






Article

AI Perspectives in Smart Cities and Communities to Enable Road Vehicle Automation and Smart Traffic Control

Cristofer Englund ^{1,2,*} , Eren Erdal Aksoy ¹ , Fernando Alonso-Fernandez ¹ , Martin Daniel Cooney ¹ , Sepideh Pashami ^{1,2}  and Björn Åstrand ¹

¹ Center for Applied Intelligent Systems Research (CAISR), Halmstad University, 301 18 Halmstad, Sweden; eren.aksoy@hh.se (E.E.A.); feralo@hh.se (F.A.-F.); martin.cooney@hh.se (M.D.C.); Sepideh.pashami@ri.se (S.P.); bjorn.astrand@hh.se (B.Å.)

² RISE Research Institutes of Sweden, Lindholmspiren 3A, 417 56 Göteborg, Sweden

* Correspondence: cristofer.englund@ri.se; Tel.: +46-708-560-227

Abstract: Smart cities and communities (SCC) constitute a new paradigm in urban development. SCC ideate a data-centered society aimed at improving efficiency by automating and optimizing activities and utilities. Information and communication technology along with Internet of Things enables data collection and with the help of artificial intelligence (AI) situation awareness can be obtained to feed the SCC actors with enriched knowledge. This paper describes AI perspectives in SCC and gives an overview of AI-based technologies used in traffic to enable road vehicle automation and smart traffic control. Perception, smart traffic control and driver modeling are described along with open research challenges and standardization to help introduce advanced driver assistance systems and automated vehicle functionality in traffic. To fully realize the potential of SCC, to create a holistic view on a city level, availability of data from different stakeholders is necessary. Further, though AI technologies provide accurate predictions and classifications, there is an ambiguity regarding the correctness of their outputs. This can make it difficult for the human operator to trust the system. Today there are no methods that can be used to match function requirements with the level of detail in data annotation in order to train an accurate model. Another challenge related to trust is explainability: models can have difficulty explaining how they came to certain conclusions, so it is difficult for humans to trust them.

Keywords: smart cities; artificial intelligence; perception; smart traffic control; driver modeling



Citation: Englund, C.; Aksoy, E.E.; Alonso-Fernandez, F.; Cooney, M.D.; Pashami, S.; Åstrand, B. AI Perspectives in Smart Cities and Communities to Enable Road Vehicle Automation and Smart Traffic Control. *Smart Cities* **2021**, *4*, 783–802. <https://doi.org/10.3390/smartcities4020040>

Academic Editor: Pierluigi Siano

Received: 31 March 2021

Accepted: 11 May 2021

Published: 18 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Smart cities and communities (SCC) is an emerging research field that covers many topics and is promoted by major advances in technology, changes in business operations, and the overall environmental challenges. This paper reviews AI in smart cities while also pointing out the research challenges for future research. Some of the enabling technologies for SCC are information and communication technologies (ICT) that connect infrastructure, resources, and services with data surveillance and asset management systems. Another technology is Internet of Things (IoT), which enables even the smallest device to connect to the Internet and share its operational status [1,2]. Devices from, e.g., the transportation system, power plants, or residential houses can be connected with the use of IoT technology [3]. Modern business environments are highly competitive, and organizations are constantly finding ways to reduce cost. In addition, businesses are exploring new ways of developing their operations, and of converting towards not only providing a product but also a service connected to the product, which is becoming popular in many domains [4]. Economics can be seen as the main driver for industry, and the environmental challenges as the main drivers for political and private actors [5]. We have a great responsibility to protect our natural resources for our descendants.

To set free the potential of SCC in any application area, the collected data need to be processed and analyzed. With the help of AI, relationships, root causes, and patterns can be found in the data. AI can then use the new information to tailor guidance and provide suggestions to users on how to improve behavior [6,7].

There still exist various challenges related to SCC, some of which are listed in the next section.

Challenges Addressed by SCC

Business operations have changed dramatically during the last 60 years. GDP changes from 1947–2009 clearly show a decrease in industry and growth in professional and business services. In the US, the GDP decrease in industry is around 50%. On the other hand, the growth in GDP for professional and business services is 400% [8]. This trend is holding on, and in the last 17 years the GDP service share has increased from 72.8% to 77.4% of GDP, whereas the industry share has decreased from 22.5% to 18.2% [9].

Digitalization, with help from ICT, is a major contributor to this trend, and with the latest technology trends in computing power, AI is becoming a key technology to make use of the data to further develop the services.

Another challenge is energy consumption, and in particular energy that comes from non-renewable sources such as oil. In Europe, the primary energy consumption was 1561 million tonnes of oil equivalent (Mtoe) in 2017, 5.3% above the EU target for 2020. In 2018, energy came from petroleum products, including crude oil (36%), natural gas (21%), renewable energy (15%), solid fossil fuel (15%), and nuclear energy (13%) [10]. The energy consumption by sector in the EU breaks down in the following way: the industry sector (31% of final energy consumption), the transport sector (28%), households (25%), services (13%), and agriculture and forestry (2%) [11].

In this paper, we primarily address transportation within SCC and how AI can be used to improve efficiency and thus reduce energy consumption. AI can be used to learn traffic behavior and to control traffic both on the micro level in, e.g., intersections [12] and on the macro level [13].

In the 28 EU Member States (EU-28), energy consumption in transport increased 32% between 1990 and 2017. In the EU-13 states, the increase was 102% during the same period. Road transport accounts for 73% of the total energy consumption in the transport sector, and road transport alone has increased 34% between 1990 and 2017 [14].

In 2020, the expectation due to the Covid-19 pandemic was that the energy demand would be 10% below the 2019 levels. This would be twice the decline experienced throughout the financial crisis of 2008–2009. CO₂ emissions in the EU declined by 8% during the first quarter of 2020 compared with the same period in 2019 [15].

Traffic safety is also a global challenge, and traffic accidents have become one of the most common causes of death among young people [16]. Although fatalities have decreased for motorists in most countries, this is not the case for vulnerable road users (VRUs) [17] including pedestrians, bicyclists, and moped riders. In Europe 22,700 people lost their lives in 2019 [18] and more than 1.4 million people were injured in 2018 [19]. Worldwide, 1.35 million people lost their lives and up to 50 million were injured in traffic accidents in 2018 [20].

Given these facts about business development, energy and traffic safety, it is clear that there is huge potential in society for improvements, both in terms of energy efficiency and traffic safety.

Using our starting point in the enabling technologies and the challenges mentioned above, this paper will focus on how AI can be used to enable energy savings and improve traffic safety within SCC. The EU has set goals for energy consumption in the transportation sector. The climate and energy goals for a competitive, secure, and low-carbon EU economy by 2030 state that the greenhouse gas (GHG) emissions should be reduced by 55% below the 1990 level and the amount of renewable energy should be at least 27% [21,22].

With the trend towards increased vehicle automation, there is great potential for reducing the effects of accidents, or, if possible, avoiding the accidents completely. This can be done by IoT—i.e., building sensor-based safety systems that can detect VRUs and give warnings or actively react to the information. Enabling the development of such systems requires knowledge of how road users behave, and how that behavior can be described so that the automated vehicle functions can make correct interpretations and decisions. Additionally, vehicular communication can help traffic coordination and reduce travel time for, e.g., emergency vehicles [13,23,24]. Consequently, the European commission has set goals for traffic safety—i.e., close to zero deaths in 2050; and an interim goal—halving the number of seriously injured by 2030 from the level of 2020 [25].

Energy efficiency along with traffic safety are the two main goals of SSC in the domain of the transportation system.

With the help of AI and data analytics, it may be possible to improve utilization of the manageable assets within the transportation system. In particular, this paper describes the on-board AI-based systems along with the infrastructure AI-based systems that constitute SSC, addressing traffic safety and efficiency. An overview is given for the research areas of perception, traffic control, and interaction.

The rest of the paper is organized as follows. Section 2 describes research initiatives, projects, and financial programs. Section 3 describes the different approaches of using AI in SCC such as perception, traffic system control, and driver monitoring. Section 4 highlights open research questions and standardization to facilitate implementation and adoption. Finally, Section 5 provides a summary and conclusions of the findings.

2. Research Initiatives within Smart Cities and Communities

In the EU, the European Innovation Partnership on Smart Cities and Communities (EIP-SCC) is an initiative supported by the European Commission that brings together cities, industry, small and medium-sized enterprises (SMEs), banks, researchers, and other smart city actors to share information and find partners for projects [26].

The EU project CITYKeys is funded by the European Union HORIZON 2020 program. In collaboration with cities, the project has developed and validated key performance indicators along with data collection procedures for common and transparent monitoring to be able to compare smart city solutions across European cities [27]. The project has divided the smart city into sub-themes that concern, e.g., diversity and social cohesion, and aim at promoting diversity, community engagement, and social cohesion within a community. Education focuses on improving the accessibility and quality of education for everyone. Safety concerns lowering the rates of crime and accidents. Health focuses on improving the quality and accessibility of public health systems, for everyone, and encouraging a healthy lifestyle. Quality of housing and the built environment promote development of mixed-income areas, ensure high quality and quantity of public spaces and recreational areas, and improve affordability and accessibility to good housing for everyone. Finally, quality of life also promotes access to (other) services that focus on providing better access to amenities and affordable services in physical and virtual spaces for everyone. CITYKeys also aims to harmonize data collection from the cities involved that can enable comparisons of the performances of the measures introduced to reach the EU energy and climate targets.

CIVITAS [28] is a pan European network of cities that are dedicated to cleaner and better transports. The network is financed by the European Commission and was launched in 2002. Since then, 85 cities have joined the network and more than 900 measures and urban transport solutions have been tested and implemented. The main goal of CIVITAS is to make it easier for cities to obtain cleaner and better connected transport solutions in Europe and beyond. The four main characteristics of CIVITAS are a living laboratory approach to bring out research projects, maintaining a network of cities for cities, facilitating a public-private partnerships, and promoting political commitment. The network gives unique opportunities for practitioners to experience innovative transport solutions and learn from

experts in the field. Sustainable mobility is the overall area and the project is divided into 10 sub areas: car-independent lifestyles; clean fuels and vehicles; collective passenger transport; demand management strategies; integrated planning; mobility management; public involvement; safety and security; transport telematics; urban freight logistics.

In Sweden where this research is carried out, there are a number of initiatives from the Swedish government to support the development of SCC. The Strategic Innovation Program (SIP) “Drive Sweden” is a Swedish cross disciplinary collaboration platform driving the development towards sustainable mobility solutions for people and goods [29]. Drive Sweden is an important stakeholder in future mobility that fertilizes national and international collaboration to encourage the development of future sustainable mobility. Drive Sweden is financed by the Swedish Innovation Agency, Vinnova, the Swedish Energy Agency, and the Swedish research council for sustainable development, Formas. Drive Sweden also provides a weekly newsletter that summarizes news within the area of smart mobility [30]. Examples of projects that has been financed by Drive Sweden, within the field of future mobility, are, for example: *Study of communication needs in interaction between trucks and surrounding traffic in platooning*, *Intelligent and self-learning traffic control with 3D and AI* and *Security for autonomous vehicles from a societal and safety perspective*.

InfraSweden2030 [31] is another Strategic Innovation Program in Sweden. Whereas Drive Sweden focuses on future mobility, InfraSweden2030 focuses on the transportation infrastructure of the future. InfraSweden2030 is also financed by Vinnova, the Swedish Energy Agency, and Formas. The aim of the program is to contribute to reduced climate and environmental impacts from the construction, operations, and maintenance of the transport infrastructure. The program organizes seminars and workshops to facilitate collaboration and innovation within the Swedish transport infrastructure sector in order to address society’s economic and social challenges. In addition, InfraSweden2030 funds research projects that address these goals and challenges. The three objectives of the project are to develop innovation for transport infrastructure; create an open, dynamic, and attractive environment; and reduce the impacts on the environment and climate. An example of a project financed by the InfraSweden2030 program is the *iBridge* project. The overall aim of the project is to automate and make available knowledge about bridges that can be used to lower maintenance cost.

Viable Cities [32] is also a Swedish Strategic Innovation Program. Viable Cities focuses on smart sustainable cities. The vision of the program is to accelerate the transition towards inclusive, climate neutral cities until 2030 with a good life for all with the help of digitalization and citizen engagement. Viable Cities is like its siblings Drive Sweden and InfraSweden2030 financed by Vinnova, the Swedish Energy Agency, and Formas. Within the Viable Cities program, a strategic initiative, the Viable Cities Transition Lab was formed to foster capabilities to handle the societal challenges within climate and environmental transitions. The Transition Lab aims at luring the full potential out of humans in the era of digitalization and automation, and thus obtaining new methods to transform society for an equal and circular economy, and responsible and ground-breaking technology to create behavioral changes toward a more sustainable and entrepreneurial society. *Xplorion—a residential mobility service in car-free accommodation* is a project financed by Viable Cities [33]. It offers a mix of mobility services such as public transport, carpooling, and bicycle pooling to households in a new residential area called Södra Brunnshög in Lund Sweden. The aim is to provide mobility as part of the residential rent, and thus allow more efficient use of transport leading to a reduction in emissions from residents’ travel. It is also expected that by connecting housing and mobility, there will be synergies that will make the resources in the system more efficient than today.

Smart City Sweden [34] is a governmental export platform for sustainable city solutions. The platform reaches out to international delegates who are interested in investing in smart and sustainable city solutions from Sweden. The platform has five focus areas: mobility; climate, energy, and environment; digitalization; social sustainability; and urban planning. Through their Web page, Smart City Sweden promotes solutions ranging from

smart production of biogas from household waste and water treatment facilities to congestion pricing solutions in Stockholm, future multi-modal transportation services, and service platforms aimed at supporting an automated transportation system.

3. Approaches

This section describes different approaches of using AI in SCC. The paper was founded from the perspectives of the authors' own research; the paper initially gathers previous research in the SCC domain, and secondly, in regard to operational, tactical, and strategic vehicle and traffic functions, the paper describes future challenges from a speculative design approach to enable road vehicle automation and smart traffic control.

Figure 1 illustrates a sample smart city scenario where information is continuously shared among different units such as smart buildings, vehicles, and infrastructures to enable road vehicle automation and smart traffic control applications. In Figure 1 a sample of a smart city is illustrated. Information is continuously shared among different buildings, infrastructures, and vehicles. Thus, autonomously operating vehicles safely react to detected VRUs.



Figure 1. A sample smart city scenario.

Table 1 shows how in-vehicle and infrastructure-based systems contribute to the different levels of control in automated driving. The driving tasks can broadly be categorized into three different levels, i.e., strategic, tactical, and operational [35]. The *strategic* tasks comprise high-level (and longer-term) planning decisions, such as route choice, traffic flow control, and fuel cost estimates, whereas *operational* tasks include low-level (short-term) and continuous routine tasks, such as lateral control based on immediate environmental input, and in-vehicle inputs such as driver monitoring. The *tactical* tasks fall in the middle of the two and are mid-level, medium-term tasks, including, but not limited to, turning, overtaking, gap adjustment, and merging, based on local awareness around the vehicle. In the following subsections we describe perception systems enabling situation awareness for autonomous vehicles, and approaches for traffic system control, and finally we give examples of driver monitoring systems.

Table 1. Overview of in-vehicle and infrastructure-based systems' contributions to road vehicle automation.

	Strategical	Tactical	Operational
In-vehicle	x	x	x
Infrastructure	x	x	

3.1. Perception

Mobility within SCC concerns several of the challenges described above, e.g., traffic safety and environmental impact. These challenges in turn drive the technological development towards improved sensor systems that can improve vehicles' perceptions to help the driver in hazardous situations with, e.g., advanced driver assistance systems (ADAS). ADAS are functions that automate vehicle functions to improve safety or comfort. Examples of such functions are lane keeping aids, automated emergency braking, and adaptive cruise control. Nevertheless, with the help of sensors and AI, the vehicles' perception systems are becoming more and more intelligent and there are now several examples of highly automated vehicular systems. Realizing road vehicle automation builds on the assumption that the vehicle can maneuver automatically by itself. This requires local awareness around the vehicle to be able to handle obstacles, hazardous situations, and unanticipated events. One way to achieve local awareness is through on-board sensors. Camera, radar, and LiDAR (light detection and ranging) sensor signals are typically fused to obtain scene understanding [36]. Such sensors operate in the range from a few centimeters to 200 m [37].

Another way to obtain awareness in traffic is to use sensors in the infrastructure [38,39] and using wireless communication to exchange information between vehicles and infrastructure [40]. This section describes perception systems enabling situation awareness in traffic.

Scene understanding is an essential prerequisite for autonomous vehicles to increase the local awareness. Semantic segmentation and object detection are two fundamental early lower-level perception components which help in gaining a rich understanding of a scene. Safety-critical systems, such as highly automated vehicles, however, require not only highly accurate but also reliable predictions with a consistent measure of uncertainty. This is because the quantitative uncertainty measures can particularly be propagated to the subsequent units, such as decision-making modules that lead to safe maneuver planning or emergency braking, which is of utmost importance in safety-critical systems. Therefore, semantic segmentation and object detection integrated with reliable confidence estimates can significantly reinforce the concept of safe mobility within SCC.

Given an image or point cloud data stream, there exist two mainstream deep learning-based AI approaches used for real-time object detection tasks [41]: two-stage and one-stage detection frameworks. Two-stage methods [42–45] initially have a preprocessing step to generate category-independent object region proposals. The output of this step is then passed to the category-specific classifier, which returns the category label for each detected proposal. On the other hand, one-stage detectors [46–49] are region proposal-free frameworks where the proposal generation step is not separated, and thus the entire framework works in a unified end-to-end fashion. Such unified approaches directly predict class probabilities together with the bounding boxes using single feed-forward networks. In contrast to unified (one-stage) models, region-based (two-stage) methods achieve relatively better detection accuracies with the cost of being computationally more expensive. Unlike two-stage networks, the detection accuracy of a one-stage model is less sensitive to false detections coming from the backbone network.

Regarding the task of semantic scene segmentation, advanced deep neural networks are heavily used to generate accurate and reliable segmentation with real-time performance. Most of these approaches, however, rely on camera images [50–54], whereas relatively fewer contributions have discussed the semantic segmentation of 3D LiDAR data [55–57].

The main reason is that unlike camera images, LiDAR point clouds are relatively sparse, unstructured, and have non-uniform sampling, although LiDAR scanners have wider fields of view and return more accurate distance measurements.

As comprehensively described in [58], there exist two mainstream deep learning approaches addressing the semantic segmentation of 3D LiDAR data only: pointwise and projection-based neural networks. The former approaches operate directly on the raw 3D points without requiring any pre-processing step [59–61], whereas the latter project the point cloud into various formats such as 2D image view [55–57,62,63] or high-dimensional volumetric representation [64,65]. There is, however, a clear split between these two approaches in terms of accuracy, runtime, and memory consumption. Projection-based approaches can achieve state-of-the-art accuracy while running significantly faster. Although point-wise networks have slightly fewer parameters, they cannot efficiently scale up to large point sets due to the limited processing capacity, and thus, they have a longer runtime.

When it comes to uncertainty estimation, Bayesian neural networks (BNNs) are intensively used, since such networks can learn approximate distribution of the weights to further generate prediction confidences. There exist two types of uncertainties: aleatoric, which can quantify the intrinsic uncertainty coming from the observed data, and epistemic, where the model uncertainty is estimated by inferring with the posterior weight distribution, usually through Monte Carlo sampling. Unlike aleatoric uncertainty, which captures the irreducible noise in the data, epistemic uncertainty can be reduced by gathering more training data. For instance, segmenting out an object that has relatively fewer training samples in the dataset may lead to high epistemic uncertainty, whereas high aleatoric uncertainty may instead occur on segment boundaries or distant and occluded objects due to noisy sensor readings which are inherent in the sensors. Bayesian modeling helps in estimating both uncertainty types.

Gal et al. [66] proved that dropout can be used as a Bayesian approximation to estimate the uncertainty in classification, regression, and reinforcement learning tasks, and this idea was also extended to semantic segmentation of RGB images by Kendall et al. [50]. Recently, both uncertainty types were applied to 3D point cloud object detection [67], optical flow estimation [68], and semantic segmentation of 3D LiDAR point cloud data [57].

Typical algorithms for sensor fusion include Kalman filters [69]. Kalman filters are recursive filters that estimate the state of the system from several noisy measurements. This technology was used, for example, in [70], where a vehicle for the Grand Cooperative Driving Challenge was developed. Additionally, in [71] Kalman filters were used for sensor fusion and scene understanding.

For autonomous vehicles to behave efficiently and safely in traffic, they require not only scene understanding derived from perceptual data, but also algorithms that can model and predict other road users' behavior. Modeling of the behavior and prediction of motions has long been of interest and is applicable, especially in domains where humans and intelligent systems coexist [72]. In a survey [73], which addresses motion-prediction applications in intelligent vehicles, the authors proposed three main categories for how agent motion is modeled: physics-based, maneuver-based, and interaction aware approaches. Work that focuses on the acceleration and deceleration behavior of different vehicle types employs physics-based methods, e.g., [74,75]. The work presented in [76] suggested an interaction-aware method for risk assessment in traffic. Maneuver-based approaches assume that the maneuver intention can be recognized early on, and future trajectory should match that maneuver. The main idea in this approach is that real-world trajectories from a road agent can be clustered into categories representing different behaviors. Based on a set of behaviors, maneuver-based motion prediction approaches employ estimation techniques—for instance, Gaussian processes [77,78] to then estimate the most probable future maneuvers. Deep-learning techniques have also been applied to cluster vehicle encounters [79].

Roundabouts play a very important role in modern traffic infrastructure. Studies have shown that roundabouts reduce injury-causing crashes (in comparison to signal-

controlled intersections), can reduce delays and improve traffic flows, and even have lower long-term costs compared to signal-controlled intersections [80]. A study that employed support vector machines to classify vehicles inside a roundabout to either stay or leave the roundabout is presented in [81]. Similarly, a study that estimated the effects of a roundabout layout on driver behavior, employing simulation data, is presented in [82]. A method for estimating reachable paths using conditional transition maps is presented in [83]. A study that employed a stereo camera setup for time-to-contact estimation is presented in [84]. This study was focused on risk assessment instead of efficiency and smoothness of driving when entering a roundabout.

Recent work by Muhammad and Åstrand [85] applied particle filters to predict road user behavior. In [86] they addressed the problem of modeling and predicting agent behavior and status in a roundabout traffic scenario. They presented three ways of modeling traffic in a roundabout based on (i) the roundabout geometry (which can be generated using drawings or satellite images, etc.); (ii) the mean path taken by vehicles inside the roundabout; and (iii) a set of reference trajectories traversed by vehicles inside the roundabout. The roundabout models were compared in terms of exit-direction classification and state (i.e., position inside the roundabout) prediction of query vehicles inside the roundabout. The results show that the model based on a set of reference trajectories is better suited, in terms of both the early and robust exit-direction classification, and a more accurate state prediction. An additional experiment was done by categorizing vehicles into classes based on vehicle size (instead of a single class). Results indicate that such a categorization can affect, and in some cases enhance the state prediction accuracy. The particle filter approach in [87] was compared to a recurrent neural network (RNN), namely, long short-term memory (LSTM) [88] to determine the specific behavior model. Additionally, the network's performance was compared with other RNN architectures, such as the Bi-LSTM and Bi-LSTM+ LSTM stacked architecture to evaluate which model has the best performance. Results showed that a LSTM network can predict the exit of the vehicle in a roundabout much sooner than the particle filter method and performs equally good when predicting the state of the vehicle in a roundabout.

Englund [38,39] used real-world trajectories to predict the intentions of cars and bicycles in an upcoming road exit. The AI methods used in [38,39] were based on support vector machines [89], random forest [90], and multi-layer perceptrons [91]. In [38] a backward elimination strategy for selecting the most important variables for predicting the behavior of the cars and bicycles in intersections was used. For bicycles the most important variable was speed, and for cars it was position. Heading was also among the six best variables for both vehicle types.

Garcia et al. [92] proposed a mix of an unscented Kalman filter and joint probabilistic data association to fuse sensor reading from a vision-based system, a laser sensor, and a global positioning system to obtain obstacle detection and object tracking. Li et al. [93] suggested combining LiDAR and vision-based sensors to obtain lane detection and extraction of an optimal drivable region. Sivaraman and Trivedi [36] reviewed on-road vision-based vehicle detection, tracking, and behavior understanding. Beside Kalman filters, other algorithms such as support vector machines [89], Adaboost [94], hidden Markov models [95] and Gaussian mixture modeling [96] were used for various fusion tasks. Recently, also deep learning in terms of generative adversarial networks (GANs) was used for the fusion of radar and camera sensor data.

In the context of multimodal object detection, most of the recent works fused RGB camera images with 3D LiDAR point clouds [97–99], whereas other works have combined regular RGB data with thermal [100] or depth images [101]. In contrast to object detection, there are relatively few contributions related to multi-model semantic segmentation: [102] fused RGB, depth, and thermal images; and [103] combined RGB and LiDAR data streams for semantic segmentation.

One of the main challenges in multi-model perception is when to fuse various sensory readouts (RGB, LiDAR, etc.) which have vast variations in time scales, dimensions (i.e., 2D

versus 3D data), and signal types (i.e., continuous versus discrete). Deep neural networks, which are good at extracting and representing features hierarchically, provide various options to fuse sensor readings at different stages, such as early, middle, and late. Early fusion [104] directly merges raw data derived from different sensor modalities, e.g., first by concatenating raw scene features derived from different sensor modalities into a single vector and then training a deep neural network on this new feature representation. Late fusion [105] combines learned unimodal sensor features at the highest network layer into a final prediction. Middle fusion [106] combines features learned at intermediate layers.

In contrast to other fusion strategies, early fusion requires less computation time and memory since the raw data readings are jointly processed. However, such methods are inflexible to changes in the network input type. For instance, when a new sensing modality is introduced [107,108], early fusion networks need to be retrained from scratch. In such cases, late fusion approaches are more flexible and ideal, since only the domain-specific network needs to be retrained while the networks processing the other sensor data types remain the same. Although middle fusion networks are also relatively flexible, the network architecture design, i.e., finding the optimal combination of intermediate features, is non-trivial. Despite having advanced fusion networks [104–106] that achieve state-of-the-art performance on challenging object detection and semantic segmentation datasets, a lack of guidelines for designing optimal fusion networks still remains as a challenge, since most networks are designed by empirical results [109].

3.2. Traffic System Control

To plan a future transportation system, simulation tools are efficient tools. Besides giving realistic visualizations of future transportation systems, the simulation models can provide valuable information on how such a system functions under different conditions. Simulation of Urban Mobility (SUMO) is a popular open source traffic micro-simulation tool [110,111]. It was used to simulate smart infrastructure that could improve traffic flow and energy efficiency [13]. One of the challenges with a simulation is to validate the results. Building infrastructure is costly and the planning horizon is 50 years. To improve the simulation models, and to adapt to future vehicles that will have different levels of automation, researchers at UC Berkeley have proposed a plug in for SUMO called Flow [112,113]. Flow is developed to take into consideration fully automated, semi-automated, and manually driven vehicles into the simulation. The automated vehicle models take into consideration information from surrounding vehicles and infrastructure to be able to optimize the traffic behavior [113].

A review on intersection management is given in [12]. The paper discusses control strategies in signalized and non-signalized intersections. Four types of strategies have been investigated. Cooperative resource reservation concerns how vehicles reserve the tiles on their planned route for certain time slots to pass the intersection. Whereas resource allocation considers time slots and space tiles in, for example, intersections and roundabouts, trajectory planning concerns the scheduling of travel routes, e.g., trajectory planning. Another strategy is to use virtual traffic lights to control traffic. The final approach is collision avoidance and is a complement to the above-mentioned ones. The resource planning or scheduling tools may have one plan; however, the vehicle may have unknown constraints or deficiencies, and therefore, collision avoidance for input control adjustments can be applied to make sure of perpetual safety, e.g., collision avoidance in both the short term and the long term.

Graph neural networks (GNNs) have shown great potential to use existing traffic data to model future transportation systems and enable performing counterfactual reasoning about factors that affect them. DeepMind in a collaboration with Google Map [114] has shown that the prediction of traffic and estimated arrival time improves by 50% once the problem is formulated using GNNs. The graph represents the road structure and the artificial neural network learns the dynamics between roads building up the traffic system. The scalability of their approach enables modeling a complex structure of the traffic with

a single model. Although GNNs have been around for several years, only today have they reached the maturity suitable for solving realistically complex problems, due to both algorithmic progress and new GPU-optimized implementations.

Forecasting any given parameter in the complex dynamics of traffic can be considered as a spatio-temporal problem. While spatial relations between road and road sections can be modeled with graph structure, the ways to model the temporal aspect can vary. Xie et al. [115] proposed a SeqGNN which combines sequence-to-sequence (Seq2Seq) models with GNNs. On the other hand, Song et al. [116] modeled the temporal dependency between the graphs by a recurrent approach and Guo et al. [117] added an attention mechanism to control which weights need to be updated. Although RNN-based approaches seem to be a more popular choice for modeling the temporal aspects, Yu et al. [118] proposed a structure with several spatio-temporal convolutional blocks. The convolution structure has been defined on the time axis to model the temporal dependencies. Their goal is to exploit a simpler structure with fast training capabilities that can handle multi-scale traffic networks. Although a lot can be gained by modeling the roads and road connections as a graph, there is no guarantee that such a graph structure models the true underlying relationships between time-series. There might be a need for using graph embeddings [119], proximity embedding [120], and walk embedding [121], which can lower the dimensional complexity of the problem. A representation that is optimized to preserve both the proximity of nodes and the structure of the traversals is intuitively very appealing. In addition, the underlying graph might vary given different circumstances throughout time. Thus, Löwe et al. proposed creating amortized graphs [122] that take advantage of the variations in the data.

3.3. Driver Monitoring

With the introduction of more intelligent infrastructure and vehicles, one might conclude that the role that humans play will become less significant. However, some people might still want to drive themselves, which means that there will be a complex mixture of vehicles operating at different levels of autonomy (e.g., in terms of the levels described by SAE International [123]). Some potential dangers include “risk compensation”—people engage in riskier behaviors because they think technology can deal with it—and lower driving ability due to less opportunities to drive. Support can target accident-prone times, such as when a vehicle transitions to a more manual mode. Moreover, when control is removed from a human driver, there is a responsibility to ensure that humans in the vehicle are comfortable and safe. Thus, an important task includes detecting the state of people within a vehicle, to avoid negative states and seek to achieve positive states. Negative states can involve sleepiness, distraction, drunkenness, health problems (e.g., epilepsy), and negative moods (angry, fearful, and embarrassed), and individual predilections (some drivers can prefer to drive more wildly or be unsure how to interpret some driving situations due to lack of experience) [124]. Positive states can involve comfort and enjoyment [125]. To infer a human’s state, some typical features detected include eye analysis (e.g., eye aspect ratio and blinking), and gaze, head pose, and posture.

In our previous work, we explored the idea of using recurrent neural networks [126] to estimate the future behavior of a car driver. A video dataset was collected describing typical (future) driver behavior/activities, e.g., driving safe, glancing, leaning, removing a hand from the wheel, reaching, grabbing, retracting, and holding. A classification network was trained to recognize the current activity, and the result was fed into a recurrent network to predict the activity in the next frame. The accuracy of predicting the activity in the next frame was 80%; nevertheless, the method is capable of predicting activity up to 20 frames ahead with an accuracy of 62% (video was captured at 30 fps). Furthermore, we have explored how social media could be used to support drivers, by leveraging insights into how they feel outside of the vehicle and interacting to reduce loneliness—a construct which has been tied to factors that increase the risk of accidents, such as depression and sleep deprivation [127]. Some open research challenges include how “wisdom of the crowd”

strategies can be incorporated to find potential dangers and anomalous driving, and how to infer the “meanings” of detected emotions by detecting what they refer to, i.e., the “emotional referent”. For example, if a passenger frowns, is this behavior a reaction to driving conditions, or to something on their cell phone?

Additionally, the need for continuous user authentication and monitoring is increasingly observable when larger fleets of professional vehicles are on the road and many drivers are brought under pressure to drive longer than what legislation allows. They can also aid, for example, in assessing whether an authorized person is driving a vehicle, or detecting driver drowsiness or distraction. Here, modalities captured with cameras (face [128,129] or eye regions [130,131]) can be complemented with sensors attached to the seat or the steering wheel that allow one to capture bio-signals such as heartbeats [132] or skin impedance [133], which correlate, for example, with sweating—stress level—but also with fitness levels [134]. There are also proofs of concept using Doppler radar for vital signs measurement [135], which has the evident advantage of not needing any type of contact. Infrared thermal imaging is another possibility, derived from subcutaneous blood flow and perspiration patterns [136], which allows one to mitigate privacy concerns that may arise from the use of regular cameras operating in the visible range. These solutions allow unobtrusive monitoring of human vital signs that goes beyond driver monitoring as well. While fatigue or distraction detection may seem the most straightforward task, other examples can include: (i) Pre-crash road safety, since abnormal vital signs can reflect the presence of drugs, alcohol, stress, or even diseases such as pre-dementia [137]; (ii) Post-crash road safety, because detected signals can be used by advance automatic crash notification (AACN) systems in order to improve alarm handling; (iii) Person identity, which can be achieved in an unobtrusive way not only with traditional facial images, but also with bio-signals [138].

Investments in driverless cars are already massive, both in public and private sectors. However, their benefits will be several orders of magnitudes higher if they are used collectively (taxis, buses, trucks feeding trains, etc.) to reduce the number of vehicles serving transportation needs, thereby producing saving time and reducing the environmental footprint. One bottle-neck will be then to secure the use of such vehicles when people who do not know each other travel together when there is no driver; another is allowing travelers without tickets or identity cards, since the identity and rights management can be done entirely by the system. In continuous biometrics, users are constantly monitored without needing active cooperation, in contrast to one-time authentication, e.g., at the beginning of a session. This may be done by using all pieces of biometric information available at a particular time, including soft-biometrics, behavior, or emotional state [139]. Accumulated evidence over time also can be used to improve accuracy. In this context, active modalities, such as fingerprints and iris scanning, are often stronger than the weaker passive modalities (e.g., facial recognition), but the latter demand no cooperation. Additionally, certain intentions, expressions, and physical states are noticeable in the footprint left only in continuous visual signals of face and body, e.g., drowsiness, stress, or irregular behavior. In these scenarios, localization of body parts and handheld objects are also important for safety and comfort, in order to detect potentially dangerous people or events. For example, holding a book, a steering wheel, or a weapon makes a big difference, as does to whom the hands belong or what the hands are engaged in.

In addition, for a vehicle with automated functions, knowing the driver to provide a pleasant experience is as important as knowing the actions and intentions of the surrounding road users. In [140] we developed AI-based algorithms that could detect the actions and intentions of pedestrians. Such information is valuable while building reliable ADAS functions.

4. Open Research Challenges and Standardization

The previous sections have outlined current research initiatives and the state of the art within perception, traffic system control, and driver monitoring. We now explore the

viable future research directions based on the previous work. Perception from an ego vehicle is typically limited by the field of view of the available sensors, which thus calls for reliable communication between road users and possibly infrastructure to improve local awareness and thereby allow for improved tactical and operational vehicle control. Standards such as Cooperative ITS [141] are put forward to ease collaboration in the traffic system. However, the current standard promotes simple, “Here I am”, messages and does not allow negotiation between road users to improve traffic safety and traffic flow. Future research should consider data formats and ontologies to enable interactions not only between vehicles and between vehicles and infrastructure, but also between vehicles and VRUs. In addition, ontologies can be used to harmonize traffic behavior as described in [142]. Other challenges are the ownership of the intelligent infrastructure and the shared data, and how to manage revenue. Consequently, to utilize the full potential of the technology, IoT, and ICT, sharing of aggregated information such as trajectories or behavior, should be enabled to further build local awareness and thus facilitate tactical decision-making in traffic.

Another area of future research is the security of the AI-based systems. Research should obstruct risks with sensor vulnerabilities. In, e.g., [143], the author highlighted the risk of hijacking a vehicle with the help of manipulated billboards. Risks with hacking [144] are also prevalent and could be remedied by time critical AI-based anomaly detection methods. Such methods would ensure safe and secure tactical and operational road vehicle automation.

As sensor technologies develop and computing power increases, the use of autonomous drones, both aerial and with wheels, will increase. With full anti-collision capabilities they will ease our lives with instant delivery, guiding, carrying, and surveillance [145]. Such AI technology would need capabilities in three domains, strategic, tactical, and operational, to be efficient.

Another field that will evolve is how we interact with the technology. As mentioned in the introduction, there have been attempts with light-based external HMI to ease the introduction of platoons in traffic, to let the surrounding traffic understand the intentions of the platooning vehicles. With a higher level of automation, where the vehicles completely handle tactical and operational tasks, new ways of interacting with the vehicles are necessary. AI will play a major role in enabling interactions with automated vehicles through gestures and speech. A vehicle should observe the state of the driver or operator and thereby enable natural interactions and control in all three levels of automation—strategic, tactical, and operational.

Pervasive intelligence is another concept that can have great impacts on both business and society. While systems become more and more digitized and start interacting more with other systems (trucks in a platoon or different administrations within a municipality, or vehicles in a multi-modal transportation service system, for example), there is always a risk of sub-optimization since the learning functions do not have access to the whole set of data. Consequently, future research should focus on how to optimize behavior on a larger scale, allowing the system to access data also outside of the own system boundaries.

One challenge of applying machine learning in vehicles is their rigorous safety requirements. The traditional standard ISO26262 [146] does not apply to a trained machine learning-based software, since the behavior of a machine learning-based software is not explicitly expressed in source code and it does not follow a certain specification. The developers rather define an algorithm and an architecture that learns the functionality. In the process, enormous amounts of data complemented with domain-specific labels are used to teach the machine to capture the relationships between input and output data. For the development of road vehicle automation, this step usually concerns the collection and preprocessing of huge amounts of data from, e.g., cameras, LiDAR, and radar sensors, along with training and evaluation using even more data.

In January 2019, ISO/PAS 21448—Safety of the Intended Functionality (SOTIF) was published containing guidance on the applicable design, verification, and validation mea-

asures that are needed to achieve the SOTIF [147]. A PAS (publicly available specification) is not an established standard, but rather a document that resembles the content of what is planned to be included in a future standard.

It is the intention that ISO 26262 and SOTIF should be complementary standards: ISO 26262 covers “absence of unreasonable risk due to hazards caused by malfunctioning behavior” [146] by mandating rigorous development and is structured under the V-model way-of-working. The focus of SOTIF is to addresses “hazards resulting from functional inefficiencies of the intended functionality” [147], e.g., classification failures in an automotive local awareness system, which are different from the types of malfunctions targeted by defect-oriented ISO 26262.

SOTIF is not structured based on the V-model but around (i) known safe states, (ii) known unsafe states, and (iii) unknown unsafe states. Note that SOTIF concerns the process of minimizing the two unsafe states, by focusing on detailing the requirements specifications for the developed functionality, where the aim is to shift the hazards from (iii) → (ii) and from (ii) → (i), which in turn are derived from hazard identification and hazard mitigation, respectively.

Another challenge is harmonizing the introduction of more advanced vehicular functions and making them socially accepted. ADAS such as forward vehicle collision mitigation systems (FVCMS), pedestrian detection and collision mitigation systems (PDCMS), and bicyclist detection and collision mitigation systems (BDCMS) are examples of vehicle functions that make use of on-board sensors that build local awareness around vehicles. These systems are described and specified by ISO, the International Organization for Standardization, i.e., FVCMS by ISO 22839:2013 [148], PDCMS by ISO 19237:2017 [149], and BDCMS by ISO 22078:2020 [150]. These standards detail concepts of functionality, and their minimum functionalities and system requirements in conjunction with interfaces and how testing of the functions should be performed. Recently, standards on external human–machine interfaces (eHMI) have been put forward. The functionality is described in the technical report *Road Vehicles—Ergonomic aspects of external visual communication from automated vehicles to other road users*, ISO/TR 23049 [151]. The document describes how automated vehicles should communicate their intentions to surrounding road users. Enabling such functionality, the vehicles need to be able to interpret the behavior of their fellow road users.

5. Summary and Conclusions

SCC refers to a cohesive concept to develop a sustainable future society. This paper highlighted how AI can support the development of SCC, and in particular within the future cooperative ITS. The two main challenges within ITS that are addressed by SCC are the traffic safety and the environmental challenges. In Europe, reducing the number of severely injured by half by 2030 (in regard to 2020 data) and the ultimate goal to have close to zero deaths in 2050 are two goals set by the European Commission [25].

Perception using computer vision and sensing is one of the enablers of road vehicle automation. Sensor fusion is typically used to obtain a robust mapping of the surroundings. For the vehicle to understand and interpret the surroundings, it uses semantic mapping that makes the vehicle aware. Another way to interpret the surroundings that are highlighted is to use classification and tracking to predict the behavior of surrounding road users.

Traffic system control is central in managing traffic flow in larger cities. To improve traffic system control, traffic simulators are often used. One can use simulators to perform planning of cities or new road infrastructure and use them to predict future scenarios or the effect of a presumptive maintenance effort. Challenges include how to incorporate the effect of future automated vehicles, since their behavior is so far unknown.

Driver monitoring is a research field that improves the user experience in vehicle automation. In ADAS the driver monitoring system helps to warn the driver if he/she becomes distracted or drowsy. The driver monitoring system also plays an important role in automated vehicles. The driver monitoring system should then know about the driver

to be able to hand over control only when the driver is capable of driving. In-vehicle, sensors may also be used for authentication. For this purpose, camera sensors can be fused with other modalities, e.g., in the steering wheel or the seat, that can capture pulse or skin impedance. Doppler radar sensors are another example of a system that with the help of AI can be used to estimate vital parameters of a driver, i.e., to improve handover or warning applications.

As reported in this paper, AI-based technology has achieved many great things and promises huge benefits in general, concerning economic growth, social development, and quality of life—and in particular, reducing environmental impacts and improving traffic safety.

Most of the AI-based technologies that are mentioned in this paper are data intensive. Not only are the volume and frequency of the data important, but also there is a need to merge a variety of data of different types and from different sources. Data from service providers, municipalities, and traffic authorities, and user data are required to create a holistic view of the situation on the scale of a city. Collecting and analyzing such data comes with privacy challenges, such as a trade-off between preserving the rights of the road users and providing personalized services. The possibilities of decentralized analysis of data or aggregation of intermediate results could also be considered.

The common challenge in real world applications of AI is how to trust the system—since the behavior of an AI is not built by explicitly expressed source code and does not follow certain specifications, but is rather built by algorithms that learn the intended functionality based on historical data. The generalizability of the algorithms in a real-world setting can be measured when they are deployed in practice and they are forced to encounter corner cases that have not been encountered before.

However, one of the main challenges is how can we set up the requirements of a system that is based on historical data. In road vehicle automation, where safety is imperative, verification and validation are crucial to benefit from the generalizability of the AI technology. Even if the training data are annotated and contain labels of the objects in the scenes, what are the guarantees that no new objects, different from the ones in the training set, will appear in future traffic situations? In addition, what is the relationship between requirements and the level of detail of the data annotation?

In addition, the low-level explainability of AI models, and data biases and data privacy pose considerable risks for users, developers, humanity, and societies. Consequently, although AI can predict, model, and sense, AI technology is not yet capable of explaining how it comes to certain conclusions, and therefore, it is difficult for humans to fully trust it.

All in all, enabling technologies for smart transport have come a long way in recent years. Now, it is time to start to connect these technological building blocks to unlock the socioeconomic benefits of smart cities. This step is going to be challenging, and it needs the collaboration of several actors, including end-users, SMEs, original equipment manufacturers (OEMs), cities, and legislators, to implement existing technologies in practice.

Author Contributions: Conceptualization, C.E.; methodology, C.E., E.E.A., F.A.-F., M.D.C., S.P. and B.Å.; writing—original draft preparation, C.E., E.E.A., F.A.-F., M.D.C., S.P. and B.Å.; writing—review and editing, C.E., E.E.A., F.A.-F., M.D.C., S.P. and B.Å.; visualization, E.E.A. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results has partially received funding from the Vinnova FFI project SHARPEN, under grant agreement no. 2018-05001 and the Vinnova FFI project SMILE III, under the grant agreement no. 2019-05871. The funding received from the Knowledge Foundation (KKS) in the framework of “Safety of Connected Intelligent Vehicles in Smart Cities—SafeSmart” project (2019–2023) is gratefully acknowledged. Finally, the authors thank the Swedish Research Council (project 2016-03497) for funding their research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Belli, L.; Cilfone, A.; Davoli, L.; Ferrari, G.; Adorni, P.; Di Nocera, F.; Dall'Olio, A.; Pellegrini, C.; Mordacci, M.; Bertolotti, E. IoT-Enabled Smart Sustainable Cities: Challenges and Approaches. *Smart Cities* **2020**, *3*, 1039–1071. [CrossRef]
2. Scuotto, V.; Ferraris, A.; Bresciani, S.; Al-Mashari, M.; Del Giudice, M. Internet of Things: Applications and challenges in smart cities. A case study of IBM smart city projects. *Bus. Process. Manag. J.* **2016**, *22*, 2. [CrossRef]
3. Arasteh, H.; Hosseinneshad, V.; Loia, V.; Tommasetti, A.; Troisi, O.; Shafie-khah, M.; Siano, P. Iot-based smart cities: A survey. In Proceedings of the 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC), Florence, Italy, 7–10 June 2016; pp. 1–6.
4. Tukker, A. Product services for a resource-efficient and circular economy—A review. *J. Clean. Prod.* **2015**, *97*, 76–91. [CrossRef]
5. Gärling, T.; Schuitema, G. Travel demand management targeting reduced private car use: Effectiveness, public acceptability and political feasibility. *J. Soc. Issues* **2007**, *63*, 139–153. [CrossRef]
6. Byttner, S.; Rögnvaldsson, T.; Svensson, M. Consensus self-organized models for fault detection (COSMO). *Eng. Appl. Artif. Intell.* **2011**, *24*, 833–839. [CrossRef]
7. Englund, C.; Verikas, A. Ink feed control in a web-fed offset printing press. *Int. J. Adv. Manuf. Technol.* **2008**, *39*, 919–930. [CrossRef]
8. GDP in US between 1947–2009. Available online: https://commons.wikimedia.org/wiki/File:Sectors_of_US_Economy_as_Percent_of_GDP_1947-2009.png (accessed on 16 May 2021).
9. Statista Research Department. Distribution of Gross Domestic Product (GDP) Across Economic Sectors in the United States from 2000 to 2017. 2020. Available online: <https://www.statista.com/statistics/270001/distribution-of-gross-domestic-product-gdp-across-economic-sectors-in-the-us/> (accessed on 13 May 2021).
10. Eurostat. *Shedding Light on Energy in the EU-A Guided Tour of Energy Statistics*; Technical Report; Eurostat: Luxembourg, 2020.
11. Eurostat. *Shedding Light on Energy in the EU-A Guided Tour of Energy Statistics*; Technical Report; Eurostat: Luxembourg, 2019.
12. Chen, L.; Englund, C. Cooperative Intersection Management: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 570–586. [CrossRef]
13. Englund, C.; Chen, L.; Voronov, A. Cooperative Speed Harmonization for Efficient Road Utilization. In Proceedings of the IEEE Nets4Cars, St. Petersburg, Russia, 6–8 October 2014; pp. 19–23.
14. Agency, E.E. *Final Energy Consumption in Europe by Mode of Transport*; Technical Report; European Environment Agency: Copenhagen, Denmark, 2019.
15. European Union 2020. Available online: <https://www.iea.org/reports/european-union-2020> (accessed on 21 April 2021).
16. World Health Organization (WHO). *Global Status Report on Road Safety 2015*; WHO Press: Geneva, Switzerland, 2015; p. 340.
17. Niska, A.; Eriksson, J. *Statistik över Cyklisters Olyckor: Faktaunderlag Till Gemensam Strategi För säker Cykling*; VTI: Linköping, Sweden, 2013.
18. Commission, E. *Data Table—Number of Road Deaths and Rate per Million Population by Country, 2010–2019*; Technical Report; CARE (Community Road Accident) Database: Brussels, Belgium, 2019.
19. Commission, E. *Annual Accident Report 2018*; Technical Report; European Commission, Directorate General for Transport: Brussels, Belgium, 2018.
20. World Health Organization. *Global Status Report on Road Safety*; Technical Report; World Health Organization: Geneva, Switzerland, 2019.
21. State of the Union: Commission raises climate ambition. Available online: https://ec.europa.eu/commission/presscorner/detail/en/IP_20_1599/ (accessed on 21 April 2021).
22. European Commission. *2030 Climate and Energy Goals for a Competitive, Secure and Low-Carbon EU Economy*; European Commission: Brussels, Belgium, 2014.
23. Barrachina, J.; Garrido, P.; Fogue, M.; Martinez, F.J.; Cano, J.C.; Calafate, C.T.; Manzoni, P. Reducing emergency services arrival time by using vehicular communications and Evolution Strategies. *Expert Syst. Appl.* **2014**, *41*, 1206–1217. [CrossRef]
24. Englund, C.; Chen, L.; Ploeg, J.; Semsar-Kazerooni, E.; Voronov, A.; Bengtsson, H.H.; Didoff, J. The grand cooperative driving challenge 2016: Boosting the introduction of cooperative automated vehicles. *IEEE Wirel. Commun.* **2016**, *23*, 146–152. [CrossRef]
25. Commission, E. *EU Road Safety Policy Framework 2021–2030-Next Steps Towards “Vision Zero”*; European Commission: Brussels, Belgium, 2019.
26. EU Smart Cities Marketplace. Available online: <https://eu-smartcities.eu/> (accessed on 21 April 2021).
27. EU project CITYkeys. Available online: <http://www.citykeys-project.eu/> (accessed on 21 April 2021).
28. EU Initiative CIVITAS. Available online: <http://civitas.eu/> (accessed on 21 April 2021).
29. Strategic Innovation Program, Drive Sweden. Available online: <http://www.drivesweden.net/> (accessed on 21 April 2021).
30. Strategic Innovation Program Newsletter, Drive Sweden. Available online: <https://www.drivesweden.net/en/newsletters> (accessed on 21 April 2021).
31. Strategic Innovation Program, Infra Sweden. Available online: <https://www.infrasweden2030.se/> (accessed on 21 April 2021).
32. Strategic Innovation Program, Viable Cities. Available online: <https://www.viablecities.se/> (accessed on 21 April 2021).

33. Xplorion - Residential mobility service in car-free accommodation. Available online: <https://en.viablecities.se/foi-projekt/xplorion/> (accessed on 21 April 2021).
34. Strategic Innovation Program, Smart City Sweden. Available online: <https://smartcitysweden.com/> (accessed on 21 April 2021).
35. Aramrattana, M.; Larsson, T.; Jansson, J.; Englund, C. Dimensions of Cooperative Driving, ITS and Automation. In Proceedings of the Intelligent Vehicles Symposium (IV), Seoul, Korea, 28 June–1 July 2015; pp. 144–149. [CrossRef]
36. Sivaraman, S.; Trivedi, M.M. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1773–1795. [CrossRef]
37. Kocić, J.; Jovičić, N.; Drndarević, V. Sensors and sensor fusion in autonomous vehicles. In Proceedings of the 2018 26th Telecommunications Forum (TELFOR), Belgrade, Serbia, 20–21 November 2018; pp. 420–425.
38. Englund, C. Aware and Intelligent Infrastructure for Action Intention Recognition of Cars and Bicycles. In Proceedings of the 6th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS), Prague, Czech Republic, 2–4 May 2020.
39. Englund, C. Action Intention Recognition of Cars and Bicycles in Intersections. In *International Journal Vehicle Design, Special Issue on: Safety and Standards for Connected and Autonomous Vehicles*; Inderscience: Geneva, Switzerland, 2020; in press.
40. Lidström, K.; Larsson, T. Act normal: Using uncertainty about driver intentions as a warning criterion. In Proceedings of the 16th World Congress on Intelligent Transportation Systems (ITS WC), Stockholm, Sweden, 21–25 September 2009; p. 8.
41. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.W.; Chen, J.; Liu, X.; Pietikäinen, M. Deep Learning for Generic Object Detection: A Survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [CrossRef]
42. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
43. Girshick, R.; Iandola, F.; Darrell, T.; Malik, J. Deformable part models are convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 437–446.
44. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
45. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
46. Szegedy, C.; Toshev, A.; Erhan, D. Deep Neural Networks for Object Detection. In Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13), Lake Tahoe, CA, USA, 5–8 December 2013; Curran Associates Inc.: Red Hook, NY, USA, 2013; Volume 2, pp. 2553–2561.
47. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the CVPR, Las Vegas, NV, USA, 27–30 June 2016.
48. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.E.; Fu, C.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the ECCV, Amsterdam, The Netherlands, 8–16 October 2016.
49. Law, H.; Deng, J. CornerNet: Detecting Objects as Paired Keypoints. In Proceedings of the ECCV, Munich, Germany, 8–14 October 2018.
50. Kendall, A.; Badrinarayanan, V.; Cipolla, R. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv* **2015**, arXiv:1511.02680.
51. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
52. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the The European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 October 2018.
53. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
54. Poudel, R.P.K.; Liwicki, S.; Cipolla, R. Fast-SCNN: Fast Semantic Segmentation Network. *CoRR* **2019**, doi:abs/1902.04502. [CrossRef]
55. Wu, B.; Wan, A.; Yue, X.; Keutzer, K. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Brisbane Convention & Exhibition Centre, Brisbane, Australia, 21–26 May 2018.
56. Milioto, A.; Vizzo, I.; Behley, J.; Stachniss, C. RangeNet++: Fast and Accurate LiDAR Semantic Segmentation. In Proceedings of the IROS, Macau, China, 3–8 November 2019.
57. Cortinhal, T.; Tzelepis, G.E.; Aksoy, E.E. SalsaNext: Fast, Uncertainty-Aware Semantic Segmentation of LiDAR Point Clouds. In *Advances in Visual Computing ISVC 2020 Lecture Notes in Computer Science*; Bebis, G., Eds.; Springer: Cham, Switzerland, 2020; Volume 12510, pp. 207–222.
58. Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; Bennamoun, M. Deep Learning for 3D Point Clouds: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. [CrossRef] [PubMed]
59. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

60. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
61. Landrieu, L.; Simonovsky, M. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. In Proceedings of the CVPR, Salt Lake City, UT, USA, 18–22 June 2018.
62. Aksoy, E.E.; Baci, S.; Cavdar, S. SalsaNet: Fast Road and Vehicle Segmentation in LiDAR Point Clouds for Autonomous Driving. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 926–932.
63. Wu, B.; Zhou, X.; Zhao, S.; Yue, X.; Keutzer, K. SqueezeSegV2: Improved Model Structure and Unsupervised Domain Adaptation for Road-Object Segmentation from a LiDAR Point Cloud. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019.
64. Zhang, C.; Luo, W.; Urtasun, R. Efficient convolutions for real-time semantic segmentation of 3D point clouds. In Proceedings of the International Conference on 3D Vision (3DV), Verona, Italy 5–8 September 2018.
65. Zhou, Y.; Tuzel, O. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA 18–23 June 2018; pp. 4490–4499.
66. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York City, NY, USA 19–24 June 2016; pp. 1050–1059.
67. Feng, D.; Rosenbaum, L.; Dietmayer, K. Towards safe autonomous driving: Capture uncertainty in the deep neural network for lidar 3d vehicle detection. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3266–3273.
68. Ilg, E.; Cicek, O.; Galesso, S.; Klein, A.; Makansi, O.; Hutter, F.; Brox, T. Uncertainty estimates and multi-hypotheses networks for optical flow. In Proceedings of the ECCV, Munich, Germany, 8–14 October 2018; pp. 652–667.
69. Welch, G.; Bishop, G. *An Introduction to the Kalman Filter*; Department of Computer Science, University of North Carolina at Chapel Hill: Chapel Hill, NC, USA, 1995.
70. Lidstrom, K.; Sjoberg, K.; Holmberg, U.; Andersson, J.; Bergh, F.; Bjade, M.; Mak, S. A modular CACC system integration and design. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1050–1061. [[CrossRef](#)]
71. Kianfar, R.; Augusto, B.; Ebadighajari, A.; Hakeem, U.; Nilsson, J.; Raza, A.; Tabar, R.S.; Irukulapati, N.V.; Englund, C.; Falcone, P.; et al. Design and experimental validation of a cooperative driving system in the grand cooperative driving challenge. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 994–1007. [[CrossRef](#)]
72. Rudenko, A.; Palmieri, L.; Herman, M.; Kitani, K.M.; Gavrila, D.M.; Arras, K.O. Human Motion Trajectory Prediction: A Survey. *arXiv* **2019**, arXiv:1905.06113.
73. Lefevre, S.; Vasquez, D.; Laugier, C. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH J.* **2014**, *1*. [[CrossRef](#)]
74. Bokare, P.S.; Maurya, A.K. Acceleration-Deceleration Behaviour of Various Vehicle Types. In Proceedings of the World Conference on Transport Research, Shanghai, China, 10–15 July 2016.
75. Maurya, A.K.; Bokare, P.S. Study of deceleration behaviour of different vehicle types. *Int. J. Traffic Transp. Eng.* **2012**, *2*, 253–270. [[CrossRef](#)]
76. Lefevre, S.; Laugier, C.; Ibanez-Guzman, J. *Intention-Aware Risk Estimation for General Traffic Situations, and Application to Intersection Safety*; Research Report; Research Centre Grenoble, HAL-Inria: Paris, France, 2013.
77. Christopher, T. Analysis of Dynamic Scenes: Application to Driving Assistance. Ph.D. Thesis, Automatic. Institut National Polytechnique de Grenoble—INPG, Grenoble, France, 2009. (In English)
78. Joseph, J.; Doshi-Velez, F.; Huang, A.S.; Roy, N. A Bayesian nonparametric approach to modeling motion patterns. *Auton. Robot.* **2011**, *31*, 383. [[CrossRef](#)]
79. Li, S.; Wang, W.; Mo, Z.; Zhao, D. Cluster Naturalistic Driving Encounters Using Deep Unsupervised Learning. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018.
80. WSDOT. Roundabout benefits—Washington State Department of Transportation. 2019. Available online: <https://www.wsdot.wa.gov/Safety/roundabouts/benefits.htm> (accessed on 16 May 2021).
81. Zhao, M.; Kathner, D.; Jipp, M.; Soffker, D.; Lemmer, K. Modeling Driver Behavior at Roundabouts: Results from a Field Study. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Redondo Beach, CA, USA, 11–14 June 2017.
82. Zhao, M.; Kathner, D.; Soffker, D.; Jipp, M.; Lemmer, K. Modeling Driving Behavior at Roundabouts: Impact of Roundabout Layout and Surrounding Traffic on Driving Behavior. 2017. Available online: <https://core.ac.uk/download/pdf/84275712.pdf> (accessed on 16 May 2021).
83. Kucner, T.; Saarinen, J.; Magnusson, M.; Lilienthal, A.J. Conditional transition maps: Learning motion patterns in dynamic environments. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013.
84. Muffert, M.; Milbich, T.; Pfeiffer, D.; Franke, U. May I Enter the Roundabout? A Time-To-Contact Computation Based on Stereo-Vision. In Proceedings of the Intelligent Vehicles Symposium (IV), Alcalá de Henares, Madrid, Spain, 3–7 June 2012.
85. Muhammad, N.; Åstrand, B. Intention estimation using set of reference trajectories as behaviour model. *Sensors* **2018**, *18*, 4423. [[CrossRef](#)] [[PubMed](#)]

86. Muhammad, N.; Åstrand, B. Predicting Agent Behaviour and State for Applications in a Roundabout-Scenario Autonomous Driving. *Sensors* **2019**, *19*, 4279. [\[CrossRef\]](#)
87. Magavi, S.A. Behaviour Modelling of Vehicles at a Roundabout. Master's Thesis, Halmstad University, Halmstad Embedded and Intelligent Systems Research (EIS), Halmstad, Sweden, 2020.
88. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Vapnik, V.N. *Statistical Learning Theory (Adaptive and Learning Systems for Signal Processing, Communications and Control Series)*; Wiley-Interscience: Hoboken, NJ, USA, 1998.
90. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
91. Bishop, M.C. *Neural Networks for Pattern Recognition*; Oxford University Press, Inc.: New York, NY, USA, 1995; Volume 92, p. 502. [\[CrossRef\]](#)
92. Garcia, F.; Martin, D.; De La Escalera, A.; Armingol, J.M. Sensor fusion methodology for vehicle detection. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 123–133. [\[CrossRef\]](#)
93. Li, Q.; Chen, L.; Li, M.; Shaw, S.L.; Nüchter, A. A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios. *IEEE Trans. Veh. Technol.* **2013**, *63*, 540–555. [\[CrossRef\]](#)
94. Freund, Y.; Schapire, R.; Abe, N. A short introduction to boosting. *J. Jpn. Soc. Artif. Intell.* **1999**, *14*, 1612.
95. Jazayeri, A.; Cai, H.; Zheng, J.Y.; Tuceryan, M. Vehicle detection and tracking in car video based on motion model. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 583–595. [\[CrossRef\]](#)
96. Wang, C.C.R.; Lien, J.J.J. Automatic vehicle detection using local features—A statistical approach. *IEEE Trans. Intell. Transp. Syst.* **2008**, *9*, 83–96. [\[CrossRef\]](#)
97. Qi, C.R.; Liu, W.; Wu, C.; Su, H.; Guibas, L.J. Frustum PointNets for 3D Object Detection from RGB-D Data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
98. Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-view 3d object detection network for autonomous driving. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1907–1915.
99. Liang, M.; Yang, B.; Chen, Y.; Hu, R.; Urtasun, R. Multi-Task Multi-Sensor Fusion for 3D Object Detection. In Proceedings of the CVPR, Long Beach, CA, USA, 16–19 June 2019.
100. Takumi, K.; Watanabe, K.; Ha, Q.; Tejero-De-Pablos, A.; Ushiku, Y.; Harada, T. Multispectral Object Detection for Autonomous Vehicles. In Proceedings of the on Thematic Workshops of ACM Multimedia, Mountain View, CA, USA, 23–27 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; Thematic Workshops '17, pp. 35–43. [\[CrossRef\]](#)
101. Mees, O.; Eitel, A.; Burgard, W. Choosing Smartly: Adaptive Multimodal Fusion for Object Detection in Changing Environments. *CoRR* **2017**. [\[CrossRef\]](#)
102. Valada, A.; Mohan, R.; Burgard, W. Self-Supervised Model Adaptation for Multimodal Semantic Segmentations. *Int. J. Comput. Vis.* **2019**, *128*, 1239–1285. [\[CrossRef\]](#)
103. Kim, D.K.; Maturana, D.; Uenoyama, M.; Scherer, S. Season-Invariant Semantic Segmentation with a Deep Multimodal Network. In *Field and Service Robotics*; Springer Proceedings in Advanced Robotics; Springer: Cham, Switzerland, 2018; Volume 5.
104. Xu, D.; Anguelov, D.; Jain, A. PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
105. Wang, Z.; Jia, K. Frustum ConvNet: Sliding Frustums to Aggregate Local Point-Wise Features for Amodal 3D Object Detection. *CoRR* **2019**. [\[CrossRef\]](#)
106. Hazirbas, C.; Ma, L.; Domokos, C.; Cremers, D. FuseNet: Incorporating Depth into Semantic Segmentation via FusionBased CNN Architecture. In Proceedings of the ACCV, Taipei, Taiwan, 20–24 November 2016.
107. Tavares de Araujo Cesariny Calafate, C.M.; Wu, C.; Natalizio, E.; Martínez, F.J. Crowdsensing and Vehicle-Based Sensing. *Mob. Inf. Syst.* **2016**, *2016*. [\[CrossRef\]](#)
108. Alvear, O.; Calafate, C.T.; Cano, J.C.; Manzoni, P. Crowdsensing in smart cities: Overview, platforms, and environment sensing issues. *Sensors* **2018**, *18*, 460. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Feng, D.; Haase-Schuetz, C.; Rosenbaum, L.; Hertlein, H.; Duffhauss, F.; Gläser, C.; Wiesbeck, W.; Dietmayer, K. Deep Multimodal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges. *CoRR* **2019**. [\[CrossRef\]](#)
110. Krajzewicz, D.; Erdmann, J.; Behrisch, M.; Bieker, L. Recent development and applications of SUMO-Simulation of Urban MObility. *Int. J. Adv. Syst. Meas.* **2012**, *5*, 128–138.
111. Krajzewicz, D.; Hertkorn, G.; Feld, C.; Wagner, P. SUMO (Simulation of Urban MObility); An open-source traffic simulation. In Proceedings of the 4th Middle East Symposium on Simulation and Modelling (MESM2002), American University, Sharjah, United Arab Emirates, 28–30 October 2002; pp. 183–187.
112. Kheterpal, N.; Parvate, K.; Wu, C.; Kreidieh, A.; Vinitzky, E.; Bayen, A. Flow: Deep reinforcement learning for control in sumo. *EPiC Ser. Eng.* **2018**, *2*, 134–151.
113. Wu, C.; Kreidieh, A.; Parvate, K.; Vinitzky, E.; Bayen, A.M. Flow: Architecture and benchmarking for reinforcement learning in traffic control. *arXiv* **2017**, arXiv:1710.05465.
114. Lange, O.; Perez, L. Traffic Prediction with Advanced Graph Neural Networks. 2020. Available online: <https://deepmind.com/blog/article/traffic-prediction-with-advanced-graph-neural-networks> (accessed on 16 May 2021).

115. Xie, Z.; Lv, W.; Huang, S.; Lu, Z.; Du, B.; Huang, R. Sequential Graph Neural Network for Urban Road Traffic Speed Prediction. *IEEE Access* **2020**, *8*, 63349–63358. [\[CrossRef\]](#)
116. Song, L.; Zhang, Y.; Wang, Z.; Gildea, D. N-ary Relation Extraction using Graph-State LSTM. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 31 October–4 November 2018; pp. 2226–2235.
117. Guo, Z.; Zhang, Y.; Lu, W. Attention Guided Graph Convolutional Networks for Relation Extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 241–251.
118. Yu, B.; Yin, H.; Zhu, Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 13–19 July 2018; pp. 3634–3640.
119. Chen, H.; Perozzi, B.; Hu, Y.; Skiena, S. Harp: Hierarchical representation learning for networks. In Proceedings of the AAAI Conference on Artificial Intelligence, 2–7 February 2018; Volume 32.
120. Wang, D.; Cui, P.; Zhu, W. Structural Deep Network Embedding. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1225–1234. [\[CrossRef\]](#)
121. Grover, A.; Leskovec, J. Node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 855–864. [\[CrossRef\]](#)
122. Lowe, S.; Madras, D.; Zemel, R.; Welling, M. Amortized Causal Discovery: Learning to Infer Causal Graphs from Time-Series Data. *arXiv* **2020**, arXiv:cs.LG/2006.10833.
123. SAE. *J3016 Levels of Driving Automation*; Technical Report; SAE International: Warrendale, PA, USA, 2018.
124. Yang, D.; Li, X.; Dai, X.; Zhang, R.; Qi, L.; Zhang, W.; Jiang, Z. All in one network for driver attention monitoring. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Barcelona, Spain, 4–8 May 2020; pp. 1–6.
125. Beggiato, M.; Rauh, N.; Krems, J. Facial Expressions as Indicator for Discomfort in Automated Driving. In Proceedings of the International Conference on Intelligent Human Systems Integration, Modena, Italy, 19–21 February 2020; pp. 932–937.
126. Torstensson, M.; Duran, B.; Englund, C. Using Recurrent Neural Networks for Action and Intention Recognition of Car Drivers. In Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods, Prague, Czech Republic, 19–21 February 2019; pp. 232–242.
127. Valle, F.; Galozy, A.; Ashfaq, A.; Etmnani, F.; Vinel, A.; Cooney, M. Lonely road: Speculative challenges for a social media robot aimed to reduce driver loneliness. In Proceedings of the MAISON2021, Virtual Conference, 7 June 2021; p. 6.
128. Guo, G.; Zhang, N. A survey on deep learning based face recognition. *Comput. Vis. Image Underst.* **2019**, *189*, 102805. [\[CrossRef\]](#)
129. Li, S.; Deng, W. Deep Facial Expression Recognition: A Survey. *IEEE Trans. Affect. Comput.* **2020**. [\[CrossRef\]](#)
130. Alonso-Fernandez, F.; Bigun, J.; Englund, C. Expression Recognition Using the Periocular Region: A Feasibility Study. In Proceedings of the 2018 14th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 26–29 November 2018; pp. 536–541.
131. Alonso-Fernandez, F.; Bigun, J. A survey on periocular biometrics research. *Pattern Recognit. Lett.* **2016**, *82*, 92–105. [\[CrossRef\]](#)
132. Wartzek, T.; Eilebrecht, B.; Lem, J.; Lindner, H.; Leonhardt, S.; Walter, M. ECG on the Road: Robust and Unobtrusive Estimation of Heart Rate. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 3112–3120. [\[CrossRef\]](#)
133. Macias, R.; García, M.A.; Ramos, J.; Bragos, R.; Fernández, M. Ventilation and Heart Rate Monitoring in Drivers using a Contactless Electrical Bioimpedance System. *J. Phys. Conf. Ser.* **2013**, *434*, 012047. [\[CrossRef\]](#)
134. Jaffrin, M.Y.; Morel, H. Body fluid volumes measurements by impedance: A review of bioimpedance spectroscopy (BIS) and bioimpedance analysis (BIA) methods. *Med Eng. Phys.* **2008**, *30*, 1257–1269.
135. Li, C.; Lubecke, V.M.; Boric-Lubecke, O.; Lin, J. A Review on Recent Advances in Doppler Radar Sensors for Noncontact Healthcare Monitoring. *IEEE Trans. Microw. Theory Tech.* **2013**, *61*, 2046–2060. [\[CrossRef\]](#)
136. Ioannou, S.; Gallese, V.; Merla, A. Thermal infrared imaging in psychophysiology: Potentialities and limits. *Psychophysiology* **2014**, *51*, 951–963.
137. Nicolini, P.; Ciulla, M.M.; Malfatto, G.; Abbate, C.; Mari, D.; Rossi, P.D.; Pettenuzzo, E.; Magrini, F.; Consonni, D.; Lombardi, F. Autonomic dysfunction in mild cognitive impairment: Evidence from power spectral analysis of heart rate variability in a cross-sectional case-control study. *PLoS ONE* **2014**, *9*. [\[CrossRef\]](#) [\[PubMed\]](#)
138. Maiorana, E.; Campisi, P. Longitudinal Evaluation of EEG-Based Biometric Recognition. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 1123–1138. [\[CrossRef\]](#)
139. Jain, A.; Nandakumar, K.; Ross, A. 50 Years of Biometric Research: Accomplishments, Challenges, and Opportunities. *Pattern Recognit. Lett.* **2016**, *79*, 80–105. [\[CrossRef\]](#)
140. Varytimidis, D.; Alonso-Fernandez, F.; Duran, B.; Englund, C. Action and intention recognition of pedestrians in urban traffic. In Proceedings of the 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 26–29 November 2018; pp. 676–682.
141. Chen, L.; Englund, C. Cooperative ITS-EU standards to accelerate cooperative mobility. In Proceedings of the 3rd International Conference on Connected Vehicles & Expo (ICCVE 2014), Vienna, Austria, 3–7 November 2014; pp. 681–686.

142. Englund, C.; Lidström, K.; Nilsson, J. On the need for standardized representations of cooperative vehicle behavior. In Proceedings of the Second International Symposium on Future Active Safety Technology toward Zero-Traffic-Accident, 22–26 September 2013; pp. 1–6.
143. Cuthbertsson, A. Hacked Billboards Could Trick Self-Driving Cars Into Suddenly Stopping. In *Independent*, Thursday 15 October; Independent Digital News & Media Ltd., Northcliffe House: Kensington, UK, 2020.
144. Greenberg, A. Hackers Remotely Kill a Jeep on the Highway—With Me in It. *WIRED* **2015**. Available online: <https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/> (accessed on 16 May 2021).
145. Ortiz, S.; Calafate, C.T.; Cano, J.C.; Manzonni, P.; Toh, C.K. A UAV-based content delivery architecture for rural areas and future smart cities. *IEEE Internet Comput.* **2018**, *23*, 29–36. [CrossRef]
146. ISO. *Road Vehicles—Functional Safety—Part 1: Vocabulary*; 2018; p. 33. Available online: <https://www.iso.org/standard/68383.html> (accessed on 16 May 2021).
147. ISO. *Road Vehicles—Safety of the Intended Functionality*; 2019; p. 54. Available online: <https://www.iso.org/standard/70939.html> (accessed on 16 May 2021).
148. ISO. *Intelligent Transport Systems—Forward Vehicle Collision Mitigation Systems—Operation, Performance, and Verification Requirements*; 2013; p. 33. Available online: <https://www.iso.org/standard/45339.html> (accessed on 16 May 2021).
149. ISO. *Intelligent transport systems—Pedestrian detection and collision mitigation systems (PDCMS)—Performance requirements and test procedures*; 2017; p. 21. Available online: <https://www.iso.org/standard/64111.html> (accessed on 17 May 2021).
150. ISO. *Intelligent Transport Systems—Bicyclist Detection and Collision Mitigation Systems (BDCMS)—Performance Requirements and Test Procedures*; 2020; p. 18. Available online: <https://www.iso.org/standard/72508.html> (accessed on 16 May 2021).
151. ISO. *Road Vehicles—Ergonomic Aspects of External Visual Communication from Automated Vehicles to Other Road Users*; 2018; p. 7. Available online: <https://www.iso.org/standard/74397.html> (accessed on 16 May 2021).