified representations of Tables 2.2 and 2.3 give the high-level structures of these tables.

In the table summarizing the indicators of the initial set of references (*i.e.*, Table A.2), the indicators are partitioned into the three columns "Driver", "Vehicle", and "Environment". Figure 2.3 shows, via the simplified representations, that Tables 2.2 and 2.3 are also partitioned in this way, with the corresponding abbreviations D, V, and E in the three megarows. In the table summarizing the sensors of the initial set of references (*i.e.*, Table A.3), the sensors are partitioned in the same way as the indicators. This is reflected in Fig. 2.3 by the partitioning of Table 2.3 into the D, V, and E megacolumns. The figure shows that Table 2.2 is partitioned into the five megacolumns corresponding to the five states, denoted here by $S_1, \ldots, S_5$, where $S_i$ stands for "State i". However, unlike in our original survey paper [115], this chapter addresses only the two states that will be further discussed in the following chapters (specifically, Chapters 3 and 4).

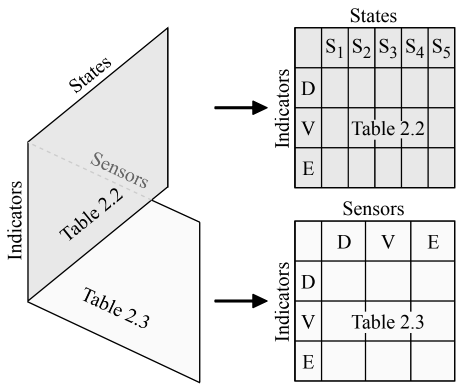



**Figure 2.3: Overview of the two interlocked tables.** The figure shows simplified representations of key Table 2.2 (states vs. indicators) and Table 2.3 (sensors vs. indicators). It also suggests that these tables can naturally be interpreted as being two views of an underlying 3D array. $S_i$, D, V, and E stand for “State i”, “Driver”, “Vehicle”, and “Environment”, respectively.

The rows and columns of Tables 2.2 and 2.3 are further divided as follows. The D-megarows of Tables 2.2 and 2.3 are subdivided as the D-megacolumns of Table A.2 are, that is, into the rows “Physiological”, “Behavioral”, and “Subjective”. The D-megacolumns of Table 2.3 are subdivided in a way that does not already appear in Table A.3, that is, into the columns “Seat”, “Steering Wheel”, “Safety Belt”, “Internal Camera”, “Internal Microphone”, and “Wearable”. Observe that the D-megarows and D-megacolumns are not subdivided in the same way, even though they correspond to the driver. The V- and E- rows and columns are also further divided as necessary.

Each lowest-level cell in both tables is destined to contain $0$, $1$, or more related references.

The pair of tables allows one to answer other questions such as: 1 If one invests in the calculation of an indicator for a particular state, what other state(s) can this indicator be useful for? 2 If one invests in a particular sensor for a particular state, what other state(s) can this sensor be useful for?

### 2.5.1 Categories of Indicators and Sensors

We give examples of the various categories of indicators and sensors that are further discussed in the next two sections. Below, we use the self-explanatory terminology of “X-based indicators” and “X-centric sensors”, where X can be replaced by driver (or D), vehicle (or V), or environment (or E).

#### Indicators

D-based indicators relate to the driver. They include physiological indicators (*e.g.*, heart activity, brain activity, EDA), behavioral indicators (*e.g.*, eye blinks, gaze direction, and hands positions), and subjective indicators (which are not suited for real-world operation, but can be used for validation at some point in the development of a DMS).

V-based indicators relate to how the driver controls his/her vehicle, for example, how he/she controls the speed, steers, and brakes.

E-based indicators relate to the environment, viewed here as consisting of three parts, that is, ① the outside environment (outside of vehicle), ② the inside environment (inside of vehicle), and ③ the contextual environment (independent of the previous two). Examples of characteristics of these parts of the environment are, respectively, ① the road type, weather conditions, and traffic density; ② the temperature and noise in the cockpit; and ③ the time of day and day of year. Each of these characteristics (*e.g.*, road type) can be used as an E-based indicator.

#### Sensors

Some D-centric sensors are placed in the seat (*e.g.*, radar for breathing activity), steering wheel (*e.g.*, electrodes for Electrocardiogram (ECG)), and safety belt (*e.g.*, magnetic induction sensors). Some D-centric sensors, in particular cameras (*e.g.*, RGB) and microphones, are appropriately placed in the cockpit to monitor the driver. We qualify these sensors of “internal”, to distinguish them from similar sensors monitoring the external environment, and qualified of “external”. Some D-centric sensors are wearables (*e.g.*, a smartwatch measuring HR and/or skin temperature). Since the aim is to monitor the state of the driver, we assume throughout this chapter that the seat, safety belt, and similar items are related to the driver.

V-centric sensors are mostly sensors—whether integrated in the vehicle or not—that allow for the acquisition of vehicle parameters such as speed, steering angle, and braking level. Such parameters are often obtained via the Controller Area Network (CAN) bus. Sensors (*e.g.*, accelerometers, gyroscopes) built into recent mobile devices can, however, also provide some of this information.

E-centric sensors are sensors that allow for the acquisition of parameters related to the environment. Cameras and radars can provide, for example, information about the driving scene.

### 2.5.2 Preview of the Next Two Sections

The next two sections successively cover the two selected states—mental workload and distraction—in detail. In general, each section defines a state, the indicators that characterize it, and the sensors that allow access to them, and progressively fills Table 2.2 and Table 2.3 with relevant references.

At the end of the last of these two sections, both tables are complete. They, together with the explanations in the two sections, constitute the main contribution of this chapter.

The structures of Tables A.1 to A.3, 2.2 and 2.3 were obtained after a significant number of iterations. This implies that the ultimate structure of Tables A.1 to A.3 was also informed by the content of Sections 2.4 to 2.7.

## 2.6 Mental Workload

We provide a detailed description of (the state of) “mental workload”, and we then present the indicators and sensors that can be used to characterize it.

#### Description

Mental workload, also known as cognitive (work)load (or simply as driver workload in the driving context), is one of the most important variables in psychology, ergonomics, and human factors for understanding performance. This psychological state is, however, challenging to monitor continuously [212]. In this section, we consider “mental” and “cognitive” to be synonyms.

A commonly-used definition of mental workload is the one proposed by Hart and Staveland [119]. They define mental workload as the cost incurred by a person to achieve a particular level of performance in the execution of a task. It is thus the portion of an individual's mental capacity—necessarily limited—that is required by the demands of this task [31, 240], that is, the ratio between the resources required to perform it and the available resources of the person doing it [286, 355].

In the literature on mental workload, one often finds references to another state called cognitive distraction. Mental workload and cognitive distraction are two different concepts, even if they can be linked when a driver performs secondary tasks while driving. Cognitive distraction increases the mental workload of a driver. An increase in mental workload is, however, not in itself an indication of cognitive distraction. First, mental workload can increase in the absence of distraction, for example, when a driver is focusing to execute the primary task of driving correctly and safely. Second, mental workload can increase significantly with an increasing complexity of the driving environment [289]. Cognitive distraction is further considered later as a particular category of (the state of) distraction.

Mental workload and stress are also linked since an increasing mental workload usually induces some stress in the driver.

### Indicators

In the driving context, visual tasks and mental tasks are closely linked. Indeed, while driving, a driver is constantly perceiving his/her driving environment and analyzing what he/she sees to make the right decisions whenever required, for example, scanning a crossroad and simultaneously judging the time and space relationships of other road users to decide when it is safe to cross an intersection. Therefore, it is logical that many researchers use eye-related parameters (*e.g.*, blinks, fixations, and pupil diameter) to assess the mental workload of a driver [211].

Among the driver-based, physiological indicators, EDA [150, 267], HR [18, 89, 267], and heart rate variability (HRV) [18, 39, 251] are often used as indicators of mental workload. HR increases as a task

**Table 2.2: Detailed "states vs. indicators" table,** introduced in simplified form in Fig. 2.3. Each cell in the heart of the table gives some references (if any) discussing how the corresponding indicator is useful for characterizing the corresponding state.

| | | | | States | | | |
|---|---|---|---|---|---|---|---|
| | | | Mental Workload | Distraction | | | |
| | | | | Manual | Visual | Auditory | Cognitive |
| Indicators | Driver | Physiological | Heart Activity | [18, 39, 82, 89, 139, 177, 251, 267, 278, 338, 352, 387] | | | | [84, 377] |
| | | | Brain Activity | [138, 139, 159, 177] | | | [292, 307] | [22, 310] |
| | | | Electrodermal Activity | [138, 139, 150, 177, 267, 338, 352, 387] | | | | [84, 113, 377] |
| | | | Pupil Diameter | [44, 89, 169, 177, 211, 255, 266, 369] | | | [99, 149] | [84] |
| | | Behavioral | Gaze Parameters | [44, 88, 174, 196, 211, 216, 387] | | [76, 87, 272, 329, 364, 370] | | [84, 108, 117, 195, 306, 311] |
| | | | Blink Dynamics | [139, 211] | | | [99, 118] | |
| | | | PERCLOS | [211] | | | | |
| | | | Body Posture | | | [86, 87] | | |
| | | | Hands Parameters | | [321] | | | |
| | | Subjective | | [119] | | | | [119] |
| | Vehicle | Wheel Steering | | [289] | [191] | [191] | | [84, 194] |
| | | Lane Discipline | | [139, 289] | | [76, 194, 377] | | [194, 271, 309] |
| | | Braking Behavior | | | | | [307] | [117] |
| | | Speed | | [139] | | [72, 76, 377] | | [84, 271] |
| | | Ambient temperature | | [339, 347, 360] | | | | |
| | Environment | Road Geometry | | | | | | |
| | | Traffic Signs | | [245] | | | | |
| | | Road Work | | | | | | |
| | | Traffic Density | | [116, 245, 352] | | | | |
| | | Obstacles | | [245] | | | | |
| | | Weather | | [245, 352] | | | | |

**Table 2.3: Detailed "sensors vs. indicators" table,** introduced in simplified form in Fig. 2.3. Each cell in the heart of the table gives some references (if any) discussing how the corresponding sensor is useful for characterizing the corresponding indicator. The indicators are identical to the ones in Table 2.2, thereby allowing one to link both tables.

| | | | | Sensors | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Driver | | | | | Vehicle | Environment | |
| | | | | Seat | Steering Wheel | Safety Belt | Internal Camera | Wearable | CAN Bus | External Camera | Radar |
| Indicators | Driver | Physiological | Heart Activity | [182, 361] | [302] | [145] | [39, 383] | [96] | | | |
| Indicators | Driver | Physiological | Brain Activity | | | | | | | | |
| Indicators | Driver | Physiological | Electrodermal Activity | | | | | [96, 387] | | | |
| Indicators | Driver | Physiological | Pupil Diameter | | | | [84, 266, 369] | | | | |
| Indicators | Driver | Behavioral | Gaze Parameters | | | | [26, 84, 87, 88, 108, 174, 227, 230, 232, 329] | | | | |
| Indicators | Driver | Behavioral | Blink Dynamics | | | | [26] | | | | |
| Indicators | Driver | Behavioral | PERCLOS | | | | [26, 226] | | | | |
| Indicators | Driver | Behavioral | Body Posture | | | | [26, 86, 87] | | | | |
| Indicators | Driver | Behavioral | Hands Parameters | | | | [19, 175, 176, 215, 365] | | | | |
| Indicators | Driver | Subjective | | | | | | | | | |
| Indicators | Vehicle | Wheel Steering | | | | | | | [38, 85, 187] | | |
| Indicators | Vehicle | Lane Discipline | | | | | | | | [16] | |
| Indicators | Vehicle | Braking Behavior | | | | | | | [38, 85, 187] | | |
| Indicators | Vehicle | Speed | | | | | | | [38, 85, 187] | | |
| Indicators | Vehicle | Ambient temperature | | | | | | | [179] | | |
| Indicators | Environment | Road Geometry | | | | | | | | [245] | |
| Indicators | Environment | Traffic Signs | | | | | | | | [245] | |
| Indicators | Environment | Road Work | | | | | | | | [245] | |
| Indicators | Environment | Traffic Density | | | | | | | | [245, 352] | |
| Indicators | Environment | Obstacles | | | | | | | | [245] | [288] |
| Indicators | Environment | Weather | | | | | | | | [245, 352] | |

gets more difficult [278] or if other tasks are added [82]. Electroencephalography (EEG) is also a valuable indicator for studying mental workload because it records the electrical activity of the brain itself, but it is complex to analyze [159]. The pupil diameter is considered to be an indicator of mental workload [44, 89, 169, 255, 266]. Indeed, Yokoyama et al. [369] indicate that the mental workload of a driver may be predicted from the slow fluctuations of the pupil diameter in daylight driving. All physiological parameters mentioned in this paragraph are, however, also influenced by other aspects of the mental and physical situation of the driver (*e.g.*, drowsiness and task-related fatigue) and by environmental situations (*e.g.*, illumination and temperature).

Among the driver-based, behavioral indicators, Fridman et al. [88] have shown that the visual scanning by a driver decreases with an increasing mental workload. Furthermore, since the interval of time between saccades has been shown to decrease as the task complexity increases, saccades may be a valuable indicator of mental workload [44, 196, 216].

Subjective measures of mental workload exist, like the NASA task load index (NASA TLX) [119], which is a workload questionnaire for self-report, and the rating scale mental effort (RSME).

Driving performance can diminish as a result of an increase in mental workload. The vehicle-based indicators which are the most sensitive to such an increase are Standard Deviation of Lane Position (SDLP) and steering wheel movement [289]. Moreover, mental workload is affected by in-vehicle ambient temperature, with increased workload observed at both low and high temperature [339, 347, 360].

Palasek et al. [245] use the driving environment to estimate the attentional demand required from the driver to drive. The features extracted from the analysis of the driving environment are thus indicators of the mental workload of the driver.

Lee et al. [177] demonstrate the benefits of combining multiple signals for mental workload estimation. Specifically, they employ a set of raw physiological signals—including EEG, ECG, EDA, electromyography (EMG), and pupil diameter—and apply functional data analysis methods. Similarly, Huang et al. [139] investigate multimodal measures encompassing physiological (heart activity, brain activity, EDA), behavioral (blink rate), and vehicle-based (speed, lateral position) indicators to assess variations in mental workload. Zhou et al. [387] also explore multimodal approaches, analyzing synchronized ECG, EDA, and gaze parameters, among others. Recent advances in deep learning have further facilitated the detection of mental workload by integrating multiple indicators [138, 338, 352].

Sándor [287] collected and analyzed the interdependencies of multimodal data streams, including vehicle dynamics (*e.g.*, speed and steering angle), driver physiological responses (heart rate), and eye-tracking metrics (blinks and fixations) to uncover patterns indicative of changes in driver state, particularly with regard to cognitive load. The author reported a negative correlation between vehicle speed and heart rate, suggesting variations in physiological arousal and cognitive load across different driving environments. Specifically, higher speeds accompanied by lower heart rates were interpreted as reduced cognitive load during highway cruising compared with city driving. Overall, this study illustrates the potential of combining vehicle-based and driver-based indicators to obtain a more comprehensive understanding of driver state and its underlying causes. However, although these two types of indicators are correlated, the driver-based ones remain more reliable, particularly as driving automation becomes more prevalent. Relying solely on vehicle-based indicators is therefore increasingly unrealistic moving forward.

The above information allows one to fill, in Table 2.2, the relevant cells of the "Mental Workload" column.

### Sensors

In a vehicle, the HR can be monitored using electrodes that can be placed at various locations, including the steering wheel (conductive electrodes [302]) and the seat (capacitive electrodes [182]). ECG monitoring using steering-wheel-based approaches is a feasible option for HR tracking, but requires both hands to touch two different conductive parts of the steering wheel.

Ballistocardiography also allows for monitoring the cardiac activity unobtrusively. The underlying sensing concept uses strain-gauge ballistocardiography sensors in the seat or in the safety belt to detect both the cardiac activity and the respiratory activity of the driver [361]. However, the vehicle vibrations make it difficult to use this sensor in real driving conditions.

Information about the cardiac activity can be obtained using a camera looking at the driver, in particular using photoplethysmography imaging [383] or using infrared imaging [39].

Radar-based methods mainly provide information about movement, which can of course be caused by both the cardiac activity and the respiratory activity. Various sensor locations are possible, including integration into the safety belt, the steering wheel, and the backrest of the seat [145].

Both heart activity and EDA are measured through electrodes placed on the skin of a person [267]. It can thus be measured through a wearable such as a smartwatch [96, 387]. Concerning the other, relevant, physiological, driver-based indicators, (1) it is challenging to get the pupil diameter in real conditions because of issues with illumination conditions and camera resolution, among others reasons, and (2) it is nearly impossible, as of this writing, to characterize brain activity in real time and in a non-intrusive, reliable way.

Beyond their use for physiological indicators such as HR [39, 383] and pupil diameter [266], cameras are widely used to assess mental workload through driver-based, behavioral indicators.

Fridman et al. [88] describe a system for characterizing, non-invasively, via a camera facing the driver, what they call his/her cognitive load. The system exploits the well-documented, experimental observation that the angular distribution of gaze direction (often characterized by the 2D pupil position) tends to become more concentrated, especially vertically, when the cognitive load increases. Using video imagery, the system classifies the cognitive load of the driver into one of the three cognitive load levels (low, medium, high), as he/she engages in activities other than the primary task of driving, such as a conversation or the adjustment of the infotainment system. The system extracts, from a $90$-frame, $6$-second video clip, via computer vision, the face and the region of one eye of the driver. It then uses one of two methods: (1) mainly active appearance models for the face, eyelids, and pupil (when visible) to produce a sequence of pupil 2D positions, and (2) one hidden Markov model for each of the three cognitive load levels. The second method uses a single 3D Convolutional Neural Network (CNN) with three output classes corresponding to these levels. The two methods thus rely on a sequence of pupil positions and on a sequence of eye images, respectively. The output of the system is one of the three cognitive load levels.

To develop this system, the authors first acquired training data in real-driving conditions while imposing on the driver a secondary task of a given cognitive load level. This imposition of a given cognitive load level while performing a primary task (here, driving) is commonly achieved in the literature through the standard "$n$-back" task, where the three values of $n$, that is, $n = 0$, $1$, and $2$, are viewed as corresponding to low, medium, and high cognitive load. For the $n$-back task, a sequence of numbers is dictated to the subject, who is asked, for each number, whether it matches the one dictated $n$ positions earlier in the sequence. For example, for $n = 2$, the subject must indicate whether the current number is the same as the one he/she heard 2 steps before, all this while he/she performs the primary task, here driving.

The authors indicate 1 that the differences in cognitive loading for the three levels have been validated using, among others, physiological measurements (*e.g.*, HR, EDA, and pupil diameter), self-report ratings, and detection-response tasks, and 2 that these levels have been found to cover the usual range of secondary tasks while driving, such as manipulating a radio or a navigation system.

It is noteworthy that the data used for building the system was acquired through real driving, during which the driver repeatedly performed $n$-back tasks, while a camera was recording his/her face and surrounding area, this by contrast with the many other developments made using a driving simulator, in highly controlled conditions, and difficult to implement in real-life conditions.

The authors indicate that, while they use the term "cognitive load", the literature often uses synonyms like "cognitive workload", "driver workload", and "workload".

Musabini and Chetitah [230] describe another system that is also based on eye-gaze dispersion. They use a camera facing the driver, produce a heatmap representing the gaze activity, and train a support vector machine (SVM) classifier to estimate the mental workload based on the features extracted from this representation.

Le et al. [174] characterize the mental workload based on the involuntary eye movements of the driver, resulting from head vibrations due to changing road conditions. They report that, as the mental workload increases, these involuntary eye movements become abnormal, resulting in a mismatch between the actual eye movements measured via an eye-tracking device and the predicted eye movements resulting from a

"VOR + OKR" model, where VOR and OKR are the abbreviations of vestibular-ocular reflex and optokinetic response. For each driver, the VOR parameters are estimated during the first $10$ s of driving in conditions of normal mental workload, whereas the OKR parameter is fixed. The hypothesis of abnormal eye movements while driving under mental workload was validated using a t-test analysis. Different levels of mental workload were induced in a driving simulator using the $n$-back task.

Palasek et al. [245] use an external camera recording the driving environment to estimate the attentional demand using attentive-driving models. Indeed, the task of driving can sometimes require the processing of large amounts of visual information from the driving environment, resulting in an overload of the perceptual systems of a human being. Furthermore, traffic density is known to increase the mental workload [116], so that urban environments lead to a higher mental workload than rural and highway environments do [371], all other conditions being equal.

Wei et al. [352] propose a multi-factor quantification and analysis method to classify the driver's mental workload by combining physiological (HR and EDA) with environment-based indicators such as traffic density and weather, the latter two being obtained from external cameras.

Vehicle-based indicators, such as wheel steering, braking behavior, and speed, can be retrieved from the vehicle's CAN bus [38, 85, 187]. In addition, in-vehicle ambient temperature is also accessible via the CAN bus [179].

The above information allows one to fill the relevant cells of Table 2.3.

TAKEAWAY MESSAGES FOR OUR OWN EXPERIMENTS: In a nutshell, Chapters 3 and 4 describe a driving simulator study in which participants were cognitively distracted through a secondary task designed to increase their mental workload. The study also manipulated traffic conditions and automation use, both of which can influence drivers' mental workload. A complete description of the user study is to be found in Sections 3.4 and 4.4. In Chapter 3, we examined the effects of our three factors (cognitive distraction, traffic conditions, and automation use) on driving performance. To this end, we measured the SDLP and speed, two vehicle-based indicators that can be captured with vehicle-centric sensors. In our case, both indicators were provided directly by the driving simulator. In Chapter 4, we analyzed driver-based physiological and behavioral indicators of mental workload, relying on driver-centric sensors. Specifically, we monitored EDA using electrodes placed on the driver's foot and gaze parameters using a camera facing the driver.

## 2.7 Distraction

By contrast with the previous section, we start with some background information (up to Section 2.7.1) on the state of distraction.

The globally accepted definition of driver distraction follows: it is a diversion of attention, away from activities critical for safe driving (the primary task) and toward a competing activity [272, 277].

Inattention, sometimes used—mistakenly—as a synonym of distraction, is defined as a diminished attention to activities that are critical for accomplishing a primary task, but not necessarily in the presence of a competing activity [277]. Therefore, driver distraction is one particular form of driver inattention [276]. Inattention is a broader term, as it can be caused, for example, by drowsiness. It indeed occurs in a wide range of situations in which the driver fails to attend to the demands of driving, such as when a desire to sleep overcomes a drowsy driver.

Driver distraction can be caused by any cognitive process, such as daydreaming, mind wandering, logical and mathematical problem solving, decision making, using any kind of in-vehicle system, for example, for entertainment, navigation, or communication (including a cell phone), and any other activity that may affect the driver's attention to driving [15]. It is helpful to distinguish between four types of distractions [65, 99]:

1. *manual* distraction (*e.g.*, manually adjusting the volume of the radio);
2. *visual* distraction (*e.g.*, looking away from the road);
3. *auditory* distraction (*e.g.*, answering a ringing cell phone); and
4. *cognitive* distraction (*e.g.*, being lost in thought).

Several distracting activities may, however, involve more than one type of distraction (*e.g.*, talking on the phone while driving creates at least an auditory distraction and a cognitive distraction, under the assumption that a hands-free system is used, thereby avoiding manual distraction).

When distracted, the driver looses awareness of the current driving situation. Being aware of a situation (whether for driving or for some other activity) is often called situation awareness. A loss of situation awareness while driving results in a reduction of vigilance and in an increase of the risk of accident. In driving, a major aspect of situation awareness is the ability to scan the driving environment and to sense dangers, challenges, and opportunities, in order to maintain the ability to drive safely. As a driver moves through the environment, he/she must—to avoid getting into an accident—identify the relevant information in rapidly changing traffic conditions (*e.g.*, distance to other vehicles, closing speed), and be prepared to react to suddenly-appearing events (*e.g.*, braking because of an obstacle, obeying a road sign). To achieve situation awareness, a driver must thus perceive his/her driving environment correctly [69], be attentive, and have a working memory [355]. It follows that any distraction that harms the driver's attention may adversely impact situation awareness [154].

Kircher and Ahlström [161] argue that existing definitions of distraction have limitations because they are difficult to operationalize, and they are either unreasonably strict and inflexible or suffering from hindsight bias, the latter meaning that one needs to know the outcome of the situation to be able (1) to tell what the driver should have paid attention to and, then, (2) to judge whether he/she was distracted or not. The authors are also concerned that distraction-detection algorithms (1) do not take into account the complexity of a situation, and (2) generally cover only Eyes-Off-Road (EOR) and engagement in Non-Driving Related Task (NDRT). They thus developed a theory, named MiRA (minimum required attention), that defines the attention of a driver in his/her driving environment, based on the notion of situation awareness. Instead of trying to assess distraction directly, one does it indirectly, by first trying to assess attention. Recall that distraction is a form of inattention.

According to the MiRA theory, a driver is considered attentive at any time when he/she samples sufficient information to meet the demands of the driving environment. This means that a driver should be classified as distracted only if he/she does not fulfill the minimum attentional requirements to have sufficient situation awareness. This occurs when the driver does not sample enough information, whether or not simultaneously performing an additional task. This theory thus acknowledges

1 that a driver has some spare capacity at his/her disposal in the less complex driving environments, and 2 that some glances toward targets other than the roadway in front of him/her may, in some situations, be needed for the driving task (like looking at, or for, a vehicle coming from each of the branches at a crossroad). This means that EOR and engagement in NDRT do not necessarily lead to driver distraction.

The MiRA theory does not conform to the traditional types of distraction (manual, visual, auditory, cognitive) as it does not prescribe what sensory channel a certain piece of information must be acquired through.

In an attempt to operationalize the MiRA theory, Ahlström et al. [7] present an algorithm for detecting driver distraction that is context dependent and uses 1 eye-tracking data registered in the same coordinate system as an accompanying model of the surrounding environment and 2 multiple buffers. Each buffer is linked to a corresponding glance target of relevance. Such targets include: windshield, left and right windows, (rearview) mirrors, and instrument cluster. Some targets and their buffers are always present (like the roadway ahead via the windshield, and behind via the mirrors), while some other targets and their buffers appear as a function of encountered traffic-regulation indications and infrastructural features. Each buffer is periodically updated, and its update rate can vary in time according to requirements that are either "static" (*e.g.*, the presence of a specific on-ramp that requires one to monitor the sides and mirrors) or "dynamic" (*e.g.*, a reduced speed that lessens the need to monitor the speedometer). At each scheduled update time, a buffer is incremented if the driver looks at the corresponding target and decremented otherwise; this is a way of quantifying the "sampling" (of the environment) performed by the driver. A buffer running empty is an indication that the driver is not sampling enough of the corresponding target; he/she is then considered to be inattentive (independently of which buffer has run empty). Until declared inattentive, he/she is considered attentive.

This completes the background information on the state of distraction. We now successively consider the four types of distraction. For each of the four corresponding substates, we provide a detailed description, and we then present the indicators and sensors that can be used to characterize it.

### 2.7.1 Manual Distraction

#### Description

Manual distraction, also called biomechanical distraction, occurs when the driver is taking one or both of his/her hands off the steering wheel. The driver may do so to answer a call or send a text message, grab food and eat, or grab a beverage and drink, all while driving. According to the National Highway Traffic Safety Administration (NHTSA), texting while driving is the most alarming distraction. It is mainly due to manual distraction, but, inevitably, it also includes both visual distraction and cognitive distraction.

#### Indicators

Unsurprisingly, the best indicator used to detect manual distraction is the behavior of the driver's hands, mainly through their positions and movements. For safe driving, these hands are expected to be, most of the time, exclusively on the steering wheel, the gearshift, or the turn-signal lever. On the contrary, a hand using a phone, adjusting the radio, or trying to grab something on the passenger seat indicates a manual distraction [321].

Vehicle-based indicators can also be used, as shown in [191]. Using naturalistic-driving data, the authors studied the correlation between (1) performance metrics linked to the steering-wheel behavior and to the vehicle speed, and (2) manual and visual driver distractions induced, for example, by texting. They found a good correlation between the steering movements and the manual-visual distraction of the driver.

The above information allows one to fill, in Table 2.2, the relevant cells of the "Manual Distraction" column.

#### Sensors

The most common solution to analyze the behavior of the driver's hands is to use a camera placed inside the vehicle, usually near the central mirror, looking down in the direction of the driver.

Le et al. [176, 175] propose an approach to detecting [176] and classifying [175] human-hand regions in a vehicle using CNNs. Their technique for hands detection is robust in difficult conditions caused, for example, by occlusions, low resolution, and/or variations of illumination.

Using deep CNNs, Yan et al. [365] classify six actions involving the driver's hands, that is, calling, eating, smoking, keeping hands on the steering wheel, operating the gearshift, and playing on the phone. Similarly, both Baheti et al. [19] and Masood et al. [215] use ten classes to detect when the driver is engaged in activities other than safe driving and to identify the cause of distraction.

Distraction caused by activities such as drinking, adjusting radio settings, or looking toward the back seat and typically characterized as manual or visual distraction is often detected using deep learning. End-to-end approaches, which identify distraction directly from images without relying on intermediate indicators, are a common choice when using deep learning models [225, 261, 262, 351]. However, these methods tend to be less explainable and provide limited insight into the underlying causes of distraction.

Vehicle-based indicators can be obtained from the CAN bus of the vehicle [38, 85, 187].

The above information allows one to fill the relevant cells of Table 2.3.

### 2.7.2 Visual Distraction

#### Description

Visual distraction occurs when the driver is looking away from the road scene, even for a split second. It is often called EOR, and is one of the most common distractions for a driver. Examples of activities causing EOR are: (1) adjusting devices in the vehicle (like a radio or navigation system); (2) looking towards other seats; (3) regarding a new message on the phone or glancing at the phone to see who is calling; and (4) looking outside when there is a distraction by the roadside. All generally result in the driver not looking straight ahead, which is what he/she needs to be doing for safe driving.

#### Indicators

The gaze is the main indicator used to detect a visual distraction of a driver. The duration of EOR is probably the most-used metric. The longer the EOR duration is, the lower the situation awareness of the driver is, and the higher the visual distraction of the driver is [364, 370].

The glance pattern and the mean glance duration are other metrics [76, 272].

Occasionally, the head direction is used to approximate the gaze direction to characterize the driver's visual distraction [86, 87]. For example, Fridman et al. [86] classify driver gaze regions on the sole basis of the head pose of the driver. Fridman et al. [87] compare classifications of driver gaze using either head pose alone or both head pose and eye gaze. They classify, based on facial images, the focus of the attention of the driver using $6$ gaze regions (road, center stack, instrument cluster, rear-view mirror, left, and right). To do so, they consecutively perform face detection, face alignment, pupil detection, feature extraction and normalization, classification, and decision pruning. Vicente et al. [329] similarly classify the driver gaze, but use $18$ regions instead of $6$.

Visual distraction can also be inferred using vehicle-based indicators such as wheel steering, braking behavior, and speed. Indeed, a driver generally slows down when distracted by a visual stimulus [72, 76, 377], and visual distraction impairs lateral control because the driver needs to compensate for errors made when taking his/her eyes off the road, which leads to larger deviations in lane positioning [76, 194, 377]. Such deviations have various causes, including drowsiness and visual distraction. This re-emphasizes the need to use as many indicators as possible. This also explains why more and more vehicles are equipped with systems that keep the vehicle within its lane whenever possible.

Beyond the commonly used indicators described above, it is important to acknowledge that gaze data can be coded in different ways, each offering distinct insights into visual distraction. Three main coding schemes exist: direction-based, target-based, and purpose-based [9, 162]. Direction-based coding focuses on *where* the driver is looking and is typically used to compute indicators such as EOR. Target-based coding, by contrast, manually assigns each glance to an object or region of interest (*e.g.*, a bicyclist, a traffic sign, or the phone), thereby providing a richer contextual understanding of visual behavior and indicating *what* objects drivers are looking at. Finally, purpose-based coding classifies glances according to *why* the driver is looking at a particular area by defining which areas must be monitored to be considered attentive [9]. From a driver-monitoring perspective, these coding schemes provide different information regarding visual distraction. Direction-based metrics capture deviations from the road scene; target-based metrics reveal the nature of competing visual stimuli; and purpose-based metrics link glances to the attentional requirements of the driving task.

The above information allows one to fill, in Table 2.2, the relevant cells of the “Visual Distraction” column.

#### Sensors

In order to monitor driver visual distraction, one mainly uses at least one camera facing the driver, thus as for manual distraction. The camera can be placed in various positions as long as the head pose and/or gaze of the driver can be obtained.

Naqvi et al. [232] use a near-infrared camera (with wavelengths of $0.75$–$1.4\,\mu$m) placed in the dashboard in conjunction with a deep-learning-based gaze-detection system, classifying the driver’s gaze into $17$ gaze zones.

Mukherjee and Robertson [227], similarly to Fridman et al. [86], present a CNN-based model to estimate human head pose and to classify human gaze direction. They use, however, low-resolution RGB-depth (RGB-D), thus with a camera providing depth information.

The above information allows one to fill the relevant cells of Table 2.3.

### 2.7.3 Auditory Distraction

#### Description

Auditory distraction occurs when some sound prevents the driver from making the best use of his/her hearing, because his/her attention is drawn to the source of the sound. Hearing a phone ringing, listening to a passenger, listening to music, and following navigation instructions can all lead to auditory distraction.

This component of driver distraction is the least studied in the literature, likely because (1) it is often accompanied by at least one other more-easily detectable source of distraction falling among the other three types, and (2) it poses lower safety risks in comparison to the other types of distraction, in particular visual distraction [305].

The literature does not appear to introduce the concept of “auditory indicators”, which would characterize (1) the sounds captured both inside and outside of the vehicle, and, preferably, (1) the distraction they create. By using several microphones (including arrays thereof) and techniques for separating audio sources [333], one could imagine breaking down and localizing the various sources of sounds both inside and outside the vehicle.

#### Indicators

When the driver appears to be auditorily distracted, there occur changes in pupil diameter [99, 149] and blink frequency [99, 118]. EEG [292] can also be used as an indicator of auditory distraction. Sonnleitner et al. [307] describe the impact of an auditory secondary task on a driver during a primary driving task, and show changes in braking reaction and brain activity.

The above information allows one to fill, in Table 2.2, the relevant cells of the "Auditory Distraction" column.

#### Sensors

Obtaining the pupil diameter is challenging in real conditions due to illumination conditions and/or camera resolution, among others. Furthermore, brain activity cannot, at this time, be measured both in real time and in a non-intrusive, reliable way. Blink frequency can, however, be monitored via a camera, and braking behavior via the CAN bus.

Although microphones and, even better, arrays thereof, both inside and outside the vehicle, would be natural sensors to provide values for auditory indicators, we did not find any references considering such sensors for characterizing auditory distraction. One can also envision using the microphone(s) of a smartphone linked to a DMS.

The above information did not lead to the addition of any reference to Table 2.3.

### 2.7.4 Cognitive Distraction

#### Description

In the context of driving, cognitive distraction is defined by the NHTSA as the mental workload associated with a task that involves thinking about something other than the (primary) driving task [237]. A driver who is cognitively distracted due to a secondary task, such as mind wandering, experiences an increase in his/her mental workload (the state discussed in Section 2.6). The characterization of his/her cognitive distraction could therefore be achieved (1) by examining how his/her mental workload evolves and (2) by finding characteristics of this evolution allowing one to decide whether it is caused by cognitive distraction. The monitoring of cognitive distraction is thus, before all, a monitoring of the mental workload and/or its time variations. Section 2.6 shows that there are (1) many ways to characterize mental workload, and (2) many indicators thereof. The challenge is to be able to pinpoint the components of, or changes in, the mental workload that are due to distraction.

Cognitive distraction occurs when a driver is thinking about something that is not related to the driving task. In the driving context, while visual distraction can be summarized by EOR, cognitive distraction can similarly be viewed as Mind-Off-Road (MOR). While it is relatively easy to monitor EOR (with a camera facing the driver), it is difficult to monitor MOR. It has, however, been shown that, when a driver is cognitively distracted, his/her visual behavior is impacted. Mind-wandering and daydreaming are two causes of cognitive distraction.

### Indicators

As cognitive distraction induces mental workload, the indicators allowing one to detect and characterize these two states are similar, if not identical. Therefore, it is difficult, if not impossible, to distinguish, in the driving context (as well as others), between these two states since they have nearly the same influences on the indicators.

Among the four types of distractions, cognitive distraction has proven to be the most difficult to detect and characterize. This is because it happens inside the brain, and, obviously, "observing" the brain of a driver is more challenging than observing his/her hands and eye(s).

As for visual distraction, cognitive distraction can be characterized by indicators of both driving performance and eye movements [196], including (1) vehicle-based indicators, such as speed [271], wheel steering [194], lane discipline [194, 271, 309], and braking behavior [117], and (2) driver-based, behavioral indicators, such as gaze parameters (*e.g.*, fixation duration, glance frequency, and gaze distribution) [108, 117, 195, 306, 311] and head orientation. A driver makes significantly fewer high-speed saccadic eye movements and spends less time looking to the relevant periphery for impending hazards with increasing complexity of the secondary task(s). He/She also spends less time checking his/her instruments and mirrors [117].

Cognitive distraction can also be measured through various driver-based, physiological indicators. Among these, brain activity [22, 310] and pupil diameter may be the most convincing. While Yusoff et al. [377] reported that EDA and HR show only weak relationships with cognitive distraction, these signals are nonetheless recognized indicators of mental workload, as discussed in Section 2.6. Moreover, Halin et al. [113] demonstrated that three EDA-based indicators are significantly influenced by cognitive distraction.

Fresta et al. [84] implemented an end-to-end, real-time, deep learning approach using vehicular (*e.g.*, speed, wheel steering), eye-tracking (*e.g.*, pupil diameter, gaze direction), and physiological signals (HR and EDA) to detect driver cognitive distraction.

Among the subjective measures, the NASA TLX [119] is commonly used in driving-distraction studies even though it is a subjective measure of mental workload and, thus, not a measure specific to cognitive distraction.

The above information allows one to fill, in Table 2.2, the relevant cells of the "Cognitive Distraction" column.

#### Sensors

Since the main indicators of cognitive distraction are driving performance and gaze parameters, the main sensors to characterize it are vehicle-centric sensors [85, 187] and cameras [84, 108].

The above information allows one to fill the relevant cells of Table 2.3.

TAKEAWAY MESSAGES FOR OUR OWN EXPERIMENTS: As similarly described at the end of the previous section, Chapters 3 and 4 report on a driving simulator study in which participants were cognitively distracted through a secondary task designed to increase their mental workload. The study also manipulated traffic conditions and automation use, both of which can influence drivers' mental workload. A complete description of the user study is provided in Sections 3.4 and 4.4. In this context, although all three factors affect mental workload, two of them (traffic conditions and automation use) are intrinsic to the driving task, whereas the third (the secondary task) is not. This distinction makes it possible to attribute specific variations in mental workload explicitly to cognitive distraction induced by the secondary task.

## 2.8 Summary

This chapter focuses on the characterization of the state of a driver, which is the first key step for driver monitoring and DMS. It surveys (in Section 2.3) the relevant scientific and technical literature on driver-state characterization, and subsequently provides a synthesis (in Sections 2.4 to 2.7) of the main, published techniques for this characterization.

The survey yielded $56$ publications in scientific/technical journals and conference proceedings. Their examination led to the conclusion that the state of a driver should be characterized according to five main dimensions—called here “(sub)states”—of drowsiness, mental workload, distraction (further subdivided into four types qualified of manual, visual, auditory, and cognitive), emotions, and under the influence.

In comparison with standard physical quantities, such as voltage and power, these states are not well defined and/or are very difficult—if at all possible—to quantify or to label, not only in a validated way, but also in real time and non-invasively, as is required in the driving context. The only reasonable approach, found almost universally in the literature, is to have recourse to indicators (of each of these states), the value of which can be obtained in a practical and validated way. Examples of indicators are the eye-blink rate, the SDLP, and the outside temperature. The values of many indicators (but not all) are obtained by applying algorithms, often complex, to data (typically signals and images) collected from sensors.

The last paragraph brings to light the three ingredients that, in our view, lie at the heart of driver monitoring and DMSs, that is, the triad of states, indicators (of these states), and sensors (providing data, which are the source of the values of these indicators). Figure 2.2 links these three ingredients.

Our survey confirmed the intuition that one should monitor, not only the driver (D), but also the (driven) vehicle (V) and the (driving) environment (E). Accordingly, we partitioned both the indicators and the sensors into D, V, and E categories, leading to the phrases “X-based indicators” and “X-centric sensors”, where X can be D, V, or E. For the D-based indicators, we further distinguished between three types: physiological, behavioral, and subjective. The three examples of indicators given earlier correspond to D, V, and E, respectively.

The major outcome of the chapter—and the paper [115] on which it is based—is the pair of interlocked tables "states vs indicators" (Table 2.2) and "sensors vs indicators" (Table 2.3), where each cell contains zero, one, or more references. These tables bring together, in an organized way, most of the useful information found in the literature, up to the time of this writing, about driver-state characterization for driver monitoring and DMSs. These tables constitute an up-to-date, at-a-glance, visual reference guide for anyone active in this field. They provide immediate answers to key questions that arise in the design of DMSs, such as the four questions posed in Section 2.5.

## 2.9 Conclusions

Back in 2021, we drew the following main conclusions from the pair of tables (Tables 2.2 and 2.3) and the references they compile, and these conclusions still hold today:

1. Each state can be inferred from several indicators (which are often far from perfect), thereby encouraging multimodal fusion.

2. The internal camera (possibly with several instances) appears to be the most-commonly-used sensor.

3. Wearable sensors (*e.g.*, smartwatches) are increasingly used to obtain driver-based, physiological indicators and vehicle-based indicators.

4. Environment-based indicators are often ignored, even though there is an agreement that they should be used.

5. Driver-based, subjective indicators, although sometimes alluded to, cannot be used in real driving but are essential for the validation of some indicators of some states.

6. Brain activity is a recognized indicator of several states, but cannot be accessed today in a non-invasive, reliable, and inexpensive way in real driving.

7. Several methods for characterizing each of the five states use, without surprise, techniques of machine learning and, especially, of deep learning.

8. The term “predict(ion)” often refers to a present state rather than to a future state, and few papers describe techniques “to tell beforehand”, for example, the future values of indicators and levels of states.

The next two paragraphs respectively elaborate on the last two points.

For driving safety, it is paramount that the processing and decisions made by any algorithm used in a vehicle, including for driver monitoring, be fully explainable (to a human being) at the time of design and certification of this algorithm. Most algorithms using machine learning do not, however, have this necessary feature of explainability or interpretability, and this is certainly the case for machine-learning-based algorithms that would learn on the fly during one or more trips. Therefore, while machine learning algorithms and, especially, deep-learning algorithms often provide stellar performances on specific datasets in comparison with other types of algorithms, they are unlikely to be deemed acceptable by an equipment provider or a vehicle manufacturer, especially if they operate as end-to-end solutions without any form of interpretability. There is, however, a trend toward designing machine learning algorithms that produce results that can be explained [199, 378]. The above remarks apply not only to machine learning but also to any approach whose operation cannot be explained simply. Our framework, which implies the use of indicators and states, supports the desired explainability. It indeed prevents any algorithm from going, in one fell swoop, from (nearly-)raw sensor data to driver characterization, by forcing it to estimate both the values of indicators and the levels of states as a stepping stone toward the ultimate characterization of the state of a driver.

The literature on driver monitoring focuses almost exclusively on characterizing the “present” state of the driver. We use quotes because the characterization is typically based on data from the recent past, for example, in a window that extends over several tens of seconds and butts against almost the present time. This results in a characterization of the “recent-past” state of the driver. If the driver is in control, a DMS using this characterization may not have sufficient lead time to take proper emergency action (to issue an alarm and/or to take back the control) and, if the vehicle is in control, such a DMS may hand the control over to the driver even though he/she might be falling asleep or getting distracted in a few tens of seconds or more. A major missing link in current DMS research and development is thus the true prediction of the future state of the driver, at least a few tens of seconds into the future.

On the one hand, Tables 2.2 and 2.3 show, at a glance, which areas of driver-state characterization have been the object of research and with what intensity (as measured by the number of references listed in each cell). On the other hand, the two tables show, also at a glance, where little or no research has been performed to date, thereby suggesting new, potentially-fruitful research areas. The two tables should thus prove to be a rich source of information for both research and product development.

In 2021, starting from a set of $56$ initial references, our exploration of the field of driver monitoring led us to examine a total of $254$ references for the five states. While our crisscrossing of the field, at several different times, led us to identify many relevant publications, our search cannot, obviously, be exhaustive.

The methodology used for this survey and synthesis of the state of the art in driver monitoring can be applied to update the tables over time to reflect new developments—and this is precisely what has been done during the writing of this chapter in 2025 for Tables 2.2 and 2.3. Updates can be made by adding and/or removing rows, columns, and/or references, as appropriate.

Back in 2021, we concluded that characterizing the state of a driver and, more generally, driver monitoring would remain important despite the progressive increase in vehicle automation. Recent European regulations mandating the use of DMSs in new vehicles further reinforce this perspective. SAE Level 3 enables vehicles to drive by themselves under certain conditions, such as on a highway and in sunny weather, but a driver must still be present and able to take back control of the vehicle at any time and in a relatively short lapse of time. To ensure that the driver is able to take back control, technologies for monitoring the state of the driver become even more critical. These technologies are also needed to monitor the driver during the time he/she is driving, and to possibly allow the vehicle to take back control if necessary.

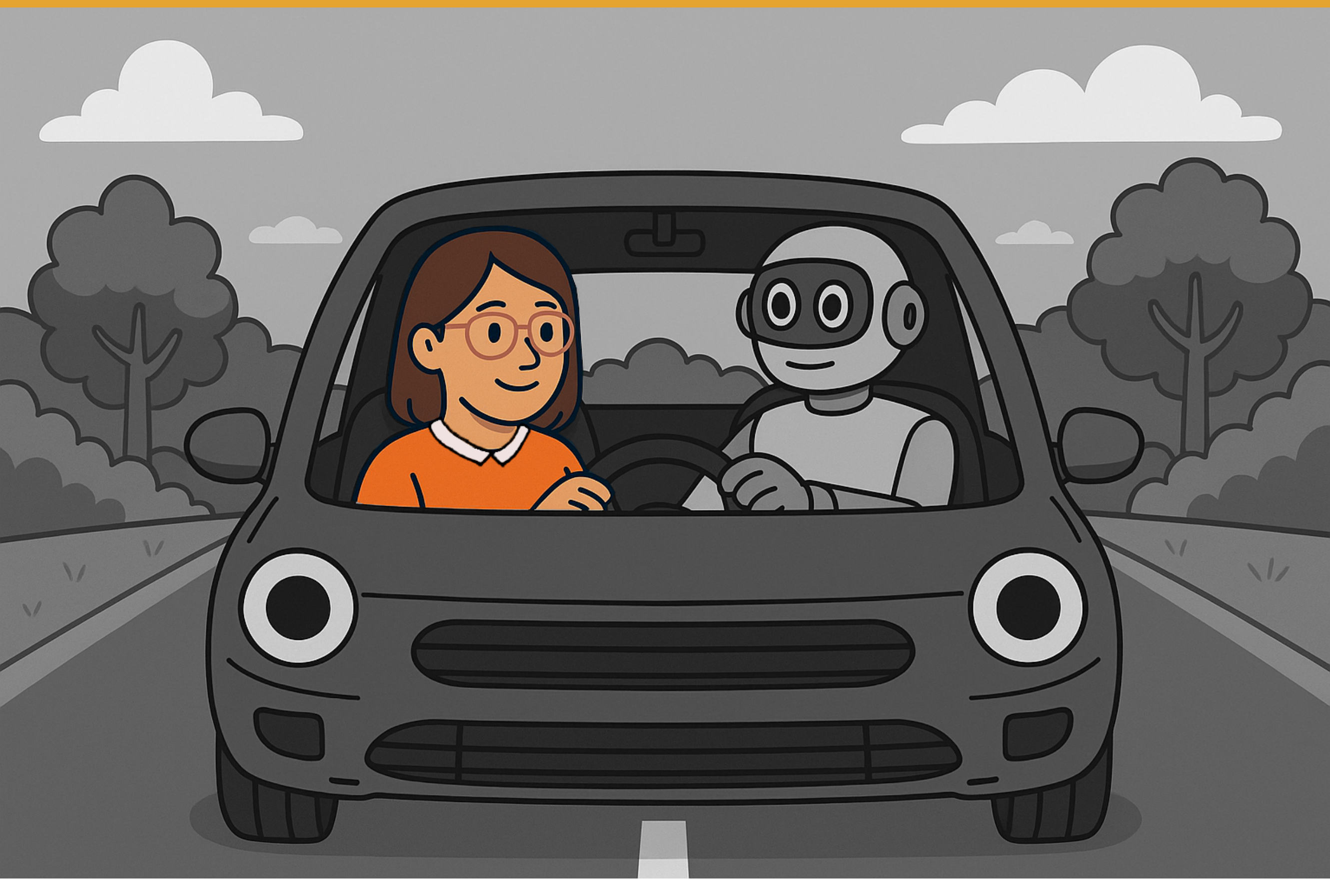
Once again, in 2021, we observed that some vehicle manufacturers were offering DMSs based on the behavior of the driver and/or the behavior of the vehicle, such as the detection of steering-wheel movements and lane deviations, respectively. We noted that although these systems could be useful in vehicles with automation up to (SAE) Level $2$, they would become obsolete at higher levels of automation. Indeed, when a vehicle drives autonomously, monitoring its behavior does not provide any information about the state of the driver. Consequently, technologies that directly monitor both the driver and the driving environment are a necessity as long as the driver is involved in the driving task, at least partially. Again, the recent European regulations imposing the use of camera-based systems to monitor driver's gaze now confirm these earlier insights.

To date, the development of DAS has moved at a faster pace than has the development of DMSs, even though interest in DMSs has been growing due, *e.g.*, to the aforementioned new regulations. This imbalance is, in major part, a consequence of the long-held belief by some automotive-industry players that they would be able to easily leapfrog Levels 3 and 4, and move on directly to Level 5, where there is no need to monitor the driver. However, most experts now agree that it will be decades before most privately owned vehicles are fully automated, if ever. Along the long and winding road to Level 5, the automotive industry will need to significantly boost the research on, and the development of, DMSs. For Levels 3 and 4, the same industry will need to develop ADS and DMSs in full synergy. The future could thus not be brighter for the field of driver monitoring and DMSs.

Most of the conclusions we drew in 2021 therefore remain valid, and practice has, in fact, confirmed several of our initial insights. As highlighted in Chapter 1, considerable attention has been devoted to DMSs since 2021, with, for example, new European regulations introduced in 2023 requiring camera-based systems to monitor driver's gaze in real time. However, even though DMSs have become mandatory in Europe, we still argue that their potential goes far beyond warning functions and that they can play a key role in fostering more effective interaction between the driver and the automation system. In this vein, we introduce the *DEV* framework at the end of this dissertation, in Chapter 7, which aims to move towards a more integrative model of driver–automation cooperation.

# Part II

## Human Studies: Understanding Driver Behavior



# 3

# ACC Use and Its Impact on Driving Performance

**Contents of this chapter**

CONTEXT. Within the scope of Part II, which investigates driver behavior through human studies, this chapter examines how drivers interact with automation. With the increasing integration of driving automation into vehicles, driving has evolved into a cooperative task between the human driver and the vehicle automation. However, to make this collaboration effective, it is essential to understand the interplay between the driver, the vehicle, and the environment, as these three elements continuously influence one another. In the study presented in this chapter, we examine whether drivers' cognitive state and the complexity of the driving environment influence reliance on ACC, and whether such reliance, in turn, affects driving performance.

**Contributions and Key Findings**

1. The complexity of the driving environment significantly influences ACC use, whereas cognitive load does not.
2. ACC use improves lateral control and affects compliance with speed limits but does not significantly influence lane-changing behavior.
3. Overall, this chapter contributes to understanding how driver cognitive distraction and traffic conditions shape driver reliance on ACC and the resulting driving performance.

RELATED PUBLICATION. This chapter is adapted with minor changes from Halin et al. [114].

## 3.1 Introduction

The Society of Automotive Engineers (SAE) [283] defines six levels of driving automation, ranging from no automation (SAE Level 0) to full automation (SAE Level 5). Vehicles currently on the market are mainly SAE Level 2 (partial driving automation), such as Tesla Autopilot, BMW Extended Traffic Jam Assistant, Ford Blue Cruise, and many more [295]. In 2022, Mercedes-Benz's system called "Drive Pilot" was the world's first fully certified SAE Level 3 driving system (conditional driving automation). Initially restricted to geo-fenced areas on the German highway at speeds up to $60$ km/h, an updated version available in 2025 extends its use to flowing traffic up to $95$ km/h under certain conditions on the entire German highway network. Despite these advancements, fully autonomous SAE Level 5 vehicles remain a distant reality, meaning that driving will remain a cooperation between humans and autonomous driving systems in the foreseeable future [363]. It is therefore essential to ensure that this cooperation is as effective as possible in terms of safety, driver comfort, and overall performance.

These SAE levels provide a useful classification for regulatory bodies and automotive manufacturers, enabling legal and technical classification of vehicles based on their level of automation. Yet, they do not fully capture how automation operates in real-world driving and offer limited practical value for drivers. Indeed, automation capabilities do not remain static throughout a drive, and specific features may switch off in some situations. For instance, in an SAE Level 3 vehicle, control may shift between the driver and the automation system. Moreover, even when the driver remains responsible for driving (SAE Levels 0–2), they may activate or deactivate specific assistance features (*e.g.*, Adaptive Cruise Control (ACC) or lane keeping assistance) at will, dynamically adjusting the level of automation.

For safe and effective human-automation interaction, drivers must accurately understand their moment-to-moment responsibilities, and the capabilities and limitations of the automation in use [28, 35, 238]. However, research has shown that drivers often lack awareness of automation limitations [24], which can result in misuse, disuse, or overreliance. The phenomenon of *autonowashing* [63], where marketing exaggerates automation capabilities, leading to misinterpretations of system reliability, further compounds this issue. Given these challenges, we argue that a *dynamic operational level of automation*, which fluctuates over time as driving automation features are engaged or disengaged, is more relevant to day-to-day driver safety than the static SAE Level. Moreover, this operational level of automation could be dynamically adapted based on both the driver state and the driving environment, a concept known as *adaptive automation* [36].

The primary motivation for integrating automation into vehicles is to enhance safety and comfort for drivers and passengers. Automation may contribute to this objective by relieving the driver from parts of the driving task. Yet, excessive automation may have unintended consequences, leading to reduced situation awareness and disengagement from the driving task, which in turn increase risks [20, 58]. Young and Stanton [373] advocate for the ‘cliff-edge’ principle, which consists in restraining automation capabilities until it can fully and reliably take over the task at hand, and favor human-centered support rather than technology-centered automated replacement. Given that drivers will still be involved in the control loop for some time, they should be actively involved rather than passively supervising. Similarly, Endsley [71] emphasizes that automation should only be used where necessary and at the lowest possible level.

Adaptive automation offers a potential solution to these concerns by dynamically adjusting the operational automation level in response to both the driving environment and the driver state. This approach seeks to balance automation and driver involvement, activating automation only when beneficial and without compromising safety. Ideally, adaptive automation should help maintain optimal driver involvement, preventing both overreliance and underuse. Furthermore, in situations where a high level of automation is temporarily feasible, it should allow for a smooth transition back to driver control, providing sufficient time for drivers to regain situation awareness. A key challenge is determining when and how automation should take over versus when it should promote driver involvement.

In this study, we investigate how driver state and driving environment complexity influence the use of driving automation features and how automation, in turn, impacts driving performance. Our findings contribute to the broader discussion on adaptive automation by providing insights into how automation can be designed to support rather than replace human drivers, ultimately enhancing safety. Specifically, we conducted a driving simulator study in which participants could freely activate or deactivate ACC while driving under different conditions across six scenarios. These scenarios varied along two dimensions: driver state, with two levels of cognitive distraction (*i.e.*, presence/absence of a secondary mental calculation task), and driving environment complexity, with three levels (*i.e.*, increasing traffic density and addition of road constructions restricting the number of traffic lanes).

## 3.2 Related Work

### 3.2.1 Driving Environment Complexity

The complexity of the driving environment is influenced by many factors, such as weather conditions, road types, or traffic density. For example, urban driving is considered more complex than rural driving [35].

The driving environment complexity influences the use of driving automation features or systems, either because of system limitations [283], regulations [295], or driver preferences [94, 358]. For instance, the use of driving automation features is usually recommended only in good weather or other beneficial driving conditions [295], and drivers tend to not activate ACC on city streets that have traffic lights [358].

The impact of driving environment complexity is particularly studied for SAE Level 3 vehicles, where the driving automation system reaches its limitations and issues a Take-Over Request (TOR), requiring the driver to quickly resume control. Research indicates that traffic density influences both take-over time and quality [67, 98, 269], whereas weather conditions, such as fog-induced low visibility, do not significantly affect take-over performance [197]. Cooperative perception of the driving environment [233] helps reduce risks associated with TORs. By combining vehicle-localized environment perception (*e.g.*, via cameras and LiDARs) with information from other road users or infrastructure, it allows TORs to be issued earlier, giving the driver more time to get back into the loop while the system is still active.

### 3.2.2 Driver State

As studied in Chapter 2, driver state is commonly categorized into five (sub-)states: drowsiness, mental workload, distraction, emotions, and under influence. Distraction is further divided into manual, visual, auditory, and cognitive distraction.

Maintaining an optimal mental workload is important for effective task performance, as both underload and overload are associated with decreased driving performance [53, 368, 372]. The impact of driving automation on driver workload remains a topic of debate. Some studies suggest that automation reduces workload [308], while Vasta and Biondi [328] conducted a meta-analysis that found no significant reduction in mental workload between manual driving (SAE Level 0) and partially automated driving (SAE Level 2). They argue that workload reduction may only occur in drivers with automation experience or after extended use during long drives. However, it is widely accepted that driving automation reduces stress [308] while also lowering situation awareness [58, 308] and increasing drowsiness [10, 335], distraction, and engagement in Non-Driving Related Task (NDRT) [328]. While much research has examined how automation affects driver state, less attention has been given to how driver state influences the use of automation.

Studies indicate that factors contributing to the complexity of the driving environment have little to no effect on driver state, while others do. For instance, Ahlström et al. [5] reported that light conditions (daylight versus darkness) had a small impact on sleepiness-related measures (*e.g.*, subjective sleepiness, lateral position, speed, or blink durations). However, a more recent study [219] found no significant effect of light conditions on driver sleepiness. Similarly, road environment (rural versus suburban) was shown to have little influence on driver sleepiness [6]. In contrast, other factors appear to have stronger effects. Stanton and Young [308] found that high traffic density significantly increases both driver mental workload and stress, highlighting the importance of considering these interactions when designing adaptive automation systems. Similarly, Park et al. [250] found that environmental factors (such as visual complexity or number of objects in the visual scene) significantly impact driver situation awareness.

### 3.2.3 Driving automation

Driving automation systems consist of a collection of features that can perform all or part of the dynamic driving task, depending on the level of automation [283]. As such systems become increasingly integrated into production vehicles, research has expanded to investigate how drivers interact with these features (for example, during transitions of control [94, 127, 128]), how cooperation between driver and automation can be supported [104, 337], and how human–machine interfaces can enhance these interactions [64].

For instance, Gershon et al. [94] analyzed how drivers leverage automation in real-world driving, focusing on SAE levels 0, 1, and 2, and identified driver- and system-initiated transfers of control. They found that drivers frequently initiated transitions between automation levels, and that these were not necessarily related to immediate risk mitigation, but often due to a mismatch between system capabilities and driver expectations or preferences. While this work sheds light on drivers' reasons for engaging or disengaging automation (*e.g.*, disengaging ACC), it does not address how these decisions are modulated by the driving environment complexity or the driver state.

The influence of driving context on automation use has been highlighted by Orlovska et al. [243], who demonstrated that the relationship between driver, system, and environment is highly interdependent. However, their study also did not incorporate the impact of driver state, such as cognitive distraction, on automation use.

## 3.3 Research Questions

In this study, we examined the relationships between driver state, driving environment complexity, and driving automation. Specifically, we aimed at answering the following four research questions (RQs):

**RQ1** How does driving environment complexity influence ACC use?

**RQ2** How does cognitive distraction influence ACC use?

**RQ3** How does ACC use influence driving performance in different driving environments?

**RQ4** How does ACC use influence driving performance in the presence or absence of a cognitive distraction?

## 3.4 User Study

To analyze the relationships between driver state, driving environment complexity, and driving automation, we conducted a within-subjects driving simulator experiment. Participants drove the same route six times, experiencing three levels of environmental complexity, once with a cognitively demanding secondary task and once without (two levels of cognitive distraction). Participants were instructed to respect traffic laws, including speed limits, and were free to activate or deactivate the ACC at any time.

We investigated whether the activation or deactivation of driving automation features depended on the driver state and/or on the driving environment complexity. We also examined whether ACC use affected driving performance.

### 3.4.1 Participants

A total of $31$ participants were recruited. However, $2$ participants did not complete the experiment due to simulator sickness, leaving a final sample of $N = 29$ participants ($22$ male, $7$ female; mean age $= 31.59$, SD $= 10.68$ years). Data from the $2$ withdrawn participants were entirely discarded for the analysis. Participants were recruited via posters, word-to-mouth or social media (*e.g.*, Facebook, LinkedIn). All participants had a valid driver's license, on average, for $11.62$ (SD $= 11.18$) years. The study was approved by the ethics committee of the Faculty of Psychology, Logopedics, and Educational Sciences of the University of Liege under the reference 2223-081. Participants all provided informed and signed consent before taking part in the experiment.

### 3.4.2 Apparatus and Materials

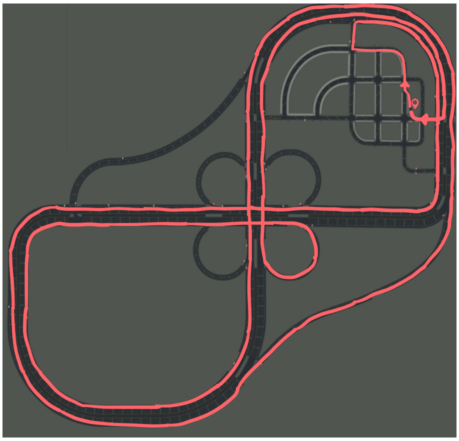
#### Driving Simulator

The experiment was conducted in a driving simulator developed by AISIN Europe, running a customized version of CARLA [66], which is an open-source simulation environment based on Unreal Engine. The setup included three large $50$-inch curved screens, an adjustable car seat, and a Fanatec system comprising a steering wheel, gear shifter, and pedals. The driving simulator featured an automatic transmission vehicle. Turn signals (blinkers) were controlled via buttons on the steering wheel. The vehicle was equipped with an ACC system that not only adapted to the speed of the preceding vehicle but also automatically slowed down in curves. The ACC implementation was inspired by [349] and based on a carrot-chasing algorithm [220, 313]. Participants were able to engage ACC at their current speed and adjust it in increments or decrements of 5 km/h using dedicated buttons on the steering wheel. ACC could only be disengaged via these buttons; pressing the brake pedal neither deactivated ACC nor initiated braking while ACC was engaged. The ACC system did not automatically adapt speed based on recognized speed limit signs, as no signs were present. Instead, speed limits were communicated to participants via a voice agent. Consequently, when using ACC, participants were responsible for manually adjusting the set speed to comply with speed limits (*e.g.*, in road construction zones). Additionally, the vehicle was equipped with a blind spot detection system. Participants also had a small screen on their right to mimic a navigation system. Experimental sessions were recorded using two 4K cameras capturing front and back views, a high-resolution infrared camera positioned behind the wheel as part of a Driver Monitoring System (DMS), and a microphone for audio recording. Figure 3.1 shows the experimental setup. A custom software suite, developed by AISIN Europe, was used to design the test scenarios, execute the simulations, and log experimental data.

### Pre-Test and Feedback Questionnaires

The pre-test questionnaire aimed at collecting demographic and study-specific data. To gain insights on participants' experience and habits, we asked them to rate their familiarity with, use of, and attitude toward ACC, driving automation features and systems on five-point Likert scales. In addition, after each scenario, participants completed a feedback questionnaire, collecting, *e.g.*, information about the reasons for engaging or not engaging the ACC. Both questionnaires are available in Chapter B of the appendix and at `https://osf.io/t3ug7`.

**(a)** Driving simulator



**(b)** Route

**Figure 3.1: Illustration of the experimental setup.** A participant is seated in the driving simulator (a) and performing six scenarios on the same standardized CARLA map named “Town 4” (b). The route consisted of a segment through a small town with pedestrians and traffic lights, followed by an infinite loop highway section. The person depicted provided permission.

### 3.4.3 Tasks

#### Driving Task

Participants operated a simulated vehicle in six sessions of approximately eight minutes each. The route was identical in all six sessions. It began at a gas station in a small town with pedestrians and multiple traffic lights, then continued onto an 8-shaped, three-lane highway before returning to the starting point. All drives took place in daylight and clear weather conditions. A top-down view of the map, including the route, is shown in Fig. 3.1b. Participants were instructed to follow the route provided by the navigation system, which included both oral directions and a visual display on a dedicated screen. They were asked to drive as they normally would in everyday life, *i.e.*, aiming to reach their destination efficiently while adhering to speed limits and maintaining safe driving behavior to minimize the risk of accidents. It was up to them to decide whether and when to engage the ACC system.

#### Secondary Task

In half of the scenarios, one for each level of driving environment complexity, participants had to respond orally to arithmetic calculations dictated by a voice agent. These calculations were triggered based on the vehicle's position along the route. While the specific calculations varied across the three levels of driving environment complexity, their overall difficulty remained comparable. Each calculation involved the addition of two two-digit numbers (*e.g.*, $83 + 42$). This secondary task served as a cognitively distracting element, akin to a mentally demanding conversation a driver might have with a passenger or over the phone.

### 3.4.4 Procedure

Participants were first equipped with electrodes on their left foot to acquire and record Electrodermal Activity (EDA) data using a BIOPAC MP160 system (these data were collected as part of another study, described in Chapter 4). Next, they filled out the pre-test questionnaire. The experimenter then introduced the driving simulator. Participants were instructed to drive as naturally as possible while respecting the speed limits of $50$ km/h in the city and $90$ km/h on the highway. They were also informed about the possibility of road construction in certain scenarios, where a $70$ km/h speed limit would apply, and that these would be announced by the navigation system.

Next, we trained participants on the operation of the driving simulator and on the secondary task through two dedicated training scenarios. The first training one served as a basic tutorial lasting approximately five minutes, conducted on the highway within the CARLA Town 4 map. In this scenario, participants drove alone, with no traffic, and were instructed by the voice agent to perform fundamental tasks such as activating turn signals to change lanes, engaging and disengaging the ACC at specified speeds. The second training scenario followed the same route as in the experiment with low traffic and included a few calculation tasks. This allowed participants to familiarize themselves with both the driving route and the cognitively demanding secondary task before the actual experiment.

Then, participants completed the six test scenarios, each representing different conditions based on a $3$ (driving environment complexity) $\times$ $2$ (cognitive distraction) design. The order of scenarios was counterbalanced across participants. However, two scenarios with the same level of complexity were always presented consecutively, once with cognitive distraction and once without. Half of the participants began with the distraction condition for each level of driving environment complexity, while the other half started without it. The three levels of driving environment complexity varied in traffic conditions (*i.e.*, number of vehicles), as shown in Table 3.1. The highest level also featured road construction zones along the route, restricting traffic to one or two lanes. However, traffic kept flowing, and no standstill sections occurred. Participants were not made aware of the specific conditions for each scenario. After each scenario, participants completed the feedback questionnaire.

**Table 3.1: Overview of the three levels of driving environment complexity (DEC),** detailing the total number of pedestrians, vehicles, and road construction zones present on the route (Fig. 3.1b) in each corresponding scenario.

| Driving environment complexity (DEC) | Nb. of pedestrians | Nb. of vehicles | Nb. of road construction zones |
|---|---|---|---|
| **Level 1 (low)** | 30 | 37 | 0 |
| **Level 2 (medium)** | 30 | 80 | 0 |
| **Level 3 (high)** | 30 | 80 | 3 |

### 3.4.5 Measurements

#### ACC Use

The use of ACC was evaluated using two measurements: the number of ACC activations and the percentage of time the ACC was engaged while the vehicle was in motion, excluding periods when the vehicle was stationary. Additionally, the feedback questionnaires allowed us to gather the reasons behind participants' decisions to engage, disengage, or refrain from engaging the ACC in each condition.

#### Driving Performance

Driving performance was evaluated based on longitudinal and lateral vehicle control. We measured the percentage of time the speed limit was adhered to while the vehicle was in motion, thus excluding periods when the vehicle was stationary. We considered both strict adherence to the speed limit and the effect of allowing a margin of error, defined as a permissible percentage exceeding the limit. For instance, if the speed limit was $90$ km/h and the margin of error was $5\%$, then speeds included in the percentage of time within the speed limit had to remain below $90 * (1 + 5\%) = 94.5$ km/h. Lateral vehicle control was analyzed using the Standard Deviation of Lane Position (SDLP). SDLP was computed after excluding data points corresponding to zero vehicle speed (*e.g.*, when the vehicle was stationary at a traffic light) and periods involving overtaking maneuvers or road changes. These periods were identified by a change in lane or road ID. To account for the full duration of these maneuvers, data were removed from $5$ seconds before to $5$ seconds after each identified lane or road change, as a $10$-second window encompasses $99\%$ of lane changes according to Li et al. [189]. In addition, we also measured the number of lane changes initiated, which serves as a complementary measurement to SDLP for assessing lateral vehicle control. Finally, feedback questionnaires allowed us to gather information on participants' subjective experience of the different driving scenarios and conditions.

## 3.5 Results

Statistical analyses were conducted using JASP (v0.19.3) [146]. Analyses of Variance (ANOVA) with repeated measures were conducted for continuous balanced dependent variables. For continuous unbalanced dependent variables, Linear Mixed Models (LMM) were used. For discrete dependent variables (*e.g.*, count data), Generalized Linear Mixed Models (GLMM) with a Poisson distribution were employed. If significant effects were found, adequate post-hoc comparisons with Bonferroni correction were performed. In such cases, adjusted p-values are systematically reported in the text. All results are reported as statistically significant at $p < .05$. We preferred partial eta squared to partial eta as a measure of effect size for within-subjects design [279].

### 3.5.1 Insights Provided by the Pre-Test Questionnaire

#### Familiarity with Driving Automation Features

All participants reported that they were familiar with the concept of cruise control before taking part in the study. One participant ($3.45\%$) had never driven a vehicle equipped with cruise control, fifteen participants ($51.72\%$) had driven a vehicle equipped with conventional cruise control, and thirteen participants ($44.83\%$) had driven a vehicle equipped with adaptive or intelligent cruise control. Sixteen participants ($55.17\%$) reported driving a vehicle equipped with cruise control daily, six ($20.69\%$) several times a week, four ($13.79\%$) several times a month, two ($6.90\%$) rarely, and one ($3.45\%$) never.

#### Use of Driving Automation Features

Eighteen participants ($62.07\%$) indicated using driving automation features regularly, in particular seven ($24.14\%$) responding 'completely' and eleven ($37.93\%$) 'rather yes' on a five-point scale from 'Not at all' to 'Completely' to the question 'Do you regularly activate driving assistance features (*e.g.*, cruise control, lane-keeping assist)?'. The main reasons given were comfort and convenience, maintaining a constant speed or adhering to speed limits, and managing fuel consumption. Nine participants ($31.03\%$) indicated not using driving automation features regularly, in particular four ($13.79\%$) responding 'not at all' and five

($17.24\%$) 'rather no'. Their main reasons were a dislike of driving automation features, a preference for being in full control of the vehicle, a belief that they did not need them, or simply forgetting to engage them. Two participants ($6.90\%$) gave a neutral response, mainly because they only use these features on highways.

#### Attitudes toward Driving Automation Features

Twenty-six participants ($89.66\%$) stated that they were in favor of driving automation features, with fourteen ($48.28\%$) responding 'completely' and twelve ($41.38\%$) 'rather yes'. The main reasons mentioned were comfort and convenience, increased safety and reduced risk of accidents, reduced fatigue, and the ability to focus more on other tasks. Three participants ($10.34\%$) responded 'neither yes nor no' stating that they prefer to remain in full control of the vehicle or do not use the features anyway.

Finally, twenty-two participants ($75.86\%$) stated that they favored (semi-)autonomous vehicles (namely, vehicles at SAE Levels 3, 4 and 5), with seven ($24.14\%$) responding 'completely' and fifteen ($51.72\%$) 'rather yes'. Similarly to driving automation features, the main reasons included comfort and convenience, increased safety and reduced risk of accidents. Additionally, participants mentioned time management—allowing them to engage in other activities while driving—and a smoother traffic flow. Four participants ($13.79\%$) were not in favor of (semi-)autonomous vehicles, with one ($3.45\%$) responding 'not at all' and three ($10.34\%$) 'rather no'. The main reasons cited were the desire to maintain full control of the vehicle, a love of driving, and the importance of retaining driving capacity. Three participants ($10.34\%$) gave a neutral response, stating that their opinion depended on the intended use of the vehicle or that they did not fully trust autonomous driving.

### 3.5.2 Effects of Driving Environment Complexity and Cognitive Distraction on ACC Use (RQ1 & RQ2)

#### Number of ACC Activations

A GLMM with a Poisson distribution and log link function was fitted to analyze the number of ACC activations based on $174$ observations from

$29$ participants each completing the $6$ scenarios $(29 \times 6 = 174)$. It included fixed effects of driving environment complexity and cognitive distraction, with participant as a random intercept to account for individual differences. Results showed no significant main effect of driving environment complexity $(\chi^2(2) = 4.348, p = .114)$ or cognitive distraction $(\chi^2(1) = .442, p = .506)$ on ACC activations. Additionally, there was no significant interaction between environment complexity and distraction $(\chi^2(2) = .544, p = .762)$. Table 3.2 gives the number of ACC activations in each scenario.

**Table 3.2: Average number of ACC activations ($\pm$ standard deviation)** for the $N = 29$ participants in each of the six scenarios, *i.e.*, in the different driving environment complexities (DEC), with and without cognitive distraction.

| DEC | No distraction | Distraction |
|---|---|---|
| **Level 1** | $3.4 \pm 1.9$ | $3.1 \pm 1.4$ |
| **Level 2** | $3.2 \pm 2.2$ | $3.3 \pm 1.1$ |
| **Level 3** | $4.0 \pm 1.9$ | $3.7 \pm 2.2$ |

### Percentage of ACC Engagement Time

A two-way repeated-measures ANOVA was conducted to examine the effects of driving environment complexity (Level 1, Level 2, Level 3) and cognitive distraction (present or absent) on the percentage of ACC engagement time. Mauchly's test of sphericity was significant for both driving environment complexity $(\chi^2(2) = 6.573, p = .037)$ and the interaction of the two factors $(\chi^2(2) = 10.654, p = .005)$, indicating that the assumption of sphericity was violated. A Huynh-Feldt correction was applied, with epsilon values of $\epsilon = .867$ for driving environment complexity and $\epsilon = .788$ for the interaction. The standard p-value without correction was used for cognitive distraction, as the assumption of sphericity applies only to factors with three or more levels.

The analysis revealed a significant main effect of driving environment complexity $(F(1.734, 48.546) = 4.391, p = .022, \eta_p^2 = .136)$, with a medium effect size (see Table 3.3). Post-hoc comparisons using the Bonferroni correction revealed that the percentages of ACC engagement time at Level 3 of driving environment complexity was significantly smaller than at Level 1 $(p = .022)$, while no significant difference was found between Level 2 and the two other levels $(p = 1.0$ for Level 1 and

$p = .248$ for Level 3). The main effect of cognitive distraction, as well as the interaction between driving environment complexity and cognitive distraction, were not significant ($F(1, 28) = .005, p = .947, \eta_p^2 = .0001$ and $F(1.576, 44.115) = .271, p = .711, \eta_p^2 = .010$, respectively). In other words, ACC engagement time is only influenced significantly by the level of driving environment complexity but not by cognitive distraction.

**Table 3.3: Average percentage of ACC engagement time (± standard deviation)** for the $N = 29$ participants in each of the six scenarios, *i.e.*, in the different driving environment complexities (DEC), with and without cognitive distraction.

| DEC | No distraction | Distraction |
|---|---|---|
| Level 1 | $51.6 \pm 18.2$ | $53.4 \pm 15.4$ |
| Level 2 | $50.3 \pm 21.9$ | $49.8 \pm 15.0$ |
| Level 3 | $43.3 \pm 19.2$ | $41.7 \pm 24.1$ |

#### Rationale for ACC Use Based on Feedback Questionnaires

Participants stated that they activated ACC in all scenarios whenever traffic conditions allowed and to assist with maintaining speed limits. They also reported using ACC during cognitively demanding tasks to help manage all tasks. Additionally, participants indicated activating ACC while driving on the highway. In contrast, participants reported not activating ACC in any scenario when traffic conditions were unsuitable. In particular, they found it challenging to activate ACC in complex driving environments, especially when cognitively distracted. Additionally, they refrained from using ACC in urban areas or situations requiring frequent braking, such as approaching traffic lights. Finally, participants indicated that they deactivated ACC in all scenarios when traffic conditions required it. They also disengaged ACC at highway exits to brake.

### 3.5.3 Impact of ACC Use on Driving Performance (RQ3 & RQ4)

#### Percentage of Time within the Speed Limit

Each of the $29$ participants completed the $6$ scenarios ($3$ driving environment complexity $\times$ $2$ cognitive distraction). Within each scenario, they were free to activate and deactivate ACC as often as they wished. Thus, we expected to process $29 \times 6 \times 2 = 348$ observations. However,

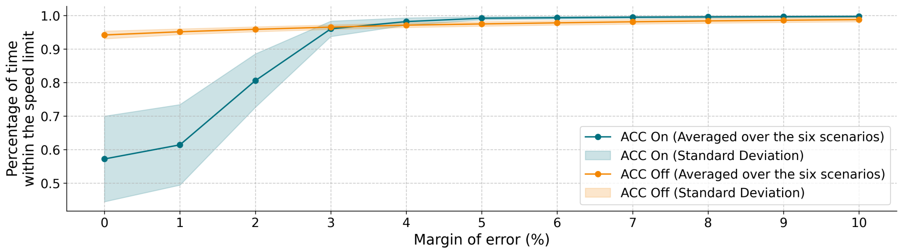



**Figure 3.2: Percentage of time within the speed limit,** averaged across the six scenarios (three levels of driving environment complexity $\times$ two levels of cognitive distraction), as a function of the allowed margin of error on the speed limit, when the ACC is engaged (ACC On, in blue) and disengaged (ACC Off, in orange). When the margin of error is $0\%$, the percentage of time within the speed limit is lower when the ACC is engaged compared to when it is not. However, as the margin of error increases, this percentage rises more notably when the ACC is engaged, eventually surpassing the percentage observed when the ACC is disengaged. This effect can be attributed to the ACC's operation, which aims to maintain the set speed but may occasionally exceed it slightly, or to instances where the ACC set speed is slightly above the speed limit.

$6$ observations were missing because participants had not engaged the ACC at all during an entire scenario. Such missing values led to unbalanced data when analyzing the effect of ACC use on the percentage of time participants were driving under the speed limit. Therefore, a LMM was fitted using Restricted Maximum Likelihood (REML), with Satterthwaite's method, with $342$ observations from $29$ participants. The model included driving environment complexity, cognitive distraction, and ACC as fixed effects, with a random intercept for participants to account for individual variability.

A margin of error was considered when comparing the vehicle's speed to the speed limit. In the driving simulator, when the ACC was set to a given speed, the system attempted to regulate the vehicle's speed accordingly, but some variations leading to higher speeds still occurred. As a result, when no margin of error was allowed, the percentage of time within the speed limit while ACC was engaged appeared unexpectedly low, as shown in Fig. 3.2. However, when a $5\%$ margin of error was permitted, the percentage of time within the speed limit was higher when ACC was engaged than when it was disengaged, meaning that vehicles with ACC engaged were more often within the speed limit than vehicles with ACC disengaged.

We fitted LMMs with margins of error of $0\%$ and $5\%$. When no margin of error was considered (see Table 3.4a), the results showed significant effects of driving environment complexity ($F(2, 308.61) = 30.308, p < .001$), cognitive distraction ($F(1, 308.58) = 11.501, p < .001$), and ACC use ($F(1, 308.67) = 583.321, p < .001$) on the percentage of time within the speed limit. The percentage of time within the speed limit increased as driving environment complexity increased. Post-hoc pairwise comparisons (using contrasts) revealed significant differences between all levels ($Level\ 1 - Level\ 2$: $b = -.064, SE = .019, z = -3.396, p = .002$; $Level\ 2 - Level\ 3$: $b = -.082, SE = .019, z = -4.344, p < .001$; $Level\ 1 - Level\ 3$: $b = -.146, SE = .019, z = -7.769, p < .001$). Moreover, participants drove more time within the speed limit when cognitively distracted compared to when they were not distracted, and when ACC was disengaged compared to when it was engaged.

**Table 3.4: Average percentage of time within the speed limit ($\pm$ standard deviation)** with (a) $0\%$ and (b) $5\%$ margins of error for the $N = 29$ participants when the ACC was engaged or disengaged across the six scenarios (three levels of driving environment complexity (DEC) $\times$ two levels of cognitive distraction). In other words, this corresponds to the conditional probability that speed is within the limit with a margin of error, given whether ACC is on or off, whether the driver is distracted or not, and whether driving environment complexity is at Level 1, Level 2, or Level 3 (*e.g.*, $P$(speed $<$ $(1 +$ margin of error$) *$ speed limit|ACC=off, distraction=without, DEC=Level 1$)$ for the top left cell of the table).

**(a)** No margin of error

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $92.1 \pm 14.8$ | $39.4 \pm 19.6$ | $95.0 \pm 9.5$ | $47.7 \pm 23.2$ |
| **Level 2** | $93.3 \pm 12.1$ | $50.1 \pm 15.3$ | $95.0 \pm 8.8$ | $61.1 \pm 26.3$ |
| **Level 3** | $94.6 \pm 10.0$ | $67.8 \pm 18.6$ | $95.3 \pm 9.6$ | $77.1 \pm 18.9$ |

**(b)** $5\%$ margin of error

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $96.3 \pm 12.5$ | $98.3 \pm 7.4$ | $97.5 \pm 7.5$ | $99.3 \pm 2.3$ |
| **Level 2** | $97.0 \pm 8.6$ | $99.3 \pm 2.3$ | $97.7 \pm 7.0$ | $99.1 \pm 3.2$ |
| **Level 3** | $98.3 \pm 6.6$ | $99.6 \pm 1.2$ | $98.1 \pm 6.6$ | $99.7 \pm 1.1$ |

When a $5\%$ margin of error was considered (see Table 3.4b), the results showed a significant main effect of ACC use ($F(1, 308.67) = 10.188, p = .002$) on the percentage of time within the speed limit. However, there was no significant main effect of driving environment complexity ($F(2, 308.66) = 1.282, p = .279$) or cognitive distraction ($F(1, 308.65) = .778, p = .379$), and no significant interaction. As observed in Fig. 3.2 and Table 3.4b, when accounting for a $5\%$ margin of error, the percentage of time within the speed limit was higher when ACC was engaged compared to when it was disengaged. Although the difference was not significant, the percentage of time within the speed limit tended to increase as driving environment complexity increased and when participants were cognitively distracted, compared to when they were not distracted.

#### Standard deviation of lateral position (SDLP)

A LMM was fitted using REML, with Satterthwaite's method, based on the same $342$ observations. The model included driving environment complexity, cognitive distraction, and ACC as fixed effects, with a random intercept for participants to account for individual variability. Results showed significant effects of driving environment complexity ($F(2, 301.87) = 5.966, p = .003$), cognitive distraction ($F(1, 301.82) = 8.002, p = .005$), and ACC use ($F(1, 302.01) = 48.776, p < .001$) on SDLP. A significant two-way interaction between driving environment complexity and ACC use was found ($F(2, 301.87) = 3.154, p = .044$). No other significant interaction was observed.

Table 3.5 indicates that SDLP was lower when participants were cognitively distracted than when they were not, and that it was also lower when ACC was engaged than when it was disengaged. Post-hoc pairwise comparisons (using contrasts) were conducted to explore the effect of driving environment complexity on SDLP, averaging over cognitive distraction and ACC use. Results showed that SDLP was significantly lower at $Level\ 1$ than at $Level\ 2$ ($b = -.090, SE = .026, z = -3.444, p = .002$) of driving environment complexity. No significant difference was however observed between $Level\ 3$ and the two other levels ($Level\ 1$: $b = -.050, SE = .026, z = -1.934, p = .159$; $Level\ 2$: $b = .039, SE = .026, z = 1.500, p = .401$). We also analyzed the significant interaction between the driving environment complexity and ACC use, comparing ACC on and ACC off within each level of environment complexity. Results showed that ACC significantly reduced SDLP in $Level\ 1$ ($b = .147, SE = .037, Z = 4.001, p < .001$) and $Level\ 2$ of driving environment complexity ($b = .216, SE = .037, z = 5.821, p < .001$). In $Level\ 3$, the effect was smaller and did not reach significance ($b =$

$.085, SE = .037, z = 2.275, p = .069$).

**Table 3.5: Average Standard Deviation of Lane Position (SDLP) (± standard deviation)** for the $N = 29$ participants when the ACC was engaged or disengaged across the six scenarios, *i.e.*, in the three levels of driving environment complexity (DEC), with and without cognitive distraction. SDLP values are expressed in meters.

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $.67 \pm .21$ | $.51 \pm .12$ | $.64 \pm .18$ | $.51 \pm .08$ |
| **Level 2** | $.86 \pm .46$ | $.61 \pm .23$ | $.70 \pm .21$ | $.52 \pm .08$ |
| **Level 3** | $.69 \pm .24$ | $.60 \pm .13$ | $.65 \pm .22$ | $.57 \pm .14$ |

### Number of Lane Changes

A GLMM with a Poisson distribution and log link function was fitted, based on the same $342$ observations. The results revealed a significant main effect of driving environment complexity ($\chi^2(2) = 57.354, p < .001$), but no significant effect of cognitive distraction ($\chi^2(1) = 1.310, p = .252$) or ACC use ($\chi^2(1) = 2.111, p = .146$) on the number of lane changes. A significant three-way interaction between driving environment complexity, cognitive distraction, and ACC use was found ($\chi^2(2) = 11.648, p = .003$). However, no significant two-way interactions were observed. We observe in Table 3.6 that the number of lane changes increased with greater driving environment complexity, likely due to higher traffic levels and road constructions requiring more overtaking maneuvers. Post-hoc pairwise comparisons (using contrasts) revealed significant differences between *Level* 1 and the two other levels (*Level* 2: $b = -1.700, SE = .350, z = -4.860, p < .001$; *Level* 3: $b = -2.590, SE = .375, z = -6.905, p < .001$), but no significant difference between *Level* 2 and *Level* 3 ($b = -.890, SE = .374, z = -2.378, p = .052$). We also analyzed the significant three-way interaction between driving environment complexity, cognitive distraction, and ACC use. Results showed that participants significantly changed lanes more often when the ACC was activated than when it was deactivated at *Level* 2 of driving environment complexity when they were not distracted. In all other conditions, ACC use had no significant effect on the number of lane changes.

**Table 3.6: Average number of lane changes (± standard deviation)** for the $N = 29$ participants when the ACC was engaged or disengaged across the six scenarios, *i.e.*, in the three levels of driving environment complexity (DEC), with and without cognitive distraction.

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $6.1 \pm 2.9$ | $7.0 \pm 5.3$ | $5.3 \pm 3.7$ | $6.3 \pm 4.2$ |
| **Level 2** | $6.7 \pm 4.3$ | $9.4 \pm 5.4$ | $8.6 \pm 5.9$ | $7.2 \pm 4.2$ |
| **Level 3** | $9.3 \pm 4.9$ | $8.3 \pm 5.2$ | $9.0 \pm 4.7$ | $8.6 \pm 6.6$ |

### Perceived Safety Based on Feedback Questionnaires

Regarding the subjective ratings to the statement 'It was complicated to drive safely' (on a seven-point scale from 1 'strongly disagree' to 7 'strongly agree'), a two-way repeated-measures ANOVA revealed a significant main effect of driving environment complexity ($F(2, 56) = 7.974, p < .001, \eta_p^2 = .222$), with a large effect size. Post-hoc comparisons with Bonferroni correction showed that Level 2 and Level 3 of driving environment complexity were rated as being more complicated to drive safely than Level 1 (respectively, $p = .042$ and $p = .002$), but there was no significant difference between Level 2 and Level 3. There was also a large main effect of cognitive distraction ($F(1, 28) = 18.06, p < .001, \eta_p^2 = .392$) with people finding it more complicated to drive safely when cognitively distracted compared to when they were not. Finally, there was no significant interaction between the two factors (see Table 3.7).

**Table 3.7: Average subjective ratings (± standard deviation) to the statement 'It was complicated to drive safely'** (on a seven-point scale from (1) 'strongly disagree' to (7) 'strongly agree') for the $N = 29$ participants in each of the six scenarios, *i.e.*, in the different driving environment complexities (DEC), with and without cognitive distraction.

| DEC | No distraction | Distraction |
|---|---|---|
| **Level 1** | $2.6 \pm 1.8$ | $3.4 \pm 1.6$ |
| **Level 2** | $3.3 \pm 1.9$ | $3.9 \pm 1.8$ |
| **Level 3** | $3.4 \pm 1.7$ | $4.9 \pm 1.7$ |

## 3.6 Discussion

Our results indicate that driving environment complexity influences the percentage of ACC engagement time, but not the number of ACC activations (**RQ1**). Regardless of cognitive distraction conditions, the percentage of time ACC is engaged was lower in the most complex driving environment compared to the less complex one. In line with survey studies [358], participants reported that they preferentially activated ACC on highways but refrained from using it in urban areas where frequent braking was required, notably at traffic lights. Overall, it seems that cognitive distraction does not influence the absolute ACC use, either in terms of the number of activations or overall engagement time, even if people may feel that the more demanding situations (*i.e.*, most complex environment coupled with a secondary task) make the activation more challenging (**RQ2**).

Additionally, our study suggests that ACC use influences driving performance (**RQ3** and **RQ4**). While ACC use does not appear to affect the number of lane changes, it impacts both the percentage of time the speed limit is adhered to and the SDLP. Specifically, regarding longitudinal control, whether ACC has a positive or negative effect on speed limit compliance depends on the margin of error allowed. Indeed, if no margin of error is permitted, and we consider only the percentage of time the vehicle's speed is strictly less than or equal to the speed limit, then driving with ACC engaged leads to lower speed limit compliance than driving without it. This effect may be explained by the ACC's implementation, which aims to maintain a set speed but may occasionally exceed it slightly, or by situations where the set speed itself is slightly above the speed limit. In contrast, if a $5\%$ margin of error is allowed (meaning that the vehicle's speed can slightly exceed the limit), then driving with ACC improves speed limit compliance. Therefore, allowing a margin of error cancels this effect, resulting in nearly total compliance within the speed limit plus the margin of error. Notably, a $5\%$ tolerance is commonly accepted by speed cameras in real-world settings. Note that participants reported in feedback questionnaires that they activated ACC to help them maintain speed limits. This was also one of the main reasons cited in the pre-test questionnaire for using driving automation features regularly in real life. In the literature, the effect of ACC on speed remains inconclusive. For instance, Hoedemaeker and Brookhuis [131] found that drivers in a simulator study were driving faster with ACC activated, while Vollrath et al. [336] observed lower maximum speed and fewer speed limit violations when driving with ACC compared to manual driving. Meanwhile, Bianchi Piccinini et al. [27] found no statistically significant effect of ACC on traveling speed. However, as shown in our study, results highly depend on the way ACC is implemented and on the measures used. Besides, when considering strict compliance (no margin of error), speed limit adherence increases with driving environment complexity and cognitive distraction, regardless of ACC use. With a $5\%$ margin of error, no significant differences are observed. Regarding lateral control, SDLP was lower when ACC was engaged compared to when it was disengaged, suggesting that ACC use improves lateral driving performance. SDLP was also lower when participants were cognitively distracted than when they were not, a finding consistent with previous studies [194, 271]. Additionally, lateral control was better (*i.e.*, lower SDLP) in the lowest driving environment complexity compared to the medium complexity level. Notably, a significant interaction was found between driving environment complexity and ACC use, suggesting that lateral control was improved by ACC use in the two lowest levels of environment complexity, but not in the most complex one.

As for lane changes, the complexity of the driving environment had a significant effect, with more lane changes occurring in environments with higher traffic density, likely due to more frequent overtaking maneuvers. Neither cognitive distraction nor ACC use had a significant effect on the number of lane changes. However, a significant three-way interaction was observed between driving environment complexity, cognitive distraction, and ACC use. More lane changes was observed when ACC was engaged compared to when it was disengaged, but only in the medium-complexity environment and in the absence of distraction. No significant differences were found in the other conditions.

Our findings underscore the importance of adopting a *dynamic operational level of automation* starting as early as SAE Level 1, as even a feature like ACC was activated and deactivated multiple times during an eight-minute route. Moreover, the study supports the idea that adaptive automation, dynamically adjusting the level of automation, should account for both driving environment complexity and driver state. While cognitive distraction, unlike driving environment complexity, did not significantly influence ACC use in this study, participants reported that both factors affected their perception of driving safety. Although participants did not significantly modify their ACC use in response to cognitive distraction, automatically adapting the automation level based on driver state might still improve driving performance.

## 3.7 Limitations

We acknowledge several limitations to our study. First, we operationalized driver state, driving environment complexity, and driving automation features by focusing specifically on cognitive distraction, traffic density, and ACC, respectively. While this constitutes an important first step, future research should explore other driver states (*e.g.*, drowsiness or visual distraction), additional aspects of driving environment complexity (*e.g.*, weather conditions or road types), and other driving automation features (*e.g.*, lane keeping assistance) to assess whether similar conclusions can be drawn. Furthermore, traffic density was constrained by the computational limits of the system running CARLA, as higher densities increased the risk of simulator crashes. Besides, it is worth noting that the ACC system used here could only be disengaged using a designated button, as pressing the brake pedal neither deactivated ACC nor initiated braking while ACC was engaged. This differs from real-world ACC implementations, where braking typically deactivates the system. Second, participants were not selected based on age, background, or driving experience. As these factors may also influence automation use and driving behavior, their impact should be considered in future studies to support the development of adaptive automation that is commensurate with experience or other drivers' characteristics. Of course, as with all simulator studies, the question of transferability, reliability, and validity of the results to real-world driving conditions arises. Although driving simulators allow for controlled experimental conditions, conclusions drawn in a simulated environment may differ from the ones drawn in the real world.

## 3.8 Conclusion

With our simulator study, we investigated whether and how driving environment complexity and driver state influence reliance on driving automation features (**RQ1** & **RQ2**), and how using such features may impact driving performance (**RQ3** & **RQ4**). $N = 29$ participants drove in six different conditions, consisting of three levels of driving environment complexity and two levels of cognitive distraction, with the ability to activate or deactivate ACC at any time.

We found that (**RQ1**) the overall time of ACC engagement was lower in complex driving environments than in less complex ones, while the number of ACC activations itself was not significantly affected by the environment. Additionally, (**RQ2**) we did not observe any significant effect of cognitive distraction on ACC use. Regarding driving performance, our study revealed that ACC use had an effect on the number of lane changes only in the medium-complexity environment and in the absence of cognitive distraction. ACC use led to lower SDLP and thus improved lateral control in the two lowest levels of environment complexity, no matter if participants were distracted or not. Furthermore, ACC use did have a negative or positive impact on compliance with speed limits, depending on whether a strict speed limit or a margin of error was considered. Additionally, we found that (**RQ3**) more lane changes occurred in more complex driving environments. The lateral control (*i.e.*, SDLP) was better in the least complex driving environment. For strict speed limit compliance, adherence was higher without ACC and increased with driving environment complexity, regardless of ACC use. However, with a $5\%$ margin of error, speed limit adherence was higher with ACC than without, with no significant effect of driving environment complexity. Finally, (**RQ4**) cognitive distraction did not affect the number of lane changes. However, lateral control was better with ACC and with cognitive distraction than in other situations, while strict speed limit compliance was better without ACC and with cognitive distraction. With a $5\%$ margin of error, speed limit adherence was higher with ACC but with no significant effect of cognitive distraction.

In conclusion, the present study suggests that *to develop adaptive automation features that vehicle users rely on, it is essential to consider both the driving environment and the driver state*. This will help ensure that driving automation systems can dynamically adjust to varying conditions and driver needs to enhance both safety and comfort.

# 4

# EDA- and Gaze-Based Indicators of Driver Cognitive Distraction

**Contents of this chapter**

CONTEXT. After examining driving performance in Chapter 3, this chapter focuses on the characterization of the driver state with physiological and behavioral indicators. Specifically, we investigate whether and how Electrodermal Activity (EDA) and gaze parameters reflect driver cognitive distraction under varying traffic conditions and Adaptive Cruise Control (ACC) use through the driving simulator study described in Chapter 3. We analyze (1) three EDA-based indicators of cognitive distraction: mean skin conductance level (SCL), mean amplitude of skin conductance responses (SCR amplitude), and rate of skin conductance responses (SCR rate); as well as (2) three gaze-based indicators: percent road center, horizontal gaze dispersion, and vertical gaze dispersion.

**Contributions and Key Findings**

(1) All three EDA-based indicators are significantly influenced by cognitive distraction and ACC use, whereas environment complexity significantly influences SCL and SCR amplitude, but not SCR rate.

(2) All three gaze-based indicators are significantly influenced by ACC use, whereas environment complexity only significantly influences vertical gaze dispersion, but not percent road center and horizontal gaze dispersion, and cognitive distraction significantly influences percent road center and vertical gaze dispersion, but not horizontal gaze dispersion.

(3) Surprisingly, cognitive distraction reduces road center gaze and increases vertical dispersion. However, complementary analyses reveal that these effects primarily occur between mental calculations periods, whereas active mental calculations phases are characterized by a temporary increase in gaze concentration.

④ Overall, this chapter provides an analysis of physiological and behavioral indicators of driver state, examining how they are jointly influenced by driver cognitive distraction, environment complexity, and ACC use.

RELATED PUBLICATION. This chapter is adapted from Halin et al. [108, 113], which have been merged into a single chapter. The data analyzed here were collected during the same simulator study presented in Chapter 3. Therefore, we do not repeat the full description of the user study but focus instead on the elements specific to this chapter.

## 4.1 Introduction

Until fully Autonomous Vehicles (AVs) populate our roads entirely, ensuring that drivers remain in an appropriate state to operate the vehicle or to supervise driving automation features is critical for safety. Monitoring the state of drivers is therefore essential to guarantee that they are ready to perform the required actions, whether they are actively controlling the vehicle, supervising driving automation features, or expected to take over control at any moment.

Cognitive distraction, a state in which drivers' mental resources are diverted from the driving task (*e.g.*, talking with a passenger or on the phone), can significantly impair performance and jeopardize safety. To detect such a state, both physiological and behavioral signals can be monitored [115]. However, these signals may also be influenced by external factors such as traffic conditions and the use of driving automation features like Adaptive Cruise Control (ACC).

In this chapter, we investigate whether and how two types of indicators reflect driver cognitive distraction under varying traffic conditions and ACC use: 1 three Electrodermal Activity (EDA)-based physiological indicators—mean skin conductance level (SCL), mean amplitude of skin conductance responses (SCR amplitude), and rate of skin conductance responses (SCR rate); and 2 three gaze-based behavioral indicators—percent road center, horizontal gaze dispersion, and vertical gaze dispersion.

The data analyzed in this chapter were collected during the driving simulator study described in Chapter 3, in which participants could freely activate or deactivate the ACC at any time while driving across six scenarios. These scenarios varied along two dimensions: 1 *driver state*, with two levels of cognitive distraction (*i.e.*, presence or absence of a mental calculation task); and 2 *driving environment complexity*, with three levels (*i.e.*, increasing traffic density and the presence or absence of road construction restricting the number of traffic lanes).

This work focuses on the characterization of the driver state and investigates whether and how both physiological (EDA-based) and behavioral (gaze-based) indicators reflect driver cognitive distraction under varying traffic conditions and ACC use.

## 4.2 Related Work

Driver state can be categorized into five (sub-)states: drowsiness, mental workload, distraction (manual, visual, auditory, and cognitive), emotions, and under influence [115]. Driver Monitoring System (DMS) aim to characterize these states using specific indicators and appropriate sensors that capture the corresponding signals.

Cognitive distraction, defined by the National Highway Traffic Safety Administration (NHTSA) [237] as the mental workload associated with a task that involves thinking about something else apart from the driving task, is commonly characterized using behavioral and physiological indicators, such as gaze parameters (*e.g.*, fixation duration, gaze distribution) [311, 211, 174], pupil diameter [369], Electrodermal Activity (EDA) [377, 150], and heart activity [251, 278]. In this chapter, we focus on EDA and gaze parameters.

Monitoring cognitive distraction fundamentally involves assessing mental workload, which is influenced not only by the complexity and demands of the driving task itself but also by any concurrent secondary task (*i.e.*, cognitive distraction). For this reason, we study both how selected indicators reflect the driver's cognitive state and how these indicators behave under varying traffic conditions and during the use of driving automation features.

### 4.2.1 EDA-Based Indicators of Cognitive Distraction

EDA refers to autonomic changes in the electrical properties of the skin, driven by sweat gland activity regulated by the sympathetic nervous system [34]. EDA is typically decomposed into 1 a tonic component, reflecting slow, baseline shifts in arousal over time, and 2 a phasic component, representing rapid, transient changes in response to discrete stimuli or events. The most commonly studied EDA signal is Skin Conductance (SC), which can be measured via skin-surface electrodes [325]. The measure of the tonic component is referred to as the Skin Conductance Level (SCL), while the abrupt changes in the phasic component, called peaks, are referred to as Skin Conductance Response (SCR).

Li et al. [186] analyzed seventeen features extracted from SC, SCL, and SCR under varying cognitive load and identified SCR rate as the most influential indicator of driver arousal, while also recommending

SCR amplitude. Radhakrishnan et al. [268] demonstrated that both automation and the environment significantly influence SCR rate. Similarly, Foy and Chapman [83] found that road type significantly affects mean SC and SCR rate. However, none of these studies examined the combined influence of cognitive distraction, driving environment complexity, and ACC use on the EDA-based indicators to assess their reliability in dynamic, real-world driving conditions. In the present study, we analyze these influences on SCL, SCR amplitude, and SCR rate.

### 4.2.2 Gaze-Based Indicators of Cognitive Distraction

In the driving context, visual and mental tasks are inherently intertwined. Drivers must continuously perceive their environment and interpret visual information to make appropriate decisions, such as, *e.g.*, scanning highway lanes while simultaneously determining when to brake or change lanes. Given this strong link between visual perception and cognitive processing, gaze parameters have often been used to assess drivers' mental workload [211].

Different approaches exist to code glance data, namely the direction-based (where are drivers looking?), the target-based (what objects are drivers looking at?), and the purpose-based (why are drivers looking where they are looking?) coding schemes of glance data [9, 162]. Direction-based coding is commonly used to compute indicators like "eyes off road" or "percent road center", which is the percentage of gaze data points that fall within a predefined road center area [330]. Target-based coding involves a manual coding of glance targets, which are categorized by the target type (*e.g.*, bicyclist, traffic sign, or mobile phone). The purpose-based approach specifically defines which areas a driver must acquire information from to be considered attentive [9].

Target-based approaches infer driver distraction from glances toward targets deemed irrelevant for driving. While effective for detecting visual distraction, they are less suited for cognitive distraction, where drivers may maintain their gaze on the road while being mentally distracted. By definition, purpose-based approaches might better capture cognitive distraction but still face similar limitations, as drivers may appear visually attentive while mentally disengaged (*i.e.*, "look but not see" situations, where drivers' eyes are on the road, but the mind is elsewhere). Moreover, their practical implementation is complex as it requires extensive contextual information about, *e.g.*, glance history, infrastructure layout, or traffic regulations. Therefore, we focus on direction-based gaze indicators of cognitive distraction, which are more accessible and widely used. Specifically, we examine percent road center, horizontal and vertical gaze dispersion.

Cognitive distraction has been associated with increased gaze concentration toward the road center, manifested as higher percent road center [330], lower horizontal and vertical gaze dispersion, and decreased glance frequency to mirrors and speedometer [117, 330]. Similar patterns have also been noted under high driving task complexity [330].

Let us note that percent road center areas can vary considerably across studies: they are sometimes circular [8, 330, 348], sometimes rectangular [117, 330], with horizontal angle of view ranging from $8$ to $20$ degrees. Furthermore, this area can be a fixed area on the windshield or can be centered around the driver's most frequent gaze angle, and thus vary for each driver [348]. While percent road center was initially computed based on gaze fixations, using raw gaze points (*i.e.*, all gaze data not clustered into fixations and saccades) was shown to give highly correlated values [8, 348].

Gaze dispersion, expressed as the standard deviation of gaze points (*a.k.a.*, gaze variability), can be computed using the driver's gaze angle or the projection of the gaze trail on a plane in front of the driver [330]. Furthermore, it can also be assessed separately for the horizontal and vertical gaze positions or angles [348], or for the combined vertical and horizontal components of gaze into a single gaze position or angle [330].

In this study, we compute 1 the percent road center using raw gaze angles and a rectangular road center area with a horizontal angle of view of $20$ degrees, 2 horizontal gaze dispersion, and 3 vertical gaze dispersion.

## 4.3 Research Questions

This work aims to answer the following overarching research question (RQ): *What are the effects of cognitive distraction on both 1 physiological (EDA-based) and 2 behavioral (gaze-based) indicators under varying traffic conditions and ACC use?* To answer this question, we investigated the six following questions organized in two sub-series:

**RQ1.1** How is the *mean skin conductance level (SCL)* affected by cognitive distraction, traffic conditions, and ACC use?

**RQ1.2** How is the *mean amplitude of skin conductance responses (SCR amplitude)* affected by cognitive distraction, traffic conditions, and ACC use?

**RQ1.3** How is the *rate of skin conductance responses (SCR rate)* affected by cognitive distraction, traffic conditions, and ACC use?

**RQ2.1** How is the *percent road center* affected by cognitive distraction, traffic conditions, and ACC use?

**RQ2.2** How is the *horizontal gaze dispersion* affected by cognitive distraction, traffic conditions, and ACC use?

**RQ2.3** How is the *vertical gaze dispersion* affected by cognitive distraction, traffic conditions, and ACC use?

## 4.4 User Study

This study is based on data collected during a driving simulator experiment designed to investigate how drivers' cognitive state and driving environment complexity influence their reliance on driving automation features. Full details regarding the materials, experimental tasks, and procedure are provided in Chapter 3.

The present within-subjects study examines how the presence of a cognitively distracting task affects EDA- and gaze-based indicators under varying traffic conditions and ACC use.

### 4.4.1 Participants

Thirty-one individuals were initially recruited. However, two did not complete the experiment due to simulator sickness, resulting in a final sample of $N = 29$ participants ($22$ male, $7$ female).

For the analysis of EDA data, one additional participant was excluded due to generally poor signal quality and outlier values. Consequently, the EDA dataset included $N = 28$ participants ($21$ male, $7$ female).

All participants were recruited through posters, word of mouth, or social media, and all held a valid driver's license. The study was approved by the ethics committee of the Faculty of Psychology, Logopedics, and Educational Sciences at the University of Liège, Belgium, under the reference 2223-081. All participants provided written informed consent before participation.

### 4.4.2 Apparatus and Materials

#### Driving Simulator

The experiment was conducted using a driving simulator developed by AISIN Europe, running a customized version of CARLA [66], which is an open-source simulation environment based on Unreal Engine. The setup included three large $50$-inch curved screens, an adjustable car seat, and a Fanatec system comprising a steering wheel, gear shifter, and pedals. The vehicle was equipped with an intelligent ACC system that adapted its speed to both the leading vehicle and the curvature of the road, adjustable in $5$ km/h increments via steering-wheel buttons. A secondary, smaller screen mimicked a navigation display. A custom software suite developed by AISIN Europe was used to design the test scenarios, execute the simulations, and log experimental data. Verbal responses to the cognitive distraction task were recorded via a microphone.

#### BIOPAC MP160 System

Participants were equipped with electrodes (BIOPAC EL507A) placed on their left foot to acquire and record EDA data using a BIOPAC MP160 system at a sampling rate of $2{,}000$ Hz (see Fig. 4.1). A gel (BIOPAC GEL101A) was applied to enhance conductivity between the skin and the electrode. To minimize driving inconvenience caused by the electrodes and cables, and to reduce motion artifacts, the electrodes were attached to the sole of the left foot, which rested on a designated area. Since the driving simulator replicated an automatic transmission vehicle, participants did not need to use their left foot to control the pedals. This electrodes' placement has been employed in previous studies [264] and is known to provide reliable EDA measurements [325].

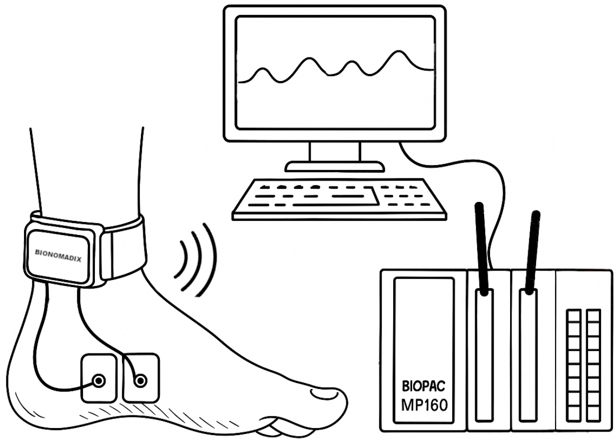



**Figure 4.1: Illustration of the setup for the acquisition of the EDA signal.** This image shows the placement of the two electrodes on the sole of the foot, as well as the setup with the BIOPAC to acquire the EDA signal.

#### Driver Monitoring System

The DMS was composed of a high-resolution infrared camera positioned behind the wheel (see Fig. 4.2), recording images at $60$ Hz. Images were processed in real time by algorithms developed by AISIN Europe. Face orientation (pitch, yaw, and roll, in degrees), eye-opening (in millimeters), and gaze direction (horizontal and vertical gaze angles) were logged.

### 4.4.3 Procedure and Tasks

Participants were first equipped with EDA electrodes and completed a pre-test questionnaire collecting demographic and study-related information. They were then familiarized with the driving simulator and tasks through a brief training session.

The experiment followed a within-subjects design, in which each participant completed six driving sessions of approximately eight minutes each. The route was identical across sessions and included both urban segments (with traffic lights) and highway segments, all under clear daylight conditions. Participants were instructed to follow navigation cues, comply with traffic rules (including speed limits), and drive as they normally would, with the option to activate or deactivate the ACC

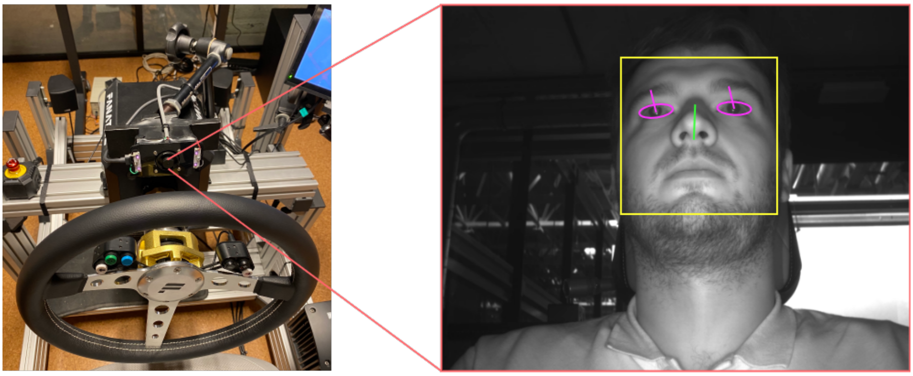


**Figure 4.2: Illustration of the setup for the driver monitoring system.** The figure shows on the left the infrared camera positioned behind the wheel and on the right an image acquired by the camera and processed by the algorithms developed by AISIN Europe. The person depicted provided permission.

at any time.

Two experimental factors were manipulated: 1 *driving environment complexity*, with three levels based on traffic density and road constructions; and 2 *cognitive distraction*, with or without a secondary mental calculation task. Level 1 of driving environment complexity corresponded to low traffic density ($37$ vehicles in the environment). Level 2 involved medium traffic density ($80$ vehicles). Level 3 maintained the same medium traffic density ($80$ vehicles) but included three road construction zones that reduced the number of lanes to one or two without causing congestion. In the cognitive distraction condition, participants performed a secondary task consisting of responding aloud to arithmetic calculations (additions of two two-digit numbers, *e.g.*, $83 + 42$) dictated by a voice agent at a frequency of approximately three per minute.

The six scenarios, combining the three levels of environment complexity and the two levels of distraction, were presented in a counterbalanced order across participants. For each level of complexity, the two corresponding scenarios (with and without distraction) were performed consecutively. Each scenario began with a brief calibration phase to synchronize all recorded data streams. After each drive, participants completed a short feedback questionnaire. Analysis of both pre-test and feedback questionnaire data is presented in Chapter 3. The experiment concluded with a five-minute relaxation period accompanied by soothing music.

### 4.4.4 Measurements

#### EDA-Based Indicators of Cognitive Distraction

Tonic (skin conductance level, SCL) and phasic (skin conductance response, SCR) components of the EDA signal were extracted using the NeuroKit2 Python toolbox [208]. We computed (1) the mean values of the SCL, (2) the SCR amplitude (excluding the tonic component), and (3) the SCR rate (number of peaks per minute). These indicators were extracted from the total duration of each session (*i.e.*, approximately $8$ minutes).

#### Gaze-Based Indicators of Cognitive Distraction

We analyzed (1) the percent road center [330], which is the percentage of gaze data points that fall within the road center area, and (2) the horizontal and (3) vertical gaze dispersions, which are computed as the standard deviations of horizontal and vertical gaze angles, respectively.

For each participant and each session, the road center point was determined as the mode, or most frequent gaze angle, and the road center area was defined as a $20$ degrees (horizontal) $\times$ $15$ degrees (vertical) rectangular area centered around the road center point, as done in [330]. In line with prior work showing high correlation between fixation-based and raw-data-based indicators [8, 348], we used all gaze data points to compute the indicators, without distinguishing between fixations and saccades. However, we analyzed these indicators only during the drive on the highway segment, which lasted approximately $5$ minutes (thus removing the driving period in the city center, where vehicles were often stopped at traffic lights).

## 4.5 Results

Statistical analyses were conducted using JASP (Version 0.19.3) [146]. Descriptive and inferential statistics are reported for each indicator. Results were considered statistically significant at $p < .05$.

### 4.5.1 EDA-Based Indicators of Cognitive Distraction (RQ1)

Each of the $28$ participants whose EDA data were analyzed completed the $6$ scenarios, resulting from the combination of $3$ levels of driving environment complexity and $2$ levels of cognitive distraction. Within each scenario, they were free to activate and deactivate ACC as often as they wished. This design was expected to yield $28 \times 6 \times 2 = 336$ observations (including data from both ACC on and ACC off conditions). However, $4$ observations were missing because participants never engaged the ACC during an entire given scenario. These missing values resulted in unbalanced data when analyzing the effect of ACC use on the different indicators. Additionally, electrodermal data from one participant were identified as outliers for the Level 2 driving environment complexity with no distraction conditions using a Z-score threshold of $3$ and were thus discarded. As a result, the final analysis was conducted on $330$ observations from $28$ participants. A Linear Mixed Models (LMM) was fitted using Restricted Maximum Likelihood (REML) estimation and Satterthwaite's approximation for degrees of freedom. The model included driving environment complexity, cognitive distraction, and ACC as fixed effects, with a random intercept for participants to account for individual variability.

#### SCL (RQ1.1)

The LMM analysis revealed significant main effects of cognitive distraction ($F(1, 291.01) = 11.982, p < .001$), driving environment complexity ($F(2, 291.01) = 7.212, p < .001$), and ACC use ($F(1, 291.00) = 3.893, p = .049$) on SCL. There was no significant interaction between cognitive distraction and driving environment complexity ($F(2, 291.01) = .160, p = .852$), cognitive distraction and ACC use ($F(1, 291.01) = .170, p = .680$), driving environment complexity and ACC use ($F(2, 291.01) = .123, p = .884$), and finally cognitive distraction, driving environment complexity and ACC use ($F(2, 291.01) = .118, p = .889$).

Table 4.1a indicates that SCL was higher when participants were cognitively distracted compared to when they were not, and that SCL was lower when ACC was engaged compared to when it was disengaged. Post-hoc pairwise comparisons (using contrasts) were conducted to explore the effect of driving environment complexity on SCL, averaging over distraction and ACC use. P-values are adjusted using Bonferroni correction. Results showed that SCL was significantly lower at $Level\ 3$ compared to both $Level\ 1$ ($b = 1.933, SE = .569, z = 3.399, p = .002$) and $Level\ 2$ ($b = 1.821, SE = .575, z = 3.166, p = .005$) of driving environment complexity. Thus, SCL decreases as the driving environment complexity increases. No significant difference was, however, observed between $Level\ 1$ and $Level\ 2$ ($b = .112, SE = .572, z = .196, p = 1.0$).

**Table 4.1: Averaged EDA-related indicators of cognitive distraction ($\pm$ standard deviation) for the $N = 28$ participants** when the ACC was engaged and disengaged across the six scenarios, *i.e.*, in the three levels of driving environment complexity (DEC), with and without cognitive distraction.

**(a)** Skin conductance level (SCL)

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $6.3 \pm 3.3$ | $6.1 \pm 3.3$ | $6.7 \pm 3.5$ | $6.4 \pm 3.5$ |
| **Level 2** | $6.4 \pm 3.5$ | $5.6 \pm 3.1$ | $6.7 \pm 3.6$ | $6.5 \pm 3.6$ |
| **Level 3** | $5.8 \pm 3.1$ | $5.7 \pm 3.1$ | $6.2 \pm 3.5$ | $6.2 \pm 3.6$ |

**(b)** Amplitude of the skin conductance responses (SCR amplitude)

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $.16 \pm .14$ | $.14 \pm .15$ | $.22 \pm .21$ | $.16 \pm .14$ |
| **Level 2** | $.20 \pm .20$ | $.11 \pm .13$ | $.23 \pm .21$ | $.16 \pm .15$ |
| **Level 3** | $.22 \pm .22$ | $.18 \pm .23$ | $.22 \pm .25$ | $.21 \pm .21$ |

**(c)** Rate of the skin conductance responses (SCR rate)

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $5.7 \pm 1.9$ | $4.5 \pm 1.8$ | $5.8 \pm 2.0$ | $5.4 \pm 2.7$ |
| **Level 2** | $5.7 \pm 2.1$ | $4.5 \pm 2.2$ | $5.4 \pm 1.7$ | $4.8 \pm 1.8$ |
| **Level 3** | $5.0 \pm 2.4$ | $4.7 \pm 2.6$ | $5.7 \pm 2.4$ | $5.5 \pm 2.8$ |

### SCR Amplitude (RQ1.2)

The analysis revealed significant main effects of cognitive distraction ($F(1, 291.03) = 7.585, p = .006$), driving environment complexity ($F(2, 291.03) = 3.198, p = .042$), and ACC use ($F(1, 291.00) = 17.304, p < .001$) on SCR amplitude, but no significant interaction between cognitive distraction and driving environment complexity ($F(2, 291.03) = .626, p = .535$), cognitive distraction and ACC use ($F(1, 291.01) = .087, p = .769$), driving environment complexity and ACC use ($F(2, 291.01) = 1.603, p = .203$), and finally cognitive distraction, driving environment complexity and ACC use ($F(2, 291.01) = 1.046, p = .353$).

Table 4.1b indicates that SCR amplitude was higher when participants were cognitively distracted compared to when they were not. SCR amplitude was lower when ACC was engaged compared to when it was disengaged. Post-hoc pairwise comparisons showed that SCR amplitude was higher at $Level\ 3$ than at $Level\ 1$ of driving environment complexity ($b = -.126, SE = .052, z = -2.442, p = .044$). Thus, SCR amplitude increases as the driving environment complexity increases. No significant difference was, however, observed between $Level\ 1$ and $Level\ 2$ ($b = -.033, SE = .052, z = -.627, p = 1.0$), and $Level\ 2$ and $Level\ 3$ ($b = -.094, SE = .052, z = -1.791, p = .220$).

#### SCR Rate (RQ1.3)

The analysis revealed significant main effects of cognitive distraction ($F(1, 291.13) = 6.303, p = .013$) and ACC use ($F(1, 291.05) = 15.323$, $p < .001$) on SCR rate. However, there was no significant effect of driving environment complexity ($F(2, 291.13) = .492, p = .612$) on SCR rate and no significant interaction between cognitive distraction and driving environment complexity ($F(2, 291.12) = 1.713, p = .182$), cognitive distraction and ACC use ($F(1, 291.06) = 2.002, p = .158$), driving environment complexity and ACC use ($F(2, 291.06) = .988, p = .373$), and finally cognitive distraction, driving environment complexity and ACC use ($F(2, 291.06) = .317, p = .728$).

Table 4.1b indicates that the SCR rate was higher when participants were cognitively distracted compared to when they were not. Similarly to SCL and SCR amplitude, SCR rate was lower when ACC was engaged compared to when it was disengaged.

### 4.5.2 Gaze-Based Indicators of Cognitive Distraction (RQ2)

#### Main Analyses

The $29$ participants completed $6$ scenarios ($3$ levels of driving complexity $\times$ $2$ levels of cognitive distraction) in which they could freely activate or deactivate the ACC, allowing for up to $348$ observations ($29 \times 6 \times 2$).

However, in $7$ of them, ACC was never used or the gaze direction was not detected at all, resulting in missing data and an unbalanced dataset for analyses involving ACC use, thus preventing us from conducting a standard Analyses of Variance (ANOVA). Therefore, we defaulted to a LMM using REML and Satterthwaite's approximation, based on the remaining $341$ observations. The model included environment complexity, cognitive distraction, and ACC use as fixed effects, with a random intercept for participants to account for individual differences.

### Percent Road Center (RQ2.1)

The analysis revealed significant main effects of cognitive distraction ($F(1, 301.02) = 14.196, p < .001$) and ACC use ($F(1, 301.04) = 27.587, p < .001$) on percent road center. However, there was no significant effect of driving environment complexity ($F(2, 301.03) = 1.197, p = .303$), and no significant cross-factor interaction. As shown in Table 4.2a, percent road center was lower when participants were cognitively distracted ($73.3 \pm 12.8$ globally), compared to when they were not ($76.1 \pm 12.8$ globally). Percent road center was also lower when ACC was off ($71.9 \pm 13.8$ globally), compared to when ACC was on ($75.8 \pm 13.4$ globally). Participants were thus looking more often at the road center area during sessions without distraction and when the ACC was activated.

### Horizontal Gaze Dispersion (RQ2.2)

For horizontal gaze dispersion, there was only a significant main effect of ACC use ($F(1, 301.08) = 65.531, p < .001$), with greater horizontal gaze dispersion when ACC was deactivated ($8.7 \pm 3.1$ globally) compared to when it was activated ($7.3 \pm 2.3$ globally) (see Table 4.2b). Participants thus had a more concentrated horizontal gaze when ACC was engaged. There were no significant effects of cognitive distraction ($F(1, 301.05) = .023, p = .880$) or driving environment complexity ($F(2, 301.06) = .226, p = .798$), and no significant interactions.

### Vertical Gaze Dispersion (RQ2.3)

For vertical gaze dispersion, significant main effects were found for cognitive distraction ($F(1, 301.04) = 9.645, p = .002$), driving environment complexity ($F(2, 301.05) = 3.974, p = .020$), and ACC use

**Table 4.2: Average gaze-based indicators of cognitive distraction ($\pm$ standard deviation) for the $N = 29$ participants (main analysis).** Main analyses report values for $12$ conditions: ACC engaged vs. disengaged across $3$ levels of driving environment complexity (DEC, from Level 1 with low complexity to Level 3 with high complexity) and $2$ levels of cognitive distraction (no distraction vs. distraction).

**(a)** Percent road center

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $73.3 \pm 13.3$ | $78.9 \pm 13.0$ | $69.1 \pm 14.3$ | $74.9 \pm 14.2$ |
| **Level 2** | $72.7 \pm 15.8$ | $78.5 \pm 11.3$ | $72.0 \pm 12.4$ | $74.8 \pm 13.0$ |
| **Level 3** | $72.7 \pm 13.1$ | $75.8 \pm 13.7$ | $71.5 \pm 14.1$ | $72.0 \pm 15.1$ |

**(b)** Horizontal gaze dispersion

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $9.0 \pm 2.9$ | $7.1 \pm 2.6$ | $8.5 \pm 3.3$ | $7.2 \pm 2.3$ |
| **Level 2** | $8.5 \pm 3.1$ | $7.1 \pm 2.4$ | $8.7 \pm 2.9$ | $7.6 \pm 2.3$ |
| **Level 3** | $9.0 \pm 3.2$ | $7.4 \pm 2.4$ | $8.9 \pm 3.3$ | $7.1 \pm 2.2$ |

**(c)** Vertical gaze dispersion

| | No distraction | | Distraction | |
|---|---|---|---|---|
| **DEC** | **ACC Off** | **ACC On** | **ACC Off** | **ACC On** |
| **Level 1** | $6.3 \pm 2.0$ | $6.0 \pm 2.2$ | $7.2 \pm 2.8$ | $6.6 \pm 2.2$ |
| **Level 2** | $6.6 \pm 2.3$ | $6.1 \pm 1.7$ | $6.8 \pm 2.4$ | $6.5 \pm 2.2$ |
| **Level 3** | $6.7 \pm 2.5$ | $6.7 \pm 2.7$ | $7.4 \pm 3.0$ | $7.1 \pm 2.7$ |

**Table 4.3: Average gaze-based indicators of cognitive distraction ($\pm$ standard deviation) for the $N = 29$ participants (complementary analysis).** Complementary analyses focus on the sessions with cognitive distraction at each driving environment complexity (DEC) level, comparing segments where participants performed mental calculations (calculations) with interleaving segments (no calculations).

**(a)** Percent road center

| | **Distraction** | |
|---|---|---|
| **DEC** | **Calculations** | **No Calculations** |
| **Level 1** | $74.7 \pm 14.9$ | $73.2 \pm 12.7$ |
| **Level 2** | $75.7 \pm 13.5$ | $73.0 \pm 12.6$ |
| **Level 3** | $72.6 \pm 14.4$ | $71.9 \pm 12.9$ |

**(b)** Horizontal gaze dispersion

| | **Distraction** | |
|---|---|---|
| **DEC** | **Calculations** | **No Calculations** |
| **Level 1** | $6.6 \pm 2.1$ | $8.1 \pm 2.5$ |
| **Level 2** | $7.2 \pm 2.7$ | $8.4 \pm 2.4$ |
| **Level 3** | $7.3 \pm 2.5$ | $8.3 \pm 2.5$ |

**(c)** Vertical gaze dispersion

| | **Distraction** | |
|---|---|---|
| **DEC** | **Calculations** | **No Calculations** |
| **Level 1** | $6.3 \pm 1.9$ | $7.1 \pm 2.5$ |
| **Level 2** | $6.2 \pm 2.3$ | $6.8 \pm 2.3$ |
| **Level 3** | $7.0 \pm 2.7$ | $7.5 \pm 2.8$ |

($F(1, 301.07) = 4.232, p = .041$), with no significant interactions. Table 4.2c shows that vertical gaze dispersion was higher under cognitive distraction (distraction: $6.9 \pm 2.4$; no distraction: $6.5 \pm 2.3$, globally) and increased with driving environment complexity (Level 1: $6.5 \pm 2.2$; Level 2: $6.5 \pm 2.1$; Level 3: $7.2 \pm 2.8$, globally). Finally, as with horizontal gaze dispersion, vertical gaze dispersion was greater when ACC was deactivated ($6.8 \pm 2.5$ globally), compared to when it was activated ($6.5 \pm 2.3$ globally).

#### Complementary Analyses of the Effects of Mental Calculation

We conducted complementary analyses to further investigate the effects of cognitive distraction on gaze-based indicators. Indeed, we observed that participants directed their gaze more toward the road center area in driving sessions without a secondary cognitive task, whereas vertical gaze dispersion increased in sessions involving mental calculations. This contrasts with previous studies, which typically report a concentration of gaze on the road center area under cognitive distraction [117, 330, 348]. We hypothesized that this discrepancy may stem from differences in the nature of the cognitive tasks. Unlike the continuous n-back tasks commonly used in earlier studies, our task involved intermittent mental calculations. Notably, addition problems are associated with upward and rightward gaze shifts [120, 285], which may account for the observed increase in vertical dispersion.

We thus segmented gaze data from the three sessions involving cognitive distraction, in each level of driving environment complexity, into 1 segments during which participants were actively solving arithmetic problems (from the onset of the voice agent's question to the participant's response, lasting approximately $7$ seconds: $2$ seconds for the question and $3.5 \pm 1.5$ seconds for the response), and 2 interleaving segments between calculations. We then computed the same gaze-based indicators as in the main analyses. This time, we were able to perform an ANOVA with repeated measures and compute effect sizes using partial eta squared [279].

For percent road center, the analysis revealed no significant main effects of either calculation phase ($F(1, 28) = 2.155, p = .153, \eta_p^2 = .071$) or driving environment complexity ($F(2, 56) = 1.553, p = .221, \eta_p^2 = .053$), and no significant interaction. In contrast, for both horizontal and vertical gaze dispersions, calculation phase ($F(1, 28) = 24.808, p <$

$.001, \eta_p^2 = .470$ and $F(1, 28) = 12.364, p = .002, \eta_p^2 = .306$, respectively) had significant effects with large effect sizes. No significant effects were found for driving environment complexity (horizontal: $F(2, 56) = 1.816, p = .172, \eta_p^2 = .061$; vertical: $F(2, 56) = 3.016, p = .057, \eta_p^2 = .097$), and there were no significant interactions.

Table 4.3 show that our initial hypothesis was incorrect. In fact, both horizontal and vertical gaze dispersions were lower during the calculation phase, indicating a more concentrated gaze toward the road center. The increased gaze dispersion observed in the main analyses for sessions with cognitive distraction stemmed from the interleaving periods between calculations, not from the calculations themselves. Interestingly, gaze dispersion (both horizontal and vertical) tended to be higher during non-calculation periods in sessions with distraction than in sessions without distraction. Conversely, horizontal gaze dispersion was lower during the calculation phase in sessions with distraction than in sessions without distraction. These results suggest that mental calculations prompted gaze concentration on the road center, while the periods between calculations were marked by more dispersed gaze patterns.

## 4.6 Discussion

This chapter investigated how cognitive distraction (mental calculation task), driving environment complexity (varying traffic densities and road construction zones), and the use of driving automation (ACC) influence physiological (EDA-based) and behavioral (gaze-based) indicators of driver state.

### 4.6.1 Effects of Cognitive Distraction

In our study, cognitive distraction was analyzed through two levels: with or without a secondary mental calculation task.

**EDA-based indicators.** We found that all three selected EDA-based indicators (SCL, SCR amplitude, and SCR rate) are significantly influenced by cognitive distraction. Specifically, all three indicators showed higher values when drivers were cognitively distracted compared to when they were not. While Li et al. [186] reported no significant effect of cognitive distraction on SCL, our findings of increased SCR rate and SCR amplitude with a cognitive load are consistent with theirs.

**Gaze-based indicators.** The results of the main analyses indicate that cognitive distraction significantly affects both percent road center and vertical gaze dispersion, but not horizontal gaze dispersion. The percent road center is lower during drives involving cognitive distraction, and the vertical gaze dispersion is higher. These findings contradict previous results reported in the literature. For instance, Victor et al. [330] reported that cognitively distracted drivers increase their road viewing time and spatially concentrate their gaze in the road center area at the expense of peripheral glances. One possible explanation for this discrepancy lies in the type of distracting task being used. Prior studies often used continuous tasks such as the n-back, which involves working memory. In contrast, our study used an intermittent mental calculation task, which is known to trigger upward gaze shifts during addition problems [120, 285].

To further investigate the origin of our findings, we conducted complementary analyses focusing on the sessions with cognitive distraction. We compared the gaze-based indicators from driving segments occurring during mental arithmetic to those from the interleaving segments between calculations. It turned out that our initial hypothesis that the increase in vertical gaze dispersion was due to cognitive distraction was incorrect. In fact, participants showed a more concentrated gaze both horizontally and vertically during the mental calculation segments. The increased gaze dispersion observed in the main analysis originated from the interleaving segments between calculations in the sessions with distraction, not the calculations themselves. Participants actually had more concentrated gaze toward the road center during calculations than in sessions without distraction, whereas between calculations, their gaze was more dispersed than in distraction-free sessions. One possible explanation is that, after a cognitively demanding episode leading to gaze concentration, drivers may engage in increased gaze dispersion to regain situation awareness of their driving surroundings (like a compensation phenomenon). This hypothesis should be investigated in future studies.

### 4.6.2 Effects of Driving Environment Complexity

In our study, driving environment complexity was analyzed through three levels: low, medium, and high traffic densities. In addition, the highest level also involved road construction zones.

**EDA-based indicators.** SCL and SCR amplitude were significantly influenced by driving environment complexity, while SCR rate was not. SCR amplitude increased when the level of complexity increased, while SCL was unexpectedly lower in the most complex environment ($Level\ 3$) compared to the least complex environment ($Level\ 1$). In a previous study, Radhakrishnan et al. [268] compared two different driving environments (rural vs. urban) and found a higher SCR rate in the rural environment compared to the urban one and explained this by the higher speed limits, narrower roads, and tighter curves associated with the rural environment. Similarly, Foy and Chapman [83] reported that SCR rate was influenced by road types, such as city center multi-lane or suburban single-lane roads. Thus, although our study did not find a significant main effect of traffic conditions on SCR rate, prior research suggests that other aspects of driving environment complexity, beyond traffic density and construction zones, may influence this indicator.

**Gaze-based indicators.** The main analysis reveals that driving environment complexity only influences vertical gaze dispersion, but not the percent road center or horizontal gaze dispersion. The vertical gaze dispersion is higher in the most complex environment than in the least complex one. In a previous study, Victor et al. [330] found that, as driving task complexity increases, drivers increase their road viewing time and spatially concentrate their gaze on the road center area. However, driving task complexity in this study was analyzed by comparing rural curves and straight sections, rural and motorway road types, simulator and field motorways.

### 4.6.3 Effects of the Use of Driving Automation

In our study, the use of driving automation was analyzed through two levels: ACC on and ACC off. Participants were free to activate or deactivate it at any time during any of the sessions.

**EDA-based indicators.** All three EDA-based indicators (SCL, SCR amplitude, and SCR rate) were significantly influenced by the use of ACC. Their values were lower when ACC was engaged than when it was not. This is consistent with the findings of Radhakrishnan et al. [268], who compared automated drives to manual drives and found a higher SCR rate during manual driving.

**Gaze-based indicators.** Similarly, we show that ACC use significantly influences all gaze-based indicators. When ACC is engaged, drivers look more often at the road center, and both horizontal and vertical gaze dispersions are lower compared to when ACC is disengaged. While Louw and Merat [203] reported more dispersed gaze (particularly horizontally) during SAE Level 2 automation than in manual driving, we may explain the discrepancies with our findings by differences in automation level. In our study, even with ACC engaged, drivers remain responsible for lateral control, which likely kept them more visually focused on the driving task. Importantly, participants were free to activate or deactivate ACC at any time, meaning that comparisons between ACC-on and ACC-off conditions do not necessarily involve the same road segments. This limits the ability to isolate the effect of ACC use from potential differences in driving context.

### 4.6.4 Implications and Perspectives

Overall, the results indicate that both physiological and behavioral measures are sensitive to variations in cognitive distraction, driving environment complexity, and automation use. These findings suggest that the EDA signal and gaze parameters can capture workload fluctuations related not only to cognitive distraction but also to traffic complexity and the use of driving automation features. This underscores the importance of jointly accounting for the driver, the environment, and the vehicle's automation to achieve a holistic understanding of the situation.

Moreover, the consistent decrease in EDA activity and gaze dispersion during ACC use supports the idea that increasing vehicle automation can help reduce drivers' mental workload. However, since automation also alters the driver's role and level of involvement, continuous monitoring of the state of drivers could support adaptive automation strategies, *i.e.*, systems capable of adjusting automation levels dynamically based on both the driver's state and the driving environment. To make such systems a reality, future work is needed to further elucidate the complex relationships between driver state, driving environment complexity, and driving automation to determine when and how to adapt the level of automation, and to ensure that adaptation strategies enhance both safety and driver trust.

Besides, building on our findings regarding the effects of our secondary cognitively distracting task on gaze behavior, future studies could more finely examine how gaze dispersion evolves during and after periods of cognitive demand across diverse distracting tasks performed while driving. Such investigations would help clarify the underlying causes and mechanisms driving these gaze dynamics.

## 4.7 Conclusion

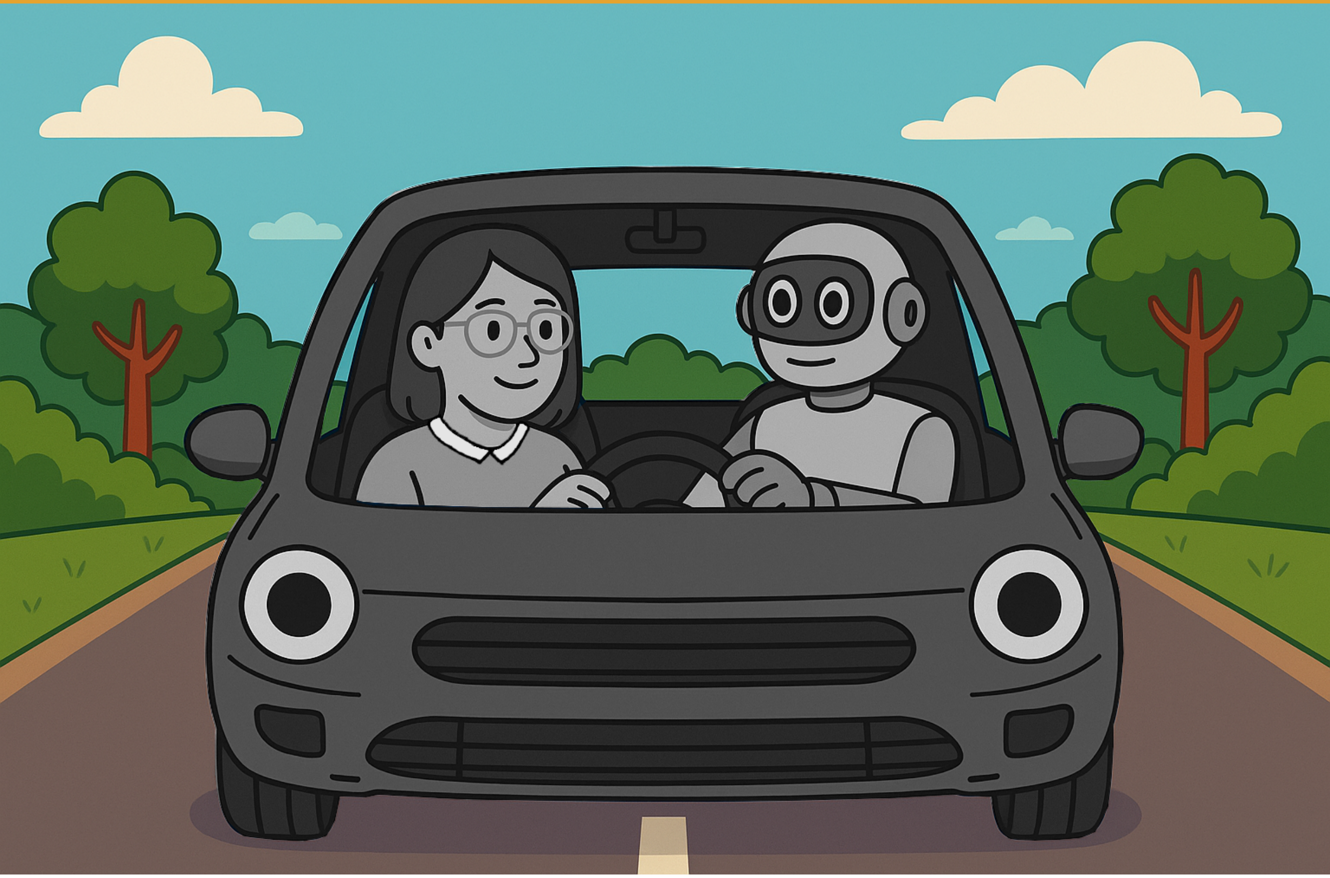
This chapter investigated how drivers' cognitive state (specifically, cognitive distraction), driving environment complexity (specifically, traffic conditions), and the use of driving automation features (specifically, ACC) influence both physiological and behavioral indicators of driver cognitive distraction. Data from $N = 29$ participants were collected in a driving simulator across six scenarios combining three levels of driving environment complexity and two levels of cognitive distraction, with the freedom to activate or deactivate ACC at any time during each scenario. We analyzed both EDA signals (*i.e.*, SCL, SCR amplitude, and SCR rate) and gaze parameters (*i.e.*, percent road center, horizontal and vertical gaze dispersions).

Overall, the results indicate that both physiological and behavioral measures are sensitive to variations in cognitive distraction, traffic conditions, and ACC use. Specifically, all three EDA-based indicators were significantly influenced by cognitive distraction and ACC use, while driving environment complexity significantly influenced SCL and SCR amplitude, but not SCR rate. Similarly, gaze analyses revealed that vertical gaze dispersion increased in more complex environments, while ACC use led to gaze concentration toward the road center, reflected by lower horizontal and vertical dispersions. In cognitively distracting conditions, drivers looked less at the road center and exhibited greater vertical dispersion. However, complementary analyses showed that gaze tended to focus on the road during mental calculations, and that increased dispersion mainly occurred during the interleaving periods between calculations.

Together, these findings suggest that both EDA signals and gaze parameters can capture workload fluctuations arising not only from cognitive distraction but also from driving environment complexity and use of automation features. Moreover, the consistent decrease in EDA activity and gaze dispersion during ACC use supports the notion that increasing automation can help reduce drivers' mental workload. These results highlight the potential of driver state monitoring to support adaptive automation systems (*i.e.*, systems capable of adjusting their level of automation dynamically based on both the driver state and the driving environment).

# Part III

## Engineering Approaches: Analyzing the Driving Environment



# 5

# Test-Time Adaptation in Dynamic Environments

Contents of this chapter



CONTEXT. Following the human studies presented in Part II, Part III shifts attention to the environment. In this chapter, we explore how perception models can adapt to rapidly changing environmental conditions, thereby enabling the analysis of the content of the driving environment. In the era of the Internet of Things (IoT), objects connect through a dynamic network, empowered by technologies like 5G, enabling real-time data sharing. However, smart objects, notably autonomous vehicles, face challenges in critical local computations due to limited resources. Lightweight AI models offer a solution but struggle when confronted with diverse and evolving data distributions. To address this limitation, we propose a novel *Multi-Stream Cellular Test-Time Adaptation* (MSC-TTA) setup where models adapt on the fly to a dynamic environment divided into cells. Then, we propose a real-time adaptive student-teacher method that leverages the multiple streams available in each cell to quickly adapt to changing data distributions. We validate our methodology in the context of autonomous vehicles navigating across cells defined based on location and weather conditions. To facilitate future benchmarking, we release a new multi-stream large-scale synthetic semantic segmentation dataset, called *DADE*, and show that our multi-stream approach outperforms a single-stream baseline. Code and data are available at github.com/ULiege-driving/MSC-TTA and github.com/ULiege-driving/DADE.

**Contributions**

1. A new Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) setup in which models adapt on the fly to a dynamic environment divided into cells.
2. A novel real-time adaptive student-teacher method to aggregate knowledge across different agents evolving in the same cell.
3. A new synthetic dataset, called *DADE*, for the semantic segmentation task on board autonomous vehicles.

RELATED PUBLICATION. This chapter is adapted from [92]. In line with my personal contributions, it focuses on the dataset [110] and describes the novel setup and method that motivated its development, without detailing the experimental results.

AUTHOR CONTRIBUTIONS. In [92], I share first co-authorship with Benoît Gérin and Anthony Cioppa. Our main individual contributions are as follows: Anthony Cioppa proposed the method, Benoît Gérin implemented the experiments, and I generated the dataset. All first co-authors contributed jointly to the analysis of the experimental results and to the writing of the original draft. More broadly, all major decisions regarding the project (such as the design of the dataset and the experiments) were made collaboratively by agreement.

## 5.1 Introduction

The Internet of Things (IoT), which enables objects to connect and exchange data over the internet or other communication networks, together with the advent of 5G technology, which offers high bandwidth and ultra-low latency [144]), open new possibilities for smart systems. In particular, it allows real-time transmission of visual data for autonomous navigation. However, despite these advances, some operations must still be executed locally, as Autonomous Vehicles (AVs) must perceive their surroundings and react safely even when disconnected from the network. This constraint requires on-board processing units, which are typically limited in both energy and computational capacity. Lightweight Artificial Intelligence (AI) models are therefore particularly suitable for enabling real-time perception under these constraints. Yet, their limited representational capacity [101] makes them vulnerable to data distribution shifts [49], such as those caused by changing environments or weather conditions.

With the help of localization technologies (*e.g.*, navigation systems, Inertial Measurement Unit (IMU)), AVs can estimate their position precisely and retrieve contextual information such as area type or weather. The environment can thus be partitioned into *cells*, each representing a specific data distribution. Furthermore, by leveraging IoT and 5G connectivity, a fleet of AVs operating within similar environments can provide *multi-stream* data that collectively capture each cell more rapidly and comprehensively.

In this work, we introduce the *Multi-Stream Cellular Test-Time Adaptation (MSC-TTA)* setup, in which a fleet of connected agents continuously adapts their models on the fly to dynamic data distributions. We further propose a real-time adaptive student–teacher method leveraging this cell-based division of the environment: each agent processes its own stream using a lightweight student model, while heavier computations (*i.e.*, teacher inference and student training) are offloaded remotely (*e.g.*, on the cloud). Aggregated data from all agents within the same cell are used to train specialized student models for each cell, as illustrated in Fig. 5.1. We evaluate our approach in the context of autonomous driving, where the environment is divided into cells based on location (*e.g.*, urban, countryside, highway) and weather conditions (*e.g.*, sunny, rainy, foggy). To support these experiments and

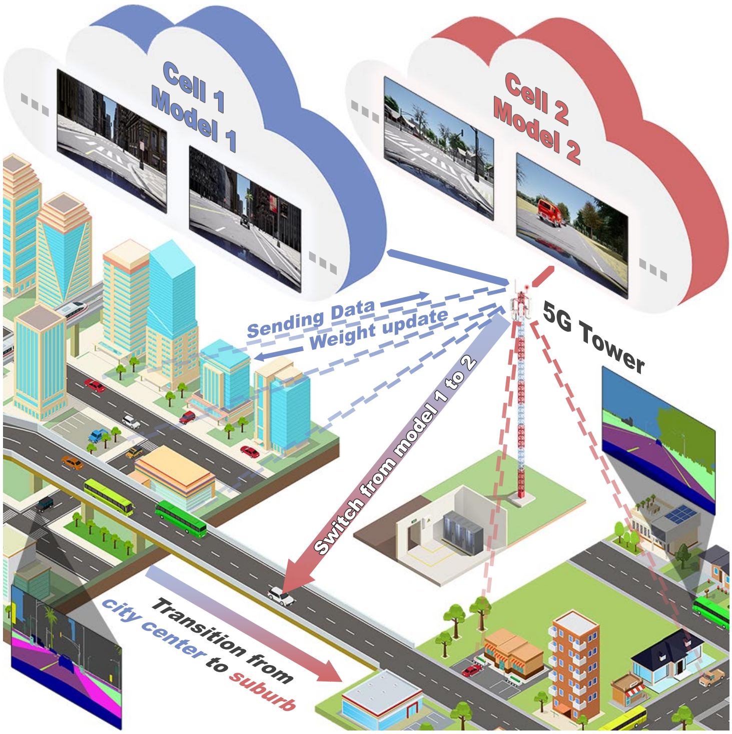



**Figure 5.1: Multi-Stream Cellular Test-time Adaptation (MSC-TTA) of real-time models.** We consider a set of agents (*e.g.*, autonomous vehicles) evolving in a dynamic environment divided into cells (*e.g.*, city center or suburb) that perform the same task (*e.g.*, semantic segmentation) in real time on their own unlabeled data stream (*e.g.*, recorded images) using an on-board model. We propose a first method in which agents share part of their data stream through an IoT network (*e.g.*, a connection to a 5G tower). Cell-based lightweight models are then trained on the fly—in our case through an adaptive student-teacher method—, and their weights are regularly broadcasted to the agents to improve their performance over time. When agents transition between cells, the agent's model is immediately switched to the one of the new cell, effectively adapting the predictions of the transiting agent.

future research, we release *DADE*, a large-scale synthetic semantic segmentation dataset generated with the CARLA simulator [66]. Our method outperforms a single-stream baseline, demonstrating the benefits of multi-stream, cell-based adaptation.

Our main contributions are as follows: (1) We introduce the *Multi-Stream Cellular Test-Time Adaptation* (MSC-TTA) setup for on-the-fly model adaptation in dynamic environments. (2) We propose a real-time adaptive student–teacher method aggregating knowledge across agents within each cell. (3) We generate and release a new synthetic dataset, *DADE*, for semantic segmentation in autonomous driving, allowing empirical validation showing improved performance over baseline methods.

## 5.2 Related Work

### 5.2.1 Online Learning

Online learning is a well-studied setup [40, 122, 133, 242, 296] formulated as an iterative game between a learner and an environment generating a continuous data stream. At each step, the learner predicts the label of the current instance based on past observations. Once the true label is revealed, it is compared to the prediction, and the learner incurs a regret score that penalizes its mistakes. The objective is to minimize future regret by exploiting previously observed data and labels. Several datasets have been introduced for benchmarking online learning, such as firehose [136] for language modeling, and CLOC [37] and CLEAR [198] for image classification involving objects whose representations evolve. In practice, online learning is particularly relevant when the true labels become available gradually, as in forecasting tasks [200, 345, 367]. In this work, we assess an upper bound of our MSC-TTA method by extending the setup to multi-stream cellular online learning (MSC-OL).

### 5.2.2 Test-Time Adaptation

Similarly, Test-Time Adaptation (TTA) aims to adapt a model on a data stream. However, the environment does not reveal the true label of previously observed data. Several setups, characterized by the data distribution of the stream, have been studied, such as Fully TTA [340], Continual TTA [342], Non-i.i.d TTA [100], or Practical TTA [376], in which the data stream contains distribution changes and correlated samples. These setups are suited for real-world applications, where the true labels are unavailable at test time. However, previous works only consider a single stream of data. In this work, we go further by proposing a setup for multiple streams and introducing prior knowledge on cross-stream data distribution through the division of the environment into cells. In addition, our methodology brings a real-time aspect, a feature often overlooked in previous setups.

To leverage the information in the data stream, multiple methods have been developed [193]. Some works adapt the model's parameters by either fitting the batch normalization layers to the target domain [190, 223, 291], training the model with auxiliary tasks [62, 316], or fine-tuning it using unsupervised objectives [275, 340, 382]. Some other works adapt the input data [91, 124, 153, 385] or weight the predictions of multiple models depending on the test distribution [78, 346]. However, few works ensure that the adaptation is real-time.

In fact, in real-world applications, the model needs to adapt within a limited time to leverage all samples of the data stream due to finite computing capabilities. Alfarra et al. [13] recently proposed an evaluation protocol to compare TTA methods under those constraints. To satisfy the real-time constraint, some works proposed a student-teacher architecture with a lightweight student model [49, 228]. Specifically, ARTHuS [49] is a real-time method in which a lightweight student model is adapted on an unlabeled data stream at test time using pseudo-labels produced by a state-of-the-art but computationally-expensive teacher model. The real-time constraint of the system is ensured by asynchronously processing the student and teacher inference and training at different frame rates. The fast, lightweight student model therefore trains online on the changing data distributions using the teacher's slow predictions. However, in the case of rapid domain shifts, the student needs several batches to adapt. Houyon et al. [134] later tackled this issue by incorporating continual learning methods in the student online training to avoid catastrophic forgetting in the case of cyclic domain shifts. Nevertheless, in the case of multiple objects (*e.g.*, autonomous vehicles), each data stream is treated independently. In this work, we extend ARTHuS [49] to multiple data streams and cell-divided environments.

### 5.2.3 Autonomous Driving

AVs rely on advanced sensor arrays, high-resolution cameras, and on-board computing power to perceive the environment and make informed decisions to navigate safely. Nowadays, perception is largely based on AI and involves several computer vision tasks such as semantic segmentation [163, 362, 374, 386] (which consists in assigning a class label to each pixel of an image), object detection [29, 185] or depth estimation [43, 79, 97]. However, the road to fully self-driving cars remains challenging. For instance, it is still complex to operate AVs in diverse environments, such as varying weather conditions, traffic patterns, and other unforeseen scenarios, and to process large amounts of data while optimizing energy consumption in Electric Vehicles (EVs).

To adapt to several environments, some methods use domain adaptation strategies [156, 256, 344] to enhance system versatility and reliability. Also, cloud computing [290] or multi-access edge computing [202, 366] provide the computational power and storage capacity for real-time data processing, enhancing energy efficiency and improving EVs mileage. Similarly to multi-access edge computing, our proposed method employs a hybrid approach. On-board processing handles immediate, low-latency operations, while resource-intensive computations are offloaded to external servers. Specifically, the heavy offloaded computations rely on the multiple streams of the fleet, while on-board, lightweight real-time perception is performed using models trained in the cloud, guaranteeing adaptability in dynamic environments.

## 5.3 Multi-Stream Cellular Test-Time Adaptation

### 5.3.1 MSC-TTA Setup

Given a finite set of $N$ agents $a_n$ forming a connected fleet $\mathbf{A}$, our proposed Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) aims to adapt over time $t \in \{0, ..., T\}$ each agent's model $f_{a_n}^t \in \mathbf{F}$, pretrained on any source domain to perform a task $\tau$, to the agent's data stream $\mathcal{X}_{a_n} \in \mathbf{X}$ of online unlabeled samples $\mathcal{X}_{a_n} = x_{a_n}^0, x_{a_n}^1, ..., x_{a_n}^t, ..., x_{a_n}^T$. As in [376], the samples are drawn from a distribution $\mathcal{P}_{a_n} \in \mathbf{P}$ shifting over time following $\mathcal{P}_{a_n}^0, \mathcal{P}_{a_n}^1, ..., \mathcal{P}_{a_n}^t, ..., \mathcal{P}_{a_n}^T$, in which consecutive samples $x_{a_n}^{t-1}, x_{a_n}^t, x_{a_n}^{t+1}$ may be highly correlated. At time $t$, the model

$f^t_{a_n}$ receives a batch of unlabeled samples $\mathcal{B}^t_{a_n} = x^t_{a_n}, x^{t+1}_{a_n}, ..., x^{t+(b-1)}_{a_n}$, where $b$ is the batch size, on which it makes predictions. Each model $f^t_{a_n}$ may be adapted to the current batch $\mathcal{B}^t_{a_n}$ by accumulating knowledge from previous samples of the multiple streams forming the following hyperspace $\cup x^{t'}_{a_n}, \forall a_n \in \mathbf{A} \times t' < (t+b)$. Let us note that samples $x^t_{a_n}$ may be unavailable for some time $t$ for some agent $a_n$. This setup describes the general case of multiple sensors recording data streams and performing the same task, *e.g.*, surveillance cameras placed in one or several cities on which crowd counting or car segmentation needs to be performed, with no assumptions on where the cameras are placed.

To include cross-stream prior knowledge on data distributions, we consider the general case in which the agents evolve inside a dynamic environment split into a non-overlapping set of $C$ cells $c \in \mathbf{E}$. We suppose that, at time $t$, each agent is located within one cell such that $e^t_{a_n} = c \in \mathbf{E}$, with agents being able to transition between cells over time. The cells $c$ are predefined by a set of rules (*e.g.*, based on the location, the weather, etc.) such that the expected data distribution of agents evolving in the same cell is similar, *i.e.*, $\mathcal{P}^t_{a_n} \approx \mathcal{P}^t_{a_m}$ if $e^t_{a_n} = e^t_{a_m}$. Our setup therefore allows the different data streams to share common data distribution properties at times that can be leveraged to effectively adapt the models. Naturally, in practice, this assumption may fail if the cells are incorrectly defined or estimated. This setup is particularly interesting in the real-world case of autonomous driving, in which vehicles evolve in different locations (*e.g.*, city centers, suburbs, highway, etc.) that they analyze through various sensors. Also, vehicles driving in the same environment may leverage the multiple streams of the fleet to better assess and adapt to the environment.

Let us note that considering the special case of $N = 1$ and $C = 1$ (*i.e.*, a single agent in a single cell) falls back to the original Practical TTA setup [376] in which a single model is adapted to its data stream. The case of $N \geq 1$ and $C = 1$ represents a multi-stream test-time adaptation setup without division of the environment. Finally, the case of $C = 1$ and $f^t_{a_n} = f^t$, *i.e.*, in which a single model is adapted for all streams without prior knowledge on the environments, corresponds to a TTA setup in which samples from multiple streams are combined in the batch. In the following, we describe our adaptive method for the general case $N \geq 1$ and $C \geq 1$. To stay close to a real-world scenario, we add an extra real-time constraint on the method, *i.e.*, no delay accumulation or sample skipping when processing the multiple data streams.

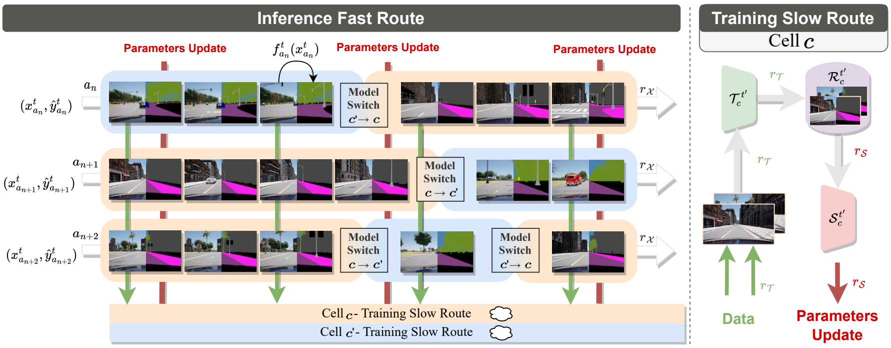



**Figure 5.2: Pipeline of our multi-stream cellular test-time adaptation of real-time models.** Our method is composed of a fast route for inference and a slow route for online training, as defined in [49, 134]. In the fast route, each agent $a_n$ processes a stream of data samples $x^t_{a_n}$ and predicts labels $\hat{y}^t_{a_n} = f^t_{a_n}(x^t_{a_n})$ in real time (*i.e.*, at the data stream rate $r_{\mathcal{X}}$). Agents located within a cell $c$ send a subset of their data samples at a slower rate $r_{\mathcal{T}}$ to a slow route operating on a remote server (*e.g.*, on the cloud) dedicated for each cell. In the slow route, a teacher model $\mathcal{T}^{t'}_c$ predicts pseudo labels on the received data and stores them in a replay buffer $\mathcal{R}^{t'}_c$. The replay buffer is then used to train on the fly a cell-specific student model $\mathcal{S}^{t'}_c$ at a rate $r_{\mathcal{S}}$. After each training epoch on the replay buffer, the parameters of $\mathcal{S}_c$ are transferred to all agent models $f_{a_n}$ located within that cell. Finally, agents transiting between two cells have their model switched instantly.

### 5.3.2 Real-Time MSC-TTA Method

Our method, illustrated in Fig. 5.2, produces a stream of predictions for every agent following $\hat{y}^t_{a_n} = f^t_{a_n}(x^t_{a_n})$, with the model $f^t_{a_n}$ operating in real time (*i.e.*, at the rate $r_{\mathcal{X}}$) on the data stream $\mathcal{X}_{a_n}$. To do so, we extend the adaptive real-time student-teacher method, ARTHuS, of Cioppa et al. [49], in which a lightweight student model $\mathcal{S}$ is adapted on the fly using pseudo labels produced by a state-of-the-art but computation-expensive teacher model $\mathcal{T}$. Particularly, we leverage the multiple streams and the division of the environment into cells. We allow agents evolving within the same cell to share their own data stream to produce a cell-specific data stream $\mathcal{X}^{t'}_c = \cup x^{t'}_{a_n}, \forall a_n \mid e^{t'}_{a_n} = c$ at a frame rate $r_{\mathcal{T}}$, producing samples $x^{t'}_c$.

Our method is composed of a fast route and a slow route. In the fast route (inference), student models for each agent produce predictions

$\hat{y}^t_{a_n} = \mathcal{S}^t_{a_n}(x^t_{a_n}) = f^t_{a_n}(x^t_{a_n})$ at the rate $r_\mathcal{X}$. In parallel in the slow route (training), a slow but high-performance teacher model $\mathcal{T}^{t'}_c$ for each cell produces pseudo-ground truths $\tilde{y}^{t'}_c = \mathcal{T}^{t'}_c(x^{t'}_c)$ at the rate $r_\mathcal{T}$ on the cell data streams. The pair of data $(x^{t'}_c, \tilde{y}^{t'}_c)$ are then stored in a replay buffer $\mathcal{R}^{t'}_c$ of size $R$ using a First-In-First-Out (FIFO) strategy. One student model per cell $\mathcal{S}^{t'}_c$ is trained on the updated replay buffer $\mathcal{R}^{t'}_c$ using a loss function

$$\mathcal{L} = \sum_{i=1}^{R} L(\mathcal{S}^{t'}_c(x^i_c), \tilde{y}^i_c), \tag{5.1}$$

where $L$ is a dissimilarity measure suited for task $\tau$. After training for one epoch on the replay buffer, the weights of students in the fast route are updated with the weight of the environment students in the slow route such that $\mathcal{S}^t_{a_n} = \mathcal{S}^t_c, \forall a_n | e^t_{a_n} = c$, at a slower rate $r_\mathcal{S}$. Since the slow route gathers information from several agents, the heavy teacher inference and student training processes can be offloaded to a dedicated server (*e.g.*, on the cloud). Hence, agents only perform the real-time inference with a lightweight model, greatly reducing computation requirements and saving precious battery power in the case of AVs. Finally, considering the special case $C = N$ with each agent defining its own cell is equivalent to the original ARTHuS method [49], serving as a baseline in our experiments.

## 5.4 Driving Agents in Dynamic Environments Dataset

To study our new Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) setup and evaluate the performance of our real-time method, we need a dataset that meets the following four criteria.

1. **Multi-agent long videos:** the dataset should consist of long video sequences captured by multiple agents operating within the same dynamic environment.

2. **Environment division:** the environment should be heterogeneous or dynamic to be spatially and/or temporally divided into cells, *e.g.*, encompassing various driving locations, such as rural, urban, and highway settings, or a broad spectrum of weather

conditions, including, *e.g.*, day, night, clear, rainy, and foggy scenarios.

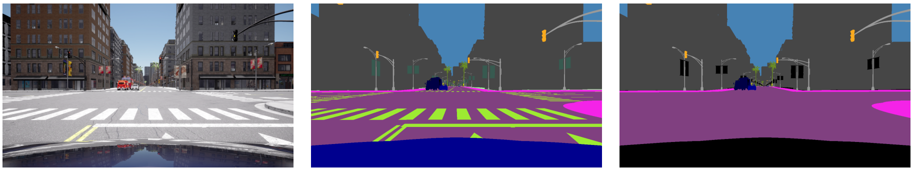
3. **Cell connection:** each agent's connection to a cell should be precisely estimated, for instance using Global Navigation Satellite System (GNSS) coordinates for the location or a weather service for the weather conditions.

4. **Available ground truths:** for evaluation purposes, we need to have access to ground-truth annotations for our semantic segmentation task.

Unfortunately, publicly available datasets do not meet these criteria. Existing datasets, such as [52, 207, 315], typically feature short video sequences, lack multi-agents, or often do not include ground-truth annotations or a diverse range of weather conditions. While the SHIFT dataset [315] contains varying weather conditions and ground truths, it is not a multi-agent dataset, and its average sequence length is under $160$ seconds, which is too short for evaluating the long-term impact of our method.

Therefore, we generated and publicly released our own Driving Agents in Dynamic Environments (*DADE*) dataset, meeting all the above criteria. To have access to ground-truth annotations and precisely control the environment, we used the CARLA simulator [66] (version 0.9.14) to generate the dataset. CARLA enables the synchronization and calibration of sensors, the collection of semantic segmentation ground truths, and fine-grained control over the weather. Furthermore, the `Town12` map offers several visually distinct locations. We deployed several agents (ego vehicles) driving through the map, each equipped with a front-facing camera capturing the forward view in a "Cityscapes"-like setup, as shown in Fig. 5.3. We collected the video sequences taken by an RGB camera, the semantic segmentation ground-truth masks, the GNSS position of each agent, as well as the overall weather information. All signals were acquired at the framerate of $1$ frame per second, with a high-resolution (*i.e.*, $1280 \times 720$ pixels) definition.

To the best of our knowledge, our dataset, large of $150$ GBytes, is unique in that it provides long videos of multiple agents evolving in diverse driving locations and weather conditions with ground-truth labels for the task of semantic segmentation. Our video sequences contain



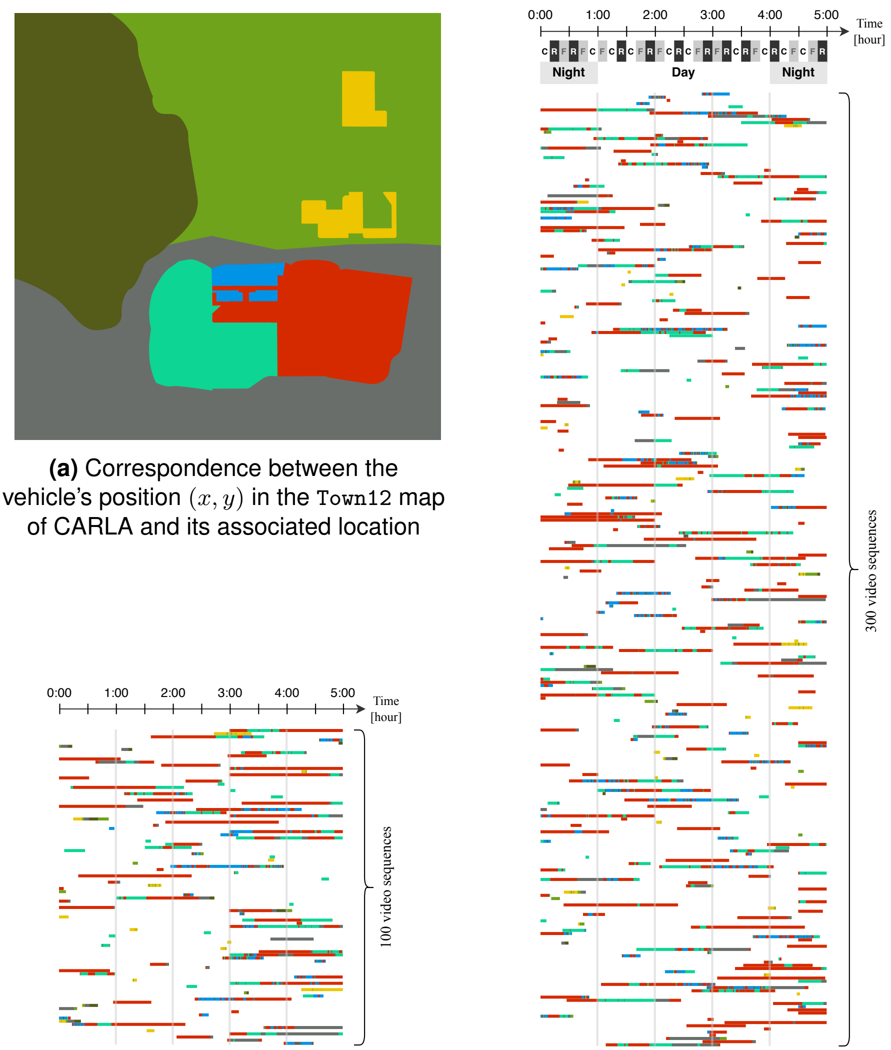
(a) (b) (c)

**Figure 5.3: Comparison between the real ground truth and the ground truth used in our experiments.** (a) An RGB image with (b) its corresponding semantic segmentation ground truth from CARLA, and (c) the semantic segmentation ground truth that we used to evaluate our method. The black pixels in image (c) correspond to ignored classes or regions, such as the hood of the ego vehicle.

between $188$ and $7{,}200$ frames acquired at $1$ frame per second (fps), with an average sequence length of $40$ minutes.

To align our dataset with current benchmarks in the semantic segmentation field, we generated two versions of the semantic segmentation ground truths in the dataset: 1 the ones directly collected from the CARLA simulator, and 2 the intersection of the semantic classes available in CARLA and the semantic classes from the Cityscapes [52] dataset. For consistency with previous works, we choose the latter version in our experiments, as most state-of-the-art models for semantic segmentation in driving environments are trained on Cityscapes. Nevertheless, the discrepancies between the two versions are minimal and could be interchanged depending on the user's preferences. Figure 5.3a shows an RGB image alongside the two versions of the semantic segmentation ground-truth masks (Fig. 5.3b and Fig. 5.3c). As can be seen, the "road line" class in the CARLA labels, visible in Fig. 5.3b, is simply merged into the "road" class in Fig. 5.3c. Also, the "car hood" is ignored (indicated by black pixels) in the second version. More details about the labels definition can be found in Chapter C of the appendix.

To study different cell divisions of the environment, our *DADE* dataset is composed of two parts. The first part, *DADE-static*, is acquired with static weather conditions (clear day) and contains $100$ video sequences, as shown in Fig. 5.4b. The second part, *DADE-dynamic*, is acquired with varying weather conditions (ranging from day to night, with clear, rainy, or foggy conditions) and contains $300$ video sequences, as shown in Fig. 5.4c. For both parts, each sequence is acquired by one agent




**(a)** Correspondence between the vehicle's position $(x, y)$ in the `Town12` map of CARLA and its associated location

**(b)** *DADE-static*

**(c)** *DADE-dynamic*

**Figure 5.4: Characteristics of the video sequences in *DADE-static* and *DADE-dynamic* datasets.** For each sequence of (b) *DADE-static* and (c) *DADE-dynamic*, the color of the line corresponds to the location of the agent at a given time. The different locations are forest, countryside, rural farmland, highway, low density residential, community buildings, and high density residential. Given the x and y coordinates of an agent, (a) provides its corresponding location. We can see that the sequences are evenly distributed across the entire 5-hour time frame. For *DADE-dynamic*, C, R, and F respectively correspond to clear, rainy, and foggy weather, and night/day represent the daylight conditions.

(one ego vehicle) running for some time within a $5$-hour time frame, amounting to a total of $990$k frames for the entire dataset. In Fig. 5.4a, we show a top view of the various locations in the `Town12` map of the CARLA simulator in which the agents evolve, namely forest, countryside, rural farmland, highway, low density residential, community buildings, and high density residential. Images captured in each location can be seen in Fig. 5.5. Finally, Fig. 5.6 illustrates the $6$ different weather conditions, in the high density residential location, encountered in the *DADE-dynamic* dataset.

Let us note that due to the limitations of the CARLA simulator running the `Town12` map, there are no pedestrians on the streets, only vehicles such as cars, motorcycles, bicycles, or trucks. Also, the quantity of vehicles (traffic) is independent on the location. The vehicles spawned in the map move randomly through the seven locations. Finally, the different sequences are collected sequentially. In the following, we provide some statistics about both parts of our *DADE* dataset.

### 5.4.1 *DADE-static*

This first part of our dataset is composed of $100$ sequences, acquired in the `Town12` map of CARLA with a static, clear, sunny weather during the day. Each sequence contains between $271$ and $7{,}200$ frames acquired at $1$ fps, for a total of $270{,}527$ frames, amounting to more than $75$ hours of video. The average sequence length is $45$ minutes (more details about the distribution of the sequence length can be found in appendix in Chapter C). Figure 5.4b indicates the locations of the $100$ agents over time. We can see that, for most sequences, the agents evolve through several locations, and that the start and end times vary significantly from one agent to another.

Figure 5.7a provides an analysis of the number of agents in each location over time. Particularly, it shows that there is a high imbalance between the locations, which is expected in real-world scenarios. For instance, it is realistic to encounter many more vehicles in city centers than in the countryside. Table 5.1 summarizes those values and splits the number of images acquired during the first two hours (used for pretraining) and the last three hours (used for testing). Interestingly, data originating from the high density residential location constitute over half of our *DADE-static* dataset. We can also see that, during the first two hours, over a thousand images are collected in each location, constituting a sufficient pretraining set.


**Figure 5.5: Locations in *DADE*.** We define 7 different locations based on the GNSS data and show some images captured by the agent in each location. From left to right, we display the name of the location, an overview of the location, and six images from *DADE-static*.

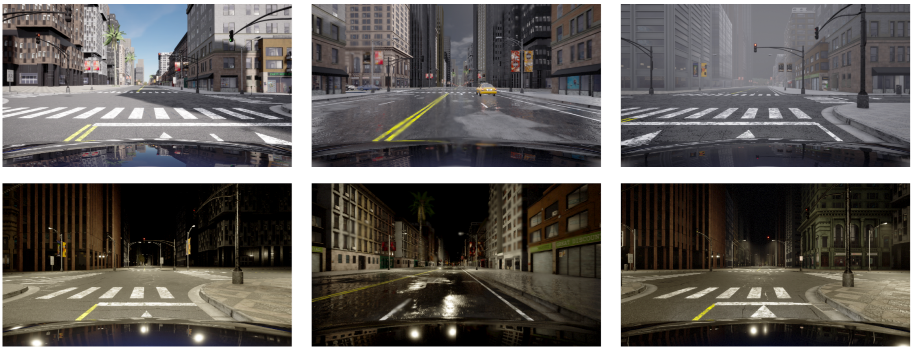
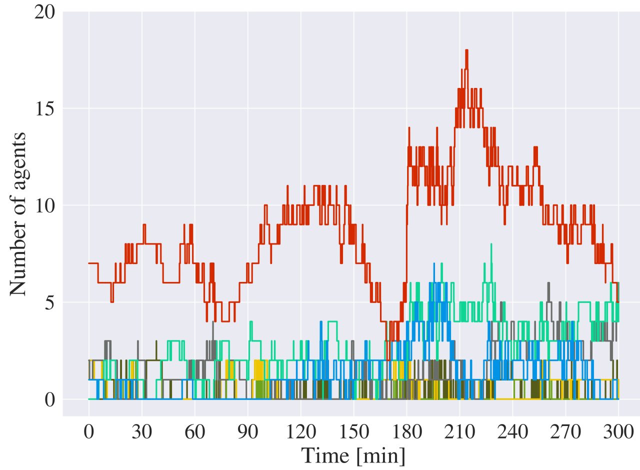


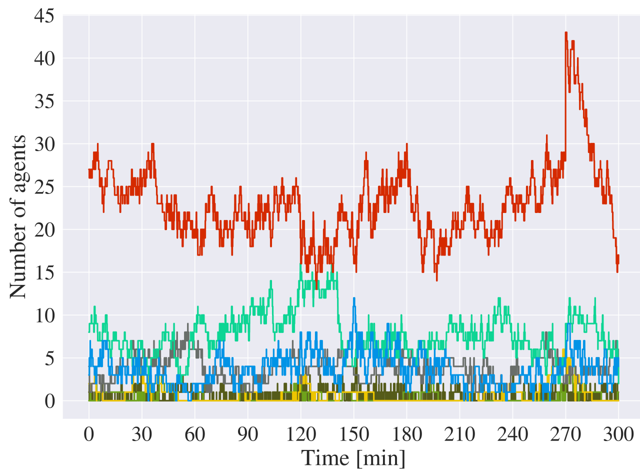
**Figure 5.6: Weather and daylight conditions in *DADE-dynamic*.** The images show the $6$ different weather and daylight conditions in the same high density residential location. The images correspond respectively, from left to right and top to bottom, to clear day, rainy day, foggy day, clear night, rainy night, and foggy night.

**Table 5.1: Number of images per location** within the *DADE-static* dataset during the pretraining time (two first hours), the test time (three last hours), and the overall time (the five hours), as well as the proportion of images originating from each location in comparison to the entire dataset.

| Location | Pretraining (2 hours) | Testing (3 hours) | Overall (5 hours) | Proportion of the entire dataset |
|---|---|---|---|---|
| Forest | 2,176 | 2,796 | 4,972 | 1.84% |
| Countryside | 2,442 | 1,215 | 3,657 | 1.35% |
| Rural farmland | 3,608 | 6,089 | 9,697 | 3.58% |
| Highway | 7,018 | 19,159 | 26,177 | 9.68% |
| Low density residential | 11,187 | 36,658 | 47,845 | 17.69% |
| Community buildings | 2,357 | 20,404 | 22,761 | 8.41% |
| High density residential | 50,034 | 105,384 | 155,418 | 57.45% |
| **Total** | **78,822** | **191,705** | **270,527** | **100%** |




**(a)** *DADE-static*




**(b)** *DADE-dynamic*

**Figure 5.7: Number of agents per location over time.** The colors of the plots correspond to the 7 locations (*cf.* Fig. 5.5). The same trends can be observed in both datasets, with three times as many agents in (b) *DADE-dynamic* as in (a) *DADE-static*. Note that there is at all time at least one agent in the high density residential location in both datasets and in the low density residential in the *DADE-dynamic*. Conversely, forest, countryside, and rural farmland locations exhibit the least agent presence, often remaining empty of agents for extended periods.

### 5.4.2 *DADE-dynamic*

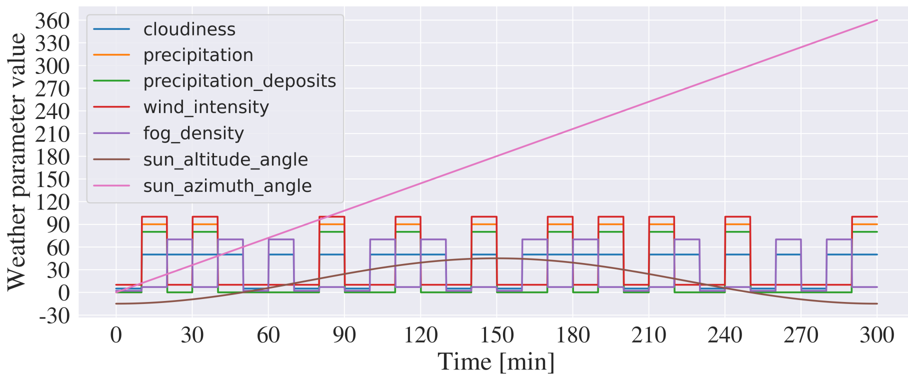
This second part of our dataset is acquired during a period of time of $5$ hours with varying weather conditions as shown in Fig. 5.4c. Particularly, it is composed of $300$ sequences containing between $188$ and $7{,}200$ frames acquired at $1$ fps, for a total of $719{,}742$ frames or $200$ hours of videos. The average sequence length is $40$ minutes (see Chapter C in appendix for more details on the distribution of the sequence length).

To generate diverse weather conditions, we dynamically varied the weather parameters over time for the entire map, *i.e.*, all agents experienced the same weather at any given time regardless of their location. The weather condition was changed every $10$ minutes, alternating arbitrarily between clear, rainy, and foggy conditions, with smooth transitions lasting $10$ seconds. For the daylight cycle, the 5-hour time frame started at night and progressed to sunrise after one hour, with sunset occurring during the final hour. The $5$ hours thus consist of approximately $2$ hours of nighttime and $3$ hours of daytime conditions. During the entire $5$ hours, there are $4$ ten-minute periods for each weather condition (*i.e.*, clear, rainy, and foggy) during the night and $6$ during the day, as shown in Fig. 5.8a. To better visualize the transitions, Fig. 5.8b provides a close-up of the first thirty minutes, during which the weather evolves from clear to rainy, and finally foggy conditions.

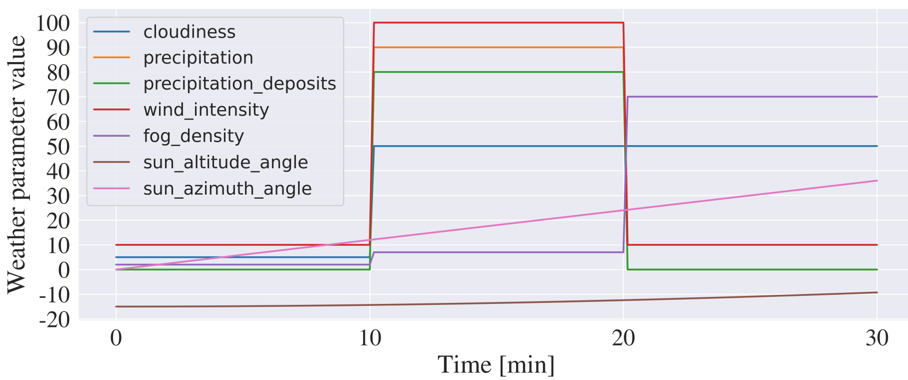
Figure 5.7b shows the number of agents in each location. Similarly to *DADE-static*, there is also a high imbalance between the different locations. Table 5.2 provides a summary of the number of images captured in each location, weather, and daylight conditions during the first two hours (used for pretraining) and the last three hours (used for testing). We can see that the proportion of images in each location is similar to the one of our *DADE-static* dataset. However, upon further division based on both weather and daylight conditions, we can see a significant decrease in the number of images for each cell. Notably, this division results in the absence of pretraining data for the clear day weather condition in the countryside location.

## 5.5 Overview of Experimental Results

This section summarizes the main results of the experiments conducted to evaluate our MSC-TTA setup and method on the proposed *DADE*




**(a)**




**(b)**

**Figure 5.8: Evolution of the weather parameters over time.** The weather switches arbitrarily every $10$ minutes between clear, rainy, and foggy weathers with a smooth transition of $10$ seconds. The parameters for the daylight conditions are related to the sun position, *i.e.*, sun altitude angle and sun azimuth angle, which vary smoothly over time, respectively between $-15$ and $45$ degrees, and between $0$ and $360$ degrees. We consider that it is nighttime during the first hour and the last hour, when the sun's altitude is below $5$ degrees, and that it is the daytime in between during three hours, *i.e.*, when the sun's altitude is over $5$ degrees. In total, the clear, rainy, and foggy weathers each occur $10$ times; $4$ times during the night and $6$ times during the day, as shown in (a). (b) zooms in on the first thirty minutes. During the first ten minutes, the weather is clear, then there is a smooth transition of $10$ seconds towards a rainy weather and finally, after $20$ minutes there is again a smooth transition of $10$ seconds towards foggy weather.

**Table 5.2: Number of images per location (Part 1)** within the *DADE-dynamic* dataset during the pretraining time (two first hours), the test time (three last hours), and the overall time (the five hours), as well as the proportion of images originating from each location in comparison to the entire dataset.

| **Location** | **Pretraining (2 hours)** | **Testing (3 hours)** | **Overall (5 hours)** | **Proportion of the entire dataset** |
|---|---|---|---|---|
| Forest | 3,174 | 5,973 | 9,147 | 1.27% |
| Clear night | 845 | 695 | 1,540 | |
| Rainy night | 303 | 477 | 780 | |
| Foggy night | 585 | 602 | 1,187 | |
| Clear day | 381 | 1,467 | 1,848 | |
| Rainy day | 572 | 1,675 | 2,247 | |
| Foggy day | 488 | 1,057 | 1,545 | |
| Countryside | 3,525 | 4,283 | 7,808 | 1.09% |
| Clear night | 279 | 1,247 | 1,526 | |
| Rainy night | 1,137 | 130 | 1,267 | |
| Foggy night | 887 | 795 | 1,682 | |
| Clear day | 0 | 194 | 194 | |
| Rainy day | 1,020 | 1,312 | 2,332 | |
| Foggy day | 202 | 605 | 807 | |
| Rural farmland | 4,605 | 9,242 | 13,847 | 1.92% |
| Clear night | 736 | 2,631 | 3,367 | |
| Rainy night | 1,134 | 265 | 1,399 | |
| Foggy night | 2,059 | 2,699 | 4,758 | |
| Clear day | 221 | 1,268 | 1,489 | |
| Rainy day | 418 | 926 | 1,344 | |
| Foggy day | 37 | 1,453 | 1,490 | |
| Highway | 27,573 | 40,275 | 67,848 | 9.43% |
| Clear night | 4,676 | 4,878 | 9,554 | |
| Rainy night | 4,809 | 4,508 | 9,317 | |
| Foggy night | 6,052 | 4,876 | 10,928 | |
| Clear day | 3,533 | 7,575 | 11,108 | |
| Rainy day | 4,235 | 9,757 | 13,992 | |
| Foggy day | 4,268 | 8,681 | 12,949 | |

**Table 5.2: Number of images per location (Part 2)** within the *DADE-dynamic* dataset during the pretraining time (two first hours), the test time (three last hours), and the overall time (the five hours), as well as the proportion of images originating from each location in comparison to the entire dataset.

| **Location** | **Pretraining (2 hours)** | **Testing (3 hours)** | **Overall (5 hours)** | **Proportion of the entire dataset** |
|---|---|---|---|---|
| Low density residential | 56,108 | 84,990 | 141,098 | 19.60% |
| Clear night | 7,348 | 9,214 | 16,562 | |
| Rainy night | 6,957 | 7,381 | 14,338 | |
| Foggy night | 7,673 | 8,190 | 15,863 | |
| Clear day | 11,486 | 22,607 | 34,093 | |
| Rainy day | 11,736 | 17,570 | 29,306 | |
| Foggy day | 10,908 | 20,028 | 30,936 | |
| Community buildings | 23,965 | 42,205 | 66,170 | 9.19% |
| Clear night | 3,648 | 4,984 | 8,632 | |
| Rainy night | 3,746 | 3,532 | 7,278 | |
| Foggy night | 3,386 | 4,121 | 7,507 | |
| Clear day | 4,838 | 9,096 | 13,934 | |
| Rainy day | 4,210 | 9,539 | 13,749 | |
| Foggy day | 4,137 | 10,933 | 15,070 | |
| High density residential | 164,708 | 249,116 | 413,824 | 57.50% |
| Clear night | 26,627 | 38,134 | 64,761 | |
| Rainy night | 31,006 | 25,618 | 56,624 | |
| Foggy night | 28,064 | 33,440 | 61,504 | |
| Clear day | 26,260 | 48,795 | 75,055 | |
| Rainy day | 26,830 | 53,676 | 80,506 | |
| Foggy day | 25,921 | 49,453 | 75,374 | |
| **Total** | **283,658** | **436,084** | **719,742** | **100%** |

dataset. The experimental settings and detailed results are described in Chapter C of the appendix.

On *DADE*-static, our experiments demonstrate the benefits of using multiple streams when adapting models, compared to using a single stream from an agent. Furthermore, we show that leveraging cellular information (*i.e.*, spatial division based on the seven defined locations) leads to improved model performance. We also highlight the advantages of adapting the model online. Indeed, adapting the models during the last three hours (testing set) yields better performance than a model trained only on the first two hours of training data and not adapted further using the testing data.

On *DADE*-dynamic, the spatial division based on locations also provides the best results, outperforming in this case temporal divisions based on weather and/or daylight conditions. We hypothesize that this outcome is due to over-dividing the environment relative to the available data per condition. Increasing the amount of data, *e.g.*, by extending the time frame or adding more sequences, could mitigate this issue.

## 5.6 Conclusion

The novel Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) setup addresses multi-stream model adaptation in dynamic environments. We focus on environments where data distribution shifts pose significant challenges. To do so, we divide the environments into cells, characterized by similar conditions such as location and weather. Then, we propose a real-time method based on an adaptive student-teacher approach, leveraging the multiple streams and cellular information. Experimental validation on autonomous vehicles illustrates the benefits of our MSC-TTA setup, showcasing better performance compared to a single-stream baseline. Our novel *DADE* dataset supports our experiments and provides a comprehensive benchmark for future studies in test-time adaptation of semantic segmentation models for autonomous vehicles. This work represents a significant step forward in the field of test-time adaptation, holding promise for substantial contributions to IoT and autonomous driving.

# 6

# Physically Interpretable Probabilistic Domain Characterization

Contents of this chapter

CONTEXT. In the previous chapter, we developed a method for analyzing urban scenes that adapts to the content of the environment. The results showed the importance of considering the specific characteristics of the environment. This is also why we developed *DADE*, a dataset illustrating different environments. In this chapter, we seek to characterize the environment through probabilistic domain representations. Characterizing domains is essential for systems operating in dynamic environments, as it enables them to adapt to changing conditions or to delegate the task to backup systems when facing conditions outside their operational domain. Existing solutions typically characterize a domain by solving a regression or classification problem, which limits their applicability as they only provide a limited summarized description of the domain. In this chapter, we propose a novel approach to domain characterization by characterizing domains as probability distributions. Specifically, we develop a method to predict the likelihood of different weather conditions from images captured by vehicle-mounted cameras by estimating distributions of physical parameters using normalizing flows. To validate our proposed approach, we conduct experiments in the context of autonomous vehicles, focusing on predicting the distribution of weather parameters to characterize the operational domain. This domain is characterized by physical parameters (absolute characterization) and arbitrarily predefined domains (relative characterization). Finally, we evaluate whether a system can safely operate in a target domain by comparing it to multiple source domains where safety has already been established.

**Contributions**

1. A novel probabilistic methodology to characterize domains in the case of autonomous vehicles driving in various weather conditions.
2. Demonstration that simulation-based inference (normalizing flows) can effectively estimate weather parameter distributions, with comparison of different backbones for features extraction.
3. A method for characterizing a new target domain as a mixture of source domains.

Related publication. This chapter is adapted with minor changes from Halin et al. [112].

Author Contributions. The work presented in this chapter results from a collaboration between researchers from four Walloon universities (ULiège, UCLouvain, ULB, and UMons) and the Multitel research and innovation center. The University of Liège led the project, and the conceptualization of the work was primarily carried out by myself, Sébastien Piérard (lead), and Renaud Vandeghen. The majority of the paper was written by the two co-first authors, myself and Sébastien Piérard. I was personally involved in the project from its definition through its development to its publication. More specifically, I contributed to the overall conceptualization and methodological design, generated the complete set of synthetic data used in all experiments across the three tasks using the CARLA simulator, and participated in several aspects of the experiments, including the posterior predictive check analysis for the first task. I also contributed to the analysis and visualization of the results and was involved in drafting the original version of the manuscript.

## 6.1 Introduction

Advances in computer vision allow widespread camera monitoring, but diverse weather conditions lead to visually different data, sometimes impacting the performance of high-level tasks like object detection and surveillance. Current solutions often lack generalizability across multi-weather scenarios, highlighting the need for adaptive methods that can process visual data under diverse conditions.

More specifically, weather conditions significantly affect the perception capabilities of autonomous driving systems [254], particularly under harsh conditions, such as heavy rain or fog, which can compromise their ability to operate safely. Therefore, it is essential to develop reliable approaches to analyze the environment, irrespective of the weather conditions. Detecting critical circumstances such as extreme weather events allows the system to respond appropriately within and outside its Operational Design Domain (ODD), defined by SAE International [283] as the "operating conditions under which a given driving automation system, or feature thereof, is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics".

However, predicting exact weather conditions is challenging due to the many ambiguous cases. A single observation of the environment (*e.g.*, recorded by a vehicle-mounted camera) could be characterized by multiple sets of weather parameters. Figure 6.1 presents two synthetic images, generated using the CARLA software [66], that appear nearly identical but were captured under different weather conditions. This shows the ambiguities inherent to characterizing the domain based on a single observation (*i.e.*, image).

Current methods predict weather conditions through regression or classification [141, 143, 188]. However, those approaches only produce a single crisp answer, lacking the insight about the ambiguities of characterizing the weather conditions. Figure 6.2 illustrates how the two scenarios from Fig. 6.1, when processed with an intermediate set of values, result in very different visual representations.

In this work, we propose a novel solution based on a probabilistic characterization of the weather. Historically, statistical inference relying on likelihood estimation was intractable when dealing with high-

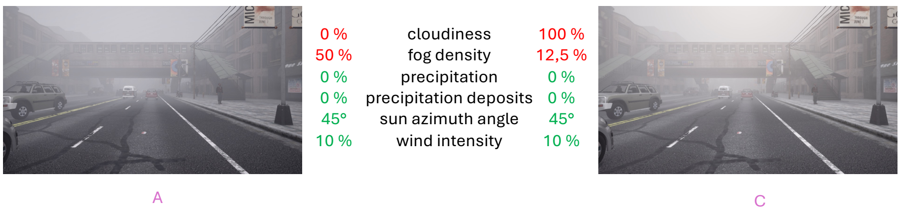



**Figure 6.1: Ambiguous cases.** The two synthetic images, A and C, generated using the CARLA software, appear almost identical, despite being acquired under very different weather conditions. This highlights the challenge of information loss from sensors like cameras when dealing with weather-related physical parameters. As a result, predictions that diverge from the ground truth in such ambiguous cases should not be penalized during evaluation.


**Figure 6.2: Regression.** Considering the images A and C from Fig. 6.1, generating an image B for the arithmetically averaged parameters, leads to an image very different from A and C. In other words, there are images $i$ such that the probability $P(I = i, W = \hat{w}(i))$ of the pair (image, estimated weather) is zero when the estimated weather is the expected value of the weather knowing the image, $\hat{w}(i) = E[W|I = i]$, as done in regression. Working with distributions of weather parameters (as shown using a contour plot on the right-hand side), as proposed in this work, rather than predicting specific values for each parameter, *i.e.*, regression (as shown on the left-hand side), avoids this problem.

dimensional data [54]. Recently, one class of density estimation techniques based on neural networks called normalizing flows [164, 246, 281] has become quite popular to solve this kind of problem. These models learn invertible transformations to go from complex distributions to more handy ones, and can therefore be used for modeling weather parameters from highly complex image and weather distributions.

These novel techniques allow us to express the domain of an image by a probabilistic distribution, *e.g.*, weather conditions, compared to current deterministic approaches. More specifically, we show on three consecutive tasks how various weather conditions can be predicted based (1) on a single color image acquired in front of a vehicle, (2) on a bag of color images (*absolute characterization*), and (3) how the current domain is related to arbitrarily chosen source domains (*relative characterization*).

Notions similar to ODD are very common in other fields where equivalent terms are used, such as *Operational Envelope* for maritime and *Operational Context* for railroad [322]. Additionally, other fields are facing domain shifts, such as in medical imagery from different acquisition devices [93]. We argue that our approach could serve a large range of practical applications. For example, determining the most suitable model for a given scenario within a fleet of lightweight AI models able to analyze the environment, as proposed in previous works [92, 209, 256], detecting significant domain transitions to collect new data for domain adaptation methods that rely on buffers or adaptable internal statistics [100, 134, 340, 342, 376, 344] or activating adaptation mechanisms based on clustering [32, 317, 380, 379].

We summarize our contributions as follows: (1) We propose a novel probabilistic methodology to characterize domains in the case of autonomous vehicles driving in various weather conditions. (2) We demonstrate that simulation-based inference (normalizing flows) is adequate to obtain distributions for weather parameters that are used to characterize the domain and compare different backbones for feature extraction. (3) Based on this weather domain characterization, we show how to characterize a new target domain as a mixture of source models.

## 6.2 Related Work

### 6.2.1 Predicting Weather Conditions from Images

Prediction of weather conditions from images was first formulated as a single-label classification task (*e.g.*, sunny, cloudy, or foggy). In 2014, Lu et al. [204] proposed a binary classification task between sunny and cloudy weathers, using features extracted from visual cues such as the sky, shadows, reflections, contrast, and haze. Later, Villarreal Guerra et al. [332] proposed a multi-class dataset extending the scope of the classification task to rain, snow, and fog. Recent works focus on optimizing Convolutional Neural Network (CNN) to obtain strong features for common and uncommon weather conditions [192, 384]. However, weather conditions can hardly be represented by crisp classification due to their continuous nature.

To reach a more realistic description of intricate relations between weather conditions, recent methods simultaneously regress several physical parameters [141, 143]. To increase interpretability, Li et al. [188] also assign cues of weather characteristics to each pixel. However, these methods still lack the ability to represent ambiguous scenarios (see Fig. 6.2). In this work, we predict the joint distribution of weather parameters by proposing a novel method based on normalizing flows.

### 6.2.2 Handling Weather Conditions for Autonomous Driving

Autonomous car driving systems need to be efficient under all weather conditions. Some methods propose to integrate a generalization step into the model to remove the environmental influences on the acquired images by introducing a style layer inspired by image style-transfer neural networks [274] or through adversarial training [184]. Generalized features obtained from alignment of source and target domains tend to be suboptimal, as they do not consider the task. To improve the estimation, a solution is to add a domain adaptation step [181]. Jeon et al. [147] further improved their estimation by adding an unsupervised domain adaptation step after domain alignment.

Many studies have also highlighted the importance of a strong ODD definition to properly assess the ability of automatic driving systems to work in given conditions regarding weather, location, other vehicles on the road, state of the car sensors, and many other environmental parameters [51, 106, 248]. Many strategies have been proposed to evaluate different situations according to specific evaluations of potential damage cost [178, 314] and define the boundaries of the ODD. In this work, we characterize the domain by a probability distribution focusing on weather parameters in an autonomous driving environment.

### 6.2.3 Probabilistic Modeling of Parameters Distributions

We propose to leverage recent observations in the Simulation-Based Inference (SBI) literature [54, 327] for domain characterization. This literature has seen a rapid expansion thanks to new density estimation techniques in problems where likelihood estimation was often intractable, especially for high-dimensional data. One class of these density estimation techniques based on neural networks is normalizing flows [246]. The principle consists in transforming an arbitrarily chosen distribution (*e.g.*, a Gaussian) into the desired distribution. Different types exist, such as the Neural Spline Flow (NSF) type [68] that can be used in two ways: 1 to determine the value of the Probability Density Function (PDF) at a given point and 2 to draw samples at random. Different techniques can be used to learn them, *e.g.* the Neural Posterior Estimation (NPE) technique [103, 206].

There are a few techniques to analyze the performance of models predicting distributions. *Coverage Plots* [126] show, objectively and quantitatively, whether the distribution prediction models are underconfident (*i.e.*, conservative), calibrated, or overconfident. Another technique, specific for parameter distributions (*e.g.*, weather parameters) predicted from an observation (*e.g.*, an image) and widespread in the field of SBI, is known as *Posterior Predictive Check* [280]. It consists of drawing parameters at random from a predicted distribution, injecting these parameters into a simulator or physical system, and comparing the resulting observations with the one from which the distribution was predicted. In this work, we leverage those analysis techniques for assessing the quality of our weather characterization models.

## 6.3 The Three Fundamental Tasks Behind the Physically Interpretable Probabilistic Domain Characterization

Our experiments are organized around three different tasks, involving a prediction of the distribution of weather conditions given some images acquired by color cameras placed in front of vehicles. Before elaborating on these tasks in Sections 6.3.2 to 6.3.4, we briefly introduce our framework in Section 6.3.1.

### 6.3.1 Framework

#### Mathematical Modeling

We denote the set of all possible values for the physical parameters (*e.g.*, weather conditions) by $\mathbb{W}$ and the set of all observations of interest (*e.g.*, images) by $\mathbb{I}$. We adopt the probability theory of Kolmogorov [165, 166] and consider a measurable space $(\Omega, \Sigma)$ as well as the (generalized) random variables $W : \Omega \to \mathbb{W}$ for the physical parameters and $I : \Omega \to \mathbb{I}$ for the observation. Following the mathematical modeling introduced in [256], we consider the set $\mathbb{D}_{(\Omega,\Sigma)}$ of domains $d$ in which there is a probability measure $P_d$ on $(\Omega, \Sigma)$. We see the ODD of a given autonomous system as the set of domains in which it can be used safely, no matter if this has been established by design or by testing. Thus, $ODD \subseteq \mathbb{D}_{(\Omega,\Sigma)}$.

#### Data

All our experiments are performed on data generated using the CARLA software [66], an open-source simulator for autonomous driving research. All images are acquired by a simulated camera placed in front of a vehicle called the *ego vehicle*. The weather is controlled through $13$ real-valued parameters (see Table 6.1). There are $6$ parameters for which we make predictions. Thus, $\mathbb{W} \subseteq \mathbb{R}^6$.

**Table 6.1: Weather parameters.** In CARLA, the weather is controlled through $13$ physical parameters. This table shows the range of values that we consider in our experiments and indicates, for each of them, if they are considered in our predictions.

| parameter | range | predicted |
|---|---|---|
| `cloudiness` (`cloud`) | $0$ to $100\%$ | yes |
| `fog density` (`fog`) | $0$ to $100\%$ | yes |
| `precipitation` (`rain`) | $0$ to $100\%$ | yes |
| `sun azimuth angle` | $0°$ to $360°$ | no |
| `sun altitude angle` (`sun`) | $-90°$ to $90°$ | yes |
| `wind intensity` (`wind`) | $0$ to $100\%$ | yes |
| `precipitation deposits` (`deposit`) | $0$ to $100\%$ | yes |
| `fog distance` | fixed to $0.75$ | no |
| `fog falloff` | fixed to $0.1$ | no |
| `mie scattering scale` | fixed to $0.03$ | no |
| `rayleigh scattering scale` | fixed to $0.033$ | no |
| `scattering intensity` | fixed to $1.0$ | no |
| `wetness` | fixed to $0.0$ | no |

### 6.3.2 Task I: Predicting Distributions of Physical Parameters

The first task (see Fig. 6.3) consists in predicting how likely various physical parameters are, jointly. As these parameters cannot, in general, be measured directly, it is necessary to estimate the likelihood, *i.e.*, the distribution of plausible values, based on an indirect observation.

#### Case Study

We study here the particular case in which the physical parameters are relative to the weather conditions and the observations are color images $i \in \mathbb{I}$ acquired by cameras placed in front of vehicles. We aim at learning, offline, a deep learning model $i \mapsto \hat{P}_{d_R}(W|I = i)$, where $d_R$ denotes a domain of reference in which the probability measure on $(\Omega, \Sigma)$ is $P_{d_R}$. This domain is arbitrarily chosen in such a way that $P_{d_R}$ has a large support.

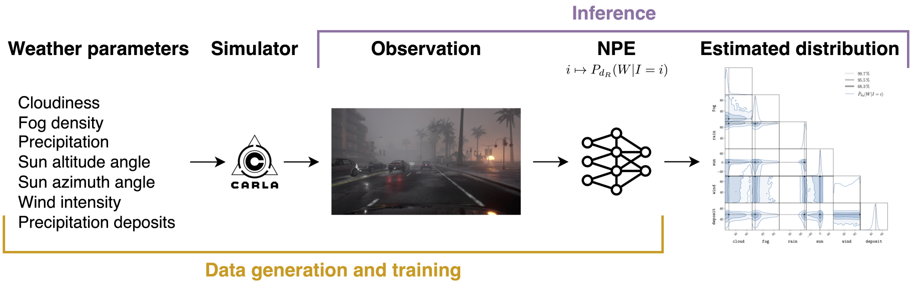



**Figure 6.3: Task I.** The aim of the first task is, based on an image, to predict the joint distribution of the weather parameters. For this purpose, (1) we generate data using the CARLA software for uniformly distributed weather parameters (offline), (2) we train a NPE model using the learning set of our generated data (offline), and (3) we infer, given an image from the test set, the estimated weather distribution (normalizing flow) and show the result on a corner plot.

**Figure 6.4: Excerpt of the images in our dataset** generated with the CARLA simulator. The ground-truth weather parameters are drawn at random for each image, following a uniform distribution with the bounds given in Table 6.1.

## Data

We use CARLA to generate a dataset with $635$k images and the corresponding ground-truth values for the weather parameters. The dataset (Fig. 6.4) is split into a learning set with $600$k images ($500$k for the training set and $100$k for the validation set) and a test set with $35$k images. Letting the model of the ego vehicle, the map, the number of pedestrians, and the number of vehicles vary brings a touch of diversity to our data.

### Method

We consider three different (frozen) backbones to extract features from the input images: ResNet-50 [123], DINOv2 [241], and CLIP [265]. We use the libraries LAMPE [282] and ZUKO [281] to learn a model (of type NPE) and to manipulate the normalizing flows (of type NSF), respectively. All predicted weather distributions are posteriors relative to the weather priors in the learning set. Note that LAMPE and ZUKO were not developed for domain characterization, but rather for SBI. Also, to the best of our knowledge, these libraries have only been used once with very high-dimensional input data [327].

### Evaluation and Results

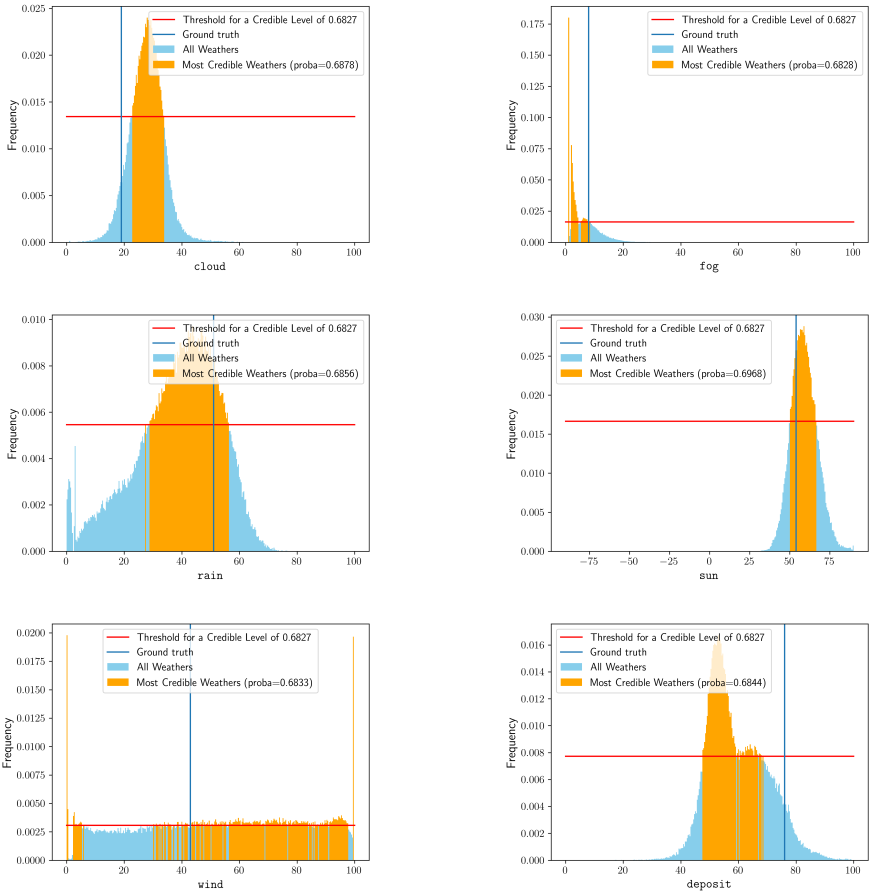
Five different analyses are carried out.

1ST ANALYSIS: HISTOGRAMS (FIG. 6.5B). Due to their physical meaning, the weather parameters are easily interpretable. This paves the way to a first, subjective, evaluation. We draw samples at random out of the predicted weather distributions $\hat{P}_{d_R}(W|I=i)$, for some images $i$ arbitrarily chosen in the test set, and conduct a visual inspection of the histograms for the $6$ marginals. The most credible weather parameters (*a.k.a.* highest density credibility sets, highest density regions, plausible sets, etc.) are highlighted for a credible level $l=68.27\%$. For any input image $i$, these weathers are those such that the predicted Probability Density Function (PDF) is above some threshold $t(i)$ and the predicted probability of the set is $l$ [140].

2ND ANALYSIS: CORNER PLOTS (FIG. 6.5C). A corner plot is a triangular array of plots. Those on the diagonal show the $n$ marginals of a $n$-D distribution. Those below the diagonal depict the highest density credibility regions delimited at some arbitrarily chosen credibility levels (in this paper: 68.27%, 95.45%, and 99.73%), for each pair of parameters. We observe on the corner plots wide distributions, meaning that there is a large uncertainty for the weather parameters given an image. However, when analyzing the results, this uncertainty is explainable, and the following analysis shows that this uncertainty is not excessive, *i.e.*, our models are not underconfident.

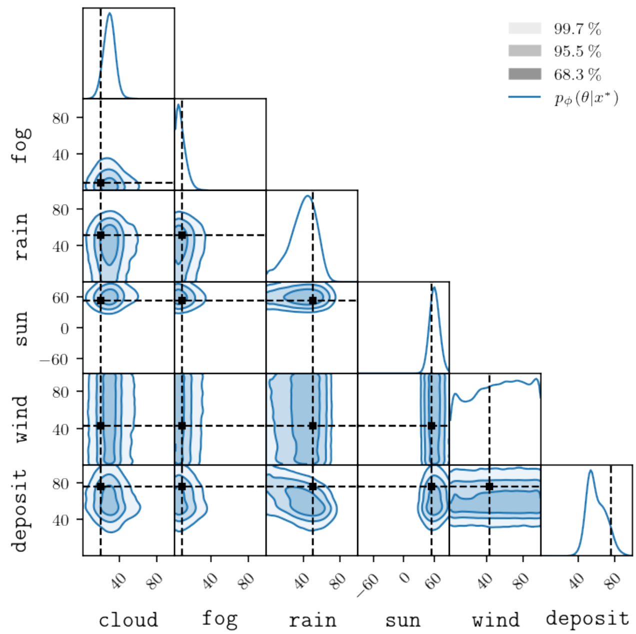
| weather parameter | ground-truth value |
|---|---|
| cloud | 19 |
| fog | 8 |
| rain | 51 |
| sun | 54 |
| wind | 43 |
| deposit | 76 |

**(a)** An input image $i$ and the corresponding ground-truth values.

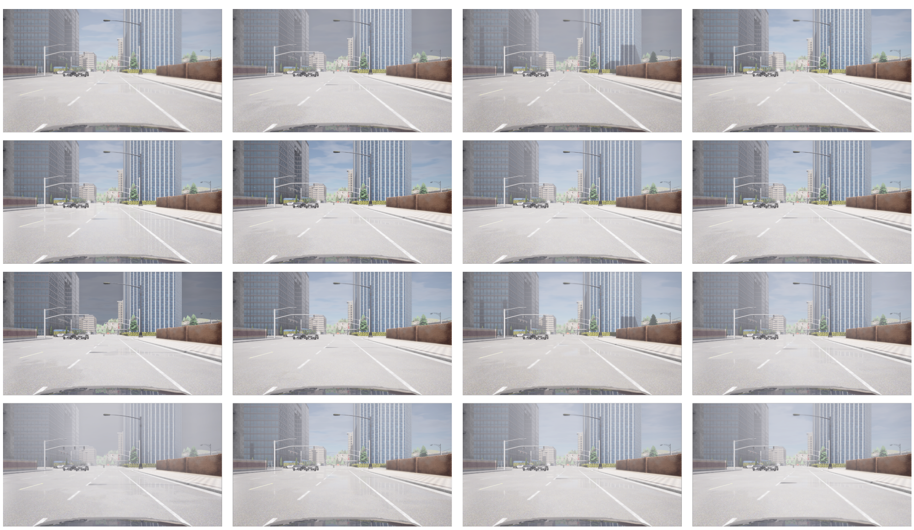



**(b)** Histograms for the 6 marginals of the predicted weather distribution for $i$.

**Figure 6.5: Task I: results obtained (Part 1),** with the model learned with the ResNet-50 backbone on $500$k learning samples, for an arbitrarily chosen input image in the test set.




**(c)** Corner plot, with the ground-truth weather pinned.



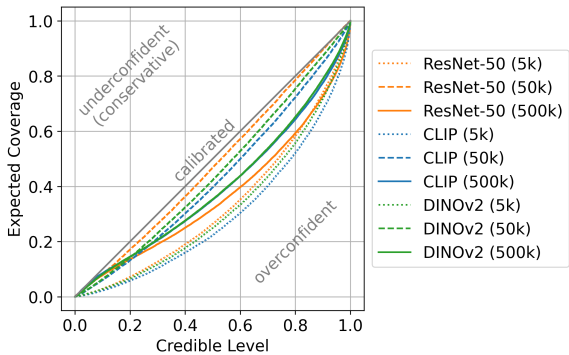
**(d)** Posterior predictive check.

**Figure 6.5: Task I: results obtained (Part 2),** with the model learned with the ResNet-50 backbone on $500$k learning samples, for an arbitrarily chosen input image in the test set.

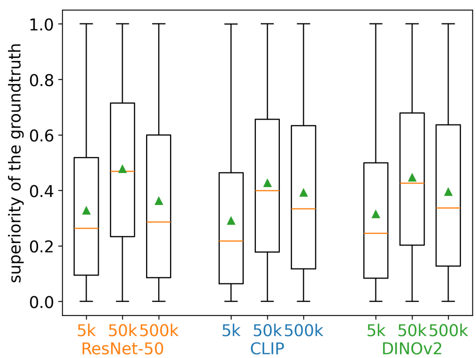
3RD ANALYSIS: POSTERIOR PREDICTIVE CHECKS (FIG. 6.5D). Once a weather distribution is predicted for an observation (input image), it is possible to 1 draw weather parameters vectors at random from it and then 2 to inject these vectors in CARLA, keeping all other parameters unchanged and immobilizing the vehicles and pedestrians, to obtain new images in order to finally 3 compare those with the initial input image. One cannot expect to obtain identical images as, in CARLA, the traffic lights continue to run, the rain continues to fall, the plants move with the wind, and pedestrians' poses are not perfectly frozen. Putting this aside, we observe that most retrieved images are very similar to the input image. We conclude that our models are not underconfident.

4TH ANALYSIS: COVERAGE PLOTS (FIG. 6.6A). Coverage plots show how the expected coverage varies with the credible level [126]. By definition, the expected coverage, at the credible level $l \in [0, 1]$, is the probability that the ground-truth weather belongs to the most credible weathers at level $l$. A method estimating the distribution of weathers based on an image is said to be underconfident (*i.e.*, conservative), calibrated, or overconfident at level $l$ when the expected coverage at level $l$ is, respectively, $> l$, $= l$, or $< l$. We observe that our models are all overconfident. The best calibrated model is the one that we obtained with $50$k learning samples and ResNet-50 as backbone.

5TH ANALYSIS: SUPERIORITY OF THE GROUND TRUTH (FIG. 6.6B). We also determine to what extent the ground truth is more credible than the other weathers. This is the proportion $\pi$ of weathers that can be drawn at random from the normalizing flow and that are predicted as having a PDF value (*i.e.*, likelihood) lower or equal to the one from the ground truth. The higher $\pi$ is, the better it is, but approaching $1.0$ is notably very challenging. We estimated $\pi$ for each image of the test set and made box-and-whisker plots for the $9$ models. This analysis is complementary to the coverage plot in the sense that a model can be perfectly calibrated while still presenting room for improvement. That being said, we are in the particular case in which this analysis leads to the same conclusion as the coverage plot: the model based on the backbone ResNet-50 and learned from $50$k samples is preferable to the others.




**(a)** Coverage plot




**(b)** Box-and-whisker plots for the $\pi$ statistic

**Figure 6.6: Task I: comparison of our 9 models on the overall test set.** Both analyses, (a) coverage plot and (b) box-and-whisker plots for the $\pi$ statistic, agree that the model learned with the ResNet-50 backbone on $50$k learning samples is preferable.

### 6.3.3 Task II: Absolute Domain Characterization

The second task (see Fig. 6.7) consists in characterizing a domain in an easy-to-interpret way. For this, we opt for a distribution of physical parameters estimated based on a bag (*a.k.a.* multiset) of observations.

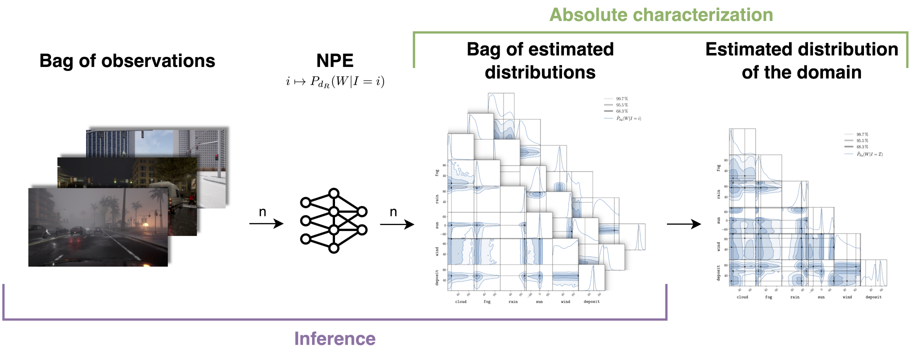



**Figure 6.7: Task II.** The aim of the second task is to obtain an absolute characterization of a domain of interest. We process with the NPE each observation of a bag of observations of the domain of interest to obtain the bag of estimated distributions corresponding to the observations, then we obtain the distribution of weather parameters for the domain of interest by averaging the individual distributions of the bag.

#### Case Study

We characterize a domain $d \in \mathbb{D}_{(\Omega,\Sigma)}$ by the estimated distribution $\hat{P}_d(W)$ of weather conditions based on a real-valued bag $b$ (multiset) of arbitrarily weighted images acquired by cameras placed in front of vehicles. In the following, we denote the weight (multiplicity) of the image $i \in b$ by $\omega(i)$. The images can either originate from a unique vehicle or from several, in the case of vehicle-to-vehicle (V2V) communications. We want to establish $b \mapsto \hat{P}_d(W|B=b)$.

#### Data

We consider a bag $b$ of $1{,}000$ equally weighted images generated with CARLA, in maps already used in Task I. The weather parameters have been drawn at random from an arbitrary distribution $P_d(W)$ for which we fixed all parameters but two, `fog density` and `precipitation`, that follow a mixture of Gaussians.

### Method

We use the NPE model developed for task I and build our solution for task II on top of it. We implement the following estimator for $P_d(W)$: $\hat{P}_d(W) = \sum_{i \in b} \omega(i) \hat{P}_{d_R}(W|I = i)$. The motivation for this estimator is threefold. 1 It is straightforward to evaluate the probability density function of $\hat{P}_d(W)$ and to draw weather conditions $w \in \mathbb{W}$ at random, following $\hat{P}_d(W)$, as we can do it with the normalizing flow $P_{d_R}(W|I = i)$. This will be valuable in our third task. 2 This estimator is useful for linear temporal filtering, *e.g.*, when one wants to weight more the recently acquired images than the old ones. 3 Finally, this estimator is fully justifiable under the assumptions that $P_d(W|I) = P_{d_R}(W|I)$ and $\omega(i) = P_d(I = i)$ as, in this case, $P_d(W) = \int_i \omega(i) P_{d_R}(W|I = i) di$.

### Evaluation and Results

Our result is shown in Fig. 6.8. We observe that the three modes of the ground-truth weather distribution $P_d(W)$ are within the highest density regions of the prediction $\hat{P}_d(W)$. We made the same observation for many other ground-truth distributions. We stress the fact that $\hat{P}_d(W)$ differs significantly from $P_d(W)$ can be explained by the inherent loss of information resulting from the use of color cameras, as already discussed in the introduction (*cf.* Fig. 6.1).

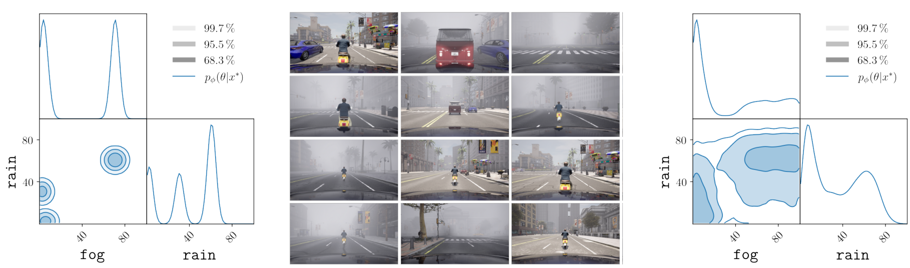



**Figure 6.8: Task II: example of result.** Left: corner plot showing an arbitrarily chosen distribution of weather conditions, $P_d(W)$ (ground truth). Center: a bag $b$ of images generated using CARLA with weather conditions drawn from $P_d(W)$. Right: corner plot showing the estimated likelihood of the weather conditions $\hat{P}_d(W)$ based on $b$.

### 6.3.4 Task III: Relative Domain Characterization

The third task (see Fig. 6.9) consists in characterizing any target domain $d_T \in \mathbb{D}_{(\Omega,\Sigma)}$, relatively, *w.r.t.* to some arbitrarily chosen source domains $d_{S_1}$, $d_{S_2}$, $\dots d_{S_{n_S}} \in \mathbb{D}_{(\Omega,\Sigma)}$. Our motivation for this task originates from the importance of knowing if a system implementing a given many-to-one domain adaptation method can operate safely in the target domain $d_T$ when it is known to operate safely in the source domains $d_{S_1}, d_{S_2}, \dots d_{S_{n_S}}$.

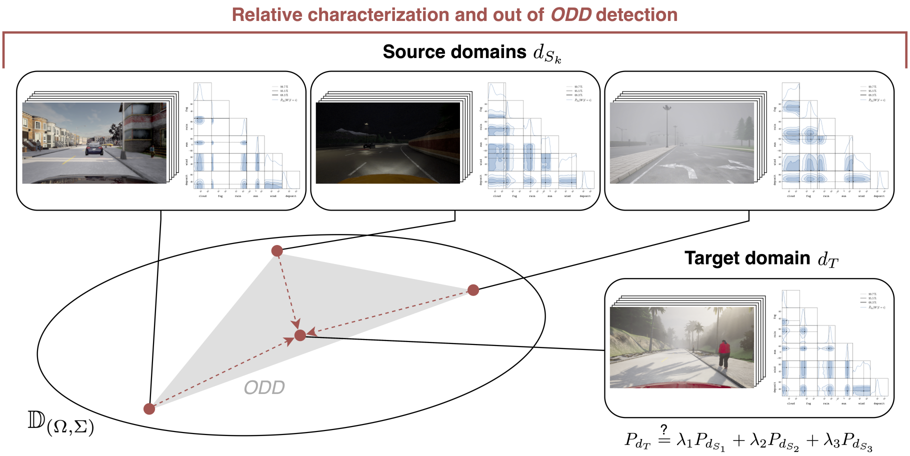



**Figure 6.9: Task III.** The aim of the third task is twofold: (1) having a relative characterization of a target domain $d_T$ based on source domains $d_{S_k}$ and (2) detecting when this target domain $d_T$ is out of the ODD.

#### Case Study

We discuss the case of *Mixture Domain Adaptation* (MDA) in which the system adapts to any target domain for which $P_{d_T} = \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}$, with $\sum_{k=1}^{n_S} \lambda_k = 1$ and $\lambda_k \geq 0 \, \forall k$ (this is the *mixture assumption*). In this framework, the ODD is the convex hull of $\{P_{d_{S_1}}, P_{d_{S_2}}, \dots P_{d_{S_{n_S}}}\}$. An autonomous car implementing MDA is expected to drive safely when $P_{d_T} \in ODD$. So, our goal in this third task is to determine if the mixture assumption holds and, if so, what the values of the mixture weights $\lambda_k$ are.

MDA supports many potential applications in the high-level tasks of the *Sense* pillar of the *Sense-Plan-Act* model [299]. While Mansour et al. [209] studied it generically, the application to the two-class classification task has been studied in [260], and the application to the semantic segmentation task of images acquired by vehicle-mounted cameras has been studied in [256]. These latter two works put an important emphasis on the on-the-fly applicability and provide mathematically proven exact solutions. However, a critical limitation of these works is the need for the mixture weights to be known at adaptation time. Here, we remove this limitation by introducing a method that determines these weights automatically.

### Data

We consider $4$ subsets of the test set that we created for Task I: $b_0$, $b_1$, $b_2$, and $b_3$, containing $176$, $27$, $21$, and $26$ images, respectively. In the bag $b_k$, the distribution of the ground-truth weather parameters follows a uniform distribution on $\mathbb{W}_k \subsetneq \mathbb{W}$. These sets are such that $\mathbb{W}_0 \cap (\mathbb{W}_1 \cup \mathbb{W}_2 \cup \mathbb{W}_3) = \emptyset$.

### Method

We characterize the source and target domains with the method presented for Task II, using the same weather distribution predictive model for all domains. We define the mean squared gap between the target domain and the mixture of the source domains as:

$$\begin{aligned}\delta(\hat{\lambda}_1, \dots \hat{\lambda}_{n_S}) &= \int_{\mathbb{W}} \left[ \hat{P}_{d_T}(W = w) - \sum_{k=1}^{n_S} \hat{\lambda}_k \hat{P}_{d_{S_k}}(W = w) \right]^2 \hat{P}_{d_T}(W = w) dw \\ &\simeq \frac{1}{n_W} \sum_{i}^{n_W} \left[ \hat{P}_{d_T}(W = w_i) - \sum_{k=1}^{n_S} \hat{\lambda}_k \hat{P}_{d_{S_k}}(W = w_i) \right]^2 \end{aligned} \tag{6.1}$$

with $\{w_i\}_{i=1}^{n_W} \sim \hat{P}_{d_T}(W)$. We aim at finding the values of $\hat{\lambda}_1, \dots \hat{\lambda}_{n_S}$ that minimize $\delta(\hat{\lambda}_1, \dots \hat{\lambda}_{n_S})$. This is a constrained least squares problem in which the constraints are $\sum_{k=1}^{n_S} \hat{\lambda}_k = 1$ and $\hat{\lambda}_k \geq 0, \forall k$. To use the CVXPY library [3, 61] to solve our problem, we converted our original problem into a convex quadratic programming problem with the same constraints.

### Evaluation and Results

The goal of this experiment is twofold.

① To show that it is possible to recover the mixture weights needed for the MDA technique presented in [256]. If the probability measure in the target domain is a mixture of the probability measures in the source domains, then we expect our characterization of the target domain to be a mixture of our characterizations for the source domains, with the same weights, as $P_{d_T} = \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}} \Rightarrow P_{d_T}(W) = \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}(W)$.

② To show that it is possible to detect when the target domain is not a mixture of the source domains, which means that the target domain is out of the ODD for the MDA technique presented in [256]. If our characterization of the target domain is not a mixture of our characterizations for the source domains, then we expect that the probability measure in the target domain is not a mixture of the probability measures in the source domains, as we have $P_{d_T}(W) \neq \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}(W) \Rightarrow P_{d_T} \neq \sum_{k=1}^{n_S} \lambda_k P_{d_{S_k}}$.

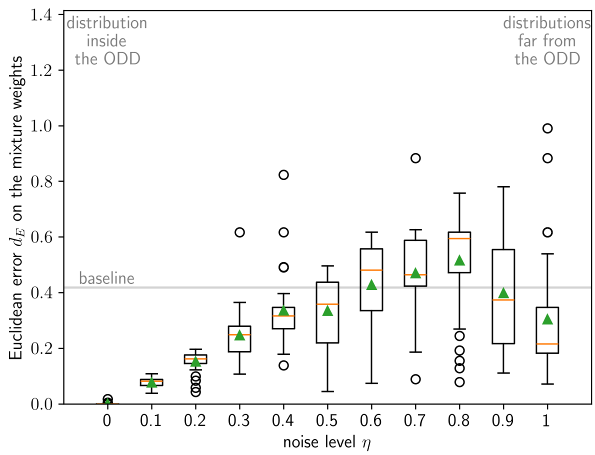
To achieve this goal, we perform experiments in which we characterize target domains relative to $n_S = 3$ source domains. However, to study the robustness of our method, we let the probability measures in the target domains $d_T$ be mixtures not only of the probability measures in the source domains, but also in some domains that are out of the ODD. We use the bags $b_1$, $b_2$, and $b_3$ for the source domains $d_{S_1}$, $d_{S_2}$, and $d_{S_3}$, respectively. We also use the bag $b_0$ for domains $d_{\notin ODD}$ that are, by construction, guaranteed to be out of the ODD. Putting this into equations, we have $P_{d_T} = (1-\eta) P_{d_{\in ODD}} + \eta P_{d_{\notin ODD}}$ with $P_{d_{\in ODD}} = \lambda_1 P_{d_{S_1}} + \lambda_2 P_{d_{S_2}} + \lambda_3 P_{d_{S_3}}$. The quantity $\eta \in [0,1]$ is interpreted as a proportion of noise, which is the keystone to assess the robustness of our method.

In each experiment, the images in $b_1$, $b_2$, and $b_3$ are equally weighted, whereas those in $b_0$ are randomly weighted following a uniform distribution. The method introduced in Task II for the absolute characterization of domains is applied for $d_{S_1}$, $d_{S_2}$, $d_{S_3}$, and $d_T$. We arbitrarily chose $(\lambda_1, \lambda_2, \lambda_3) = (0.2, 0.3, 0.5)$. The mixture weights are optimized on $n_W = 16$ points.

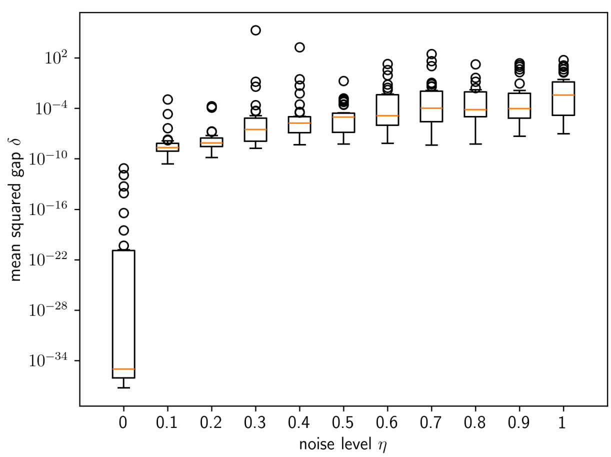
The overall experiment consists of (1) choosing a target domain at random, as described here-above, (2) computing its absolute characterization with the method of Task II, (3) executing the weight estimation algorithm, and (4) reporting both the mean square gap $\delta(\hat{\lambda}_1, \dots \hat{\lambda}_{n_S})$ and the Euclidean distance $d_E$ between $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$ and $(\lambda_1, \lambda_2, \lambda_3)$. This experiment has been performed $330$ times ($30$ times for $11$ values of $\eta$). The results are shown in Fig. 6.10. The mean square gap achievable by chance is shown (baseline). Note that the target domain belongs to the ODD only when $\eta = 0$. We observe that $d_E$ is negligible when $d_T \in ODD$ and that the mean square gap is effective to detect when $d_T \notin ODD$.

## 6.4 Conclusion

In this work, we present a novel approach to characterize domains by estimating the distribution of physical parameters. Our method enhances interpretability, facilitates domain adaptation, and provides safeguards for systems operating outside their ODD. Our experiments, organized around three tasks, are performed in the particular, but very important, case of autonomous vehicles and demonstrate how to obtain an absolute characterization of a domain by predicting a distribution of weather parameters (1) using a single image acquired by a vehicle-mounted camera and (2) using a bag of images. Our experiments also demonstrate (3) how to obtain a relative characterization of a target domain based on arbitrarily chosen source domains. Our approach includes two types of domain characterization: absolute and relative. The relative characterization is particularly valuable for domain adaptation, allowing the expression of a target domain in terms of source domains and verifying whether the current domain is part of the ODD. This is important for autonomous driving systems as well as for other fields requiring interpretable and trustworthy domain adaptation.




**(a)** Box-and-whisker plots showing the distributions of $d_E$

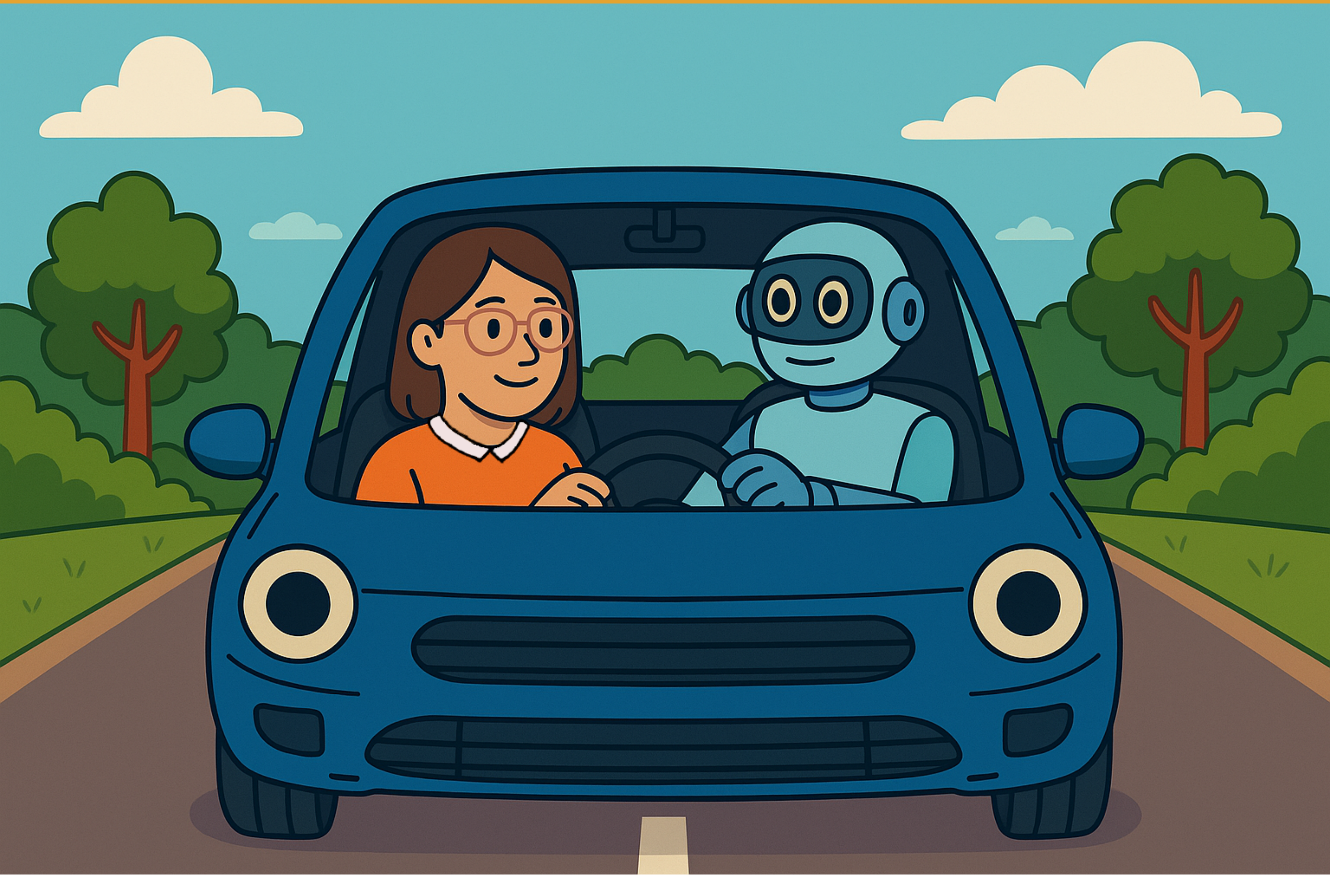



**(b)** Box-and-whisker plots showing the distributions of $\delta(\hat{\lambda}_1, \dots \hat{\lambda}_{n_S})$

**Figure 6.10: Task III: results obtained** when the noise level $\eta$ sweeps the $[0, 1]$ interval. These box-and-whisker plots of (a) the distributions of $d_E$ and (b) the distributions of $\delta(\hat{\lambda}_1, \dots \hat{\lambda}_{n_S})$ show (1) that we can recover the relationship between the target domain and the source domains, under the mixture assumption, when the target domain belongs to the ODD and (2) that the $\delta(\hat{\lambda}_1, \dots \hat{\lambda}_{n_S})$ can be used to determine when the target domain belongs to the ODD.

# Part IV

## Toward an Integrated Framework for Adaptive Driving Automation



# 7

# A Framework for Risk-Aware Adaptive Automation of Driving

**Contents of this chapter**



CONTEXT. Following the analyses conducted in Parts II and III, which examined the driver and the environment separately, Part IV brings these elements together by proposing an integrated framework for adaptive driving automation. The increasing integration of automation in vehicles aims to enhance both safety and comfort, but it also introduces new risks, including driver disengagement, reduced situation awareness, and mode confusion. In this work, we propose the *DEV* framework, a closed-loop framework for risk-aware adaptive driving automation that captures the dynamic interplay between the driver, the environment, and the vehicle. The framework promotes continuously adjusting the operational level of automation based on a risk management strategy. The real-time risk assessment supports smoother transitions and effective cooperation between the driver and the automation system. Furthermore, we introduce a nomenclature of indexes corresponding to each core component, namely driver involvement, environment complexity, and vehicle engagement, and discuss how their interaction influences driving risk. The *DEV* framework offers a comprehensive perspective to align multidisciplinary research efforts and guide the development of dynamic, risk-aware driving automation systems.

**Contributions**

1. A closed-Loop framework, called *DEV*, for risk-aware adaptive automation of driving integrating the driver, the environment, and the vehicle.
2. A nomenclature of indexes corresponding to each core component of the framework: driver involvement, environment complexity, and vehicle engagement.

Related publication. This chapter is adapted with minor changes from Halin et al. [109].

## 7.1 Introduction

Vehicles are becoming increasingly automated so that driving has truly become a cooperation between the human driver and the machine driver, commonly referred to as the driving automation system. But for this cooperation to be truly effective and to take place under optimal conditions, we need a human-machine interface that is well designed [235, 341], that is, an interface that allows drivers to clearly understand their responsibilities at all times, to form an accurate mental model, and to grasp what the driving automation system is doing and why it is doing it. Only then can drivers trust the system and play their role in the best conditions.

Currently, there are many issues where drivers lack awareness of automation limitations [24], overtrust the system (*e.g.*, caused by phenomena like autonowashing [63]), or have mode confusion [356]. On the other hand, some drivers do not benefit from the advantages of driving automation at all because they never activate the available driving automation features [243].

The SAE levels [283] of driving automation do little to help drivers better understand their responsibilities or to support effective real-time cooperation between the human and the automation in driving [33, 137, 142]. At best, the SAE levels serve legal or manufacturer-related purposes.

Even if we assume the existence of an ideal communication interface between the human and the machine, one that enables perfect cooperation between the two, the question remains: how should the human and the machine cooperate to mitigate the risk? Drivers have limited capacities that fluctuate depending on their state and their willingness to engage in the driving task. Similarly, the vehicle is equipped with a driving automation system that includes more or fewer automation features (depending on its SAE level), which can be activated or deactivated. The driver state and the availability or appropriateness of automation both depend on the driving environment. For example, the literature tells us that a monotonous environment increases driver fatigue [77], while a complex environment leads to cognitive overload [319]. The vehicle, for its part, operates within an Operational Design Domain (ODD) that restricts when and where automation features can be used [283]. Additionally, some driving automation features reduce the driver's mental workload by helping with certain tasks, sometimes even too much, to the point where the driver disengages entirely from the driving task [20, 58]. Others, such as warning systems, may actually increase the driver workload [167].

In short, there is a highly complex interaction between the driver, the driving automation system, and the driving environment. Understanding the full scope of this interaction is crucial to defining the optimal cooperation between the human and the automation system, one that enables safe driving and minimizes risk.

We therefore introduce a closed-loop framework that accounts for the driver, the vehicle, and the environment, and that enables dynamic adaptation of the operational level of automation to mitigate risk. We discuss how the driver, the vehicle, and the environment each influence risk. This framework highlights the key questions that still need to be addressed to implement such a system.

## 7.2 The *DEV* Closed-Loop Framework for Risk-Aware Adaptive Automation

In this section, we introduce the *DEV* closed-loop framework, which supports risk-aware adaptive automation by dynamically adjusting the operational level of vehicle automation (see Fig. 7.1). This adjustment is based on three core components: 1 driver involvement, 2 environment complexity, and 3 vehicle engagement. We define these components below and give some leads to assess risk to guide the adaptation process.

### 7.2.1 Driver Involvement

> **Definition 1.** Driver involvement *is defined as the proportion of driver (cognitive, perceptual, and physical) resources that are available to perform the driving task.*

High driver involvement indicates that the driver is fully fit and ready to drive, while low driver involvement suggests that the driver is in a state where taking control of the vehicle would be undesirable or unsafe. Driver involvement depends on the driver's state (*e.g.*, drowsi-

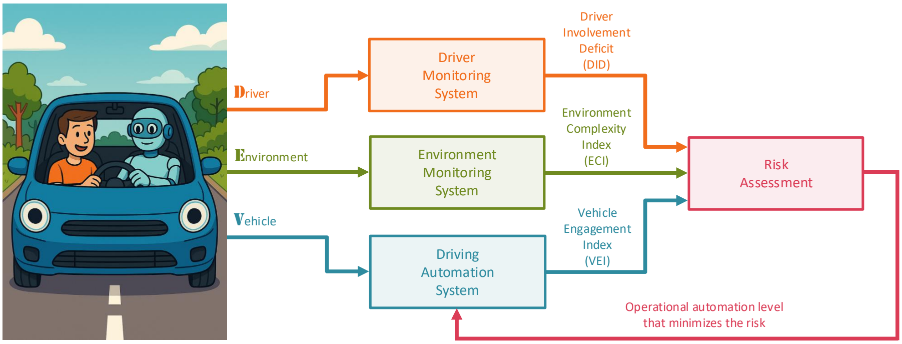



**Figure 7.1: The DEV closed-loop framework for risk-aware adaptive automation.** At all SAE levels except level 5, the driver and the vehicle collaborate to perform the driving task. The operational automation level of the driving automation system can be dynamically adjusted to mitigate the risk by accounting for three core dimensions: the driver, the environment, and the vehicle. A driver involvement deficit (DID) is computed using a driver monitoring system, based on the driver's state and fitness to drive. An environment complexity index (ECI) is derived from an environment monitoring system using data such as traffic conditions, road type, and visibility. The vehicle engagement index (VEI) reflects the extent to which the vehicle, via the driving automation system, is involved in the driving task. Risk is assessed based on DID, ECI, and VEI, and the system adjusts the operational level of automation accordingly.

ness, distraction) and, to some extent, on their willingness to engage in the driving task. For instance, a driver may be fully awake but choose to disengage out of laziness or disinterest, thereby reducing their effective involvement.

While driver involvement cannot be measured directly, it can be estimated. We define two complementary numbers, both ranging from $0$ to $1$: the driver involvement index (DII) and its complement, the driver involvement deficit (DID), such that DID $= 1 -$ DII. A driver involvement index of $1$ (*i.e.*, a deficit of $0$) indicates that the driver is fully capable of performing the entire driving task. Conversely, an index of $0$ (*i.e.*, a deficit of $1$) indicates that the driver is entirely unable to contribute to the driving task. Note that an index or a deficit should have various properties, such as being measurable, increasing (or decreasing), on a scale (which can be discrete), and reproducible, similarly to how an indicator is defined in Section 2.4.

The driver involvement deficit DID can be estimated by assessing the various components affecting the driver's state, such as drowsiness, mental workload, distraction, emotions, and under influence [115]. One possible approach is to compute a weighted average of these components to obtain an overall estimate. Another approach is to directly assess a global estimate of the driver involvement deficit through the implementation of the Minimum Required Attention (MiRA) theory [7, 161]. According to MiRA, drivers are considered attentive if they sample sufficient information from the environment to meet the demands of the driving task, or, in other words, if they fulfill the preconditions needed to form and maintain a good enough mental representation of the situation. Drivers are thus considered inattentive only when information sampling is not sufficient, regardless of whether they are drowsy, concurrently executing an additional task or not.

### 7.2.2 Environment Complexity

> **Definition 2.** Environment complexity *is defined as the proportion of (cognitive, perceptual, and physical) resources that the driver would need to safely perform the driving task in a given context.*

The more complex the environment, the greater the demands placed on the driver, making the driving task more challenging. Environment complexity is conceptualized from the perspective of the driver and should ideally reflect perceptual and cognitive factors, such as visual masking (*i.e.*, the difficulty in perceiving a target stimulus when it is rapidly followed or spatially overlapped by another stimulus) or joint attention with other road users (*i.e.*, the coordination of attentional focus between the driver and other road users, which plays a critical role in interpreting intentions and ensuring safe interactions) [170, 273].

For the environment complexity, we also define an index, named environment complexity index (ECI), which ranges from $0$ (minimal complexity) to $1$ (maximal complexity).

The environment complexity depends not only on the number of elements present but also on their behavioral relevance [375]. Simply counting objects in the driving scene is not sufficient to estimate the environment complexity index. For instance, a group of pedestrians conversing on the sidewalk may contribute less to complexity than a single pedestrian approaching a crosswalk. Other factors, such as road type, road geometry, and weather conditions, also play a role in shaping the complexity of the driving environment. However, how to accurately compute the environment complexity index in a way that is meaningful for risk assessment remains an open question that current research has largely overlooked. The work presented in Chapter 6 constitutes a first step toward characterizing the environment with the aim of assessing risk by determining the extent to which systems operate outside their intended ODD.

### 7.2.3 Vehicle Engagement

> **Definition 3.** Vehicle engagement *is defined as the proportion of vehicle automation resources that are available to perform the driving task.*

Vehicle engagement reflects the set of driving automation features that can be activated to assist the driver in performing the driving task. It varies depending on the driving environment and the ODD, which defines the limitation of the driving automation system. For instance, some features could only be activated on highways and in clear weather conditions (avoiding adverse conditions like heavy rain). Besides, the maximum achievable vehicle engagement is actually limited by the SAE level of the vehicle.

To quantify this, we define the vehicle engagement index (VEI), which ranges from $0$ (no automation) to $1$ (full automation). We also define its complement, the vehicle engagement deficit (VED) as VED $= 1 -$ VEI.

Vehicle engagement depends on the availability of driving automation features such as (adaptive) cruise control, lane centering assistance, lane keeping assistance, automatic emergency braking, rear automatic braking, pedestrian automatic emergency braking, blind spot intervention, and others. The scale is absolute, with VEI $= 1$ denoting full automation, attainable exclusively by a SAE Level 5 vehicle. Using an absolute rather than a relative scale facilitates interpretation and comparison between vehicles.

### 7.2.4 Risk Assessment

> **Definition 4.** *In the context of our framework,* risk *refers to the likelihood of adverse outcomes, such as collisions, traffic violations, or even near-misses, resulting from the interaction between the driver, the environment, and the vehicle.*

Risk results from the real-time ability (or inability) to safely perform the driving task. It is influenced by the driver involvement, the environment complexity, the vehicle engagement, and by how they interact (*e.g.*, the nature of the cooperation between the human driver and the driving automation system of the vehicle, but also the design of the Human-Machine Interface (HMI)).

The goal of risk assessment is to enable dynamic adjustment of the operational level of automation by activating or deactivating specific driving automation features in a closed-loop fashion, thereby mitigating the computed risk. By continuously adapting the level of automation based on driver involvement, environmental complexity, and vehicle engagement, we aim to support smooth transitions of control between the human driver and the automation system.

Such gradual transitions offer several advantages. First, they allow the driver to progressively adapt their mental model, avoiding abrupt changes in system behavior. Second, they help maintain an appropriate level of driver involvement relative to the demands of the current driving context and associated risk.

However, assessing risk based on the intricate interplay between the driver, the environment, and the vehicle remains a major challenge. This requires a deep understanding of how these three components influence one another. For instance, how specific driving automation features affect driver's state under varying environment complexities. The work presented in Chapter 3 takes a first step in this direction by analyzing the influence of cognitive load and traffic conditions on the use of an Adaptive Cruise Control (ACC), and how its use in turn impacts driving performance.

To begin formalizing this interplay, we first consider a simplified case in which the environment complexity is held constant. This allows us to focus on how the relationship between driver and vehicle resources contributes to risk. Specifically, we examine the two-dimensional space defined by the vehicle engagement deficit (VED) and the driver involvement deficit (DID). We define the cumulative deficit (CD) by CD $=$ VED $+$ DID $- 1$ such that CD $= 1$ when DID $=$ VED $= 1$, CD $= 0$ when DID $=$ VED $= 0.5$, and CD $= -1$ when DID $=$ VED $= 0$.

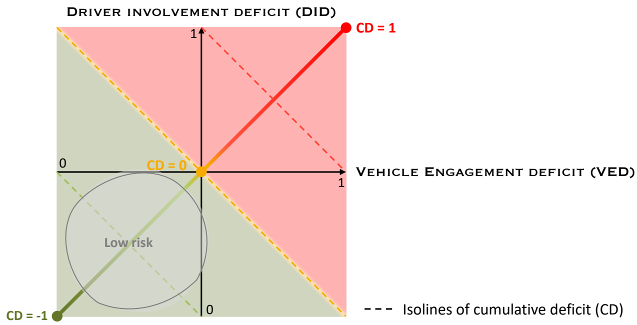



**Figure 7.2: Cumulative deficit in the two-dimensional space defined by VED and DID, for a fixed ECI.** The main diagonal (top-left to bottom-right) and its parallels (dashed lines) are isolines of the cumulative deficit, meaning that CD remains constant along each of these lines. The cumulative deficit ranges from $-1$ to $1$ along the direction of the minor diagonal (solid line running from the bottom-left corner to the top-right corner). The main diagonal delineates two regions: one where the cumulative deficit is negative (green area), indicating redundancy in the combined resources of the driver and the vehicle to perform the driving task; and one where it is positive (red area), indicating that the combination of the driver and vehicle resources is insufficient to fully carry out the driving task. The objective of the *DEV* framework is to manage the risk through adaptive automation to remain in a low-risk area (in gray-tinted green).

Figure 7.2 illustrates the two-dimensional space defined by VED and DID, and shows how the cumulative deficit evolves within this space. By definition, the cumulative deficit ranges from $-1$ to $1$ along the minor diagonal. In contrast, in the direction of the main diagonal, the space is structured by isolines, each corresponding to a constant value of the cumulative deficit. On the main diagonal, the cumulative deficit is zero, meaning that combining the resources of the driver and the vehicle barely allows performing the entire driving task, assuming perfect cooperation between the two agents. The main diagonal separates the space into two regions: one where the cumulative deficit is positive and one where it is negative. In the positive cumulative deficit region, the combined resources of the driver and the vehicle are insufficient to carry out the full driving task. This region is thus inherently associated with high risk and should be avoided by any means. As this region is approached, proactive measures must be taken to prevent accidents. One such measure is for the vehicle to initiate a safety stop, similar to SAE level 4 vehicles, which achieve a minimal risk condition automatically when reaching their ODD limits. In the negative cumulative deficit region, there is redundancy. Combining the available resources of both the driver and the vehicle provides more resources than necessary to perform the driving task. In this region, there is a real interest in selecting the most appropriate operational level of automation to foster effective cooperation between the driver and the vehicle to lower the risk. Below the minor diagonal, where the vehicle engagement deficit is higher than the driver involvement deficit, the driver can perform a higher proportion of the driving task than the driving automation system of the vehicle. Conversely, above the minor diagonal, the driving automation system of the vehicle can perform a higher proportion of the driving task than the driver.

When in this negative cumulative deficit region, effective cooperation between the driver and the automation is expected. However, how this cooperation should be operationalized and which operational level of automation (*i.e.*, which combination of driving automation features) should be selected to minimize risk remains an open question. Several authors (*e.g.*, [71, 373]) suggest that the driver should be utilized at their highest level of involvement. If the driver is in a suitable state and willing to engage in the driving task, they should be allowed to do so. However, when driving automation features are available, they should also be leveraged to reduce risk. Caution is needed, though: increasing the level of automation may negatively impact driver involvement. The operational level of automation can be selected within the range $[0, \text{VEI}]$. This choice should account not only for the driver's current level of involvement but also for the complexity of the driving environment, so that the selected operational level of automation mitigates risk. In low-complexity driving environments, allowing the driver to be less involved and delegating more responsibility to the vehicle (*i.e.*, increasing the operational level of automation) may not significantly increase risk. However, as the complexity of the driving environment rises, greater driver involvement may become necessary.

Finally, it is important to note that assessing risk and adjusting the operational level of automation cannot be done without also considering the mode of cooperation between the human driver and the driving automation system. Cooperation can take the form of traded control (*i.e.*, working on the same task at different times [298]; *a.k.a.* control shifts [337]), shared control (*i.e.*, working on the same task at the same time [298]), for example, through haptic feedback mechanisms [1, 104], or a combination of both. A recent taxonomy for human-vehicle cooperation has been proposed to aid in structuring cooperative control concepts [253]. According to de Winter et al. [59], whether shared control reduces risk depends on the environment complexity. This highlights the importance of integrating considerations about cooperation modes into the risk assessment process of our *DEV* framework. However, how to incorporate different cooperation modes effectively remains an open area for future research.

## 7.3 Discussion and Future Work

Researchers have long studied human-machine interaction. Indeed, over the last two decades, numerous European research projects have investigated the cooperation and interaction between the driver and the vehicle, such as HAVE-it [121, 132], D3COS [55, 205], HFAuto [129], ADAS&ME [2], HADRIAN [107], or MEDIATOR [217]. Yet, many challenges remain, particularly in the automotive domain. Interest in this area continues to grow, driven by the complexity of designing systems that must integrate insights from multiple disciplines. This includes, for example, developing DMSs that meaningfully support risk mitigation or designing HMIs that enable effective cooperation. While a wide variety of studies have emerged from these research efforts and continue to be produced across these domains, building fully functional systems requires a shared understanding and coordinated direction. This is why a framework that captures the big picture is crucial, as it can help guide research efforts and align them toward a common goal.

In this work, we proposed the *DEV* framework for risk-aware adaptive automation of driving, which aims to dynamically and continuously adjust the operational level of automation based on the complex interplay between the driver, the surrounding environment, and the vehicle. This framework is motivated by the observation that the static definition of SAE levels is inadequate for real-time driving, as the level of automation may need to change during a single drive. Moreover, operational levels should be more fine-grained to enable smooth, dynamic transitions, depending, for example, on the driver's level of involvement, and to account for different cooperation modes between the driver and the automation system.

To move beyond a purely theoretical proposal, several key research directions must be pursued. The first step is to refine the definitions and develop methods for estimating the driver involvement deficit, the environment complexity index, and the vehicle engagement index. While these parameters cannot be directly measured, they can be estimated. Developing reliable and interpretable estimation methods for each component is a prerequisite for implementation.

The next step is to empirically establish a model for risk assessment, either through the theoretical development of a mathematical formulation or by leveraging data-driven approaches, such as deep learning. For example, Herzberger et al. [127] presented the concept of confidence horizons, aiming at quickly identifying whether safety-critical transitions of control (takeovers or handovers) should be performed. Then, Herzberger et al. [128] presented the theoretical concept of the diagnostic Take-Over Request (TOR), which predicts risky takeovers based on a driver's response to a TOR. These works are important first steps toward cooperative driving but currently only allow to gaining more reaction time during critical transitions.

Once the risk model is defined, it becomes possible to design, validate, and implement strategies to mitigate that risk (*i.e.*, remain in the gray-tinted green area in Fig. 7.2). Within the *DEV* framework, we propose modulating the operational level of automation by activating or deactivating driving automation features, as the primary strategy for risk mitigation. However, interventions can also target the driver or the environment. Actions on the driver, such as reducing distraction or increasing situation awareness (*e.g.*, [102, 359]), should only be implemented when adjusting the automation level alone cannot adequately compensate for low driver involvement. Drivers should retain the freedom to rest or disengage from the driving task, so long as doing so does not significantly increase the risk.

Actions on the environment to reduce its complexity may include, *e.g.*, improving road infrastructure and design, for instance, following the principles of *Self-Explaining Roads* [320], which aim to encourage safe behavior through intuitive road layouts. However, such interventions can only be implemented occasionally and require substantial time and financial resources.

Finally, once such a system is implemented, its acceptability and usability must be thoroughly investigated. These aspects hinge largely on the design of the HMI, which should support efficient cooperation between the driver and the vehicle. Drivers must clearly understand their role in the human-machine partnership and be able to form accurate mental models of how the system operates and why it adapts in real time. Although recent work has addressed these challenges [64, 353], much remains to be explored.

# 8 Insights and Perspectives of the Thesis

Completing a PhD thesis is a long journey. Few candidates, I believe, can foresee the precise direction their research will take from the outset—and that was certainly not my case. This conclusion looks back on the path followed throughout the thesis, describing what was accomplished at each stage and reflecting on the ideas and events that ultimately shaped its final content and form.

## 8.1 Part I

At the beginning, my initial focus was on driver monitoring and, more precisely, on characterizing the state of drivers from camera images. The field was gaining importance, but it was completely new to me. Therefore, I began by conducting a survey on driver monitoring, which proved to be a valuable way to enter the topic, understand its foundations, and fuel reflection on its future directions.

This work resulted in a publication [115], on which Chapter 2 is based. In this work, we identified five key driver states relevant to driving safety: drowsiness, mental workload, distraction, emotions, and under the influence. In Chapter 2, we focus on mental workload and distraction, and synthesize our findings using a pair of interlocked tables that relate these states to their indicators and the indicators to the sensors providing access to them. The chapter also discusses the role of driver monitoring across the six SAE levels of driving automation.

Many conclusions, insights, and perspectives emerged from this first work. However, two of them had a decisive impact on the rest of my PhD. First, I realized that driver monitoring plays a crucial role in ensuring safety—at all levels of driving automation except the very last one—but that its potential extends far beyond issuing warnings, which was essentially its primary purpose at that time. Second, the survey also revealed that elements from the driver, the vehicle, and the environment could serve as indicators to characterize the driver's state. Hence, all three should ideally be monitored and considered for safety. These two reflections ultimately led to the *DEV* framework introduced in Chapter 7.

After completing the survey, our discussions thus naturally evolved toward broader questions: *How can driver monitoring be integrated with driving automation to mitigate risk?* and *How involved should the driver be in driving automation?* These questions guided the remainder of the thesis and led to the idea of a closed-loop system, in which driver monitoring would continuously inform and adapt the system behavior. Naturally, such a system would have to consider the driver, the vehicle, and the environment, as well as their dynamic interplay.

## 8.2 Part II

The question of this interplay between the driver, the vehicle, and the environment became central. To explore it, we designed a human study to analyze how drivers interact with automation depending on their cognitive state and the complexity of the driving environment. Obviously, it was not possible within a single study to investigate all five identified driver states, all driving automation levels and/or features, and all factors influencing the complexity of the driving environment. We therefore made deliberate choices, selecting one specific element for each of the three components. We investigated, through a controlled driving simulator experiment, how and when drivers activate or deactivate Adaptive Cruise Control (ACC) depending on their cognitive load and traffic conditions. This study, presented in Chapter 3, provided a better understanding of how the driver's cognitive state and the driving environment complexity influence reliance on Adaptive Cruise Control (ACC) and how such reliance affects driving performance.

Yet, even as my interests broadened toward driving automation and the driving environment, I remained convinced that the key to safer driving lies in understanding drivers and the ability to characterize their state. Therefore, I took the opportunity to further analyze the data collected during the simulator study to examine physiological and behavioral indicators of driver cognitive distraction, such as Electrodermal Activity (EDA) and gaze parameters, under varying environment complexity and ACC use. This work, described in Chapter 4, deepened the understanding of the relationships between driver state, automation, and environment.

The studies presented in these two chapters represent important steps toward understanding the interplay between the driver, the vehicle, and the environment. However, much remains to be done. Other elements of the driver state (*e.g.*, drowsiness, visual distraction, emotions), aspects of vehicle automation (*e.g.*, different features or levels), and factors of driving environment complexity (*e.g.*, weather, road types) still need to be analyzed and better understood.

## 8.3 Part III

In parallel, as we soon realized that accounting for the driving environment was essential to the driver-vehicle cooperation, I focused on its analysis from a computer vision perspective—an area well aligned with my research team's expertise. In Chapter 5, I first devoted time to generating a synthetic dataset for the task of semantic segmentation, composed of long video sequences captured from vehicles driving in dynamic environments and featuring both spatial diversity and varying weather conditions. The task of semantic segmentation of driving scenes, which consists of assigning a class label to every pixel in an image, provides valuable knowledge about the content of the driving environment. Furthermore, such data are essential to ensure that perception models embedded in vehicles can adapt and remain robust under dynamic conditions. The generated dataset, called the *Driving Agents in Dynamic Environments (*DADE*)* dataset, was used in [92] to evaluate a new *Multi-Stream Cellular Test-Time Adaptation (MSC-TTA)* method, in which models adapt on the fly to dynamic environments divided into cells.

Then, in Chapter 6, I explored, together with my collaborators, how to characterize the driving environment. Specifically, we focused on predicting the likelihood of different weather conditions from images captured by vehicle-mounted cameras. The ability to characterize the environment not only supports the design of context-aware driving automation but also provides essential information for driver-vehicle cooperation and the allocation of control. It can help determine whether it is safer to let the driver drive, to rely on automation, or, for instance, to enable a cooperative mode between the two.

Note that several recent research projects (*e.g.*, AI-SEE [11], Hi-Drive [130], and EVENTS [75]) aimed at improving perception models in adverse weather conditions to extend the ODD of driving automation systems.

However, the environment remains the most overlooked component in the interplay among the driver, the vehicle, and the environment. For instance, it is still unclear how to quantify the environment complexity to support adaptive automation.

## 8.4 Part IV

Finally, this body of work converged in the *Driver-Environment-Vehicle (*DEV*)* framework, presented in Chapter 7, which integrates the three key components of the driving task into a closed-loop approach for risk-aware adaptive driving automation. This framework addresses the question of how driver and environment monitoring can be used to adapt automation and support safe and efficient cooperation between the human driver and the vehicle automation. It highlights how a better understanding of the interplay between the driver, the vehicle, and the environment, together with risk modeling, can enable more informed decisions on how the driving task can be shared between the driver and the automation.

It is worth noting that the question of *how to navigate the intermediate levels of vehicle automation* is particularly timely. In this regard, the very recent work of van Nes et al. [326] adopts a perspective closely aligned with the one developed in this thesis, advocating for human-centered approaches to automation. They introduce concepts—such as the *Driver/Automation Fitness Plane*, which shares similarities with our Fig. 7.2, and the *mediator approach*, which resonates with the principles underlying our *DEV* framework—designed to foster safer and more intuitive interactions between humans and automated systems. This convergence of ideas across independent research efforts underscores the relevance and timeliness of human-centered adaptive automation as a promising direction for the future of driving.

Of course, much remains to be done to operationalize the *DEV* framework. It opens numerous avenues for future research: from refining driver state estimation and environment characterization to implementing adaptive control strategies in real vehicles. Ultimately, I hope that this work contributes to shaping a future of human-centered, adaptive driving automation, where technology enhances our ability to drive safely.

The road toward safe and cooperative driving automation is still long, but each contribution brings us closer to it. By integrating the driver, the vehicle, and the environment into a unified perspective, I hope this thesis helps pave the way for systems that adapt to humans, rather than the other way around.

# A
# Supplementary material for Chapter 2

Chapter 2 reviews the state of the art in driver monitoring, with a particular focus on existing approaches for characterizing the driver's mental workload and distraction. It synthesizes these approaches to provide a unique, structured, polychotomous view of the diverse techniques used for driver state characterization. This perspective is illustrated through a pair of interlocked tables that relate these states to their indicators and the sensors that can access each of these indicators. The structure of these tables was mainly informed by conclusions drawn from a preliminary analysis of an initial set of references, summarized in Section A.1.

## A.1 More Details on the Survey of Literature on Driver Monitoring

Following the strategy described in Section 2.3 to build an initial set of references, the screening process yielded $56$ references. A summary of each reference, detailing the states, indicators, and sensors considered therein, is provided in Tables A.1 to A.3, respectively. The test conditions are also reported in Table A.3. For clarity, the information is organized into three separate tables.

### A.1.1 Structure and Content of the Table Organizing the Initial Set of References

We used the conclusions drawn in Section 2.3 to design the structure of Tables A.1 to A.3 for organizing the $56$ initial references in a useful way.

The $56$ references are listed, in each table, in the first column, labeled "References", by alphabetical order of first author. Following the first column, we have for each table one of the three key items, namely states (Table A.1), indicators (Table A.2), and sensors (Table A.3). The last column of Table A.3, labeled "Tests", indicates whether the technique or system described in a reference was tested in the laboratory, in real conditions ("in the wild"), or both.

In Table A.1, the $5$ columns next to the references correspond to the $5$ (sub)states of interest. In Tables A.2 and A.3, the indicators and the sensors are divided into $3$ categories corresponding to the $3$ previously-listed items that a Driver Monitoring System (DMS) should ideally monitor, that is, the driver, vehicle, and environment. The column corresponding to the indicators for the driver is further divided into $3$ subcolumns corresponding to the qualifiers "physiological", "behavioral", and "subjective". Some other columns could be further subdivided, such as for "Distraction" in Table A.1, but the table deals with such additional subdivisions in a different way.

In the following, we successively describe the content of the three tables.

#### States

For each of the $56$ papers, we indicate which particular (sub)state(s) it addresses. If a paper addresses drowsiness, we place the checkmark "V" in the corresponding column, and similarly for mental workload. For the three other states, we either use a general "V" or give more specific information, often via an abbreviation. There are four types of distraction, that is, manual, visual, auditory, and cognitive, respectively abbreviated via man, vis, aud, and cog. These types are self-explanatory, but they are addressed later. For emotions, we indicate the type, that is, stress or anger (ang). For under the influence, we also indicate the type; in all cases, it turns out to be alcohol (alc).

As an example, the second paper, by Alluhaibi and Moyaid [14], addresses drowsiness, distraction, and the emotion of anger.

All abbreviations used in Table A.1 and the two other tables are defined in Table A.4.

### Indicators

The indicator(s) used by a paper is (are) indicated in the same way as above, using the abbreviations defined in Table A.4.

### Sensors

The sensor(s) used by a paper is (are) indicated in a similar, but not identical, way. If a sensor is embedded in a mobile device (typically a smartphone), rather than in the vehicle, we add a "*", leading to "cam*"/"mic*" for a camera/microphone of a mobile device. In the vehicle column, "V" indicates that the sensor is integrated in the vehicle, whereas "V*" indicates that it is part of a mobile device. For example, the vehicle speed can be obtained via the Controller Area Network (CAN) system/bus or a mobile device.

**Table A.1: Initial set of references: States (Part 1).** This table presents the state(s) addressed by each reference. For readability, the indicators, sensors, and test conditions are detailed in separate tables. The first column lists, in alphabetical order of the first author, the $56$ references identified in our survey on driver monitoring and related systems (DMSs). The next five columns correspond to the five states identified during the preliminary analysis of the initial set of references.

| | References | Drowsiness | Mental Workload | Distraction | Emotions | Under the Influence |
|---|---|---|---|---|---|---|
| 1 | Ahir and Gohokar [4] | V | | | | |
| 2 | Alluhaibi and Moyaid [14] | V | | V | ang | |
| 3 | Arun et al. [17] | | | vis, cog | | |
| 4 | Balandong et al. [21] | V | | | | |
| 5 | Begum [25] | V | | V | stress | |
| 6 | Chacon-Murguia and Prieto-Resendiz [41] | V | | | | |
| 7 | Chan et al. [42] | V | | | | |
| 8 | Chhabra et al. [45] | V | | V | | alc |
| 9 | Chowdhury et al. [46] | V | | | | |
| 10 | Chung et al. [47] | | | | stress | |
| 11 | Coetzer and Hancke [50] | V | | | | |
| 12 | Dababneh and Gindy [56] | V | | | | |
| 13 | Dahiphale and Rao [57] | V | | V | | |
| 14 | Dong et al. [65] | V | | V | | |
| 15 | El Khatib et al. [70] | V | | man, vis, cog | | |
| 16 | Ghandour et al. [95] | | | man, vis, aud, cog | stress | |
| 17 | Hecht et al. [125] | V | V | V | | |
| 18 | Kang [151] | V | | V | | |
| 19 | Kaplan et al. [152] | V | | V | | |
| 20 | Kaye et al. [155] | V | | | stress | |
| 21 | Khan and Lee [158] | V | | man, vis, aud, cog | | |
| 22 | Kumari and Kumar [171] | V | | | | |
| 23 | Lal and Craig [172] | V | | | | |
| 24 | Laouz et al. [173] | V | | | | |
| 25 | Leonhardt et al. [183] | | | | | |
| 26 | Liu et al. [201] | V | | | | |
| 27 | Marquart et al. [211] | | V | | | |
| 28 | Marina Martinez et al. [210] | | | | ang | |
| 29 | Mashko [213] | V | | | | |
| 30 | Mashru and Gandhi [214] | V | | | | |
| 31 | Melnicuk et al. [218] | V | V | cog | stress, ang | |
| 32 | Mittal et al. [224] | V | | | | |
| 33 | Murugan et al. [229] | V | | | | |
| 34 | Nair et al. [231] | V | | V | | alc |
| 35 | Němcová et al. [239] | V | | | stress | |

**Table A.1: Initial set of references: States (Part 2).** This table presents the state(s) addressed by each reference. For readability, the indicators, sensors, and test conditions are detailed in separate tables. The first column lists, in alphabetical order of the first author, the $56$ references identified in our survey on driver monitoring and related systems (DMSs). The next five columns correspond to the five states identified during the preliminary analysis of the initial set of references.

| | References | Drowsiness | Mental Workload | Distraction | Emotions | Under the Influence |
|---|---|---|---|---|---|---|
| 36 | Ngxande et al. [236] | V | | | | |
| 37 | Oviedo-Trespalacios et al. [244] | | V | V | | |
| 38 | Papantoniou et al. [247] | | V | V | | |
| 39 | Pratama et al. [263] | V | | | | |
| 40 | Ramzan et al. [270] | V | | | | |
| 41 | Sahayadhas et al. [284] | V | | | | |
| 42 | Scott-Parker [293] | | | | stress, ang | |
| 43 | Seth [294] | V | | | | |
| 44 | Shameen et al. [297] | V | | | | |
| 45 | Sigari et al. [300] | V | | | | |
| 46 | Sikander and Anwar [301] | V | | | | |
| 47 | Singh and Kathuria [303] | V | V | V | V | |
| 48 | Subbaiah et al. [312] | V | | | | |
| 49 | Tu et al. [323] | V | | | | |
| 50 | Ukwuoma and Bo [324] | V | | | | |
| 51 | Vilaca et al. [331] | V | | V | | |
| 52 | Vismaya and Saritha [334] | | | V | | |
| 53 | Wang et al. [343] | V | | | | |
| 54 | Welch et al. [354] | | | | stress, ang | |
| 55 | Yusoff et al. [377] | | | vis, cog | | |
| 56 | Zhang et al. [381] | V | | | | |

**Table A.2: Initial set of references: Indicators (Part 1).** This table presents the indicators addressed by each reference. For readability, the states, sensors and test conditions are detailed in separate tables. The first column lists, in alphabetical order of the first author, the $56$ references identified in our survey on driver monitoring and related systems (DMSs). The next columns correspond to the indicators.

| | References | Driver | | | Vehicle | Environment |
|---|---|---|---|---|---|---|
| | | Physiological | Behavioral | Subjective | | |
| 1 | Ahir and Gohokar [4] | HR, brain | gaze, blink, PERCLOS, facial, body | | wheel, lane, speed | |
| 2 | Alluhaibi and Moyaid [14] | | speech | | wheel, lane, brake, speed | |
| 3 | Arun et al. [17] | HR, brain, EDA, pupil | gaze, blink, body | V | wheel, lane, brake, speed | |
| 4 | Balandong et al. [21] | HR, brain | gaze, blink, PERCLOS, body | V | wheel, lane, brake, speed | |
| 5 | Begum [25] | HR, brain | | | | |
| 6 | Chacon-Murguia and Prieto-Resendiz [41] | HR, brain, EDA | gaze, blink, body | | wheel, lane, brake, speed | |
| 7 | Chan et al. [42] | HR, brain | blink, PERCLOS, facial, body | | wheel, brake, speed | |
| 8 | Chhabra et al. [45] | breath | gaze, PERCLOS, facial, body | | wheel | road |
| 9 | Chowdhury et al. [46] | HR, brain, EDA | blink, PERCLOS | | | |
| 10 | Chung et al. [47] | HR, breath, brain, EDA, pupil | speech | V | wheel, lane, brake, speed | |
| 11 | Coetzer and Hancke [50] | brain | gaze, PERCLOS, facial, body | | wheel, lane, speed | |
| 12 | Dababneh and Gindy [56] | brain, EDA, pupil | blink, PERCLOS, body | | wheel, lane, speed | road |
| 13 | Dahiphale and Rao [57] | | gaze, blink, facial, body | | wheel | |
| 14 | Dong et al. [65] | HR, brain, pupil | gaze, blink, PERCLOS, facial, body | V | wheel, lane, speed | road, wea |
| 15 | El Khatib et al. [70] | HR, breath, brain, EDA, pupil | gaze, blink, PERCLOS, facial, body, hands | | wheel, lane, speed | |
| 16 | Ghandour et al. [95] | HR, breath, brain, EDA | gaze, facial, body, speech | V | wheel, brake, speed | |
| 17 | Hecht et al. [125] | HR, brain, EDA, pupil | gaze, blink, PERCLOS, facial, body | V | | |
| 18 | Kang [151] | HR, breath, brain, EDA | gaze, blink, facial, body | | wheel, lane, brake, speed | |
| 19 | Kaplan et al. [152] | HR, brain | gaze, blink, PERCLOS, facial, body, speech | | wheel, lane, brake, speed | |
| 20 | Kaye et al. [155] | HR, breath, brain, EDA | | V | | |

**Table A.2: Initial set of references: Indicators (Part 2).** This table presents the indicators addressed by each reference. For readability, the states, sensors and test conditions are detailed in separate tables. The first column lists, in alphabetical order of the first author, the $56$ references identified in our survey on driver monitoring and related systems (DMSs). The next columns correspond to the indicators.

| | References | Driver | | | Vehicle | Environment |
|---|---|---|---|---|---|---|
| | | Physiological | Behavioral | Subjective | | |
| 21 | Khan and Lee [158] | HR, brain, EDA | gaze, PERCLOS, body | | wheel, lane, brake, speed | |
| 22 | Kumari and Kumar [171] | HR, brain | gaze, blink, PERCLOS, body | V | wheel, lane | |
| 23 | Lal and Craig [172] | HR, brain, EDA | PERCLOS, facial | | | |
| 24 | Laouz et al. [173] | HR, brain, EDA | blink, PERCLOS, facial, body | V | wheel, speed | |
| 25 | Leonhardt et al. [183] | HR, breath | | | | |
| 26 | Liu et al. [201] | HR, brain, pupil | gaze, blink, PERCLOS, body | | wheel, lane, speed | |
| 27 | Marquart et al. [211] | pupil | gaze, blink, PERCLOS | V | | |
| 28 | Marina Martinez et al. [210] | | | | brake, speed | |
| 29 | Mashko [213] | HR, brain, EDA | gaze, blink, body | | wheel, lane, brake, speed | |
| 30 | Mashru and Gandhi [214] | HR, breath | blink, PERCLOS, facial, body | V | wheel, lane | |
| 31 | Melnicuk et al. [218] | HR, brain | blink, PERCLOS, facial | | wheel, brake, speed | road, traf, wea |
| 32 | Mittal et al. [224] | HR, brain, pupil | blink, PERCLOS, body | V | wheel, lane, brake, speed | |
| 33 | Murugan et al. [229] | HR, breath, brain, EDA, pupil | blink, PERCLOS, body | V | wheel, lane, speed | |
| 34 | Nair et al. [231] | | gaze, PERCLOS, facial, body | | lane | |
| 35 | Němcová et al. [239] | HR, breath, brain, EDA | gaze, blink, PERCLOS, facial, body | | wheel, lane, brake, speed | |
| 36 | Ngxande et al. [236] | | blink, PERCLOS, facial, body | | | |
| 37 | Oviedo-Trespalacios et al. [244] | | gaze | | wheel, lane, brake, speed | |
| 38 | Papantoniou et al. [247] | HR, breath, brain | gaze, blink, speech | V | wheel, lane, speed | |
| 39 | Pratama et al. [263] | HR, brain, EDA | gaze, blink, PERCLOS, facial, body, hands | V | wheel, lane | |
| 40 | Ramzan et al. [270] | HR, breath, brain | blink, PERCLOS, facial, body | | wheel, lane, speed | |

**Table A.2: Initial set of references: Indicators (Part 3).** This table presents the indicators addressed by each reference. For readability, the states, sensors and test conditions are detailed in separate tables. The first column lists, in alphabetical order of the first author, the $56$ references identified in our survey on driver monitoring and related systems (DMSs). The next columns correspond to the indicators.

| | References | Driver | | | Vehicle | Environment |
|---|---|---|---|---|---|---|
| | | Physiological | Behavioral | Subjective | | |
| 41 | Sahayadhas et al. [284] | HR, brain, pupil | gaze, blink, PERCLOS, body | V | wheel, lane | |
| 42 | Scott-Parker [293] | HR, brain, EDA | gaze, facial | V | wheel, lane, brake, speed | traf |
| 43 | Seth [294] | | | | | |
| 44 | Shameen et al. [297] | brain | gaze, blink | | | |
| 45 | Sigari et al. [300] | | gaze, blink, PERCLOS, facial, body | | | |
| 46 | Sikander and Anwar [301] | HR, brain, pupil | gaze, blink, PERCLOS, body | V | wheel, lane | |
| 47 | Singh and Kathuria [303] | pupil | gaze, blink, PERCLOS, facial | | wheel, brake, speed | road, traf |
| 48 | Subbaiah et al. [312] | HR, brain, pupil | blink, PERCLOS, facial, body | | | |
| 49 | Tu et al. [323] | HR, brain | blink, PERCLOS, facial, body | | wheel, lane, speed | |
| 50 | Ukwuoma and Bo [324] | HR, breath, brain | blink, PERCLOS, facial, body | | wheel, lane, brake | |
| 51 | Vilaca et al. [331] | brain | gaze, body | | wheel, lane, brake, speed | |
| 52 | Vismaya and Saritha [334] | | gaze, blink, PERCLOS, body | | | |
| 53 | Wang et al. [343] | brain, pupil | gaze, blink, PERCLOS, body | | lane | |
| 54 | Welch et al. [354] | HR, breath, brain, EDA | blink, facial, speech | | wheel, brake, speed | |
| 55 | Yusoff et al. [377] | HR, brain, EDA, pupil | gaze, body | V | lane, speed | |
| 56 | Zhang et al. [381] | HR, brain | gaze, blink, PERCLOS, body | | lane, speed | |

**Table A.3: Initial set of references: Sensors (Part 1).** This table presents the sensors and test conditions addressed by each reference. For readability, the states and indicators are detailed in separate tables. The first column lists, in alphabetical order of the first author, the $56$ references identified in our survey on driver monitoring and related systems (DMSs). The three following columns correspond to the sensors and the last column corresponds to the test conditions.

| | References | Driver | Vehicle | Environment | Tests |
|---|---|---|---|---|---|
| 1 | Ahir and Gohokar [4] | cam, elec | | ext cam | real, sim |
| 2 | Alluhaibi and Moyaid [14] | cam*, mic* | V* | | |
| 3 | Arun et al. [17] | cam, wea d, eye t | V | | sim |
| 4 | Balandong et al. [21] | elec | | | sim |
| 5 | Begum [25] | seat, ste w, saf b, wea d | | | real, sim |
| 6 | Chacon-Murguia and Prieto-Resendiz [41] | ste w, cam | | radar | real |
| 7 | Chan et al. [42] | cam*, mic* | | | real |
| 8 | Chhabra et al. [45] | seat, cam*, mic* | V* | | real, sim |
| 9 | Chowdhury et al. [46] | | | | sim |
| 10 | Chung et al. [47] | cam, wea d | V | | real, sim |
| 11 | Coetzer and Hancke [50] | cam | V | | real, sim |
| 12 | Dababneh and Gindy [56] | cam, wea d | | radar | real, sim |
| 13 | Dahiphale and Rao [57] | cam | | | real |
| 14 | Dong et al. [65] | cam | V | | real |
| 15 | El Khatib et al. [70] | cam | V* | ext cam, radar | real, sim |
| 16 | Ghandour et al. [95] | cam, wea d | | | real, sim |
| 17 | Hecht et al. [125] | elec, eye t | | | real, sim |
| 18 | Kang [151] | seat, ste w, cam | V | | real, sim |
| 19 | Kaplan et al. [152] | ste w, cam*, mic*, wea d | V | | real, sim |
| 20 | Kaye et al. [155] | | | | real, sim |
| 21 | Khan and Lee [158] | wea d | | | real |
| 22 | Kumari and Kumar [171] | cam | | | |
| 23 | Lal and Craig [172] | cam | | | sim |
| 24 | Laouz et al. [173] | seat, cam, wea d | | ext cam | real |
| 25 | Leonhardt et al. [183] | seat, ste w, saf b, cam | | | real |
| 26 | Liu et al. [201] | cam | V | | real |
| 27 | Marquart et al. [211] | eye t | | | real, sim |
| 28 | Marina Martinez et al. [210] | | V* | | |
| 29 | Mashko [213] | cam, wea d | V | ext cam, radar | real, sim |
| 30 | Mashru and Gandhi [214] | seat, ste w, cam, wea d | | | sim |
| 31 | Melnicuk et al. [218] | seat, ste w, saf b, cam*, wea d | V* | | real |
| 32 | Mittal et al. [224] | cam, elec | V | ext cam | real |
| 33 | Murugan et al. [229] | cam, elec | V | | sim |
| 34 | Nair et al. [231] | seat, cam* | V | radar | |
| 35 | Němcová et al. [239] | seat, ste w, cam, wea d, eye t | V | | real, sim |

**Table A.3: Initial set of references: Sensors (Part 2).** This table presents the sensors and test conditions addressed by each reference. For readability, the states and indicators are detailed in separate tables. The first column lists, in alphabetical order of the first author, the $56$ references identified in our survey on driver monitoring and related systems (DMSs). The three following columns correspond to the sensors and the last column corresponds to the test conditions.

| | References | Driver | Vehicle | Environment | Tests |
|---|---|---|---|---|---|
| 36 | Ngxande et al. [236] | cam | | | |
| 37 | Oviedo-Trespalacios et al. [244] | | | | real, sim |
| 38 | Papantoniou et al. [247] | cam | | ext cam, radar | real, sim |
| 39 | Pratama et al. [263] | cam, wea d, elec | | ext cam | real, sim |
| 40 | Ramzan et al. [270] | cam, wea d, elec | V | | real, sim |
| 41 | Sahayadhas et al. [284] | seat, ste w, cam, wea d | V | | real, sim |
| 42 | Scott-Parker [293] | eye t | | ext cam | real, sim |
| 43 | Seth [294] | cam | V | | real |
| 44 | Shameen et al. [297] | elec | | | sim |
| 45 | Sigari et al. [300] | cam | | | real |
| 46 | Sikander and Anwar [301] | seat, ste w, saf b, cam, wea d, elec | | | real |
| 47 | Singh and Kathuria [303] | cam, wea d | V | ext cam, radar | real |
| 48 | Subbaiah et al. [312] | cam | | | real, sim |
| 49 | Tu et al. [323] | cam*, wea d, elec | V | | real, sim |
| 50 | Ukwuoma and Bo [324] | cam, wea d, elec | | | real |
| 51 | Vilaca et al. [331] | cam, mic | V | ext cam | |
| 52 | Vismaya and Saritha [334] | cam, eye t | | | real, sim |
| 53 | Wang et al. [343] | cam, wea d | | | real, sim |
| 54 | Welch et al. [354] | seat, ste w, cam, wea d | V | | real, sim |
| 55 | Yusoff et al. [377] | eye t | | | |
| 56 | Zhang et al. [381] | cam | | ext cam | real, sim |

**Table A.4: List of abbreviations.** The table defines the abbreviations used in Table A.1, Table A.2, and Table A.3. They are listed in alphabetical order and organized according to the different columns.

| States | | Indicators | | Sensors | | Tests | |
|---|---|---|---|---|---|---|---|
| ***Distraction*** | | ***Driver*** | | ***Driver*** | | real | real conditions |
| aud | auditory | blink | blink dynamics | cam | camera | sim | simulated conditions |
| cog | cognitive | body | body posture | elec | electrode(s) | | |
| man | manual | brain | brain activity | eye t | eye tracker | | |
| vis | visual | breath | breathing activity | mic | microphone | | |
| ***Emotions*** | | EDA | electrodermal activity | saf b | safety belt | | |
| ang | anger | facial | facial expressions | ste w | steering wheel | | |
| ***Under the Influence*** | | hands | hands parameters | ***Environment*** | | | |
| alc | alcohol | HR | heart rate/activity | ext cam | external camera | | |
| | | pupil | pupil diameter | | | | |
| | | ***Vehicle*** | | | | | |
| | | brake | braking behavior | | | | |
| | | lane | lane discipline | | | | |
| | | wheel | wheel steering | | | | |
| | | ***Environment*** | | | | | |
| | | road | road geometry | | | | |
| | | traf | traffic density | | | | |
| | | wea | weather | | | | |

# B
# Supplementary material for Chapter 3

Chapter 3 presents a simulator study aiming at examining whether drivers' cognitive state and the complexity of the driving environment influence reliance on Adaptive Cruise Control (ACC), and whether such reliance, in turn, affects driving performance. During this study, participants completed two types of questionnaires: a pre-test questionnaire at the beginning and a feedback questionnaire after each drive in the simulator. Both questionnaires are respectively provided in Sections B.1 and B.2.

## B.1 Pre-Test Questionnaire

The pre-test questionnaire aimed at collecting demographic and study-specific data. To gain insights on participants' experience and habits, we asked them to rate their familiarity with, use of, and attitude toward ACC, driving automation features and systems on five-point Likert scales.

The pre-test questionnaire is available on the next pages.

## Pre-Test Questionnaire

Participant ID:

### Demographic data

What is your age?

What is your gender?

How long have you had your driver's license?

What is your manual preference (motor laterality)?

| RIGHT-HANDED | LEFT-HANDED | AMBIDEXTROUS[1] | AMBIMANUAL[2] |
|---|---|---|---|
| O | O | O | O |

### Study-specific data

Do you sometimes confuse left and right?

| YES | NO |
|---|---|
| O | O |

How often do you drive?

| EVERY DAY | SEVERAL TIMES A WEEK | SEVERAL TIMES A MONTH |
|---|---|---|
| O | O | O |

How many kilometers do you drive on average per year?

| LESS THAN 5,000 KM/YEAR | BETWEEN 5,000 AND 10,000 KM/YEAR | MORE THAN 10,000 KM/YEAR |
|---|---|---|
| O | O | O |

What type(s) of vehicle(s) do you usually drive?

| GASOLINE | DIESEL | ELECTRIC | HYBRID |
|---|---|---|---|
| O | O | O | O |

| MANUAL TRANSMISSION | AUTOMATIC TRANSMISSION |
|---|---|
| O | O |

[1] An ambidextrous person is someone who can use either hand interchangeably for all activities and is equally skilled with both hands.
[2] An ambimanual person is someone who performs certain activities with one hand and others with the other (e.g., writing with the right hand and playing tennis with the left hand).

How many different vehicles do you regularly drive?

| ONLY ONE | TWO | MORE THAN TWO |
|---|---|---|
| O | O | O |

Before participating in the study, did you know what cruise control is?

| YES | NO |
|---|---|
| O | O |

Have you ever driven a vehicle equipped with cruise control?

| NO, NEVER. | YES, BUT ONLY A STANDARD CRUISE CONTROL SYSTEM. | YES, AN ADAPTIVE OR INTELLIGENT CRUISE CONTROL SYSTEM. |
|---|---|---|
| O | O | O |

Do you regularly drive a vehicle equipped with cruise control?

| NEVER | RARELY | SEVERAL TIMES A MONTH | SEVERAL TIMES A WEEK | EVERY DAY |
|---|---|---|---|---|
| O | O | O | O | O |

Do you regularly drive a vehicle equipped with adaptive cruise control?

| NEVER | RARELY | SEVERAL TIMES A MONTH | SEVERAL TIMES A WEEK | EVERY DAY |
|---|---|---|---|---|
| O | O | O | O | O |

How long have you been driving a vehicle equipped with cruise control?

| LESS THAN A YEAR | MORE THAN A YEAR | NOT APPLICABLE |
|---|---|---|
| O | O | O |

Do you regularly activate driving assistance features (e.g., cruise control, lane-keeping assist)?

| NOT AT ALL | RATHER NOT | NEITHER YES NOR NO | RATHER YES | COMPLETELY |
|---|---|---|---|---|
| O | O | O | O | O |

For what reason(s)?

Are you in favor of driving assistance features (e.g., cruise control, lane-keeping assist)?

| NOT AT ALL | RATHER NOT | NEITHER YES NOR NO | RATHER YES | COMPLETELY |
|---|---|---|---|---|
| O | O | O | O | O |

For what reason(s)?

Are you in favor of (semi-)autonomous vehicles?

| NOT AT ALL | RATHER NOT | NEITHER YES NOR NO | RATHER YES | COMPLETELY |
|---|---|---|---|---|
| O | O | O | O | O |

For what reason(s)?

| |
|---|
| |

Is there any other information you would like to share?

| |
|---|
| |
| |
| |
| |

## B.2 Feedback Questionnaire

After each scenario, participants completed a feedback questionnaire, collecting, *e.g.*, information about the reasons for engaging or not engaging the ACC.

The feedback questionnaire is available on the next pages.

**Feedback Questionnaire**

Participant ID:
Session:

**During this session…**

| | STRONGLY DISAGREE | DISAGREE | SOMEWHAT DISAGREE | NEITHER AGREE NOR DISAGREE | SOMEWHAT AGREE | AGREE | STRONGLY AGREE |
|---|---|---|---|---|---|---|---|
| **Driving safely was complicated.** | O | O | O | O | O | O | O |
| **I was distracted.** | O | O | O | O | O | O | O |
| **I was tired.** | O | O | O | O | O | O | O |
| **The driving environment was complex (traffic, roadwork, etc.).** | O | O | O | O | O | O | O |

**When I activated driving assistance features during this session, it was because…**

| | STRONGLY DISAGREE | DISAGREE | SOMEWHAT DISAGREE | NEITHER AGREE NOR DISAGREE | SOMEWHAT AGREE | AGREE | STRONGLY AGREE | I DON'T KNOW |
|---|---|---|---|---|---|---|---|---|
| **The traffic conditions allowed it.** | O | O | O | O | O | O | O | O |
| **It helped me manage all the driving tasks.** | O | O | O | O | O | O | O | O |
| **I was too tired.** | O | O | O | O | O | O | O | O |
| **It helped me comply with speed limits.** | O | O | O | O | O | O | O | O |
| **Other reason(s).** | O | O | O | O | O | O | O | O |

**If other, please specify:**

**When I did not activate driving assistance features during this session, it was because…**

| | STRONGLY DISAGREE | DISAGREE | SOMEWHAT DISAGREE | NEITHER AGREE NOR DISAGREE | SOMEWHAT AGREE | AGREE | STRONGLY AGREE | I DON'T KNOW |
|---|---|---|---|---|---|---|---|---|
| The traffic conditions did not allow it. | O | O | O | O | O | O | O | O |
| My condition allowed me to drive without assistance. | O | O | O | O | O | O | O | O |
| I did not feel like activating driving assistance. | O | O | O | O | O | O | O | O |
| I wanted to stay active /engaged. | O | O | O | O | O | O | O | O |
| I did not think about it. | O | O | O | O | O | O | O | O |
| I was too busy to activate it. | O | O | O | O | O | O | O | O |
| The driving assistance did not behave as expected. | O | O | O | O | O | O | O | O |
| Other reason(s). | O | O | O | O | O | O | O | O |

**If other, please specify:**

**When I deactivated driving assistance features during this session, it was because…**

| | STRONGLY DISAGREE | DISAGREE | SOMEWHAT DISAGREE | NEITHER AGREE NOR DISAGREE | SOMEWHAT AGREE | AGREE | STRONGLY AGREE | I DON'T KNOW |
|---|---|---|---|---|---|---|---|---|
| The traffic conditions required it. | O | O | O | O | O | O | O | O |
| I no longer needed assistance. | O | O | O | O | O | O | O | O |
| The driving assistance did not behave as expected. | O | O | O | O | O | O | O | O |
| Other reason(s). | O | O | O | O | O | O | O | O |

**If other, please specify:**

**Compared to the previous session…**

| | MUCH EASIER | EASIER | NEITHER EASIER NOR HARDER | HARDER | MUCH HARDER |
|---|---|---|---|---|---|
| How was driving in the simulator? | O | O | O | O | O |
| How was using the driving assistance features? | O | O | O | O | O |

# C
# Supplementary material for Chapter 5

Chapter 5 presents a new multi-stream large-scale synthetic semantic segmentation dataset, called *DADE*, motivated by a novel Multi-Stream Cellular Test-Time Adaptation (MSC-TTA) setup in which models adapt on the fly to a dynamic environment divided into cells. *DADE* was used to evaluate a real-time adaptive student-teacher method that leverages the multiple streams available in each cell, obtained by dividing the environment based on location and weather conditions, to quickly adapt to changing data distributions. Additional details regarding *DADE* (Section C.1) and the experiments (Section C.2) are provided hereafter.

## C.1 More Details on *DADE*

### C.1.1 Ground-Truth Labels

Table C.1 provides the complete comparison between the semantic segmentation classes of the Cityscapes dataset, the CARLA simulator, and our *DADE* dataset. The Cityscapes dataset contains $33$ different semantic classes, CARLA $29$, and *DADE* $26$. Our *DADE* dataset takes the intersection of the class definition between Cityscapes and CARLA. The classes not included in the intersection are projected to the “un-

**Table C.1: Comparison of class definition** between Cityscapes [52], CARLA [66], and our *DADE* dataset.

| Cityscapes | | | CARLA | DADE | | |
|---|---|---|---|---|---|---|
| **name** | **training** | **evaluation** | **name** | **name** | **training** | **evaluation** |
| unlabeled | | | unlabeled | unlabeled | | |
| static | | | static | static | | |
| dynamic | | | dynamic | dynamic | | |
| ground | | | ground | ground | | |
| road | ✓ | ✓ | road | road | ✓ | ✓ |
| sidewalk | ✓ | ✓ | sidewalk | sidewalk | ✓ | ✓ |
| rail track | | | rail track | rail track | | |
| building | ✓ | ✓ | building | building | ✓ | ✓ |
| wall | ✓ | ✓ | wall | wall | ✓ | ✓ |
| fence | ✓ | ✓ | fence | fence | ✓ | ✓ |
| guard rail | | | guard rail | guard rail | | |
| bridge | | | bridge | bridge | | |
| pole | ✓ | ✓ | pole | pole | ✓ | ✓ |
| traffic light | ✓ | ✓ | traffic light | traffic light | ✓ | ✓ |
| traffic sign | ✓ | ✓ | traffic sign | traffic sign | ✓ | ✓ |
| vegetation | ✓ | ✓ | vegetation | vegetation | ✓ | ✓ |
| terrain | ✓ | ✓ | terrain | terrain | ✓ | ✓ |
| sky | ✓ | ✓ | sky | sky | ✓ | ✓ |
| person | ✓ | ✓ | person | person | ✓ | ✓ |
| rider | ✓ | ✓ | rider | rider | ✓ | ✓ |
| car | ✓ | ✓ | car | car | ✓ | ✓ |
| truck | ✓ | ✓ | truck | truck | ✓ | ✓ |
| bus | ✓ | ✓ | bus | bus | ✓ | ✓ |
| motorcycle | ✓ | ✓ | motorcycle | motorcycle | ✓ | ✓ |
| bicycle | ✓ | ✓ | bicycle | bicycle | ✓ | ✓ |
| ego vehicle | | | ego vehicle | ego vehicle | | |
| rectification border | | | other | | | |
| out of roi | | | road line | | | |
| parking | | | water | | | |
| tunnel | | | | | | |
| caravan | | | | | | |
| trailer | | | | | | |
| train | | | | | | |

labeled" class, except for "road line" which is projected to "road". The classes used in training and evaluation for *DADE* are the same as the ones of Cityscapes.

### C.1.2 Sequence Lengths

In Fig. C.1a, we show the distribution of the sequence length for the $100$ sequences of *DADE-static*. As can be seen, our dataset contains a lot of short and long sequences, with an average sequence length of $45$ minutes. Similarly, Fig. C.1b shows the distribution of the sequence length for the $300$ sequences of *DADE-dynamic*. It can be noted that the distribution follows the same trend as for *DADE-static*, with a similar average sequence length of $40$ minutes.

## C.2 More Details on the Experiments

### C.2.1 Experimental Settings

#### Environment Division

We consider six scenarios based on the division of the environment into cells.

1. The *Baseline* scenarios correspond to multiple independent streams on which independent agents adapt ($C = N$), *i.e.*, ARTHuS [49] and Houyon et al. [134].

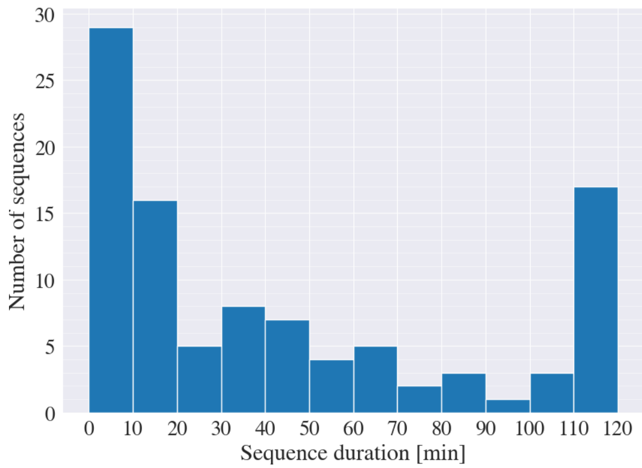
2. The *Common* scenario aggregates the multiple data streams into a single stream, on which one common model adapts ($C = 1$).

3. The *Spatial* scenario leverages the different locations of our dataset to split the environment into cells ($C = 7$).

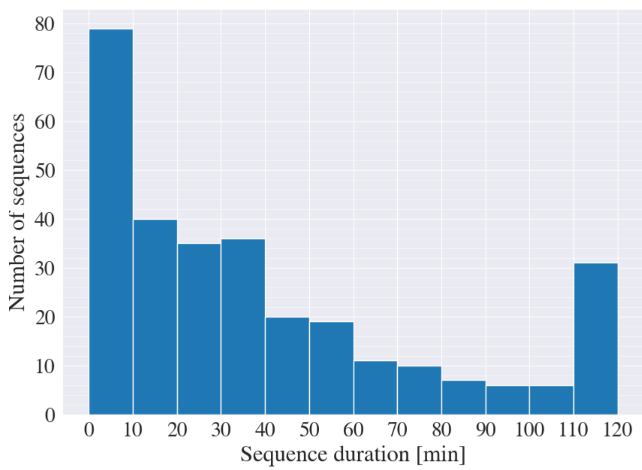
4. The *Weather* scenario temporally divides the environment based on the weather ($C = 3$: clear, rainy, foggy).

5. The *Daylight* scenarios temporally divide the environment based on the time period ($C = 2$: day, night).

6. The *Specific* scenario considers each combination of location, weather condition, and time period ($C = 42$).




**(a)** *DADE-static*




**(b)** *DADE-dynamic*

**Figure C.1: Distribution of sequence lengths.** The (a) *DADE-static* and (b) *DADE-dynamic* datasets have respectively an average sequence length of $45$ and $40$ minutes, with durations ranging from a few minutes to two hours.

#### Pretraining

We choose the same model architecture for all agent models $f_{a_n}$ and cell-specific student models $\mathcal{S}_c$. Following the work of Cioppa et al. [49], we select TinyNet: a lightweight semantic segmentation model operating in real time. The pretraining set is divided into a training set and a validation set using a $90$-$10\%$ split. For each scenario, we evaluate $3$ pretraining techniques. The *General* pretraining consists in training the student model on all samples of the training set, irrespective of the division into cells. The *Cell* pretraining considers a separate model for each cell $c$, trained on cell-specific samples. Finally, *Scratch* assigns random weights (*i.e.*, no pretraining). The models are pretrained with a learning rate of $10^{-4}$ using the Adam optimizer, and the best performing model over the validation set is selected. The number of epochs is set to $3$ for the *General* pretraining and scaled for each cell for the *Cell* pretraining to match the number of backward passes and ensure a fair comparison.

#### Testing

For a given scenario and pretraining procedure, we compare the online performance (*i.e.*, our adaptive method) with the offline performance (*i.e.*, a frozen pretrained model). We choose the teacher model as a frozen state-of-the-art SegFormer [362] model trained on Cityscapes [52] that produces pseudo labels at a rate of $r_{\mathcal{T}} = 1/3\,[\text{Hz}]$. The replay buffers are chosen as FIFO buffers with a size $R = 100$, updated at the same rate $r_{\mathcal{T}}$. Finally, the cell-specific student models are trained online at a rate $r_{\mathcal{S}} = 1/30\,[\text{Hz}]$, with a learning rate of $10^{-4}$, batch size of $25$ with the Adam optimizer, and the cross-entropy loss. The model is only trained if the buffer contains new samples to prevent overfitting.

For the online evaluation, we aggregate the confusion matrices over a sliding window of $30\,[\text{s}]$ (imminent performance) for every agent in every cell and compute the mean-Intersection-over-Union (mIoU) as defined in Houyon et al. [134]. Additionally, we propose to evaluate the current model $5$ minutes in the future (future performance) to assess the capacity of the model to generalize to future samples. Finally, we also compute the overall mIoU for the entire test set ($3$ hours) and for the last hour to assess the long-term performance. As an upper bound, we also evaluate our method in a Multi-Stream Cellular Online Learning (MSC-OL) setup by replacing the teacher pseudo labels with the true ground-truth labels. We also compare our approach to the best method proposed in Houyon et al. [134], which is equivalent to our *Baseline* with a Maximal Interfered Retrieval (MIR) buffer, and report the offline performances of the frozen teacher and student models, both trained on Cityscapes [52].

### C.2.2 Results

#### Quantitative Performances

Table C.2 shows the mean performance of the fleet in the different settings on *DADE-static*. We observe that the baseline setups [49, 134] are outperformed by our method for every scenario and pretraining, highlighting the benefits of using multiple streams when adapting the models. For no pretraining (*Scratch*), the *Spatial* division of the environment leads to the best results, indicating that leveraging cellular information improves the models. For *General* pretraining, the adapted *Common* and *Spatial* scenarios show better performance than the frozen pretrained model, highlighting the benefits of adapting the model online. In the MSC-TTA setup, the *Cell* pretraining outperforms the *General* one, while it is the opposite in the MSC-OL setup, indicating that clean generic labels compensate for cell-specific ones.

We also provide the mean performance on *DADE-dynamic* in Table C.3. As can be seen, our method still outperforms the baselines. Again, from *Scratch*, the *Spatial* scenario brings the best results, followed by the *Common* scenario. However, temporal divisions such as *Weather*, *Daylight*, and *Specific* lead to lower performances. While *DADE* includes at least one vehicle in almost every location over time, the same weather and daylight are applied to all locations simultaneously, leading to discontinuities in the availability of samples for time-based cells. This temporarily stops the adaptation and slows down model convergence. Longer sequences would allow the models to better explore those cells. Finally, the *Cell* pretraining shows the best overall performance for the MSC-OL and MSC-TTA setups, showing the advantage of dividing the environment into cells.

**Table C.2: Mean IoU performance on *DADE-static* weather.** (a) The MSC-OL setup leverages the CARLA segmentation masks as pseudo labels, while (b) the MSC-TTA setup leverages pseudo labels from the teacher model. We compare several pretraining scenarios and adaptive (✓) versus frozen (❄) models. For each pretraining, the best score is shown in **bold** and the second is underlined.

**(a)** MSC-OL setup

| Pretraining | Scenario | Adapt | mIoU imminent 3 hours | mIoU imminent Last hour | mIoU future 3 hours | mIoU future Last hour |
|---|---|---|---|---|---|---|
| Cityscapes [52] | Student | ❄ | .214 | .218 | .214 | .218 |
| | Teacher | ❄ | .668 | .671 | .668 | .671 |
| *Scratch* | *Baseline* [49] | ✓ | .223 | .249 | .208 | .231 |
| | *Baseline*+MIR [134] | ✓ | .173 | .194 | .164 | .188 |
| | *Common* | ✓ | <u>.338</u> | <u>.483</u> | <u>.316</u> | <u>.461</u> |
| | *Spatial* | ✓ | **.353** | **.513** | **.328** | **.485** |
| *General* | *Baseline* [49] | ✓ | .435 | .442 | .415 | .446 |
| | *Baseline*+MIR [134] | ✓ | .650 | .656 | .614 | .626 |
| | *Common* | ✓ | **.702** | <u>.696</u> | **.673** | <u>.692</u> |
| | *Spatial* | ✓ | <u>.700</u> | **.701** | <u>.660</u> | **.701** |
| | *Common* | ❄ | .650 | .658 | .650 | .658 |
| *Cell* | *Spatial* | ✓ | **.658** | **.681** | <u>.597</u> | **.682** |
| | *Spatial* | ❄ | <u>.634</u> | <u>.660</u> | **.634** | <u>.660</u> |

**(b)** MSC-TTA setup

| Pretraining | Scenario | Adapt | mIoU imminent 3 hours | mIoU imminent Last hour | mIoU future 3 hours | mIoU future Last hour |
|---|---|---|---|---|---|---|
| Cityscapes [52] | Student | ❄ | .214 | .218 | .214 | .218 |
| | Teacher | ❄ | .668 | .671 | .668 | .671 |
| *Scratch* | *Baseline* [49] | ✓ | .274 | .309 | .244 | .285 |
| | *Baseline*+MIR [134] | ✓ | .181 | .201 | .171 | .195 |
| | *Common* | ✓ | <u>.340</u> | <u>.363</u> | <u>.327</u> | <u>.373</u> |
| | *Spatial* | ✓ | **.368** | **.440** | **.351** | **.413** |
| *General* | *Baseline* [49] | ✓ | .422 | .442 | .397 | .425 |
| | *Baseline*+MIR [134] | ✓ | .417 | .432 | .401 | .423 |
| | *Common* | ✓ | **.474** | **.517** | <u>.461</u> | <u>.501</u> |
| | *Spatial* | ✓ | <u>.470</u> | **.517** | **.462** | **.505** |
| | *Common* | ❄ | .454 | <u>.450</u> | .454 | .450 |
| *Cell* | *Spatial* | ✓ | **.552** | <u>.567</u> | <u>.522</u> | <u>.556</u> |
| | *Spatial* | ❄ | <u>.544</u> | **.572** | **.544** | **.572** |

**Table C.3: Mean IoU performance on *DADE-dynamic* weather (Part 1).** (a) The MSC-OL setup leverages the CARLA segmentation masks as pseudo labels, while (b) the MSC-TTA setup leverages pseudo labels from the teacher model. We compare several pretraining scenarios and adaptive (✓) versus frozen (❄) models. For each pretraining, the best score is shown in **bold** and the second is underlined.

**(a)** MSC-OL setup

| | | | **mIoU imminent** | | **mIoU future** | |
|---|---|---|---|---|---|---|
| **Pretraining** | **Scenario** | **Adapt** | 3 hours | Last hour | 3 hours | Last hour |
| Cityscapes [52] | Student | ❄ | .159 | .130 | .159 | .130 |
| | Teacher | ❄ | .611 | .542 | .611 | .542 |
| *Scratch* | *Baseline* [49] | ✓ | .204 | .197 | .167 | .167 |
| | *Baseline*+MIR [134] | ✓ | .144 | .137 | .125 | .118 |
| | *Common* | ✓ | <u>.278</u> | <u>.352</u> | <u>.249</u> | <u>.323</u> |
| | *Spatial* | ✓ | **.307** | **.397** | **.269** | **.358** |
| | *Weather* | ✓ | .226 | .295 | .199 | .279 |
| | *Daylight* | ✓ | .245 | .279 | .176 | .259 |
| | *Specific* | ✓ | .22 | .204 | .203 | .187 |
| *General* | *Baseline* [49] | ✓ | .581 | .546 | .502 | .502 |
| | *Baseline*+MIR [134] | ✓ | .567 | .531 | .527 | .480 |
| | *Common* | ✓ | <u>.644</u> | .595 | <u>.613</u> | .565 |
| | *Spatial* | ✓ | **.654** | **.622** | .606 | **.589** |
| | *Weather* | ✓ | .641 | .586 | .611 | .562 |
| | *Daylight* | ✓ | .636 | <u>.603</u> | .572 | <u>.585</u> |
| | *Specific* | ✓ | .632 | .602 | .596 | .559 |
| | *Common* | ❄ | .618 | .581 | **.618** | .581 |
| *Cell* | *Spatial* | ✓ | **.662** | **.642** | .609 | <u>.590</u> |
| | *Weather* | ✓ | .634 | .580 | .607 | .551 |
| | *Daylight* | ✓ | <u>.645</u> | .592 | <u>.620</u> | .577 |
| | *Specific* | ✓ | .612 | .582 | .589 | .554 |
| | *Spatial* | ❄ | .642 | <u>.606</u> | **.642** | **.606** |
| | *Weather* | ❄ | .565 | .528 | .565 | .528 |
| | *Daylight* | ❄ | .563 | .485 | .563 | .485 |
| | *Specific* | ❄ | .447 | .400 | .447 | .400 |

**Table C.3: Mean IoU performance on *DADE-dynamic* weather (Part 2).** (a) The MSC-OL setup leverages the CARLA segmentation masks as pseudo labels, while (b) the MSC-TTA setup leverages pseudo labels from the teacher model. We compare several pretraining scenarios and adaptive (✓) versus frozen (❄) models. For each pretraining, the best score is shown in **bold** and the second is <u>underlined</u>.

**(b)** MSC-TTA setup

| Pretraining | Scenario | Adapt | mIoU imminent 3 hours | mIoU imminent Last hour | mIoU future 3 hours | mIoU future Last hour |
|---|---|---|---|---|---|---|
| Cityscapes [52] | Student | ❄ | .159 | .130 | .159 | .130 |
| | Teacher | ❄ | .611 | .542 | .611 | .542 |
| *Scratch* | *Baseline* [49] | ✓ | .212 | .190 | .173 | .173 |
| | *Baseline*+MIR [134] | ✓ | .147 | .133 | .129 | .110 |
| | *Common* | ✓ | <u>.278</u> | <u>.257</u> | <u>.253</u> | <u>.243</u> |
| | *Spatial* | ✓ | **.312** | **.300** | **.278** | **.276** |
| | *Weather* | ✓ | .227 | .216 | .202 | .197 |
| | *Daylight* | ✓ | .182 | .198 | .150 | .184 |
| | *Specific* | ✓ | .233 | .186 | .218 | .166 |
| *General* | *Baseline* [49] | ✓ | .471 | .406 | .409 | <u>.409</u> |
| | *Baseline*+MIR [134] | ✓ | .455 | .386 | .427 | .347 |
| | *Common* | ✓ | .506 | .427 | <u>.483</u> | .405 |
| | *Spatial* | ✓ | **.516** | **.442** | .473 | .405 |
| | *Weather* | ✓ | <u>.507</u> | .429 | **.484** | .408 |
| | *Daylight* | ✓ | .498 | .430 | .477 | **.413** |
| | *Specific* | ✓ | .500 | <u>.437</u> | .471 | .393 |
| | *Common* | ❄ | .476 | .403 | .476 | .403 |
| *Cell* | *Spatial* | ✓ | **.527** | **.461** | .484 | **.423** |
| | *Weather* | ✓ | <u>.509</u> | .427 | .483 | .409 |
| | *Daylight* | ✓ | .507 | .432 | **.488** | <u>.415</u> |
| | *Specific* | ✓ | .500 | <u>.438</u> | <u>.485</u> | .412 |
| | *Spatial* | ❄ | .488 | .409 | **.488** | .409 |
| | *Weather* | ❄ | .443 | .384 | .443 | .384 |
| | *Daylight* | ❄ | .421 | .362 | .421 | .362 |
| | *Specific* | ❄ | .349 | .298 | .349 | .298 |

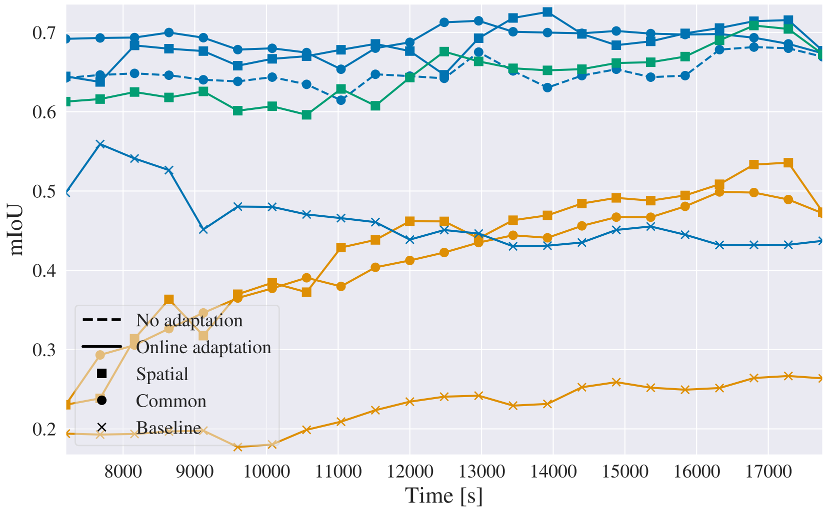
### Evolution of the Fleet Performances

The evolution of the fleet performance over time is shown in Fig. C.2. For visualization purposes, we aggregate the confusion matrices in sliding windows of $8$ minutes to compute the mIoU. Regarding the static weather (top row), the *Baseline* is outperformed by all settings of our method. Interestingly, even if the baseline starts from pretrained weights and our method from scratch, we outperform the baseline in the MSC-OL setup and reach similar performances in the MSC-TTA setup. Additionally, *Cell* pretraining with the *Spatial* scenario reaches the best performance in the MSC-TTA setup for the whole duration, keeping steady performance. This is crucial for autonomous vehicles that need to operate similarly in all conditions.

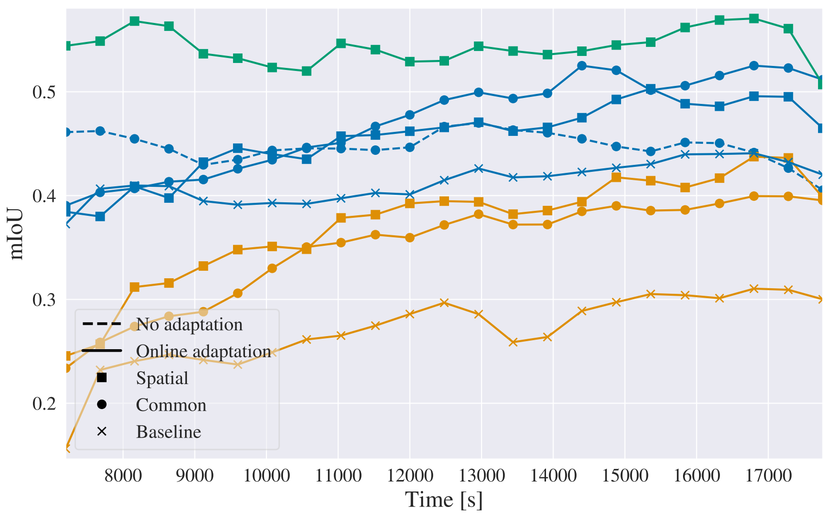
We also show the performance for *DADE-dynamic* in Fig. C.2 (bottom row). For both the MSC-OL and MSC-TTA setups, the *Daylight* scenario with *Cell* pretraining produces the best performance before nightfall, after which it drops while other scenarios, such as *Spatial* with *Cell* pretraining, become better options. This is because the night models are not updated before nightfall, while the location-based models are constantly updated during day, dusk, and night. Nevertheless, it can be seen that the performance drops regardless of the scenario or pretraining during nightfall, leaving room for improvement in future works.

### Qualitative Results

We qualitatively show the improvement of our multi-stream cellular method over the ARTHuS [49] baseline. To do so, we display in Fig. C.3 the segmentation masks predicted by our method in two scenarios: the *Common* scenario with *General* pretraining and the *Spatial* scenario with *Cell* pretraining, and compare them to the masks predicted by the *Baseline* and the ground truth labels. On the top row, we show a vehicle driving in the countryside under static (clear) weather at the end of the online training. We can see that the baseline confuses some buildings with poles, and a car is misclassified as being part of the road, while our method is able to correctly segment it. The *Spatial* model produces the most accurate segmentation masks as it is able to precisely segment the city and vegetation in the background and the cars on the left. On the bottom row, we show a vehicle driving in the

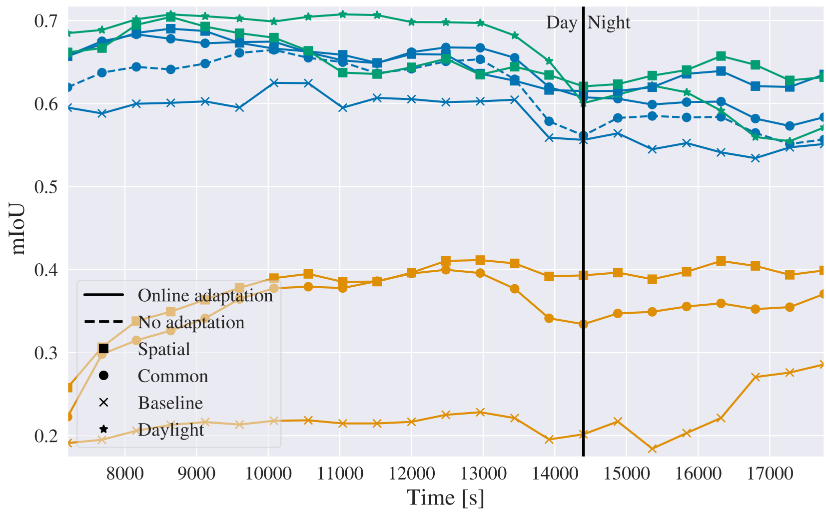



**(a)** *DADE-static* in the MSC-OL setup

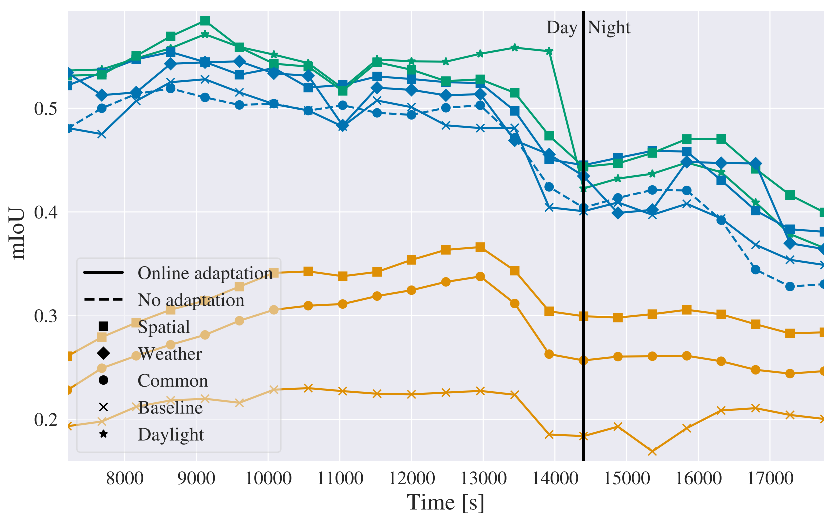



**(b)** *DADE-static* in the MSC-TTA setup

**Figure C.2: Evolution of the fleet performance over time (Part 1).** Comparison of the performance in the MSC-OL setup ((a) and (c)) and the MSC-TTA ((b) and (d)) setup of the best adaptive settings along with the baseline for each pretraining (*Scratch*, *General*, and *Cell*) using *DADE-static* weather ((a) and (b)) and *DADE-dynamic* weather ((c) and (d)).




**(c)** *DADE-dynamic* in the MSC-OL setup




**(d)** *DADE-dynamic* in the MSC-TTA setup

**Figure C.2: Evolution of the fleet performance over time (Part 2).** Comparison of the performance in the MSC-OL setup ((a) and (c)) and the MSC-TTA ((b) and (d)) setup of the best adaptive settings along with the baseline for each pretraining (*Scratch*, *General*, and *Cell*) using *DADE-static* weather ((a) and (b)) and *DADE-dynamic* weather ((c) and (d)).

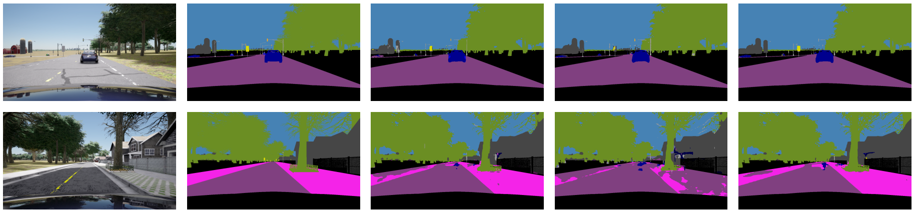


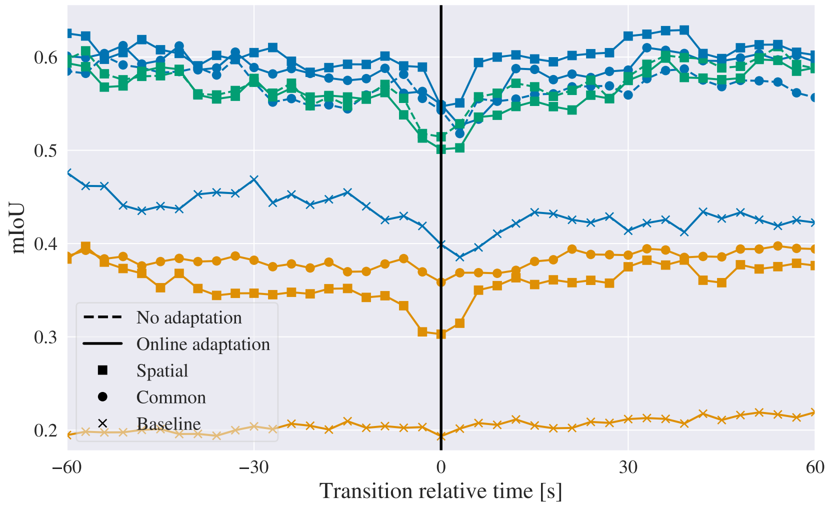
**Figure C.3: Qualitative results.** Comparison of different segmentation masks. From left to right: RGB image, ground truth, *Baseline*, *Common* scenario with *General* pretraining, and *Spatial* scenario with *Cell* pretraining. Black areas correspond to non-evaluated classes.

low density residential location under static (clear) weather also at the end of the online training. As can be seen, the *Common* model fails in this cell because it needs to learn a broader data distribution and loses accuracy due to its limited learning capacity. Contrarily, the *Spatial* model is able to better learn that particular cell data distribution and therefore produces the best results.

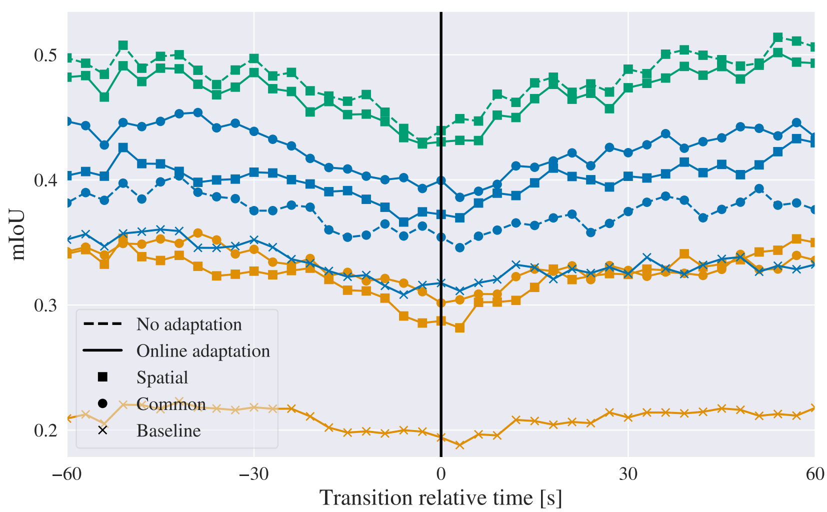
### C.2.3 Additional Results

#### Analysis of Transiting Agents

In this section, we provide insights about the transition of agents between cells. Particularly, we study the evolution of the performance of the models around cell transitions. Figure C.4 shows the mean performance of the agents transiting from one cell to another, *e.g.*, from a specific location to another, or from one weather condition to another. As can be seen, after the transition, the baseline method experiences a decrease in performance, which remains low for a long period of time. Contrarily, our method is able to recover much faster thanks to the switch between the cell-specific models. It is also interesting to observe that for both the MSC-OL and MSC-TTA setups, a temporary drop of mIoU score occurs right before a transition. This is probably because a vehicle approaching another cell may already see content from an adjacent cell while performing the task with the previous cell's model. For instance, a vehicle approaching the city center may record an image with its frontal camera showing the city center while still being registered in another location (*e.g.*, the countryside or the highway). This means that the agent will use the wrong cell model to analyze the

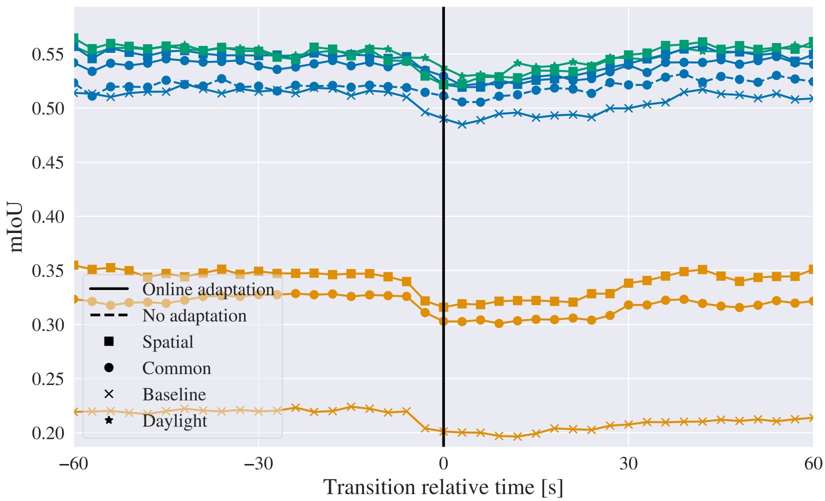



**(a)** *DADE-static* in the MSC-OL setup

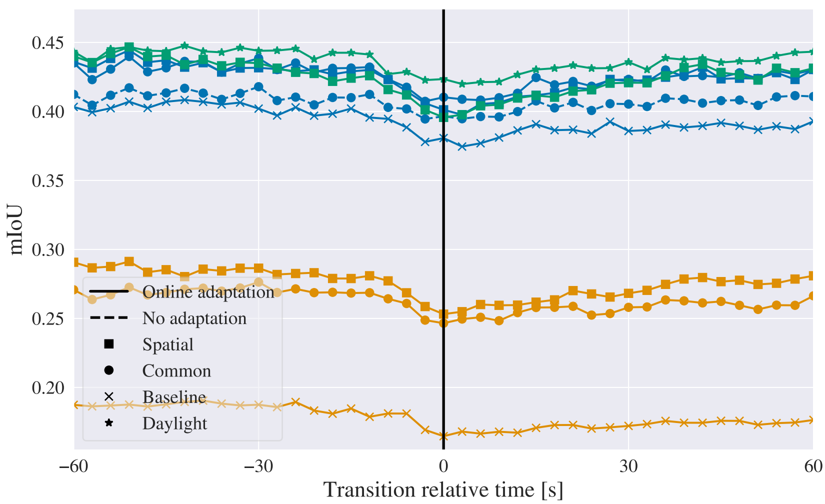



**(b)** *DADE-static* in the MSC-TTA setup

**Figure C.4: Fleet performance around cell transitions (Part 1).** Comparison of the performance in the MSC-OL setup ((a) and (c)) and the MSC-TTA setup ((b) and (d)) of our method (best settings) with the baseline for each pretraining (*Scratch*, *General*, and *Cell*) using *DADE-static* dataset ((a) and (b)) and *DADE-dynamic* ((c) and (d)) dataset. Confusion matrices of each frame are aggregated using a sliding window of $3$ seconds. The results are shown $30$ seconds before and after any cell transition that the agents encounter during the $3$ hours of testing.




**(c)** *DADE-dynamic* in the MSC-OL setup




**(d)** *DADE-dynamic* in the MSC-TTA setup

**Figure C.4: Fleet performance around cell transitions (Part 2).** Comparison of the performance in the MSC-OL setup ((a) and (c)) and the MSC-TTA setup ((b) and (d)) of our method (best settings) with the baseline for each pretraining (*Scratch*, *General*, and *Cell*) using *DADE-static* dataset ((a) and (b)) and *DADE-dynamic* ((c) and (d)) dataset. Confusion matrices of each frame are aggregated using a sliding window of $3$ seconds. The results are shown $30$ seconds before and after any cell transition that the agents encounter during the $3$ hours of testing.

environment. This issue could be addressed using the method developed in Chapter 6, which enables the automatic recognition of a cell rather than relying on predefined rules.

### Cyclic Domain Shifts

In Table C.4, we present additional experiments on the dataset and the best method proposed by Houyon et al. [134], namely *Baseline*+MIR, alongside the performance of the frozen teacher and student trained on the same set (namely, Cityscapes). The dataset and method [134] are specifically tailored for cyclic domain shifts. The first two columns are reported from Tables C.2 and C.3, for the 3 hours test sets.

Notably, our method demonstrates superior performance on *DADE-static*, *DADE-dynamic*, and the cyclic dataset of Houyon et al. [134]. Furthermore, we see that the *Baseline*+MIR performs worse than the baseline for *DADE-static* and *DADE-dynamic*, while it performs better on the cyclic dataset of Houyon et al. [134].

**Table C.4: Additional results for cyclic domain shifts.** Comparison of the MSC-TTA method with a frozen teacher, a frozen student, the *Baseline* [49], and *Baseline*+MIR [134], on our *DADE* datasets and the dataset of Houyon *et al.* [134].

| *mIoU-I* | *DADE-static* | *DADE-dynamic* | Houyon [134] |
|---|---|---|---|
| Teacher ❄ | .668 | .611 | / |
| Student ❄ | .214 | .159 | / |
| *Baseline* [49] | .274 | .212 | .234 |
| *Baseline*+MIR [134] | .181 | .147 | .256 |
| **Ours** | **.362** | **.312** | **.277** |

The results also demonstrate that our method exhibits an expected performance deficit relative to the teacher while consistently outperforming the student. The teacher is a state-of-the-art semantic segmentation model (namely, SegFormer [362] trained on Cityscapes [52]) and thus exhibits great performance on our *DADE* datasets. However, the emphasis on achieving the best possible performance often comes with increased complexity and overlooks the critical real-time aspect. The frame rate of SegFormer is approximately $2$ frames per second; it is thus far from being real time. Our proposed method aims at mimicking the performance of available teacher models while reducing computational power and battery usage, thereby bringing state-of-the-art performance at a higher frame rate.

The frozen student is trained on the Cityscapes [52] dataset with an initial learning rate of $10^{-4}$ using the Adam optimizer for $45$ epochs, reducing the learning rate by a factor of $10$ every $15$ epochs, a cross-entropy loss function, and a batch size of $8$. To match the dimension of the images in the *DADE* datasets, images from Cityscapes were resized to $720 \times 1440$ to keep the same ratio, then cropped to $720 \times 1280$.

# References


[1] David A. Abbink, Mark Mulder, and Erwin R. Boer. 2012. Haptic shared control: smoothly shifting control authority? *Cognition, Technology & Work* 14, 1 (March 2012), 19–28. doi:10.1007/s10111-011-0192-5 185

[2] ADASANDME. 2016. Adaptive ADAS to support incapacitated drivers Mitigate Effectively risks through tailor made HMI under automation. doi:10.3030/688900 185

[3] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. 2018. A rewriting system for convex optimization problems. *Journal of Control and Decision* 5, 1 (January 2018), 42–60. doi:10.1080/23307706.2017.1397554 169

[4] Archana Ahir and Vinaya Gohokar. 2019. Driver Inattention Monitoring System: A Review. In *International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET)*. IEEE, Shegoaon, India, 188–194. doi:10.1109/icitaet47105.2019.9170249 198, 200, 203

[5] Christer Ahlström, Anna Anund, Carina Fors, and Torbjörn Åkerstedt. 2018. The effect of daylight versus darkness on driver sleepiness: a driving simulator study. *Journal of Sleep Research* 27, 3 (June 2018), 1–9. doi:10.1111/jsr.12642 76

[6] Christer Ahlström, Anna Anund, Carina Fors, and Torbjörn Åkerstedt. 2018. Effects of the road environment on the development of driver sleepiness in young male drivers. *Accident Analysis & Prevention* 112 (March 2018), 127–134. doi:10.1016/j.aap.2018.01.012 76

[7] Christer Ahlström, George Georgoulas, and Katja Kircher. 2022. Towards a Context-Dependent Multi-Buffer Driver Distraction Detection Algorithm. *IEEE Transactions on Intelligent Transportation Systems* 23, 5 (May 2022), 4778–4790. doi:10.1109/tits.2021.3060168 55, 180

[8] Christer Ahlström, Katja Kircher, and Albert Kircher. 2009. Considerations When Calculating Percent Road Centre From Eye Movement Data in Driver Distraction Monitoring. In *International Driving Symposium on*

*Human Factors in Driver Assessment, Training and Vehicle Design*. University of Iowa Libraries Publishing, Big Sky, Montana, USA„ 132–139. doi:10.17077/drivingassessment.1313 104, 109

[9] Christer Ahlström, Katja Kircher, Marcus Nyström, and Benjamin Wolfe. 2021. Eye Tracking in Driver Attention Research – How Gaze Data Interpretations Influence What We Learn. *Frontiers in Neuroergonomics* 2 (December 2021), 1–34. doi:10.3389/fnrgo.2021.778043 58, 103

[10] Christer Ahlström, Raimondas Zemblys, Herman Jansson, Christian Forsberg, Johan Karlsson, and Anna Anund. 2021. Effects of partially automated driving on the development of driver sleepiness. *Accident Analysis & Prevention* 153 (April 2021), 1–9. doi:10.1016/j.aap.2021.106058 76

[11] AI-SEE. 2021. Artificial Intelligence enhancing vehicle vision in low visibility conditions. https://penta-eureka.eu/project-overview/penta-call-5/ai-see/ https://www.ai-see.eu. 192

[12] Eugene Aidman, Carolyn Chadunow, Kayla Johnson, and John Reece. 2015. Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accident Analysis & Prevention* 81 (August 2015), 8–13. doi:10.1016/j.aap.2015.03.041 29

[13] Motasem Alfarra, Hani Itani, Alejandro Pardo, Shyma Alhuwaider, Merey Ramazanova, Juan C. Pérez, Zhipeng Cai, Matthias Müller, and Bernard Ghanem. 2024. Evaluation of Test-Time Adaptation Under Computational Time Constraints. In *International Conference on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 235)*. ML Research Press, Vienna, Austria, 976–991. https://proceedings.mlr.press/v235/alfarra24a.html 132

[14] Munaf S. Najim Alluhaibi, Sarah Kadhim an Al-Din and Aiman Moyaid. 2018. Driver behavior detection techniques: A survey. *International Journal of Applied Engineering Research* 13, 11 (2018), 8856–8861. https://www.ripublication.com/Volume/ijaerv13n11.htm 37, 197, 198, 200, 203

[15] Hossam Almahasneh, Weng-Tink Chooi, Nidal Kamel, and Aamir Saeed Malik. 2014. Deep in thought while driving: An EEG study on drivers' cognitive distraction. *Transportation Research Part F: Traffic Psychology and Behaviour* 26 (September 2014), 218–226. doi:10.1016/j.trf.2014.08.001 53

[16] Nicholas Apostoloff and Alexander Zelinsky. 2003. Robust vision based lane tracking using multiple cues and particle filtering. In *IEEE Intelligent*

*Vehicles Symposium*. IEEE, Columbus, Ohio, USA, 558–563. doi:10.1109/ivs.2003.1212973 47

[17] Suman Arun, Kenneth Sundaraj, and Murugappan Murugappan. 2012. Driver inattention detection methods: A review. In *IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT)*. IEEE, Kuala Lumpur, Malaysia, 1–6. doi:10.1109/student.2012.6408351 198, 200, 203

[18] Karina Rollandovna Arutyunova, Anastasiia Vladimirovna Bakhchina, Daniil Igorevich Konovalov, Mane Margaryan, Andrei Viktorovich Filimonov, and Ivan Sergeevich Shishalov. 2024. Heart rate dynamics for cognitive load estimation in a driving simulation task. *Scientific Reports* 14, 1 (December 2024), 1–13. doi:10.1038/s41598-024-79728-x 45, 46

[19] Bhakti Baheti, Suhas Gajre, and Sanjay Talbar. 2018. Detection of Distracted Driver Using Convolutional Neural Network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Institute of Electrical and Electronics Engineers (IEEE), Salt Lake City, Utah, USA, 1145–11456. doi:10.1109/cvprw.2018.00150 47, 57

[20] Xiaolu Bai and Jing Feng. 2025. Awakening the Disengaged: Can Driving-Related Prompts Engage Drivers in Partial Automation? *Human Factors: The Journal of the Human Factors and Ergonomics Society* 67, 7 (January 2025), 731–752. doi:10.1177/00187208251314248 74, 178

[21] Rodney Petrus Balandong, Rana Fayyaz Ahmad, Mohamad Naufal Mohamad Saad, and Aamir Saeed Malik. 2018. A Review on EEG-Based Automatic Sleepiness Detection Systems for Driver. *IEEE Access* 6 (2018), 22908–22919. doi:10.1109/access.2018.2811723 198, 200, 203

[22] Ghazal Bargshady, Hakki Gokalp Ustun, Yasaman Baradaran, Houshyar Asadi, Ravinesh C Deo, Jeroen Van Boxtel, and Raul Fernandez Rojas. 2025. NeuroSafeDrive: An Intelligent System Using fNIRS for Driver Distraction Recognition. *Sensors* 25, 10 (May 2025), 1–20. doi:10.3390/s25102965 46, 62

[23] Massimo Bassani, Luca Catani, Asif Hazoor, Arlinda Hoxha, Angelo Lioi, Anna Portera, and Lumturi Tefa. 2023. Do driver monitoring technologies improve the driving behaviour of distracted drivers? A simulation study to assess the impact of an auditory driver distraction warning device on driving performance. *Transportation Research*

*Part F: Traffic Psychology and Behaviour* 95 (May 2023), 239–250. doi:10.1016/j.trf.2023.04.013 12

[24] Matthias Beggiato and Josef F. Krems. 2013. The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour* 18 (May 2013), 47–57. doi:10.1016/j.trf.2012.12.006 73, 177

[25] Shahina Begum. 2013. Intelligent driver monitoring systems based on physiological sensor signals: A review. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, The Hague, Netherlands, 282–289. doi:10.1109/ITSC.2013.6728246 198, 200, 203

[26] Luis M. Bergasa, Jaime Nuevo, Miguel Á. Sotelo, Rafael Barea, and M. Elena López. 2006. Real-Time System for Monitoring Driver Vigilance. *IEEE Transactions on Intelligent Transportation Systems* 7, 1 (March 2006), 63–77. doi:10.1109/tits.2006.869598 47

[27] Giulio Francesco Bianchi Piccinini, Carlos Manuel Rodrigues, Miguel Leitão, and Anabela Simões. 2014. Driver's behavioral adaptation to Adaptive Cruise Control (ACC): The case of speed and time headway. *Journal of Safety Research* 49 (June 2014), 77–84. doi:10.1016/j.jsr.2014.02.010 94

[28] Katja Blömacher, Gerhard Nöcker, and Markus Huff. 2020. The evolution of mental models in relation to initial information while driving automated. *Transportation Research Part F: Traffic Psychology and Behaviour* 68 (January 2020), 198–217. doi:10.1016/j.trf.2019.11.003 73

[29] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* abs/2004.10934 (2020), 17 pages. arXiv:2004.10934 doi:10.48550/arXiv.2004.10934 133

[30] Nils Ivar Bohlin. 1962. Safety belt. United States Patent Office, US 3,043,625 A, 3 pages. https://patents.google.com/patent/US3043625A 5

[31] Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews* 44 (July 2014), 58–75. doi:10.1016/j.neubiorev.2012.10.003 45

[32] Malik Boudiaf, Romain Mueller, Ismail Ben Ayed, and Luca Bertinetto. 2022. Parameter-free Online Test-time Adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), New Orleans, Louisiana, USA, 8334–8343. doi:10.1109/cvpr52688.2022.00816 155

[33] Linda Ng Boyle, Nicole van Nes, Klaus Bengler, and John D Lee. 2024. Workshop to Go Beyond Levels of Automation. In *International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Stanford, California, USA, 247–248. doi:10.1145/3641308.3677402 177

[34] Jason Ji. Braithwaite, Derrick G. Watson, Robert Jones, and Mickey Rowe. 2015. *A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments*. Technical Report. Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, University of Birmingham, UK. 1–43 pages. 102

[35] Claudia Buchner, Chantal Himmels, Jan Schmitz, Tanja Stoll, and Martin Baumann. 2024. Exploring Urban Challenges: Understanding Advanced Driver Assistance Systems in Different Situational Contexts. In *International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Stanford, CA, USA, 1–12. doi:10.1145/3640792.3675709 73, 75

[36] Christopher D. D. Cabrall, Nico M. Janssen, and Joost C. F. de Winter. 2018. Adaptive automation: automatically (dis)engaging automation during visually distracted driving. *PeerJ Computer Science* 4 (October 2018), 1–27. doi:10.7717/peerj-cs.166 13, 74

[37] Zhipeng Cai, Ozan Sener, and Vladlen Koltun. 2021. Online Continual Learning with Natural Distribution Shifts: An Empirical Study with Visual Data. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Montreal, QC, Canada, 8261–8270. doi:10.1109/iccv48922.2021.00817 131

[38] Kenneth L. Campbell. 2012. The SHRP 2 Naturalistic Driving Study. *TR News* 282 (September 2012), 30–35. https://insight.shrp2nds.us/documents/shrp2_background.pdf 47, 52, 57

[39] Daniela Cardone, David Perpetuini, Chiara Filippini, Lorenza Mancini, Sergio Nocco, Michele Tritto, Sergio Rinella, Alberto Giacobbe, Giorgio Fallica, Fabrizio Ricci, Sabina Gallina, and Arcangelo Merla. 2022. Classification of Drivers' Mental Workload Levels: Comparison of Machine Learning Methods Based on ECG and Infrared Thermal Signals.

*Sensors* 22, 19 (September 2022), 1–24. doi:10.3390/s22197300 45, 46, 47, 50

[40] Nicolò Cesa-Bianchi and Francesco Orabona. 2021. Online Learning Algorithms. *Annual Review of Statistics and Its Application* 8, 1 (March 2021), 165–190. doi:10.1146/annurev-statistics-040620-035329 131

[41] Mario I. Chacon-Murguia and Claudia Prieto-Resendiz. 2015. Detecting Driver Drowsiness: A survey of system designs and technology. *IEEE Consumer Electronics Magazine* 4, 4 (October 2015), 107–119. doi:10.1109/mce.2015.2463373 198, 200, 203

[42] Teck Kai Chan, Cheng Siong Chin, Hao Chen, and Xionghu Zhong. 2020. A Comprehensive Review of Driver Behavior Analysis Utilizing Smartphones. *IEEE Transactions on Intelligent Transportation Systems* 21, 10 (October 2020), 4444–4475. doi:10.1109/tits.2019.2940481 37, 198, 200, 203

[43] Jia-Ren Chang and Yong-Sheng Chen. 2018. Pyramid Stereo Matching Network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Salt Lake City, Utah, USA, 5410–5418. doi:10.1109/cvpr.2018.00567 133

[44] Weiya Chen, Tetsuo Sawaragi, and Toshihiro Hiraoka. 2022. Comparing eye-tracking metrics of mental workload caused by NDRTs in semi-autonomous driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 89 (August 2022), 109–128. doi:10.1016/j.trf.2022.05.004 46, 48

[45] Rishu Chhabra, Seema Verma, and C. Rama Krishna. 2017. A survey on driver behavior detection techniques for intelligent transportation systems. In *International Conference on Cloud Computing, Data Science & Engineering – Confluence*. IEEE, Noida, India, 36–41. doi:10.1109/confluence.2017.7943120 37, 198, 200, 203

[46] Anuva Chowdhury, Rajan Shankaran, Manolya Kavakli, and Md. Mokammel Haque. 2018. Sensor applications and physiological features in drivers' drowsiness detection: A review. *IEEE Sensors Journal* 18, 8 (April 2018), 3055–3067. doi:10.1109/JSEN.2018.2807245 198, 200, 203

[47] Wan-Young Chung, Teak-Wei Chong, and Boon-Giin Lee. 2019. Methods to detect and reduce driver stress: A review. *International Journal of Automotive Technology* 20, 5 (October 2019), 1051–1063. doi:10.1007/s12239-019-0099-3 198, 200, 203

[48] Jessica B. Cicchino. 2017. Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis & Prevention* 99 (Feb. 2017), 142–152. doi:10.1016/j.aap.2016.11.009 6

[49] Anthony Cioppa, Adrien Deliège, Maxime Istasse, Christophe De Vleeschouwer, and Marc Van Droogenbroeck. 2019. ARTHuS: Adaptive Real-Time Human Segmentation in Sports Through Online Distillation. In *IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), CVsports*. Institute of Electrical and Electronics Engineers (IEEE), Long Beach, California, USA, 2505–2514. doi:10.1109/cvprw.2019.00306 129, 132, 135, 136, 217, 219, 220, 221, 222, 223, 224, 230

[50] Roger Coetzer and Graham Hancke. 2009. Driver fatigue fetection: A survey. In *AFRICON*. IEEE, Nairobi, Kenya, 1–6. doi:10.1109/AFRCON.2009.5308101 198, 200, 203

[51] Ian Colwell, Buu Phan, Shahwar Saleem, Rick Salay, and Krzysztof Czarnecki. 2018. An Automated Vehicle Safety Concept Based on Runtime Restriction of the Operational Design Domain. In *IEEE Intelligent Vehicles Symposium*. IEEE, Changshu, China, 1910–1917. doi:10.1109/ivs.2018.8500530 157

[52] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), Las Vegas, Nevada, USA, 3213–3223. doi:10.1109/cvpr.2016.350 137, 138, 216, 219, 220, 221, 222, 223, 230, 231

[53] Joseph F. Coughlin, Bryan Reimer, and Bruce Mehler. 2011. Monitoring, managing, and motivating driver safety and well-being. *IEEE Pervasive Computing* 10, 3 (2011), 14–21. doi:10.1109/MPRV.2011.54 13, 76

[54] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. 2020. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences (PNAS)* 117, 48 (May 2020), 30055–30062. doi:10.1073/pnas.1912789117 155, 157

[55] D3COS. 2011. Designing Dynamic Distributed Cooperative Human-Machine Systems. https://cordis.europa.eu/project/id/269336 185

[56] Laith Dababneh and Moustafa El Gindy. 2015. Driver vigilance level detection systems: a literature survey. *International Journal of Vehicle Performance* 2, 1 (2015), 1–29. doi:10.1504/ijvp.2015.074120 198, 200, 203

[57] Varsha E. Dahiphale and Sathyanarayana Rao. 2015. A review paper on portable driver monitoring system for teal time fatigue. In *International Conference on Computing Communication Control and Automation*. IEEE, Pune, India, 558–560. doi:10.1109/ICCUBEA.2015.115 198, 200, 203

[58] Joost C. F. de Winter, Riender Happee, Marieke H. Martens, and Neville A. Stanton. 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour* 27 (November 2014), 196–217. doi:10.1016/j.trf.2014.06.016 74, 76, 178

[59] Joost C. F. de Winter, Bastiaan Petermeijer, and David Abbink. 2022. Shared control versus traded control in driving: a debate around automation pitfalls. *Ergonomics* 66, 10 (December 2022), 1–27. doi:10.1080/00140139.2022.2153175 185

[60] Hanwen Deng, Guoliang Xiang, Jiandong Pan, Xianhui Wu, Chaojie Fan, Kui Wang, and Yong Peng. 2024. How to design driver takeover request in real-world scenarios: A systematic review. *Transportation Research Part F: Traffic Psychology and Behaviour* 104 (July 2024), 411–432. doi:10.1016/j.trf.2024.06.012 33

[61] Steven Diamond and Stephen Boyd. 2016. CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *Journal of Machine Learning Research* 17, 83 (2016), 1–5. http://jmlr.org/papers/v17/15-408.html 169

[62] Antonio D'Innocente, Francesco Cappio Borlino, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. 2020. One-Shot Unsupervised Cross-Domain Detection. In *European Conference on Computer Vision (ECCV) (Lecture Notes in Computer Science, Vol. 12361)*. Springer International Publishing, Glasgow, United Kingdom, 732–748. doi:10.1007/978-3-030-58517-4_43 132

[63] Liza Dixon. 2020. Autonowashing: The Greenwashing of Vehicle Automation. *Transportation Research Interdisciplinary Perspectives* 5 (May 2020), 1–9. doi:10.1016/j.trip.2020.100113 73, 177

[64] Liza Dixon, Norbert Schneider, Dominik Mühlbacher, Dennis Befelein, Frank O. Flemisch, and Martin Baumann. 2025. Explaining Authoritative Control Interventions in Automated Driving to Support Driver Understanding, Trust, and Reliance. *Transportation Research Part F: Traffic Psychology and Behaviour* 113 (August 2025), 194–212. doi:10.1016/j.trf.2025.04.013 77, 187

[65] Yanchao Dong, Zhencheng Hu, Keiichi Uchimura, and Nobuki Murayama. 2011. Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *IEEE Transactions on Intelligent Transportation Systems* 12, 2 (June 2011), 596–614. doi:10.1109/tits.2010.2092770 37, 53, 198, 200, 203

[66] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. CARLA: An Open Urban Driving Simulator. In *Annual Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 78)*. ML Research Press, Mountain View, California, USA, 1–16. https://proceedings.mlr.press/v78/dosovitskiy17a.html 78, 106, 131, 137, 153, 158, 216

[67] Na Du, Jinyong Kim, Feng Zhou, Elizabeth Pulver, Dawn M. Tilbury, Lionel P. Robert, Anuj K. Pradhan, and X. Jessie Yang. 2020. Evaluating Effects of Cognitive Load, Takeover Request Lead Time, and Traffic Density on Drivers' Takeover Performance in Conditionally Automated Driving. In *International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Virtual conference, 66–73. doi:10.1145/3409120.3410666 75

[68] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. 2019. Neural Spline Flows. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 32. Curran Associates, Inc., Vancouver, Canada, 1–12. https://papers.nips.cc/paper_files/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html 157

[69] Frank Dursoi and Sarah Gronlund. 1999. *Situation awareness*. John Wiley & Sons, Chichester, England, 283–314. 54

[70] Alaa El Khatib, Chaojie Ou, and Fakhri Karray. 2020. Driver Inattention Detection in the Context of Next-Generation Autonomous Vehicles Design: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 21, 11 (November 2020), 4483–4496. doi:10.1109/tits.2019.2940874 33, 37, 198, 200, 203

[71] Mica R. Endsley. 2016. From Here to Autonomy. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 59, 1 (December 2016), 5–27. doi:10.1177/0018720816681350 74, 184

[72] Johan Engström and Gustav Markkula. 2007. Effects of visual and cognitive distraction on lane change test performance. In *International Driving Symposium on Human Factors in Driver Assessment*. University of Iowa, Public Policy Center, Stevenson, Washington, USA, 199–205. doi:10.17077/drivingassessment.1237 46, 58

[73] European Parliament. 2019. Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users, and amending Regulation (EU) 2018/858 of the European Parliament and of the Council. Official Journal of the European Union, 40 pages. https://eur-lex.europa.eu/eli/reg/2019/2144/oj OJ L 325, 16.12.2019. 11, 12

[74] European Parliament. 2023. Commission Delegated Regulation (EU) 2023/2590 of 13 July 2023 supplementing Regulation (EU) 2019/2144 of the European Parliament and of the Council with regard to the detailed rules concerning the advanced driver distraction warning system, and amending Annex II to that Regulation. Official Journal of the European Union, 25 pages. https://eur-lex.europa.eu/eli/reg_del/2023/2590/oj OJ L 2023/2590, 17.11.2023. 11

[75] EVENTS. 2022. Reliable in-vehicle perception and decision-making in complex environmental conditions. doi:10.3030/101069614 192

[76] Roja Ezzati Amini, Christelle Al Haddad, Debapreet Batabyal, Isidora Gkena, Bart De Vos, Ariane Cuenen, Tom Brijs, and Constantinos Antoniou. 2023. Driver distraction and in-vehicle interventions: A driving simulator study on visual attention and driving performance. *Accident Analysis & Prevention* 191 (October 2023), 1–12. doi:10.1016/j.aap.2023.107195 46, 58

[77] Behzad Farahmand and Amin Mirza Boroujerdian. 2018. Effect of road geometry on driver fatigue in monotonous environments: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour* 58 (October 2018), 640–651. doi:10.1016/j.trf.2018.06.021 13, 177

[78] Simon Föll, Alina Dubatovka, Eugen Ernst, Siu Lun Chau, Martin Maritsch, Patrik Okanovic, Gudrun Thäter, Joachim M. Buhmann, Felix Wortmann, and Krikamol Muandet. 2023. Gated Domain Units for Multi-source Domain Generalization. *Transactions on Machine Learning Research* 9 (2023), 1–30. https://openreview.net/pdf?id=V7BvYJyTmM 132

[79] Michaël Fonder, Damien Ernst, and Marc Van Droogenbroeck. 2022. Parallax Inference for Robust Temporal Monocular Depth Estimation in Unstructured Environments. *Sensors* 22, 23 (December 2022), 1–22. doi:10.3390/s22239374 133

[80] Tiago Fonseca and Sara Ferreira. 2025. Drowsiness Detection in Drivers: A Systematic Review of Deep Learning-Based Models. *Applied Sciences* 15, 16 (August 2025), 1–43. doi:10.3390/app15169018 12

[81] Yannick Forster, Nadja Schoemig, Christina Kremer, Katharina Wiedemann, Sebastian Gary, Frederik Naujoks, Andreas Keinath, and Alexandra Neukum. 2024. Attentional warnings caused by driver monitoring systems: How often do they appear and how well are they understood? *Accident Analysis & Prevention* 205 (September 2024), 1–12. doi:10.1016/j.aap.2024.107684 12

[82] Lisa R. Fournier, Glenn F. Wilson, and Carolyne R. Swain. 1999. Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology* 31, 2 (January 1999), 129–145. doi:10.1016/s0167-8760(98)00049-x 46, 48

[83] Hannah J. Foy and Peter Chapman. 2018. Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics* 73 (November 2018), 90–99. doi:10.1016/j.apergo.2018.06.006 103, 119

[84] Matteo Fresta, Francesco Bellotti, Igor Bochenko, Luca Lazzaroni, Gaëtan Merlhiot, Fabio Tango, and Riccardo Berta. 2025. Deep Learning-Based Real-Time Driver Cognitive Distraction Detection. *IEEE Access* 13 (2025), 26589–26607. doi:10.1109/access.2025.3539392 46, 47, 62

[85] Lex Fridman, Daniel E. Brown, Michael Glazer, William Angell, Spencer Dodd, Benedikt Jenik, Jack Terwilliger, Aleksandr Patsekin, Julia Kindelsberger, Li Ding, Sean Seaman, Alea Mehler, Andrew Sipperley, Anthony Pettinato, Bobbie D. Seppelt, Linda Angell, Bruce Mehler, and Bryan Reimer. 2019. MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation. *IEEE Access* 7 (2019), 102021–102038. doi:10.1109/access.2019.2926040 47, 52, 57, 62

[86] Lex Fridman, Philipp Langhans, Joonbum Lee, and Bryan Reimer. 2016. Driver Gaze Region Estimation without Use of Eye Movement. *IEEE*

*Intelligent Systems* 31, 3 (May 2016), 49–56. doi:10.1109/mis.2016.47 46, 47, 58, 59

[87] Lex Fridman, Joonbum Lee, Bryan Reimer, and Trent Victor. 2016. ‘Owl’ and ‘Lizard’: patterns of head pose and eye pose in driver gaze classification. *IET Computer Vision* 10, 4 (April 2016), 308–313. doi:10.1049/iet-cvi.2015.0296 46, 47, 58

[88] Lex Fridman, Bryan Reimer, Bruce Mehler, and William T. Freeman. 2018. Cognitive Load Estimation in the Wild. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Montréal, Canada, 1–9. doi:10.1145/3173574.3174226 46, 47, 48, 50

[89] Thomas M. Gable, Andrew L. Kun, Bruce N. Walker, and Riley J. Winton. 2015. Comparing heart rate and pupil size as objective measures of workload in the driving context. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Nottingham, England, UK, 20–25. doi:10.1145/2809730.2809745 45, 46, 48

[90] Marco Galvani. 2019. History and future of driver assistance. *IEEE Instrumentation & Measurement Magazine* 22, 1 (Feb. 2019), 11–16. doi:10.1109/mim.2019.8633345 5

[91] Jin Gao, Jialing Zhang, Xihui Liu, Trevor Darrell, Evan Shelhamer, and Dequan Wang. 2023. Back to the Source: Diffusion-Driven Adaptation to Test-Time Corruption. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Vancouver, Canada, 11786–11796. doi:10.1109/cvpr52729.2023.01134 132

[92] Benoît Gérin, Anaïs Halin, Anthony Cioppa, Maxim Henry, Bernard Ghanem, Benoît Macq, Christophe De Vleeschouwer, and Marc Van Droogenbroeck. 2024. Multi-Stream Cellular Test-Time Adaptation of Real-Time Models Evolving in Dynamic Environments. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vol. 33. Institute of Electrical and Electronics Engineers (IEEE), Seattle, Washington, USA, 4472–4482. doi:10.1109/cvprw63382.2024.00450 20, 126, 127, 155, 191

[93] Benoît Gérin, Maxime Zanella, Maxence Wynen, Saïd Mahmoudi, Benoît Macq, and Christophe De Vleeschouwer. 2024. Exploring viability of Test-Time Training: Application to 3D segmentation in Multiple Sclerosis. In *IEEE Conference on Artificial Intelligence (CAI)*, Vol. 34. Institute of Electrical and Electronics Engineers (IEEE), Singapore, 557–562. doi:10.1109/cai59869.2024.00110 155

[94] Pnina Gershon, Sean Seaman, Bruce Mehler, Bryan Reimer, and Joseph Coughlin. 2021. Driver behavior and the use of automation in real-world driving. *Accident Analysis & Prevention* 158 (August 2021), 1–9. doi:10.1016/j.aap.2021.106217 75, 77

[95] Raymond Ghandour, Bilel Neji, Ali M. El-Rifaie, and Zaher Al Barakeh. 2020. Driver Distraction and Stress Detection Systems: A Review. *International Journal of Engineering and Applied Sciences (IJEAS)* 7, 4 (April 2020), 39–46. doi:10.31873/ijeas.7.04.10 198, 200, 203

[96] Andrea Giorgi, Vincenzo Ronca, Alessia Vozzi, Nicolina Sciaraffa, Antonello Di Florio, Luca Tamborra, Ilaria Simonetti, Pietro Aricò, Gianluca Di Flumeri, Dario Rossi, and Gianluca Borghini. 2021. Wearable technologies for mental workload, stress, and emotional state assessment during working-like tasks: A comparison with laboratory technologies. *Sensors* 21, 7 (2021), 1–21. doi:10.3390/s21072332 47, 50

[97] Clement Godard, Oisin Mac Aodha, Michael Firman, and Gabriel Brostow. 2019. Digging Into Self-Supervised Monocular Depth Estimation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Seoul, South Korea, 3827–3837. doi:10.1109/iccv.2019.00393 133

[98] Christian Gold, Moritz Körber, David Lechner, and Klaus Bengler. 2016. Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, 4 (March 2016), 642–652. doi:10.1177/0018720816634226 75

[99] Joel Gonçalves and Klaus Bengler. 2015. Driver State Monitoring Systems- Transferable Knowledge Manual Driving to HAD. *Procedia Manufacturing* 3 (2015), 3011–3016. doi:10.1016/j.promfg.2015.07.845 46, 53, 60

[100] Taesik Gong, Jongheon Jeong, Taewon Kim, Yewon Kim, Jinwoo Shin, and Sung-Ju Lee. 2022. NOTE: Robust Continual Test-time Adaptation Against Temporal Correlation. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 35. Curran Associates, Inc., New Orleans, Louisiana, USA, 27253–27266. https://openreview.net/forum?id=E9HNxrCFZPV 132, 155

[101] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT Press, Cambridge, Massachusetts, USA. 129

[102] Thomas Alexander Goodge, Frank Pollick, and Stephen Anthony Brewster. 2024. Can You Hazard a Guess?: Evaluating the Effect of Augmented Reality Cues on Driver Hazard Prediction. In *Proceedings of*

*the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, Hawaii, USA, 1–28. doi:10.1145/3613904.3642300 186

[103] Daniel Greenberg, Marcel Nonnenmacher, and Jakob H. Macke. 2019. Automatic Posterior Transformation for Likelihood-Free Inference. In *International Conference on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 97)*. ML Research Press, New York City, New York, USA, 2404–2414. https://proceedings.mlr.press/v97/greenberg19a.html 157

[104] Paul G. Griffiths and R. Brent Gillespie. 2005. Sharing Control Between Humans and Automation Using Haptic Interface: Primary and Secondary Task Performance Benefits. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 47, 3 (September 2005), 574–590. doi:10.1518/001872005774859944 77, 185

[105] Jesús Gutiérrez, Víctor Rodríguez, and Sergio Martin. 2021. Comprehensive Review of Vision-Based Fall Detection Systems. *Sensors* 21, 3 (Feb. 2021), 1–50. doi:10.3390/s21030947 34

[106] Magnus Gyllenhammar, Rolf Johansson, Fredrik Warg, DeJiu Chen, Hans-Martin Heyn, Martin Sanfridson, Jan Söderberg, Anders Thorsén, and Stig Ursing. 2020. Towards an operational design domain that supports the safety argumentation of an automated driving system. In *European congress on embedded real time systems (ERTS)*. Society of Electricity, Electronics and Information and Communication Technologies (SEE), Toulouse, France, 1–10. 157

[107] HADRIAN. 2019. Holistic Approach for Driver Role Integration and Automation Allocation for European Mobility Needs. doi:10.3030/875597 185

[108] Anaïs Halin, Adrien Deliège, Christel Devue, and Marc Van Droogenbroeck. 2025. Gaze-Based Indicators of Driver Cognitive Distraction: Effects of Different Traffic Conditions and Adaptive Cruise Control Use. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Brisbane, Australia, 151–156. doi:10.1145/3744335.3758497 21, 46, 47, 61, 62, 99

[109] Anaïs Halin, Christel Devue, and Marc Van Droogenbroeck. 2025. DEV: A Driver-Environment-Vehicle Closed-Loop Framework for Risk-Aware Adaptive Automation of Driving. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Brisbane, Australia, 47–52. doi:10.1145/3744335.3758480 21, 176

[110] Anaïs Halin, Benoît Gérin, Anthony Cioppa, Maxim Henry, Bernard Ghanem, Benoît Macq, Christophe De Vleeschouwer, and Marc Van Droogenbroeck. 2023. Driving Agents in Dynamic Environments DADE dataset. doi:10.58119/ULG/H5SP5P 126

[111] Anaïs Halin, Sébastien Piérard, Anthony Cioppa, and Marc Van Droogenbroeck. 2024. A Hitchhiker's Guide to Understanding Performances of Two-Class Classifiers. *arXiv* abs/2412.04377 (2024), 35 pages. arXiv:2412.04377 doi:10.48550/arXiv.2412.04377 22

[112] Anaïs Halin, Sébastien Piérard, Renaud Vandeghen, Benoît Gérin, Maxime Zanella, Martin Colot, Jan Held, Anthony Cioppa, Emmanuel Jean, Gianluca Bontempi, Saïd Mahmoudi, Benoît Macq, and Marc Van Droogenbroeck. 2025. Physically Interpretable Probabilistic Domain Characterization. In *Asian Conference on Computer Vision Workshops (ACCV Workshops) (Lecture Notes in Computer Science, Vol. 15482)*. Springer Nature Singapore, Hanoi, Vietnam, 17–35. doi:10.1007/978-981-96-2641-0_2 21, 151

[113] Anaïs Halin, Marc Van Droogenbroeck, and Christel Devue. 2025. Are Electrodermal Activity-Based Indicators of Driver Cognitive Distraction Robust to Varying Traffic Conditions and Adaptive Cruise Control Use?. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Brisbane, Australia, 78–82. doi:10.1145/3744335.3758485 21, 46, 62, 99

[114] Anaïs Halin, Marc Van Droogenbroeck, and Christel Devue. 2025. Effects of Cognitive Distraction and Driving Environment Complexity on Adaptive Cruise Control Use and Its Impact on Driving Performance: A Simulator Study. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Brisbane, Australia, 231–241. doi:10.1145/3744333.3747822 21, 72

[115] Anaïs Halin, Jacques G. Verly, and Marc Van Droogenbroeck. 2021. Survey and synthesis of state of the art in driver monitoring. *Sensors* 21, 16 (August 2021), 1–48. doi:10.3390/s21165558 20, 27, 36, 39, 41, 64, 101, 102, 179, 189

[116] Xueqin Hao, Zhiguo Wang, Fan Yang, Ying Wang, Yanru Guo, and Kan Zhang. 2007. The Effect of Traffic on Situation Awareness and Mental Workload: Simulator-Based Study. In *International Conference on Engineering Psychology and Cognitive Ergonomics (EPCE) (Lecture Notes in Computer Science, Vol. 4562)*. Springer Berlin Heidelberg, Beijing, China, 288–296. doi:10.1007/978-3-540-73331-7_31 46, 52

[117] Joanne L. Harbluk, Y. Ian Noy, Patricia L. Trbovich, and Moshe Eizenman. 2007. An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention* 39, 2 (March 2007), 372–379. doi:10.1016/j.aap.2006.08.013 46, 61, 62, 104, 116

[118] Veronika Hargutt and Harald-Peter Krüger. 2000. Eyelid movements and their predictive value for fatigue stages. In *International Conference on Traffic and Transport Psychology (ICTTP)*. National Academy of Sciences, Bern, Switzerland, 1–10. 46, 60

[119] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload* 52 (1988), 139–183. doi:10.1016/s0166-4115(08)62386-9 45, 46, 48, 62

[120] Matthias Hartmann, Fred W. Mast, and Martin H. Fischer. 2015. Spatial biases during mental arithmetic: evidence from eye movements on a blank screen. *Frontiers in Psychology* 6 (January 2015), 1–8. doi:10.3389/fpsyg.2015.00012 116, 118

[121] HAVE-it. 2008. Highly Automated Vehicles for Intelligent Transport. https://cordis.europa.eu/project/id/212154 185

[122] Elad Hazan. 2016. Introduction to Online Convex Optimization. *Foundations and Trends in Optimization* 2, 3-4 (2016), 157–325. doi:10.1561/2400000013 131

[123] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Las Vegas, Nevada, USA, 770–778. doi:10.1109/CVPR.2016.90 161

[124] Yufan He, Aaron Carass, Lianrui Zuo, Blake E. Dewey, and Jerry L. Prince. 2020. Self Domain Adapted Network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (Lecture Notes in Computer Science, Vol. 12261)*. Springer International Publishing, Lima, Peru, 437–446. doi:10.1007/978-3-030-59710-8_43 132

[125] Tobias Hecht, Anna Feldhütter, Jonas Radlmayr, Yasuhiko Nakano, Yoshikuni Miki, Corbinian Henle, and Klaus Bengler. 2018. A Review of Driver State Monitoring Systems in the Context of Automated Driving. In *Congress of the International Ergonomics Association (Advances in Intelligent Systems and Computing, Vol. 823)*. Springer International Publishing, Florence, Italy, 398–408. doi:10.1007/978-3-319-96074-6_43 198, 200, 203

[126] Joeri Hermans, Arnaud Delaunoy, François Rozet, Antoine Wehenkel, Volodimir Begy, and Gilles Louppe. 2022. A Crisis In Simulation-Based Inference? Beware, Your Posterior Approximations Can Be Unfaithful. *Transactions on Machine Learning Research* 11 (November 2022), 2835–8856. https://openreview.net/forum?id=LHAbHkt6Aq 157, 164

[127] Nicolas Herzberger, Marcel Usai, and Frank Flemisch. 2022. Confidence Horizon for a Dynamic Balance between Drivers and Vehicle Automation: First Sketch and Application. *Human Factors in Transportation* 60 (2022), 33–39. doi:10.54941/ahfe1002431 77, 186

[128] Nicolas Herzberger, Marcel Usai, Maximilian Schwalm, and Frank Flemisch. 2024. *Cooperation Between Vehicle and Driver: Predicting the Driver's Takeover Capability in Cooperative Automated Driving Based on Orientation Patterns*. Springer International Publishing, Cham, Switzerland, 509–523. doi:10.1007/978-3-031-60494-2_17 77, 186

[129] HFAuto. 2013. Human Factors of Automated Driving. https://cordis.europa.eu/project/id/605817 185

[130] Hi-Drive. 2021. Addressing challenges toward the deployment of higher automation. doi:10.3030/101006664 192

[131] Marika Hoedemaeker and Karel A. Brookhuis. 1998. Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F: Traffic Psychology and Behaviour* 1, 2 (December 1998), 95–106. doi:10.1016/s1369-8478(98)00008-4 94

[132] Reiner Hoeger, Angelos Amditis, Martin Kunert, Alfred Hoess, F. Flemisch, Hans-Peter Krueger, Arne Bartels, Achim Beutner, and K. Pagle. 2008. Highly automated vehicles for intelligent transport: HAVEit approach. In *World Congress on Intelligent Transport Systems*. ITS World Congress, New York City, New York, USA, 1–10. 185

[133] Steven C.H. Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. 2021. Online learning: A comprehensive survey. *Neurocomputing* 459 (October 2021), 249–289. doi:10.1016/j.neucom.2021.04.112 131

[134] Joachim Houyon, Anthony Cioppa, Yasir Ghunaim, Motasem Alfarra, Anaïs Halin, Maxim Henry, Bernard Ghanem, and Marc Van Droogenbroeck. 2023. Online Distillation with Continual Learning for Cyclic Domain Shifts. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vol. abs 2211 16234. Institute of

Electrical and Electronics Engineers (IEEE), Vancouver, Canada, 2437–2446. doi:10.1109/cvprw59228.2023.00242 22, 132, 135, 155, 217, 219, 220, 221, 222, 223, 230

[135] Mark E Howard and Jennifer M Cori. 2024. Time for sleep science to wake up to drowsy driver monitoring. *Sleep* 47, 2 (Feb. 2024), 1–3. doi:10.1093/sleep/zsad324 12

[136] Hexiang Hu, Ozan Sener, Fei Sha, and Vladlen Koltun. 2023. Drinking from a Firehose: Continual Learning with Web-scale Natural Language. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 5 (May 2023), 5684–5696. doi:10.1109/tpami.2022.3218265 131

[137] Gaojian Huang, Yantong Jin, and Wei-Hsiang Lo. 2025. Beyond Levels of Driving Automation: A Triadic Framework of Human-AI Collaboration in On-Road Mobility. *arXiv* abs/2504.19120 (2025), 4 pages. arXiv:2504.19120 doi:10.48550/arXiv.2504.19120 177

[138] Jing Huang, Xinyu Huang, Yang Peng, and Lin Hu. 2024. Driver state recognition with physiological signals: Based on deep feature fusion and feature selection techniques. *Biomedical Signal Processing and Control* 93 (July 2024), 1–13. doi:10.1016/j.bspc.2024.106204 46, 49

[139] Jiaqi Huang, Qiliang Zhang, Tingru Zhang, Tieyan Wang, and Da Tao. 2024. Assessment of Drivers' Mental Workload by Multimodal Measures during Auditory-Based Dual-Task Driving Scenarios. *Sensors* 24, 3 (Feb. 2024), 1–19. doi:10.3390/s24031041 46, 48

[140] Rob J. Hyndman. 1996. Computing and Graphing Highest Density Regions. *The American Statistician* 50, 2 (May 1996), 120–126. doi:10.1080/00031305.1996.10474359 161

[141] Mohamed Ibrahim, James Haworth, and Tao Cheng. 2019. WeatherNet: Recognising Weather and Visual Conditions from Street-Level Images Using Deep Residual Learning. *ISPRS International Journal of Geo-Information* 8, 12 (November 2019), 1–18. doi:10.3390/ijgi8120549 153, 156

[142] Toshiyuki Inagaki and Thomas B. Sheridan. 2019. A critique of the SAE conditional driving automation definition, and analyses of options for improvement. *Cognition, Technology & Work* 21, 4 (2019), 569–578. doi:10.1007/s10111-018-0471-5 177

[143] Marco Introvigne, Andrea Ramazzina, Stefanie Walz, Dominik Scheuble, and Mario Bijelic. 2024. Real-time Environment Condition Classification for Autonomous Vehicles. In *IEEE Intelligent Vehicles*

*Symposium (IV)*. IEEE, Jeju Island, Republic of Korea, 1527–1533. doi:10.1109/iv55156.2024.10588692 153, 156

[144] ITU. 2015. *IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond*. Technical Report. International Telecommunication Union. Recommendation ITU-R M.2083-0. 129

[145] Shintaro Izumi. 2017. A contact-less heart rate sensor system for driver health monitoring. https://api.semanticscholar.org/CorpusID:32159958 47, 50

[146] JASP Team. 2025. JASP (Version 0.19.3): Computer software. https://jasp-stats.org. 84, 109

[147] Minsoo Jeon, Jaemin Seo, and Jaehyuk Min. 2024. DA-RAW: Domain Adaptive Object Detection for Real-World Adverse Weather Conditions. In *IEEE International Conference on Robotics and Automation (ICRA)*, Vol. 17. Institute of Electrical and Electronics Engineers (IEEE), Yokohama, Japan, 2013–2020. doi:10.1109/ICRA57147.2024.10611219 156

[148] Mishel Johns, Srinath Sibi, and Wendy Ju. 2014. Effect of cognitive load in autonomous vehicles on driver performance during transfer of control. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Seattle, Washington, USA, 1–4. doi:10.1145/2667239.2667296 33

[149] Daniel Kahneman, Bernard Tursky, David Shapiro, and Andrew Crider. 1969. Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology* 79, 1 (January 1969), 164–167. doi:10.1037/h0026952 46, 60

[150] Susumu Kajiwara. 2014. Evaluation of driver's mental workload by facial temperature and electrodermal activity under simulated driving conditions. *International Journal of Automotive Technology* 15, 1 (Feb. 2014), 65–70. doi:10.1007/s12239-014-0007-9 45, 46, 102

[151] Hang-Bong Kang. 2013. Various Approaches for Driver and Driving Behavior Monitoring: A Review. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, Sydney, New South Wales, Australia, 616–623. doi:10.1109/iccvw.2013.85 37, 198, 200, 203

[152] Sinan Kaplan, Mehmet Amac Guvensan, Ali Gokhan Yavuz, and Yasin Karalurt. 2015. Driver Behavior Analysis for Safe Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 16, 6 (December 2015), 3017–3032. doi:10.1109/tits.2015.2462084 37, 198, 200, 203

[153] Neerav Karani, Ertunc Erdil, Krishna Chaitanya, and Ender Konukoglu. 2021. Test-time adaptable neural networks for robust medical image segmentation. *Medical Image Analysis* 68 (Feb. 2021), 1–14. doi:10.1016/j.media.2020.101907 132

[154] Steven J. Kass, Kerstan S. Cole, and Claudia J. Stanny. 2007. Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 10, 4 (July 2007), 321–329. doi:10.1016/j.trf.2006.12.002 54

[155] Sherrie-Anne Kaye, Ioni Lewis, and James Freeman. 2018. Comparison of self-report and objective measures of driving behavior and road safety: A systematic review. *Journal of Safety Research* 65 (June 2018), 141–151. doi:10.1016/j.jsr.2018.02.012 198, 200, 203

[156] Mikhail Kennerley, Jian-Gang Wang, Bharadwaj Veeravalli, and Robby T. Tan. 2023. 2PCNet: Two-Phase Consistency Training for Day-to-Night Unsupervised Domain Adaptive Object Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Vancouver, Canada, 11484–11493. doi:10.1109/cvpr52729.2023.01105 133

[157] Charles F. Kettering. 1916. Engine-starting device. United States Patent Office, US 1,196,243 A, 9 pages. ttps://patents.google.com/patent/US1196243A 4

[158] Muhammad Qasim Khan and Sukhan Lee. 2019. A Comprehensive Survey of Driving Monitoring and Assistance Systems. *Sensors* 19, 11 (June 2019), 1–32. doi:10.3390/s19112574 198, 201, 203

[159] Jae Young Kim, Cheol Ho Jeong, Min Jae Jung, Jung Hyun Park, and Dae Hyun Jung. 2013. Highly reliable driving workload analysis using driver electroencephalogram (EEG) activities during driving. *International Journal of Automotive Technology* 14, 6 (November 2013), 965–970. doi:10.1007/s12239-013-0106-z 46, 48

[160] Soyeon Kim and Oscar Oviedo-Trespalacios. 2025. Disuse of advanced driver assistance systems (ADAS). *Journal of Safety Research* 95 (December 2025), 180–188. doi:10.1016/j.jsr.2025.09.004 13

[161] Katja Kircher and Christer Ahlström. 2016. Minimum Required Attention: A Human-Centered Approach to Driver Inattention. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 59, 3 (October 2016), 471–484. doi:10.1177/0018720816672756 54, 180

[162] Katja Kircher and Christer Ahlström. 2024. A comparison of glance coding approaches for driver attention assessment. *Transportation Research Part F: Traffic Psychology and Behaviour* 100 (January 2024), 243–253. doi:10.1016/j.trf.2023.12.003 58, 103

[163] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. 2020. PointRend: Image Segmentation As Rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), Seattle, Washington, USA, 9796–9805. doi:10.1109/cvpr42600.2020.00982 133

[164] Ivan Kobyzev, Simon J. D. Prince, and Marcus A. Brubaker. 2021. Normalizing Flows: An Introduction and Review of Current Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 11 (November 2021), 3964–3979. doi:10.1109/tpami.2020.2992934 155

[165] Andrey Nikolaevich Kolmogorov. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik und Ihrer Grenzgebiete, Vol. 2. Springer Berlin Heidelberg, Berlin, Germany. doi:10.1007/978-3-642-49888-6 158

[166] Andrey Nikolaevich Kolmogorov. 1950. *Foundations of the theory of probability*. Chelsea Publishing Company, New York City, New York, USA. https://archive.org/details/foundationsofthe00kolm 158

[167] Ina Marie Koniakowsky, Sara Jimenez Hamm, Yannick Forster, and Andreas Keinath. 2024. From Warnings to Workload: The Impact of Driver Monitoring System Warnings on Drivers' Self-Reported Workload. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Stanford, California, USA, 68–71. doi:10.1145/3641308.3685025 12, 178

[168] Moritz Körber, Andrea Cingel, Markus Zimmermann, and Klaus Bengler. 2015. Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving. *Procedia Manufacturing* 3 (2015), 2403–2409. doi:10.1016/j.promfg.2015.07.499 13

[169] Thomas Kosch, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. 2018. Look into my Eyes. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Montréal, Canada, 1–6. doi:10.1145/3170427.3188643 46, 48

[170] Iuliia Kotseruba, Amir Rasouli, and John K. Tsotsos. 2016. Joint Attention in Autonomous Driving (JAAD). *arXiv* abs/1609.04741 (2016), 1–10. arXiv:1609.04741 doi:10.48550/arXiv.1609.04741 180

[171] B. M. Kusuma Kumari and P. Ramakanth Kumar. 2017. A survey on drowsy driver detection system. In *International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*. IEEE, Chirala, Andhra Pradesh, India, 272–279. doi:10.1109/icbdaci.2017.8070847 198, 201, 203

[172] Saroj K. L. Lal and Ashley Craig. 2001. A critical review of the psychophysiology of driver fatigue. *Biological Psychology* 55, 3 (Feb. 2001), 173–194. doi:10.1016/s0301-0511(00)00085-5 198, 201, 203

[173] Hamed Laouz, Soheyb Ayad, and Labib Sadek Terrissa. 2020. Literature Review on Driver's Drowsiness and Fatigue Detection. In *International Conference on Intelligent Systems and Computer Vision (ISCV)*. IEEE, Fez, Morocco, 1–7. doi:10.1109/iscv49265.2020.9204306 198, 201, 203

[174] Anh Son Le, Tatsuya Suzuki, and Hirofumi Aoki. 2020. Evaluating driver cognitive distraction by eye tracking: From simulator to driving. *Transportation Research Interdisciplinary Perspectives* 4 (March 2020), 1–7. doi:10.1016/j.trip.2019.100087 46, 47, 51, 102

[175] T. Hoang Ngan Le, Kha Gia Quach, Chenchen Zhu, Chi Nhan Duong, Khoa Luu, and Marios Savvides. 2017. Robust Hand Detection and Classification in Vehicles and in the Wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Honolulu, Hawaii, USA, 1203–1210. doi:10.1109/CVPRW.2017.159 47, 56

[176] T. Hoang Ngan Le, Chenchen Zhu, Yutong Zheng, Khoa Luu, and Marios Savvides. 2016. Robust hand detection in Vehicles. In *International Conference on Pattern Recognition (ICPR)*. IEEE, Cancun, Mexico, 573–578. doi:10.1109/ICPR.2016.7899695 47, 56

[177] Chaeyoung Lee, MinJu Shin, David Eniyandunmo, Alvee Anwar, Eunsik Kim, Kyongwon Kim, Jae Keun Yoo, and Chris Lee. 2024. Predicting Driver's mental workload using physiological signals: A functional data analysis approach. *Applied Ergonomics* 118 (July 2024), 1–12. doi:10.1016/j.apergo.2024.104274 46, 48

[178] Chung Won Lee, Nasif Nayeer, Danson Evan Garcia, Ankur Agrawal, and Bingbing Liu. 2020. Identifying the Operational Design Domain for

an Automated Driving System through Assessed Risk. In *IEEE Intelligent Vehicles Symposium*, Vol. IV. IEEE, Las Vegas, Nevada, USA, 1317–1322. doi:10.1109/iv47402.2020.9304552 157

[179] Hanyong Lee, Woonki Na, and Seongkeun Park. 2026. Development Virtual Sensors for Vehicle In-Cabin Temperature Prediction Using Deep Learning. *Applied Sciences* 16, 1 (2026), 1–23. doi:10.3390/app16010300 47, 52

[180] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (January 2004), 50–80. doi:10.1518/hfes.46.1.50_30392 13

[181] Seungmin Lee, Dongwan Kim, Namil Kim, and Seong-Gyun Jeong. 2019. Drop to Adapt: Learning Discriminative Features for Unsupervised Domain Adaptation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Seoul, South Korea, 91–100. doi:10.1109/iccv.2019.00018 156

[182] Lennart Leicht, Erik Skobel, Marcel Mathissen, Steffen Leonhardt, Soren Weyer, Tobias Wartzek, Sebastian Reith, Werner Möhler, and Daniel Teichmann. 2015. Capacitive ECG recording and beat-to-beat interval estimation after major cardiac event. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Institute of Electrical and Electronics Engineers (IEEE), Milan, Italy, 7614–7617. doi:10.1109/EMBC.2015.7320155 47, 49

[183] Steffen Leonhardt, Lennart Leicht, and Daniel Teichmann. 2018. Unobtrusive Vital Sign Monitoring in Automotive Environments — A Review. *Sensors* 18, 9 (September 2018), 1–38. doi:10.3390/s18093080 198, 201, 203

[184] Jiaqi Li, Rui Xu, Jian Ma, Qiang Zou, Jun Ma, and Hao Yu. 2023. Domain Adaptive Object Detection for Autonomous Driving under Foggy Weather. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Institute of Electrical and Electronics Engineers (IEEE), Waikoloa, Hawaii, USA, 612–622. doi:10.1109/WACV56688.2023.00068 156

[185] Peixuan Li and Jieyu Jin. 2022. Time3D: End-to-End Joint Monocular 3D Object Detection and Tracking for Autonomous Driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New Orleans, Louisiana, USA, 3875–3884. doi:10.1109/cvpr52688.2022.00386 133

[186] Penghui Li, Yibing Li, Yao Yao, Changxu Wu, Bingbing Nie, and Shengbo Eben Li. 2022. Sensitivity of Electrodermal Activity Features for Driver Arousal Measurement in Cognitive Load: The Application in Automated Driving Systems. *IEEE Transactions on Intelligent Transportation Systems* 23, 9 (September 2022), 14954–14967. doi:10.1109/tits.2021.3135266 102, 117

[187] Renjun Li, Chu Liu, and Feng Luo. 2008. A design for automotive CAN bus monitoring system. In *IEEE Vehicle Power and Propulsion Conference*. IEEE, Harbin, China, 1–5. doi:10.1109/vppc.2008.4677544 47, 52, 57, 62

[188] Xuelong Li, Zhigang Wang, and Xiaoqiang Lu. 2017. A Multi-Task Framework for Weather Recognition. In *ACM International Conference on Multimedia (MM)*. ACM, Mountain View, California, USA, 1318–1326. doi:10.1145/3123266.3123382 153, 156

[189] Yang Li, Linbo Li, Daiheng Ni, and Yue Zhang. 2021. Comprehensive survival analysis of lane-changing duration. *Measurement* 182 (September 2021), 1–15. doi:10.1016/j.measurement.2021.109707 83

[190] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. 2017. Revisiting Batch Normalization For Practical Domain Adaptation. In *International Conference on Learning Representations (ICLR)*. International Conference on Learning Representations (ICLR), Toulon, France, 1–10. https://openreview.net/forum?id=Hk6dkJQFx 132

[191] Zhaojian Li, Shan Bao, Ilya V. Kolmanovsky, and Xiang Yin. 2018. Visual-Manual Distraction Detection Using Driving Performance Indicators With Naturalistic Driving Data. *IEEE Transactions on Intelligent Transportation Systems* 19, 8 (August 2018), 2528–2535. doi:10.1109/TITS.2017.2754467 46, 56

[192] Zhiqiang Li, Yingxiang Li, Jiandan Zhong, and Yongqiang Chen. 2020. Multi-class weather classification based on multi-feature weighted fusion method. In *Earth and Environmental Science (Journal of Physics: Conference Series, Vol. 558)*. IOP Publishing, Dalian, China, 13 pages. doi:10.1088/1755-1315/558/4/042038 156

[193] Jian Liang, Ran He, and Tieniu Tan. 2024. A Comprehensive Survey on Test-Time Adaptation Under Distribution Shifts. *International Journal of Computer Vision* 133, 1 (July 2024), 31–64. doi:10.1007/s11263-024-02181-w 132

[194] Yulan Liang and John D. Lee. 2010. Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention* 42, 3 (May 2010), 881–890. doi:10.1016/j.aap.2009.05.001 46, 58, 61, 94

[195] Yulan Liang, Michelle L. Reyes, and John D. Lee. 2007. Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines. *IEEE Transactions on Intelligent Transportation Systems* 8, 2 (June 2007), 340–350. doi:10.1109/tits.2007.895298 46, 61

[196] Yuan Liao, Shengbo Eben Li, Wenjun Wang, Ying Wang, Guofa Li, and Bo Cheng. 2016. Detection of Driver Cognitive Distraction: A Comparison Study of Stop-Controlled Intersection and Speed-Limited Highway. *IEEE Transactions on Intelligent Transportation Systems* 17, 6 (June 2016), 1628–1637. doi:10.1109/tits.2015.2506602 46, 48, 61

[197] Zijian Lin and Feng Chen. 2024. How various urgencies and visibilities influence drivers' takeover performance in critical car-following conditions? A driving simulation study. *Transportation Research Part F: Traffic Psychology and Behaviour* 104 (July 2024), 303–317. doi:10.1016/j.trf.2024.06.007 75

[198] Zhiqiu Lin, Jia Shi, Deepak Pathak, and Deva Ramanan. 2021. The CLEAR Benchmark: Continual LEArning on Real-World Imagery. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, Vol. 1. Curran Associates, Inc., Virtual conference, 1–13. https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/2838023a778dfaecdc212708f721b788-Abstract-round2.html 131

[199] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. 2020. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* 23, 1 (December 2020), 1–45. doi:10.3390/e23010018 65

[200] Chenghao Liu, Steven C.H. Hoi, Peilin Zhao, and Jianling Sun. 2016. Online ARIMA Algorithms for Time Series Prediction. In *AAAI Conference on Artificial Intelligence*, Vol. 30. Association for the Advancement of Artificial Intelligence (AAAI), Phoenix, Arizona USA, 1867–1873. doi:10.1609/aaai.v30i1.10257 131

[201] Fan Liu, Xueyi Li, Tanyue Lv, and Feng Xu. 2019. A Review of Driver Fatigue Detection: Progress and Prospect. In *IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, Las Vegas, Nevada, USA, 1–6. doi:10.1109/icce.2019.8662098 198, 201, 203

[202] Shaoshan Liu, Liangkai Liu, Jie Tang, Bo Yu, Yifan Wang, and Weisong Shi. 2019. Edge Computing for Autonomous Driving: Opportunities and Challenges. *Proc. IEEE* 107, 8 (August 2019), 1697–1716. doi:10.1109/jproc.2019.2915983 133

[203] Tyron Louw and Natasha Merat. 2017. Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies* 76 (March 2017), 35–50. doi:10.1016/j.trc.2017.01.001 120

[204] Cewu Lu, Di Lin, Jiaya Jia, and Chi-Keung Tang. 2017. Two-Class Weather Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 12 (December 2017), 2510–2524. doi:10.1109/tpami.2016.2640295 156

[205] Andreas Lüdtke, Denis Javaux, Fabio Tango, Rainer Heers, Klaus Bengler, and Catherine Ronfle-Nadaud. 2012. Designing dynamic distributed cooperative Human-Machine Systems. *WORK: A Journal of Prevention, Assessment & Rehabilitation* 41, S1 (January 2012), 4250–4257. doi:10.3233/wor-2012-0719-4250 185

[206] Jan-Matthis Lueckmann, Pedro J. Gonçalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H. Macke. 2017. Flexible statistical inference for mechanistic models of neural dynamics. In *Advances in Neural Information Processing Systems (NeurIPS)*. Curran Associates, Inc., Long Beach, California, USA, 1289–1299. https://papers.nips.cc/paper/6728-flexible-statistical-inference-for-mechanistic-models-of-neural-dynamics 157

[207] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 2017. 1 year, 1000 km: The Oxford RobotCar dataset. *International Journal of Robotics Research* 36, 1 (January 2017), 3–15. doi:10.1177/0278364916679498 137

[208] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods* 53, 4 (Feb. 2021), 1689–1696. doi:10.3758/s13428-020-01516-y 109

[209] Yishay Mansour, Mehryar Mohri, and Rostamizadeh Rostamizadeh. 2008. Domain Adaptation with Multiple Sources. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 21. Curran Associates, Inc., Vancouver, Canada, 1041–1048. https://papers.nips.cc/paper_files/paper/2008/hash/0e65972dce68dad4d52d063967f0a705-Abstract.html 155, 169

[210] Clara Marina Martinez, Mira Heucke, Fei-Yue Wang, Bo Gao, and Dongpu Cao. 2018. Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 19, 3 (March 2018), 666–676. doi:10.1109/tits.2017.2706978 37, 198, 201, 203

[211] Gerhard Marquart, Christopher Cabrall, and Joost de Winter. 2015. Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing* 3 (2015), 2854–2861. doi:10.1016/j.promfg.2015.07.783 45, 46, 102, 103, 198, 201, 203

[212] Gerhard Marquart and Joost de Winter. 2015. Workload assessment for mental arithmetic tasks using the task-evoked pupillary response. *PeerJ Computer Science* 1 (August 2015), 1–20. doi:10.7717/peerj-cs.16 45

[213] Alina Mashko. 2015. Review of approaches to the problem of driver fatigue and drowsiness. In *Smart Cities Symposium Prague (SCSP)*. IEEE, Prague, Czech Republic, 1–5. doi:10.1109/SCSP.2015.7181569 198, 201, 203

[214] Dishita Mashru and Vaibhav Gandhi. 2018. Detection of a Drowsy state of the Driver on road using wearable sensors: A survey. In *International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, Coimbatore, India, 691–695. doi:10.1109/icicct.2018.8473245 198, 201, 203

[215] Sarfaraz Masood, Abhinav Rai, Aakash Aggarwal, M.N. Doja, and Musheer Ahmad. 2020. Detecting distraction of drivers using Convolutional Neural Network. *Pattern Recognition Letters* 139 (November 2020), 79–85. doi:10.1016/j.patrec.2017.12.023 47, 57

[216] James G. May, Robert S. Kennedy, Mary C. Williams, William P. Dunlap, and Julie R. Brannan. 1990. Eye movement indices of mental workload. *Acta Psychologica* 75, 1 (October 1990), 75–89. doi:10.1016/0001-6918(90)90067-p 46, 48

[217] MEDIATOR. 2019. Mediating between Driver and Intelligent Automated Transport systems on Our Roads. doi:10.3030/814735 185

[218] Vadim Melnicuk, Stewart Birrell, Elizabeth Crundall, and Paul Jennings. 2016. Towards hybrid driver state monitoring: review, future perspectives and the role of consumer electronics. In *IEEE Intelligent Vehicles Symposium*, Vol. IV. IEEE, Gothenburg, Sweden, 1392–1397. doi:10.1109/IVS.2016.7535572 37, 198, 201, 203

[219] Amanda Meyerson, Johanna Eklind, Florian Fischer, Maytheewat Aramrattana, Ingemar Fredriksson, and Christer Ahlström. 2024. Effects of daylight and darkness at daytime versus nighttime on driver sleepiness: A driving simulator study. *Transportation Research Interdisciplinary Perspectives* 24 (March 2024), 101087. doi:10.1016/j.trip.2024.101087 76

[220] Alain Micaelli and Claude Samson. 1993. *Trajectory tracking for unicycle-type and two-steering-wheels mobile robots*. Technical Report. INRIA. https://inria.hal.science/inria-00074575 Research report RR-2097. 79

[221] Eva Michelaraki, Christos Katrakazas, Susanne Kaiser, Tom Brijs, and George Yannis. 2023. Real-time monitoring of driver distraction: State-of-the-art and future insights. *Accident Analysis & Prevention* 192 (November 2023), 1–15. doi:10.1016/j.aap.2023.107241 12

[222] David Miller, Annabel Sun, Mishel Johns, Hillary Ive, David Sirkin, Sudipto Aich, and Wendy Ju. 2016. Distraction Becomes Engagement in Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59, 1 (2016), 1676–1680. doi:10.1177/1541931215591362 33

[223] M. Jehanzeb Mirza, Jakub Micorek, Horst Possegger, and Horst Bischof. 2022. The Norm Must Go On: Dynamic Unsupervised Domain Adaptation by Normalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New Orleans, Louisiana, USA, 14745–14755. doi:10.1109/cvpr52688.2022.01435 132

[224] Ajay Mittal, Kanika Kumar, Sarina Dhamija, and Manvjeet Kaur. 2016. Head movement-based driver drowsiness detection: A review of state-of-art techniques. In *Parul International Conference on Engineering and Technology (PICET)*. IEEE, Coimbatore, India, 903–908. doi:10.1109/icetech.2016.7569378 198, 201, 203

[225] Adam A. Q. Mohammed, Xin Geng, Jing Wang, and Zafar Ali. 2024. Driver distraction detection using semi-supervised lightweight vision transformer. *Engineering Applications of Artificial Intelligence* 129 (March 2024), 1–16. doi:10.1016/j.engappai.2023.107618 57

[226] Zalhan Mohd Zin, Anis Hazirah Rodzi, and Norazlin Ibrahim. 2018. Vision based eye closeness classification for driver's distraction and drowsiness using PERCLOS and support vector machines. In *International Conference on Machine Vision (ICMV) (Proceedings of SPIE, Vol. 11041)*. SPIE, Munich, Germany, 57. doi:10.1117/12.2522949 47

[227] Sankha S. Mukherjee and Neil Martin Robertson. 2015. Deep Head Pose: Gaze-Direction Estimation in Multimodal Video. *IEEE Transactions on Multimedia* 17, 11 (November 2015), 2094–2107. doi:10.1109/tmm.2015.2482819 47, 59

[228] Ravi Teja Mullapudi, Steven Chen, Keyi Zhang, Deva Ramanan, and Kayvon Fatahalian. 2019. Online Model Distillation for Efficient Video Inference. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Seoul, South Korea, 3572–3581. doi:10.1109/iccv.2019.00367 132

[229] Suganiya Murugan, Jerritta Selvaraj, and Arun Sahayadhas. 2019. Analysis of Different Measures to Detect Driver States: A Review. In *IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*. IEEE, Pondicherry, India, 1–6. doi:10.1109/icscan.2019.8878844 198, 201, 203

[230] Antonyo Musabini and Mounsif Chetitah. 2020. Heatmap-Based Method for Estimating Drivers' Cognitive Distraction. In *International Conference on Cognitive Informatics & Cognitive Computing (ICCICC)*. IEEE, Beijing, China, 179–186. doi:10.1109/iccicc50026.2020.9450216 47, 51

[231] Indu Nair, N. Ebrahimkutty, B. Priyanka, M. Sreeja, and D. Gopu. 2016. A survey on driver fatigue-drowsiness detection system. *International Journal of Engineering and Computer Science* 5, 11 (November 2016), 19237–19240. doi:10.18535/ijecs/v5i11.92 37, 198, 201, 203

[232] Rizwan Naqvi, Muhammad Arsalan, Ganbayar Batchuluun, Hyo Yoon, and Kang Park. 2018. Deep Learning-Based Gaze Detection System for Automobile Drivers Using a NIR Camera Sensor. *Sensors* 18, 2 (Feb. 2018), 1–34. doi:10.3390/s18020456 47, 59

[233] Frederik Naujoks, Yannick Forster, Katharina Wiedemann, and Alexandra Neukum. 2016. A Human-Machine Interface for Cooperative Highly Automated Driving. In *Advances in Human Aspects of Transportation (Advances in Intelligent Systems and Computing, Vol. 484)*. Springer International Publishing, Orlando, Florida, USA, 585–595. doi:10.1007/978-3-319-41682-3_49 75

[234] Frederik Naujoks, Simon Höfling, Christian Purucker, and Kathrin Zeeb. 2018. From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. *Accident Analysis & Prevention* 121 (December 2018), 28–42. doi:10.1016/j.aap.2018.08.018 33

[235] Frederik Naujoks, Katharina Wiedemann, Nadja Schömig, Sebastian Hergeth, and Andreas Keinath. 2019. Towards guidelines and verification methods for automated vehicle HMIs. *Transportation Research Part F: Traffic Psychology and Behaviour* 60 (January 2019), 121–136. doi:10.1016/j.trf.2018.10.012 177

[236] Mkhuseli Ngxande, Jules-Raymond Tapamo, and Michael Burke. 2017. Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques. In *Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*. IEEE, Bloemfontein, South Africa, 156–161. doi:10.1109/robomech.2017.8261140 199, 201, 204

[237] NHTSA. 2010. *Overview of the National Highway Traffic Safety Administration's Driver Distraction Program*. Technical Report. National Highway Traffic Safety Administration, Washington, District of Columbia, USA. DOT HS 811 299. 60, 102

[238] Fjollë Novakazi, Mikael Johansson, Helena Strömberg, and MariAnne Karlsson. 2021. Levels of What? Investigating Drivers' Understanding of Different Levels of Automation in Vehicles. *Journal of Cognitive Engineering and Decision Making* 15, 2-3 (April 2021), 116–132. doi:10.1177/15553434211009024 73

[239] Andrea Němcová, Veronika Svozilová, Katerina Bucsuházy, Radovan Smišek, Martin Mézl, Branislav Hesko, Michal Belák, Martin Bilik, Pavel Maxera, Martin Seitl, Tomas Dominik, Marek Semela, Matus Šucha, and Radim Kolář. 2021. Multimodal Features for Detection of Driver Stress and Fatigue: Review. *IEEE Transactions on Intelligent Transportation Systems* 22, 6 (June 2021), 3214–3233. doi:10.1109/tits.2020.2977762 198, 201, 203

[240] Roger D. O'Donnel and Fred Thomas Eggemeier. 1986. *Workload assessment methodology*. John Wiley & Sons, New York City, New York, USA, Chapter 42, 1–49. 45

[241] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2024. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research* 1 (2024), 1–32. https://openreview.net/forum?id=a68SUt6zFt 161

[242] Francesco Orabona. 2019. A Modern Introduction to Online Learning. *arXiv* abs/1912.13213 (2019), 225. arXiv:1912.13213 doi:10.48550/arXiv.1912.13213 131

[243] Julia Orlovska, Fjollë Novakazi, Bligård Lars-Ola, MariAnne Karlsson, Casper Wickman, and Rikard Söderberg. 2020. Effects of the driving context on the usage of Automated Driver Assistance Systems (ADAS) – Naturalistic Driving Study for ADAS evaluation. *Transportation Research Interdisciplinary Perspectives* 4 (March 2020), 1–16. doi:10.1016/j.trip.2020.100093 77, 177

[244] Oscar Oviedo-Trespalacios, Md. Mazharul Haque, Mark King, and Simon Washington. 2016. Understanding the impacts of mobile phone distraction on driving performance: A systematic review. *Transportation Research Part C: Emerging Technologies* 72 (November 2016), 360–380. doi:10.1016/j.trc.2016.10.006 199, 201, 204

[245] Petar Palasek, Nilli Lavie, and Luke Palmer. 2019. Attentional demand estimation with attentive driving models. In *British Machine Vision Conference (BMVC)*. British Machine Vision Association (BMVA), Cardiff, Wales, 1–13. 46, 47, 48, 52

[246] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. 2021. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research* 22, 57 (2021), 1–64. 155, 157

[247] Panagiotis Papantoniou, Eleonora Papadimitriou, and George Yannis. 2017. Review of driving performance parameters critical for distracted driving research. *Transportation Research Procedia* 25 (2017), 1796–1805. doi:10.1016/j.trpro.2017.05.148 199, 201, 204

[248] Giuseppe Pappalardo, Raffaele Caponetto, Rosaria Varrica, and Salvatore Cafiso. 2022. Assessing the Operational Design Domain of Lane Support System for Automated Vehicles in Different Weather and Road Conditions. *Journal of Traffic and Transportation Engineering* 9, 4 (August 2022), 631–644. doi:10.1016/j.jtte.2021.12.002 157

[249] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, 2 (June 1997), 230–253. doi:10.1518/001872097778543886 13

[250] Sami Park, Yilun Xing, Kumar Akash, Teruhisa Misu, and Linda Ng Boyle. 2022. The Impact of Environmental Complexity on Drivers' Situation Awareness. In *International Conference on Automotive User In-*

*terfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Seoul, South Korea, 131–138. doi:10.1145/3543174.3546831 76

[251] Julie Paxion, Edith Galy, and Catherine Berthelon. 2014. Mental workload and driving. *Frontiers in Psychology* 5 (December 2014), 1–11. doi:10.3389/fpsyg.2014.01344 45, 46, 102

[252] William Payre, Julien Cestac, and Patricia Delhomme. 2015. Fully Automated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, 2 (December 2015), 229–241. doi:10.1177/0018720815612319 13

[253] Jakob Peintner, Bengt Escher, Henrik Detjen, Carina Manger, and Andreas Riener. 2024. How to Design Human-Vehicle Cooperation for Automated Driving: A Review of Use Cases, Concepts, and Interfaces. *Multimodal Technologies and Interaction* 8, 3 (Feb. 2024), 1–30. doi:10.3390/mti8030016 185

[254] Adriaan Perrels, Athanasios Votsis, Veikko Nurmi, and Kari Pilli-Sihvola. 2015. Weather Conditions, Weather Information and Car Crashes. *ISPRS International Journal of Geo-Information* 4, 4 (November 2015), 1–23. doi:10.3390/ijgi4042681 153

[255] Bastian Pfleging, Drea K. Fekety, Albrecht Schmidt, and Andrew L. Kun. 2016. A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Santa Clara, California, USA, 5776–5788. doi:10.1145/2858036.2858117 46, 48

[256] Sébastien Piérard, Anthony Cioppa, Anaïs Halin, Renaud Vandeghen, Maxime Zanella, Benoît Macq, Saïd Mahmoudi, and Marc Van Droogenbroeck. 2023. Mixture Domain Adaptation to Improve Semantic Segmentation in Real-World Surveillance. In *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*. Institute of Electrical and Electronics Engineers (IEEE), Waikoloa, Hawaii, USA, 22–31. doi:10.1109/wacvw58289.2023.00007 22, 133, 155, 158, 169, 170

[257] Sébastien Piérard, Adrien Deliège, Anaïs Halin, and Marc Van Droogenbroeck. 2025. A Methodology to Evaluate Strategies Predicting Rankings on Unseen Domains. In *IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Workshop on Big Surveillance Data Analysis and Processing (BIG-Surv)*. IEEE, Nantes, France, 1–6. doi:10.1109/icmew68306.2025.11152102 22

[258] Sébastien Piérard, Anaïs Halin, Anthony Cioppa, Adrien Deliège, and Marc Van Droogenbroeck. 2024. The Tile: A 2D Map of Ranking Scores for Two-Class Classification. *arXiv* abs/2412.04309 (2024), 29 pages. arXiv:2412.04309 doi:10.48550/arXiv.2412.04309 22

[259] Sébastien Piérard, Anaïs Halin, Anthony Cioppa, Adrien Deliège, and Marc Van Droogenbroeck. 2025. Foundations of the Theory of Performance-Based Ranking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Nashville, Tennessee, USA, 14293–14302. doi:10.1109/cvpr52734.2025.01333 22

[260] Sébastien Piérard, Alejandro Marcos Alvarez, Antoine Lejeune, and Marc Van Droogenbroeck. 2014. On-the-fly domain adaptation of binary classifiers. In *Proceedings of the BENELEARN Conference*. Belgian-Dutch Conference on Machine Learning (BENELEARN), Brussels, Belgium, 20–28. 169

[261] Peng Ping, Cong Huang, Weiping Ding, Yongkang Liu, Miyajima Chiyomi, and Takeda Kazuya. 2023. Distracted driving detection based on the fusion of deep learning and causal reasoning. *Information Fusion* 89 (January 2023), 121–142. doi:10.1016/j.inffus.2022.08.009 57

[262] Chittathuru Himala Praharsha and Alwin Poulose. 2024. CBAM VGG16: An efficient driver distraction classification using CBAM embedded VGG16 architecture. *Computers in Biology and Medicine* 180 (September 2024), 13 pages. doi:10.1016/j.compbiomed.2024.108945 57

[263] Bagus G. Pratama, Igi Ardiyanto, and Teguh B. Adji. 2017. A review on driver drowsiness based on image, bio-signal, and driver behavior. In *International Conference on Information Science and Technology (ICIST)*. IEEE, Yogyakarta, Indonesia, 70–75. doi:10.1109/icstc.2017.8011855 199, 201, 204

[264] Marie-Anne Pungu Mwange, Fabien Rogister, and Luka Rukonic. 2022. Measuring driving simulator adaptation using EDA. *Human Factors and Simulation* 30 (2022), 35–43. doi:10.54941/ahfe1001489 106

[265] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 139)*. ML Research Press, Virtual conference, 8748–8763. https://proceedings.mlr.press/v139/radford21a.html 161

[266] Vishnu Radhakrishnan, Tyron Louw, Rafael Cirino Gonçalves, Guilhermina Torrao, Michael G. Lenné, and Natasha Merat. 2023. Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 94 (April 2023), 254–267. doi:10.1016/j.trf.2023.02.015 46, 47, 48, 50

[267] Vishnu Radhakrishnan, Natasha Merat, Tyron Louw, Rafael Cirino Gonçalves, Guilhermina Torrao, Wei Lyu, Pablo Puente Guillen, and Michael G. Lenné. 2022. Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 87 (May 2022), 149–163. doi:10.1016/j.trf.2022.04.002 45, 46, 50

[268] Vishnu Radhakrishnan, Natasha Merat, Tyron Louw, Michael G. Lenné, Richard Romano, Evangelos Paschalidis, Foroogh Hajiseyedjavadi, Chongfeng Wei, and Erwin R. Boer. 2020. Measuring Drivers' Physiological Response to Different Vehicle Controllers in Highly Automated Driving (HAD): Opportunities for Establishing Real-Time Values of Driver Discomfort. *Information* 11, 8 (August 2020), 1–15. doi:10.3390/info11080390 103, 119

[269] Jonas Radlmayr, Christian Gold, Lutz Lorenz, Mehdi Farid, and Klaus Bengler. 2014. How Traffic Situations and Non-Driving Related Tasks Affect the Take-Over Quality in Highly Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, 1 (September 2014), 2063–2067. doi:10.1177/1541931214581434 75

[270] Muhammad Ramzan, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Mahwish Ilyas, and Ahsan Mahmood. 2019. A Survey on State-of-the-Art Drowsiness Detection Techniques. *IEEE Access* 7 (2019), 61904–61919. doi:10.1109/access.2019.2914373 199, 201, 204

[271] Thomas A. Ranney. 2008. *Driver distraction: A review of the current state-of-knowledge*. Technical Report. National Highway Traffic Safety Administration. 46, 61, 94

[272] Thomas A. Ranney, Elizabeth Mazzae, Riley Garrott, and Michael J. Goodman. 2000. *NHTSA driver distraction research: Past, present, and future*. Technical Report. SAE. 46, 53, 58

[273] Amir Rasouli and John K. Tsotsos. 2020. Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Transactions on Intelligent Transportation Systems* 21, 3 (March 2020), 900–918. doi:10.1109/tits.2019.2901817 180

[274] Julien Rebut, Andrei Bursuc, and Patrick Perez. 2021. StyleLess Layer: Improving Robustness for Real-World Driving. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Institute of Electrical and Electronics Engineers (IEEE), Prague, Czech Republic, 8992–8999. doi:10.1109/IROS51168.2021.9636204 156

[275] Nikhil Reddy, Abhinav Singhal, Abhishek Kumar, Mahsa Baktashmotlagh, and Chetan Arora. 2022. Master of All: Simultaneous Generalization of Urban-Scene Segmentation to All Adverse Weather Conditions. In *European Conference on Computer Vision (ECCV) (Lecture Notes in Computer Science, Vol. 13699)*. Springer Nature Switzerland, Milan, Italy, 51–69. doi:10.1007/978-3-031-19842-7_4 132

[276] Michael A. Regan, Charlene Hallett, and Craig P. Gordon. 2011. Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention* 43, 5 (September 2011), 1771–1781. doi:10.1016/j.aap.2011.04.008 53

[277] Mary A. Regan, John D. Lee, and Thomas Young. 2008. *Driver distraction: theory, effects, and mitigation*. CRC Press, Boca Raton, Florida, USA. 53

[278] Bryan Reimer, Bruce Mehler, Joseph F. Coughlin, Kathryn M. Godfrey, and Chuanzhong Tan. 2009. An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers. In *International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Essen, Germany, 115–118. doi:10.1145/1620509.1620531 46, 48, 102

[279] John T. E. Richardson. 2011. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6, 2 (January 2011), 135–147. doi:10.1016/j.edurev.2010.12.001 84, 116

[280] Colin Robertson, Jed A. Long, Farouk S. Nathoo, Trisalyn A. Nelson, and Cameron C. F. Plouffe. 2014. Assessing Quality of Spatial Models Using the Structural Similarity Index and Posterior Predictive Checks. *Geographical Analysis* 46, 1 (January 2014), 53–74. doi:10.1111/gean.12028 157

[281] François Rozet, Felix Divo, and Simon Schnake. 2022. Zuko: Normalizing flows in PyTorch. Software. doi:10.5281/ZENODO.7625672 155, 161

[282] François Rozet, Benjamin Kurt Miller, and Arnaud Delaunoy. 2021. LAMPE: Likelihood-free Amortized Posterior Estimation. Software. doi:10.5281/ZENODO.8405782 161

[283] SAE International. 2021. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Technical Report SAE Standard J3016 202104. Society of Automobile Engineers, Warrendale, PA, USA. 1–41 pages. doi:10.4271/J3016_202104 6, 12, 31, 73, 75, 76, 153, 177

[284] Arun Sahayadhas, Kenneth Sundaraj, and Murugappan Murugappan. 2012. Detecting Driver Drowsiness Based on Sensors: A Review. *Sensors* 12, 12 (December 2012), 16937–16953. doi:10.3390/s121216937 37, 199, 202, 204

[285] Samuel Salvaggio, Nicolas Masson, Alexandre Zénon, and Michael Andres. 2022. The predictive role of eye movements in mental arithmetic. *Experimental Brain Research* 240, 5 (March 2022), 1331–1340. doi:10.1007/s00221-022-06329-3 116, 118

[286] Mark S. Sanders and Ernest J. McCormick. 1992. *Human Factors in Engineering and Design*. Mcgraw-Hill Book Company, New York City, New York, USA. 45

[287] Ágoston Pál Sándor. 2025. Analysis Of The Relation Between Vehicle And Physiological Data. *Transportation Research Procedia* 91 (2025), 369–376. doi:10.1016/j.trpro.2025.10.048 49

[288] Sergio Saponara, Maria Sabrina Greco, and Fulvio Gini. 2019. Radar-on-Chip/in-Package in Autonomous Driving Vehicles and Intelligent Transport Systems: Opportunities and Challenges. *IEEE Signal Processing Magazine* 36, 5 (September 2019), 71–84. doi:10.1109/msp.2019.2909074 47

[289] Noud Schaap, Richard A. van der Horst, Bart van Arem, and Karel A. Brookhuis. 2013. *The relationship between driver distraction and mental workload*. Vol. 1. shgate Publishing, Farnham, UK, 147–166. 45, 46, 48

[290] Peter Schafhalter, Sukrit Kalra, Le Xu, Joseph E. Gonzalez, and Ion Stoica. 2023. Leveraging Cloud Computing to Make Autonomous Vehicles Safer. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Detroit, MI, USA, 5559–5566. doi:10.1109/iros55552.2023.10341821 133

[291] Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. 2020. Improving robustness against common corruptions by covariate shift adaptation. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33. Curran Associates, Inc., Virtual conference, 11539–11551. 132

[292] Erich Schröger, M.-H Giard, and Ch Wolff. 2000. Auditory distraction: event-related potential and behavioral indices. *Clinical Neurophysiology* 111, 8 (August 2000), 1450–1460. doi:10.1016/s1388-2457(00)00337-0 46, 60

[293] Bridie J. Scott-Parker. 2017. Emotions, behaviour, and the adolescent driver: A literature review. *Transportation Research Part F: Traffic Psychology and Behaviour* 50 (October 2017), 1–37. doi:10.1016/j.trf.2017.06.019 199, 202, 204

[294] Ishita Seth. 2020. A Survey on Driver Behavior Detection Techniques. *International Journal of Scientific Research in Science and Technology* 7, 3 (June 2020), 401–404. doi:10.32628/ijsrst207384 199, 202, 204

[295] Tina Sever and Giuseppe Contissa. 2024. Automated driving regulations - where are we now? *Transportation Research Interdisciplinary Perspectives* 24 (March 2024), 1–19. doi:10.1016/j.trip.2024.101033 73, 75

[296] Shai Shalev-Shwartz. 2011. Online Learning and Online Convex Optimization. *Foundations and Trends®in Machine Learning* 4, 2 (2011), 107–194. doi:10.1561/2200000018 131

[297] Z. Shameen, Mohd Zuki Yusoff, Mohamad Saad, Aamir Malik, and Muhammad Muzammel. 2018. Electroencephalography (EEG) based drowsiness detection for drivers: A review. *ARPN Journal of Engineering and Applied Sciences* 13, 4 (Feb. 2018), 1458–1464. 199, 202, 204

[298] Thomas B. Sheridan, William L. Verplank, and Thurston L. Brooks. 1978. Human and Computer Control of Undersea Teleoperators. In *NASA Annual Conference on Manual Control*. National Aeronautics and Space Administration (NASA), Los Angeles, California, USA, 343–357. https://ntrs.nasa.gov/citations/19790007441 185

[299] Ali Shoker, Rehana Yasmin, and Paulo Esteves-Verissimo. 2024. WIP: Savvy: Trustworthy Autonomous Vehicles Architecture. In *Symposium on Vehicle Security Privacy (VehicleSec)*. Internet Society, San Diego, California, USA, 1–7. doi:10.14722/vehiclesec.2024.23058 168

[300] Mohamad-Hoseyn Sigari, Muhammad-Reza Pourshahabi, Mohsen Soryani, and Mahmood Fathy. 2014. A Review on Driver Face Monitoring Systems for Fatigue and Distraction Detection. *International Journal of Advanced Science and Technology* 64 (March 2014), 73–100. doi:10.14257/ijast.2014.64.07 199, 202, 204

[301] Gulbadan Sikander and Shahzad Anwar. 2019. Driver Fatigue Detection Systems: A Review. *IEEE Transactions on Intelligent Transportation Systems* 20, 6 (June 2019), 2339–2352. doi:10.1109/tits.2018.2868499 199, 202, 204

[302] Hugo Silva, André Lourenço, and Ana Fred. 2012. In-vehicle driver recognition based on hand ECG signals. In *Proceedings of the ACM international conference on Intelligent User Interfaces (IUI)*. ACM, Lisbon, Portugal, 25–28. doi:10.1145/2166966.2166971 47, 49

[303] Harpreet Singh and Ankit Kathuria. 2021. Analyzing driver behavior under naturalistic driving conditions: A review. *Accident Analysis & Prevention* 150 (Feb. 2021), 1–21. doi:10.1016/j.aap.2020.105908 199, 202, 204

[304] S. Singh. 2018. *Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey*. Technical Report. National Highway Traffic Safety Administration, Washington, District of Columbia, USA. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812506 29

[305] Jaka Sodnik, Christina Dicke, Sašo Tomažič, and Mark Billinghurst. 2008. A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies* 66, 5 (May 2008), 318–332. doi:10.1016/j.ijhcs.2007.11.001 59

[306] Le Anh Son, Tatsuya Suzuki, and Hirofumi Aoki. 2018. Evaluation of Cognitive Distraction in a Real Vehicle Based on the Reflex Eye Movement. *International Journal of Automotive Engineering* 9, 1 (2018), 1–8. doi:10.20485/jsaeijae.9.1_1 46, 61

[307] Andreas Sonnleitner, Matthias Sebastian Treder, Michael Simon, Sven Willmann, Arne Ewald, Axel Buchner, and Michael Schrauf. 2014. EEG alpha spindles and prolonged brake reaction times during auditory distraction in an on-road driving study. *Accident Analysis & Prevention* 62 (January 2014), 110–118. doi:10.1016/j.aap.2013.08.026 46, 60

[308] Neville A. Stanton and Mark S. Young. 2005. Driver behaviour with adaptive cruise control. *Ergonomics* 48, 10 (August 2005), 1294–1313. doi:10.1080/00140130500252990 76

[309] David L. Strayer, Joel M. Cooper, Jonna Turrill, James R. Coleman, Nathan Medeiros-Ward, and Francesco Biondi. 2013. *Measuring cognitive distraction in the automobile*. Technical Report. AAA, Foundation for Traffic Safety, Washington, District of Columbia, USA. 46, 61

[310] David L. Strayer and Frank A. Drews. 2007. Cell-Phone-Induced Driver Distraction. *Current Directions in Psychological Science* 16, 3 (June 2007), 128–131. doi:10.1111/j.1467-8721.2007.00489.x 46, 62

[311] David L. Strayer, Jonna Turrill, Joel M. Cooper, James R. Coleman, Nathan Medeiros-Ward, and Francesco Biondi. 2015. Assessing Cognitive Distraction in the Automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, 8 (November 2015), 1300–1324. doi:10.1177/0018720815575149 46, 61, 102

[312] D. Venkata Subbaiah, Prasad Reddy, and K. Venkata Rao. 2019. Driver Drowsiness Detection Methods: A Comprehensive survey. *International Journal of Research in Advent Technology* 7, 3 (March 2019), 992–997. doi:10.32622/ijrat.73201918 199, 202, 204

[313] Pb Sujit, Srikanth Saripalli, and Joao Borges Sousa. 2014. Unmanned Aerial Vehicle Path Following: A Survey and Analysis of Algorithms for Fixed-Wing Unmanned Aerial Vehicles. *IEEE Control Systems Magazine* 34, 1 (Feb. 2014), 42–59. doi:10.1109/MCS.2013.2287568 79

[314] Chen Sun, Zejian Deng, Wenbo Chu, Shen Li, and Dongpu Cao. 2022. Acclimatizing the Operational Design Domain for Autonomous Driving Systems. *IEEE Intelligent Transportation Systems Magazine* 14, 2 (March 2022), 10–24. doi:10.1109/mits.2021.3070651 157

[315] Tao Sun, Mattia Segu, Janis Postels, Yuxuan Wang, Luc Van Gool, Bernt Schiele, Federico Tombari, and Fisher Yu. 2022. SHIFT: A Synthetic Driving Dataset for Continuous Multi-Task Domain Adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), New Orleans, Louisiana, USA, 21339–21350. doi:10.1109/cvpr52688.2022.02068 137

[316] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei A. Efros, and Moritz Hardt. 2020. Test-Time Training with Self-Supervision for Generalization under Distribution Shifts. In *International Conference on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 119)*. ML Research Press, Virtual conference, 9229–9248. http://proceedings.mlr.press/v119/sun20b.html 132

[317] Hui Tang, Ke Chen, and Kui Jia. 2020. Unsupervised Domain Adaptation via Structurally Regularized Deep Clustering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), Seattle, WA, USA, 8722–8732. doi:10.1109/cvpr42600.2020.00875 155

[318] Ralph R. Teetor. 1950. Speed control device for resisting operation of the accelerator. United States Patent Office, US 2,519,859 A, 9 pages. https://patents.google.com/patent/US2519859A 4

[319] Evona Teh, Samantha Jamson, Oliver Carsten, and Hamish Jamson. 2014. Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance. *Transportation Research Part F: Traffic Psychology and Behaviour* 22 (January 2014), 207–217. doi:10.1016/j.trf.2013.12.005 13, 177

[320] Jan Theeuwes and Hans Godthelp. 1995. Self-explaining roads. *Safety Science* 19, 2-3 (June 1995), 217–225. doi:10.1016/0925-7535(94)00022-u 186

[321] Louis Tijerina. 2000. Issues in the evaluation of driver distraction associated with in-vehicle information and telecommunications systems. *Transportation Research Inc* 12 (2000), 54–67. 46, 56

[322] Abhimanyu Tonk and Abderraouf Boussif. 2022. Operational Design Domain or Operational Envelope; Seeking a Suitable Concept for Autonomous Railway Systems. In *Proceedings of the European Safety and Reliability Conference (ESREL)*. Research Publishing Services, Dublin, Ireland, 2104–2111. doi:10.3850/978-981-18-5183-4_s06-08-245-cd 155

[323] Wei Tu, Lei Wei, Wenyan Hu, Zhengguo Sheng, Hasen Nicanfar, Xiping Hu, Edith C.-H. Ngai, and Victor C. M. Leung. 2016. A Survey on Mobile Sensing Based Mood-Fatigue Detection for Drivers. In *Smart City* $360^0$. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 166. Springer International Publishing, Toronto, Ontario, Canada, 3–15. doi:10.1007/978-3-319-33681-7_1 37, 199, 202, 204

[324] Chiagoziem C. Ukwuoma and Chen Bo. 2019. Deep Learning Review on Drivers Drowsiness Detection. In *Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*. IEEE, Bangkok, Thailand, 1–5. doi:10.1109/times-icon47539.2019.9024642 199, 202, 204

[325] Marieke van Dooren, Gert-Jan de Vries, and Joris H. Janssen. 2012. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior* 106, 2 (May 2012), 298–304. doi:10.1016/j.physbeh.2012.01.020 102, 106

[326] Nicole van van Nes, Michiel Christoph, and Ingrid van Schagen. 2025. How to navigate the intermediate levels of vehicle automation? *Trans-*

*portation Research Interdisciplinary Perspectives* 34 (November 2025), 1–10. doi:10.1016/j.trip.2025.101664 192

[327] Madhura Vasist, François Rozet, Olivier Absil, Paul Mollière, Evgeny Nasedkin, and Gilles Louppe. 2023. Neural Posterior Estimation for Exoplanetary Atmospheric Retrieval. *Astronomy & Astrophysics* 672 (April 2023), 1–10–. doi:10.1051/0004-6361/202245263 157, 161

[328] Nicola Vasta and Francesco Biondi. 2025. Effect of Partially Automated Driving on Mental Workload, Visual Behavior and Engagement in Nondriving-Related Tasks: A Meta-Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 67, 9 (March 2025), 962–991. doi:10.1177/00187208251323132 76

[329] Francisco Vicente, Zehua Huang, Xuehan Xiong, Fernando De la Torre, Wende Zhang, and Dan Levi. 2015. Driver Gaze Tracking and Eyes Off the Road Detection System. *IEEE Transactions on Intelligent Transportation Systems* 16, 4 (August 2015), 2014–2027. doi:10.1109/tits.2015.2396031 46, 47, 58

[330] Trent W. Victor, Joanne L. Harbluk, and Johan A. Engström. 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 2 (March 2005), 167–190. doi:10.1016/j.trf.2005.04.014 103, 104, 109, 116, 118, 119

[331] Afonso Vilaca, Pedro Cunha, and Andre L. Ferreira. 2017. Systematic literature review on driving behavior. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Yokohama, Japan, 1–8. doi:10.1109/itsc.2017.8317786 199, 202, 204

[332] Jose Carlos Villarreal Guerra, Zeba Khanam, Shoaib Ehsan, Rustam Stolkin, and Klaus McDonald-Maier. 2018. Weather Classification: A New Multi-Class Dataset, Data Augmentation Approach and Comprehensive Evaluations of Convolutional Neural Networks. In *NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*. IEEE, Edinburgh, UK, 305–310. doi:10.1109/AHS.2018.8541482 156

[333] Emmanuel Vincent, Réùi Gribonval, and Cédric Févotte. 2006. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing* 14, 4 (July 2006), 1462–1469. doi:10.1109/tsa.2005.858005 60

[334] Usha Kumari Vismaya and E. Saritha. 2020. A Review on Driver Distraction Detection Methods. In *International Conference on Communications and Signal Processing (ICCSP)*. IEEE, Chennai, India, 483–487. doi:10.1109/iccsp48568.2020.9182316 199, 202, 204

[335] Tobias Vogelpohl, Matthias Kühn, Thomas Hummel, and Mark Vollrath. 2019. Asleep at the automated wheel — Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention* 126 (May 2019), 70–84. doi:10.1016/j.aap.2018.03.013 76

[336] Mark Vollrath, Susanne Schleicher, and Christhard Gelau. 2011. The influence of Cruise Control and Adaptive Cruise Control on driving behaviour – A driving simulator study. *Accident Analysis & Prevention* 43, 3 (May 2011), 1134–1139. doi:10.1016/j.aap.2010.12.023 94

[337] Marcel Walch, Mark Colley, and Michael Weber. 2019. Driving-task-related human-machine interaction in automated driving. In *Adjunct Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Utrecht, The Netherlands, 427–433. doi:10.1145/3349263.3351527 77, 185

[338] Ange Wang, Jiyao Wang, Wenxin Shi, and Dengbo He. 2024. Cognitive Workload Estimation in Conditionally Automated Vehicles Using Transformer Networks Based on Physiological Signals. *Transportation Research Record: Journal of the Transportation Research Board* 2678, 12 (June 2024), 1183–1196. doi:10.1177/03611981241250023 46, 49

[339] Chao Wang, John Elson, Yingzi Lin, and Shichao Liu. 2025. The influence of temperature and interior ambient lighting color on nighttime driving performance. *Building and Environment* 275 (May 2025), 1–14. doi:10.1016/j.buildenv.2025.112815 46, 48

[340] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno A. Olshausen, and Trevor Darrell. 2021. Tent: Fully Test-Time Adaptation by Entropy Minimization. In *International Conference on Learning Representations (ICLR)*. International Conference on Learning Representations (ICLR), Virtual conference, 1–15. https://openreview.net/forum?id=uXl3bZLkr3c 131, 132, 155

[341] Meng Wang, Shashank Mehrotra, Nicholas Wong, Jah'inaya Parker, Shannon C. Roberts, Woon Kim, Alicia Romo, and William J. Horrey. 2024. Human-Machine interfaces and vehicle Automation: A review of the literature and recommendations for system Design, Feedback, and alerts. *Transportation Research Part F: Traffic Psychology and Behaviour* 107 (November 2024), 549–561. doi:10.1016/j.trf.2024.08.014 177

[342] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. 2022. Continual Test-Time Domain Adaptation. In *IEEE/CVF Conference on Computer*

*Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), New Orleans, Louisiana, USA, 7191–7201. doi:10.1109/cvpr52688.2022.00706 132, 155

[343] Qiong Wang, Jingyu Yang, Mingwu Ren, and Yujie Zheng. 2006. Driver fatigue detection: A survey. In *World Congress on Intelligent Control and Automation*, Vol. 2. IEEE, Dalian, China, 8587–8591. doi:10.1109/WCICA.2006.1713656 199, 202, 204

[344] Wei Wang, Zhun Zhong, Weijie Wang, Xi Chen, Charles Ling, Boyu Wang, and Nicu Sebe. 2023. Dynamically Instance-Guided Adaptation: A Backward-free Approach for Test-Time Domain Adaptive Semantic Segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), Vancouver, Canada, 24090–24099. doi:10.1109/cvpr52729.2023.02307 133, 155

[345] Xinying Wang and Min Han. 2014. Online sequential extreme learning machine with kernels for nonstationary time series prediction. *Neurocomputing* 145 (December 2014), 90–97. doi:10.1016/j.neucom.2014.05.068 131

[346] Xinyi Wang, Yulia Tsvetkov, Sebastian Ruder, and Graham Neubig. 2021. Efficient Test Time Adapter Ensembling for Low-resource Language Varieties. In *Findings of the Association for Computational Linguistics: EMNLP*. Association for Computational Linguistics, Punta Cana, Dominican Republic, 730–737. doi:10.18653/v1/2021.findings-emnlp.63 132

[347] Xianglin Wang, Qian Yang, Yingni Zhai, Haobo Niu, and Xinta Wang. 2023. Effects of Vehicle Air Temperature on Drivers' Cognitive Abilities Based on EEG. *Sustainability* 15, 2 (January 2023), 1–19. doi:10.3390/su15021673 46, 48

[348] Ying Wang, Bryan Reimer, Jonathan Dobres, and Bruce Mehler. 2014. The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour* 26 (September 2014), 227–237. doi:10.1016/j.trf.2014.08.003 104, 109, 116

[349] Daniel Watzenig and Martin Horn. 2017. *Automated Driving: Safer and More Efficient Future Driving*. Springer, Cham, Switzerland. doi:10.1007/978-3-319-31895-0 79

[350] Bradley W. Weaver and Patricia R. DeLucia. 2020. A Systematic Review and Meta-Analysis of Takeover Performance During Conditionally Automated Driving. *Human Factors: The Journal of the Human*

*Factors and Ergonomics Society* 64, 7 (December 2020), 1227–1260. doi:10.1177/0018720820976476 33

[351] Chuheng Wei, Chuanshi Liu, and Shaocui Chi. 2022. A Driver Distraction Detection Method Based on Convolutional Neural Network. In *International Conference on Computer and Communication Systems (ICCCS)*. Institute of Electrical and Electronics Engineers (IEEE), Wuhan, China, 272–277. doi:10.1109/icccs55155.2022.9846062 57

[352] Wanyu Wei, Xinsha Fu, Shiyu Zhong, and Hongcheng Ge. 2023. Driver's mental workload classification using physiological, traffic flow and environmental factors. *Transportation Research Part F: Traffic Psychology and Behaviour* 94 (April 2023), 151–169. doi:10.1016/j.trf.2023.02.004 46, 47, 49, 52

[353] Paul Weiser, Michael Preutenborbeck, Marcel Usai, and Frank Flemisch. 2025. From Human Computer Interaction to Human Systems Migration: Internal Interaction in Automated Road Vehicles. In *Intelligent Human Systems Integration (IHSI): Integrating People and Intelligent Systems*, Vol. 160. AHFE International, Rome, Italy, 520–527. doi:10.54941/ahfe1005857 187

[354] Karla Conn Welch, Cindy Harnett, and Yi-Ching Lee. 2019. A Review on Measuring Affect with Practical Sensors to Monitor Driver Behavior. *Safety* 5, 4 (October 2019), 1–18. doi:10.3390/safety5040072 199, 202, 204

[355] Christopher D. Wickens, Justin G. Hollands, Simon Banbury, and Raja Parasuraman. 2015. *Engineering Psychology and Human Performance*. Psychology Press, East Sussex, UK. doi:10.4324/9781315665177 45, 54

[356] Kyle M. Wilson, Shiyan Yang, Trey Roady, Jonny Kuo, and Michael G. Lenné. 2020. Driver trust & mode confusion in an on-road study of level-2 automated vehicle technology. *Safety Science* 130 (October 2020), 1–11. doi:10.1016/j.ssci.2020.104845 13, 177

[357] Peter I.J. Wouters and John M.J. Bos. 2000. Traffic accident reduction by monitoring driver behaviour with in-car data recorders. *Accident Analysis & Prevention* 32, 5 (September 2000), 643–650. doi:10.1016/s0001-4575(99)00095-0 29

[358] Yuqing Wu and Linda Ng Boyle. 2015. Drivers' engagement level in Adaptive Cruise Control while distracted or impaired. *Transportation Research Part F: Traffic Psychology and Behaviour* 33 (August 2015), 7–15. doi:10.1016/j.trf.2015.05.005 75, 93

[359] Yue Wu, Linhao Ye, Xianzhe Zheng, Hanfei Zhu, and Wei Xiang. 2023. Dangerous Slime: A Game for Improving Situation Awareness in Automated Driving. In *International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*. ACM, Ingolstadt, Germany, 136–144. doi:10.1145/3580585.3607178 186

[360] Zhengxin Wu, Tao Jin, Chunpeng Chen, Xinyu Liu, and Jiaqi Yan. 2023. How do different ambient temperatures and vehicle speeds affect the cognitive performance of male drivers? Evidence from ERP. *Traffic Injury Prevention* 24, 3 (Feb. 2023), 271–278. doi:10.1080/15389588.2023.2181078 46, 48

[361] Grace Wusk and Hampton Gabler. 2018. Non-Invasive Detection of Respiration and Heart Rate with a Vehicle Seat Sensor. *Sensors* 18, 5 (May 2018), 1–11. doi:10.3390/s18051463 47, 49

[362] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. 2021. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 34. Curran Associates, Inc., Virtual conference, 12077–12090. 133, 219, 230

[363] Yang Xing, Chen Lv, Dongpu Cao, and Peng Hang. 2021. Toward human-vehicle collaboration: Review and perspectives on human-centered collaborative automated driving. *Transportation Research Part C: Emerging Technologies* 128 (July 2021), 1–26. doi:10.1016/j.trc.2021.103199 73

[364] Linli Xu, Lie Guo, Pingshu Ge, and Xu Wang. 2024. Transition in Different Critical Situations: How Non-Driving Related Tasks Affect Drivers' Physiological Response and Takeover Behavior After Partial Automation Silent Failures. *IEEE Transactions on Intelligent Transportation Systems* 25, 11 (November 2024), 16642–16652. doi:10.1109/tits.2024.3438450 46, 57

[365] Shiyang Yan, Yuxuan Teng, Jeremy S. Smith, and Bailing Zhang. 2016. Driver behavior recognition based on deep convolutional neural networks. In *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, Changsha, China, 636–641. doi:10.1109/fskd.2016.7603248 47, 57

[366] Bo Yang, Xuelin Cao, Kai Xiong, Chau Yuen, Yong Liang Guan, Supeng Leng, Lijun Qian, and Zhu Han. 2021. Edge Intelligence for Autonomous Driving in 6G Wireless System: Design Challenges and Solutions. *IEEE Wireless Communications* 28, 2 (April 2021), 40–47. doi:10.1109/mwc.001.2000292 133

[367] Haimin Yang, Zhisong Pan, and Qing Tao. 2019. Online Learning for Time Series Prediction of AR Model with Missing Data. *Neural Processing Letters* 50, 3 (Feb. 2019), 2247–2263. doi:10.1007/s11063-019-10007-x 131

[368] Robert M. Yerkes and John D. Dodson. 1908. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology* 18, 5 (November 1908), 459–482. doi:10.1002/cne.920180503 13, 76

[369] Hiroshi Yokoyama, Koma Eihata, Junya Muramatsu, and Yusuke Fujiwara. 2018. Prediction of Driver's Workload from Slow Fluctuations of Pupil Diameter. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Maui, HI, USA, 1775–1780. doi:10.1109/itsc.2018.8569279 46, 47, 48, 102

[370] Kristie Young and Micheal Regan. 2007. *Driver distraction: A review of the literature*. Australasian College of Road Safety, Sydney, New South Wales, Australia, 379–405. 46, 57

[371] Kristie L. Young, Micheal A. Regan, and John D. Lee. 2009. *Measuring the effects of driver distraction: direct driving performance methods and measures*. CRC Press, Milton Park, Abingdon, UK, Chapter 7, 85–105. 52

[372] Mark S. Young and Neville A. Stanton. 2002. Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science* 3, 2 (January 2002), 178–194. doi:10.1080/14639220210123789 76

[373] Mark S. Young and Neville A. Stanton. 2023. To automate or not to automate: advocating the 'cliff-edge' principle. *Ergonomics* 66, 11 (October 2023), 1695–1701. doi:10.1080/00140139.2023.2270786 74, 184

[374] Changqian Yu, Changxin Gao, Jingbo Wang, Gang Yu, Chunhua Shen, and Nong Sang. 2021. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *International Journal of Computer Vision* 129, 11 (September 2021), 3051–3068. doi:10.1007/s11263-021-01515-2 133

[375] Rongjie Yu, Yin Zheng, and Xiaobo Qu. 2021. Dynamic driving environment complexity quantification method and its verification. *Transportation Research Part C: Emerging Technologies* 127 (June 2021), 1–29. doi:10.1016/j.trc.2021.103051 180

[376] Longhui Yuan, Binhui Xie, and Shuang Li. 2023. Robust Test-Time Adaptation in Dynamic Scenarios. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), Vancouver, Canada, 15922–15932. doi:10.1109/cvpr52729.2023.01528 132, 133, 134, 155

[377] Norhasliza M. Yusoff, Rana Fayyaz Ahmad, Christophe Guillet, Aamir Saeed Malik, Naufal M. Saad, and Frederic Merienne. 2017. Selection of Measurement Method for Detection of Driver Visual Cognitive Distraction: A Review. *IEEE Access* 5 (September 2017), 22844–22854. doi:10.1109/access.2017.2750743 46, 58, 62, 102, 199, 202, 204

[378] Éloi Zablocki, Hédi Ben-Younes, Patrick Pérez, and Matthieu Cord. 2021. Explainability of deep vision-based autonomous driving systems: Review and challenges. *arXiv* abs/2101.05307 (2021), 31 pages. arXiv:2101.05307 doi:10.48550/arXiv.2101.05307 65

[379] Maxime Zanella and Ismail Ben Ayed. 2024. On the Test-Time Zero-Shot Generalization of Vision-Language Models: Do we Really need Prompt Learning?. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Seattle, Washington, USA, 23783–23793. doi:10.1109/cvpr52733.2024.02245 155

[380] Maxime Zanella, Benoît Gérin, and Ismail Ben Ayed. 2024. Boosting vision-language models with transduction. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 37. Curran Associates, Inc., Vancouver, Canada, 62223–62256. 155

[381] Jian Yong Zhang, Wen Wei Qiu, Hui Jian Fu, Meng Ting Zhang, and Qing Guo Ma. 2013. Review of Techniques for Driver Fatigue Detection. *Applied Mechanics and Materials* 433-435 (October 2013), 928–931. doi:10.4028/www.scientific.net/amm.433-435.928 199, 202, 204

[382] Marvin Zhang, Sergey Levine, and Chelsea Finn. 2022. MEMO: Test Time Robustness via Adaptation and Augmentation. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 35. Curran Associates, Inc., New Orleans, Louisiana, USA, 38628–38642. https://dl.acm.org/doi/10.5555/3600270.3603069 132

[383] Qi Zhang, Qingtian Wu, Yimin Zhou, Xinyu Wu, Yongsheng Ou, and Huazhang Zhou. 2017. Webcam-based, non-contact, real-time measurement for the physiological parameters of drivers. *Measurement* 100 (March 2017), 311–321. doi:10.1016/j.measurement.2017.01.007 47, 50

[384] Yiming Zhang, Jiaqi Sun, Ming Chen, Qi Wang, Yifan Yuan, and Rui Ma. 2021. Multi-Weather Classification Using Evolutionary Algorithm on EfficientNet. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops)*. Institute of Electrical and Electronics Engineers (IEEE), Kassel, Germany, 546–551. doi:10.1109/PerComWorkshops51409.2021.9430939 156

[385] Xingchen Zhao, Chang Liu, Anthony Sicilia, Seong Jae Hwang, and Yun Fu. 2022. Test-time Fourier Style Calibration for Domain Generalization. In *International Joint Conference on Artificial Intelligence (IJCAI)*. International Joint Conference on Artificial Intelligence (IJCAI), Vienna, Austria, 1721–1727. https://yonsei.elsevierpure.com/en/publications/test-time-fourier-style-calibration-for-domain-generalization/ 132

[386] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip H.S. Torr, and Li Zhang. 2021. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), Nashville, Tennessee, USA, 6877–6886. doi:10.1109/cvpr46437.2021.00681 133

[387] Yang Zhou, Yunxing Chen, and Yixi Zhang. 2025. Driver Distraction Detection in Conditionally Automated Driving Using Multimodal Physiological and Ocular Signals. *Electronics* 14, 19 (September 2025), 1–18. doi:10.3390/electronics14193811 46, 47, 48, 50