Autonomous Vehicles: From Vision to Reality

Olivia Lee, Daniel Zamoshchin

Stanford University

November 20, 2020

Abstract

Autonomous Vehicles are undoubtedly a vision of the future—a vision recognizable in almost any science-fiction novel or movie. After years of development, how close is that vision to reality? In the last decade, researchers and technology firms worldwide have made remarkable breakthroughs in robotics, artificial intelligence, and perceptive sensory technologies. For example, new research in deep learning and improvements in data collection methods, like simulated driving time, have made autonomous technology more advanced than ever before.

While the advancements in AI and self-driving technology thus far have been impressive, just as many challenges lie ahead in enhancing the security of autonomous systems and implementation of the system at scale. Challenges include consumer safety and securing autonomous systems against unwanted exploitation, which will require tackling not only technological hurdles but also challenges on ethical and public policy fronts. This paper seeks to provide a high-level overview of current developments in the field, as well as a look into the challenges that lie ahead through a technological, ethical, and policy perspective, in order to make projections about the future of autonomous vehicles.

Contents

1 Introduction

An Automated Driving System (ADS) is a hardware-software system that can execute dynamic driving tasks on a sustainable basis [1]. Alternative vernacular terms include autonomous vehicles, autonomous driving, and self-driving cars, but ADS is the most scientifically precise. ADSs have great potential to generate positive benefits for society. Given that 94% of road casualties are caused by human error [2] as compared to only 2% caused by vehicle failure [3], ADSs are being developed with the promise of reducing road casualties, as they are not subject to fatigue or other causes of human error. ADSs can also help reduce road congestion, and by extension, reduce energy consumption and emissions, as many ADSs are used for car-sharing services and being powered by alternative energy sources. ADSs will vastly facilitate the transportation of the elderly and mobility-impaired. Lastly, ADSs could help reduce driving-related stress and increase productivity by reallocating the time spent on the road to other activities, among other benefits [4]. In sum, the annual social benefits of ADSs are projected to reach $800 billion by 2050 [5].

Research and development conducted to make ADSs more robust occurs at the intersection of vehicle dynamics, computer vision (which has been advanced by the evolution of new deep learning architectures), and the availability of new sensor modalities. The rapid development of these fields in turn facilitated a boom in this area, spanning both industry and academia. The recent activity in the field of ADSs was accelerated by the DARPA Grand Challenges that started in 2004, organized by the US Department of Defense [6]. As we will see shortly, within the short span of 3 years, the number of teams that successfully completed the challenge increased from 0 to 6. Two members from the winning and runner-up teams of the 2005 Challenge, Professors Sebastian Thrun and Chris Urmson, led Google's Self-Driving Car Project. The project became Waymo in 2016 [7], considered to be one of the leading companies in the field of ADS. Universities around the world have dedicated research laboratories and teams working to improve the accuracy and security of ADSs. By 2013, several large automotive companies including General Motors, Ford, Toyota, Mercedes-Benz, Volkswagen, and BMW were developing their own ADS technologies, and continue to invest heavily in ADSs today. Other enterprise companies including Uber, Tesla, Cruise, Waymo, and Argo AI have also pursued extensive ADS projects with varying degrees of automation.

Before reviewing the current state-of-the-art systems and making future projections about the direction of the industry, it is important to reflect on the key milestones on which the current research is premised. The concept of autonomous vehicles was already beginning to take shape around 100 years ago, but even today, we still have a long way to go before it can be fully realized. Research and development has ramped up significantly, especially in the last 10 years, as artificial intelligence, robotics, and sensing modalities have vastly improved during this period. The timeline below summarizes some of the key milestones and breakthroughs in the development of ADSs:

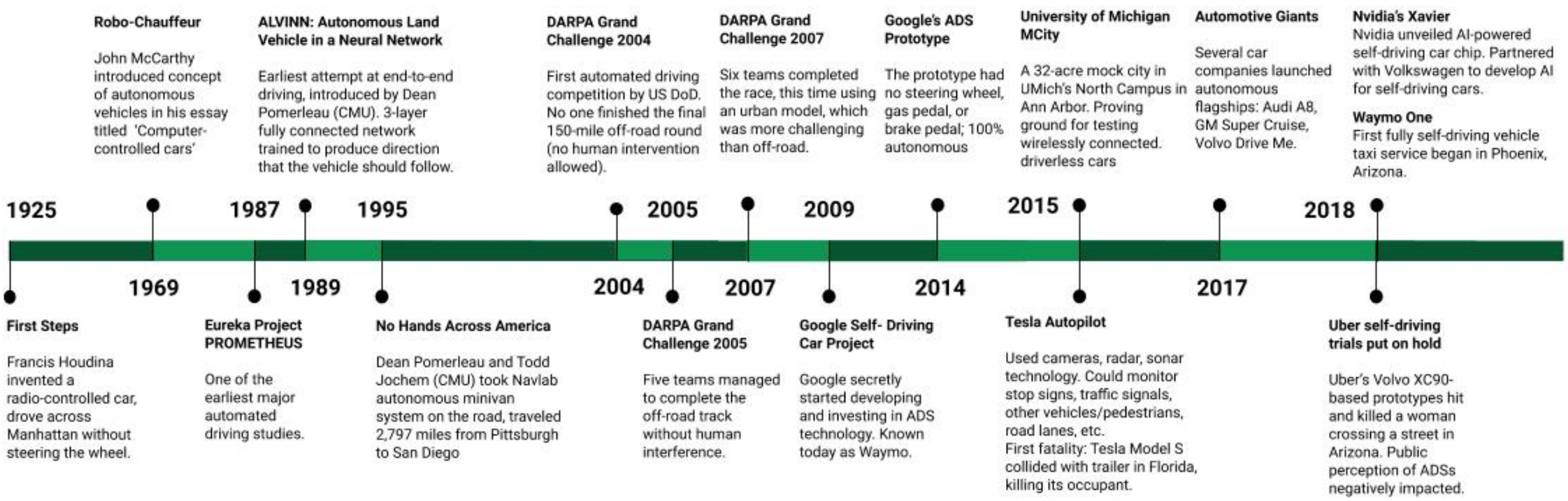

Figure 1: A timeline of key milestones and developments in the field of ADS. [8] [9] [10]

In terms of system architecture, many ADSs employ a modular, divide-and-conquer approach: dividing the huge and complex task of automated driving into smaller modules, each tackled with a unique combination of sensors and algorithms. In more recent-projects, end-to-end architectures have been explored as an alternative for modular approaches. A variety of different sensor technologies for 2D/3D object perception, such as cameras, radar, and LiDAR, have been explored and benchmarked as well. Several deep learning architectures have been developed and employed to process sensor data, especially for image processing and object detection. Automation levels, shown in Figure 2, characterize the progress and abilities of these systems. The majority of current automated driving systems are Levels 2 and 3, with a few state-of-the-art system claiming to have achieved Level 4 automation.

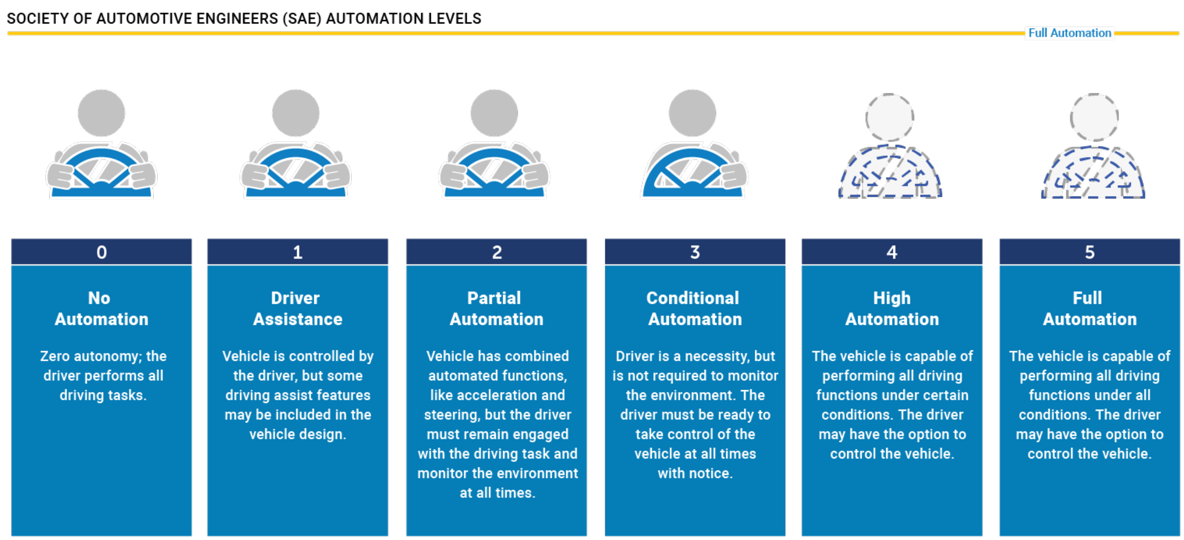

Figure 2: A brief description of the different automation levels.

This paper will begin with an overview of the state-of-the-art techniques and practices presently employed in the development of ADSs, specifically the computational and mathematical principles behind these techniques. We continue with an analysis of the technical challenges that remain to be addressed through further research and development. We then proceed to discuss the ethical and policy implications of the development of ADSs. We conclude with future projections about the field and its broader implications for society.

2 Techniques and Practices

The Automated Driving problem is an exceedingly complex and difficult problem. With an almost infinite number of driving scenarios and possible driving maneuvers, solving the Automated Driving problem will not just require breakthrough discoveries in computer science and AI but will also likely demand new research work in mathematics. In this section, we will discuss the state of current techniques and practices and delve deeper into a few areas where interesting work is being done.

2.1 Perception

The primary perception modules on ADSs are radar, camera, LiDAR, and ultrasonic sensors. The data from these sensors allow ADSs to "understand" the surrounding environment, including predicting the movement of vehicles, pedestrians, bicycles, and other dynamic elements. Other sensor data includes Global Positioning System (GPS) and speedometer and odometer data. ADSs navigate their surroundings based on this information. Each type of sensor has clear advantages and disadvantages and are used in many different applications.

Radio

Radar, or Radio Detection and Ranging, uses radio waves to detect obstacles and measure distances. Radar is especially useful at night where low-light conditions make it difficult for other sensors like cameras to pick up on objects. Radar is also advantageous because it has a long range and is unaffected by weather. However, compared to LiDAR, radar has a lower accuracy, especially in short ranges.

LiDAR

LiDAR, or Light Detection and Ranging, is another sensor module widely used in many modern ADSs. LiDAR is used to determine distances in three-dimensional spaces by measuring the reflection of pulsed infrared laser light. LiDAR sensors in the front, back, and on top of a car can create a 3D cloud of points of the surroundings with very high precision

[11]. LiDAR has a medium range and high accuracy within 200 meters, but high relative cost and is affected by elements of weather like fog and snow.

Three-dimensional LiDAR was introduced after the DARPA Grand Challenge in 2005. The introduction of 3D LiDAR has played a crucial role in accelerating the development of ADSs. At the time, each spinning unit cost $75,000. Since 2005, however, the cost of LiDAR has dramatically decreased. In 2020, Velodyne Lidar will sell a LiDAR sensor for $100 [12].

Camera

Several types of cameras are used in a wide variety of applications for automated driving. Monocular cameras, which can detect color, is important for determining the status of traffic lights. Omnidirectional cameras, cameras with 360° vision, can be used in mapping and localization tasks [13]. In some systems, thermal cameras are used to ensure pedestrians are detecting in any lighting conditions [14]. Event cameras, which are gaining interest, only transmit changes in light intensity between frames [15]. This data can be especially useful for convolutional neural networks. Cameras are critical for many automated driving tasks and are small and cheap. However, all types of cameras are affected by weather.

2.2 System Architectures

Ego-only systems and connected systems, shown in Figure 3, are the two primary architectures in consideration for ADSs. Each system has different possible algorithmic designs and utilize many different technologies. We will discuss the advantages and disadvantages of both systems in Section 3.1.

2.2.1 Ego-only Systems

In ego-only systems, the vehicle operates independently based on its own sensor data and autonomous driving design. In other words, all necessary autonomous capabilities are performed on the vehicle without communicating with other vehicles or receiving external data streams. Ego-only systems, as opposed to connected systems, are overwhelmingly the most common approaches to current ADSs. There are two main types of ego-only systems.

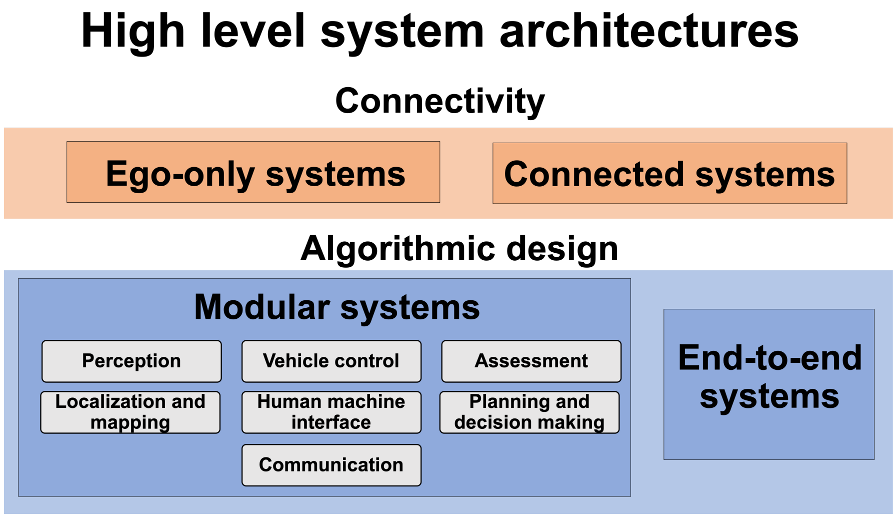

Figure 3: A high level overview of different autonomous architectures [13]

2.2.1.1 End-to-End System

End-to-end systems determine their own tasks and rules for automated driving based heavily on input data [13]. Some end-to-end systems learn how humans drive by training on human driving data. We will discuss various possible implementations in Section 2.3.1.

2.2.1.2 Modular Systems

Modular Systems break down the autonomous driving problem into several key sub-problems or modules in hopes that the sub-tasks of autonomous driving are easier to solve. Ego-only modular systems are the most common implementation for ADSs.

- **Moving Object Tracking:** Classifying objects including vehicles and pedestrians, localizing their movements, and predicting their future movements.
- **Traffic Signalization Detection:** Detecting and interpreting road signs, speed signs, and the status of traffic lights.
- **Route Planning:** Planning the optimal route to the destination, including which lanes to take depending on traffic conditions.
- **Motion Planning:** Deciding short-term movements to avoid obstacles, switch lanes, and more driving maneuvers.

- **Vehicle Control:** Translating decisions to vehicle control, including calculating the proper acceleration and and steering degree.

- **Assessment and Decision-Making:** Deciding which task to execute at the appropriate time.

2.2.2 Connected Systems

Connected Systems are multi-agent systems where autonomous vehicles may communicate with each other as well as with sensors in their environment such as connected traffic lights [13]. In the future, as a greater percentage of self-driving cars take the road, we may see connected architectures become the primary architecture for ADSs. Although this is not feasible in the short-term, we will discuss the benefits of such a system in Section 5.1 because it is a promising vision for the future of ADSs.

2.3 Computational Principles

2.3.1 End-to-End Systems

There are three primary computational approaches to end-to-end systems: deep supervised learning, deep reinforcement learning, and neuroevolution.

Deep Supervised Learning

Deep supervised learning is a deep convolutional neural network trained with human driving data [16]. The system is designed to perform like the training data, essentially attempting to learn how humans drive. The significant drawback of this approach is that, although it learns how the training data acts, the system poorly generalizes to other driving scenarios. In order for this model to perform well it needs an enormous amount of training data. But even with more data, there is an almost infinite number of complex driving maneuvers and situations that would likely prove challenging for the model. For example, navigating construction zones would be difficult for this method. The system is heavily reliant on observing human driving maneuvers and cannot "think" of its own.

Deep Reinforcement Learning

The goal of deep reinforcement learning for end-to-end systems is to choose driving actions that result in maximal future rewards [17]. As opposed to deep supervised learning, deep reinforcement learning does not mimic human driving. It aims to find the most optimal method of driving. This approach is relatively new and has not yet been tested in urban settings.

Neuroevolution

Neuroevolution uses evolutionary algorithms in order to train neural networks [18]. The advantage of this type of training is that backpropogation is not used for training. This means the system could better learn how to generalize. Although it has not been implemented in a real-world solution yet, neuroevolution has been used to train prototype autonomous system.

2.3.2 Modular Systems

Problem	Current Approaches
Moving Object Tracking	Model Based MOT, Deep Learning Based MOT
Traffic Signalization Detection	SVMs, CNNs
Route Planning	A*, Dijkstra's Algorithm
Motion Planning	Interpolating Curve Based Techniques
Vehicle Control	Model Predictive Control
Assessment and Decision-Making	FSM, Markov Decision Processes

Table 1: An overview of the computational principles at work in modular ADSs [19].

Moving Object Tracking

There are several approaches to Moving Object Tracking (MOT). Deep Learning Based MOT uses deep neural networks to detect the position of moving objects and then makes predictions on their movements. The Overfeat Convolutional Neural Network is used to detect moving objects in three steps [20].

1. **Classification:** What is the object?

   A convolutional neural network (CNN) is used to classify images. The neural network is trained using supervised learning. With a labeled image dataset, backpropogation is used to adjust the weights of the model to get the expected results.

2. **Localization:** Where is the object in the image?

   In order to answer this quesion, the Overfeat CNN uses a technique called Sliding Window Detection. A grid patch of a certain size is chosen. The classifier is run on the grid patch. The grid patch is then slid over. This is repeated until the end of the image. At the end of the image, we increase the size of the patch and do it again. Unfortunately, this is a time consuming process, but on the road, situations can change rapidly, and the vehicle cannot wait for this processing. By implementing a convolutional approach to this algorithm, we can make it fast.

   You Only Look Once (YOLO)

   Instead of running the Sliding Window Detection, which requires looking at the same image multiple times in multiple places (with various sizes of patches), we can look at the image only once [21]. First, we divide the entire image up into certain size grid patches. Then, we use the classifier to predict class probabilities and run bounding box regression for each grid cell. If there are multiple objects in a single grid, we use anchor boxes that represent the ideal shape and size of the object that was classified [21]. As a result, YOLO is an astoundingly fast and accurate algorithm. Small improvements have been made since it was first introduced, but it is still the primary algorithm for real-time object detection.

   For better localization, the intersection over union or IoU is calculated for the actual bounding box and the predicted bounding box [22]. The algorithm is trained to maximize the IoU, essentially minimizing the distance between the expected and predicted bounding boxes. The localization step also merges bounding boxes with an algorithm that takes classification in consideration. This means that it only merges bounding boxes which have a high probability of being the same object.

3. **Detection:** Do this for all objects in the image.

    This step is the same as classification and localization. However, there is an extra step because algorithm must also distinguish the background from the objects.

After using the Overfeat CNN, the next step is to predict the future movements. The estimated positions of these objects are recorded from the LiDAR and camera data over time [19]. Given this data, a recurrent neural network can predict the future movement of objects like pedestrians [23].

Traffic Signalization Detection

Traffic Signalization Detection is similar to the object tracking problem because the YOLO algorithm is used to localize objects like stop signs, speed signs, and traffic lights. In some implementations, Support Vector Machines (SVMs) are used to classify road signs and interpret the status of traffic lights using color data from cameras [24].

Motion and Route Planning

ADSs need to plan an optimal route to the destination. Similar to services like Google Maps, the mosst common method ADSs use to plan routes are graph search techniques like A* or Dijkstra's algorithm. Dijkstra's algorithm finds the shortest path between an initial node and a goal node [25]. A* improves on the efficiency of Dijkstra's in certain cases by using heuristics to prune certain paths.

Motion Planning uses similar techniques in order to reach short-term goals like avoiding an obstacle. Another technique for this is curve interpolation which results in a smooth path but is slow to compute. To avoid an obstacle, this technique interpolates a set of points that are collision-free to find a smooth curve [13]. Services like Zoox, Waymo, and Cruise rely on pre-mapped areas to make decisions like which lane to take.

Vehicle Control

Based on actions from other modules, Model Predictive Control (MPC) computes the steering, throttle, and braking needed to perform the required action [26]. MPC also models the ADS's behavior and attempts to correct errors. This was the method used by Stanford's robotic car Junior from the 2007 DARPA Urban Challenge [19].

Assessment and Decision-Making

Combining all of these possible routines, the ADS must make decisions about what the current goal should be. For the DARPA Urban Challenge, the primary techniques used to decide behavior were Finite State Machines, heuristics, and decision trees [19]. In more advanced ADSs, Markov Decision Processes are used to decide behavior. For example, a Partially Observable Markov Decision Process was used to determine whether a car can change lanes based on a graph of distances and velocities of objects around the car [27].

3 Challenges

We will now discuss the current challenges that need to be overcome through further research and development before large-scale implementation of ADSs can be achieved.

3.1 Complexity of Driving Maneuvers

The first and most significant challenge from a technical standpoint is how the system should handle complex driving maneuvers. Significant progress in AI is needed for more complex driving maneuvers to be carried out. The specific challenges differ slightly for different system architectures, which we will analyze in turn.

3.1.1 End-to-End Systems

For end-to-end systems, the most significant issue is the lack of hard-coded safety measures, because the system learns the optimal method of driving on its own. This would be present in the modular architecture as one of its components. There is also the issue of interpretability [28]. Deep learning models are often described as complicated, non-linear "black boxes," meaning their predictions, though possibly highly accurate, are not understandable by humans. As safety is the highest priority for ADSs, it is crucial for users to fully understand the predictive models underlying their vehicle. An important body of research is therefore dedicated to visualizing deep networks. In the case of ADSs, users would ideally be able to visualize and easily interpret the image regions that crucially influence the final behavior of the vehicle [29].

The different deep learning approaches to end-to-end systems each have their own merits and drawbacks. Deep reinforcement learning and neuroevolution both require online interaction with the environment. They also fail repeatedly in order to learn the desired behavior [30] [31]. This contrasts with direct supervised networks that can be trained offline with human driving data, and once that's done the system is not expected to fail during operation [32]. However, for direct supervised networks, it requires not only a significant amount of labeled data, but it also uses human behavior as a basis for learning. This raises the question of whether we should train ADSs to drive like humans, or if it should learn a more optimal driving method on its own using deep reinforcement learning and neuroevolution. As a result of these unresolved challenges, end-to-end systems have not yet been realized in real-world urban scenes.

3.1.2 Modular Systems

Moving onto the challenges facing modular architectures, we shall analyze the challenges in each of the modules or subproblems in Figure 3.

3.1.2.1 Perception

Perception refers to the collection of information from sensors and the derivation of contextual inferences about the environment. Referring to the components of Modular Systems in Section 2.2.1.2, the components Moving Object Tracking and Traffic Signalization Detection are included in this category. With regards to perception, we have discussed some exteroceptive sensors used by ADSs, each with their own advantages and disadvantages. Starting with monocular cameras, one significant advantage is color sensing, which is important for traffic light and sign recognition. Many sophisticated deep learning algorithms, such as convolutional neural networks, have been developed for 2D image recognition and object detection.

However, the limited field of vision of monocular cameras is an issue for higher levels of automation. One option is to use arrays of monocular cameras and perform external calibration to effectively stitch images together, but a promising alternative is a smaller number of omnidirectional cameras (360° or panoramic cameras). Omnidirectional perception is especially important to obtain inputs passed on to other modules involving navigation, localization, and mapping. One challenge, however, is that there are difficulties calibrating the images due to the spherical distortions of the images [33].

Furthermore, research is still ongoing to improve depth perception using cameras [34]. LiDAR and radar sensors are especially useful for depth perception, as precise distances between objects can be measured effectively by emitting infrared light (for LiDAR) or radio waves (for radar). However, the emission of energy in the form of electromagnetic (EM) waves can interfere with other signal-dependent hardware in the ADS, unlike cameras, which are passive sensors that perceive EM waves in the environment. Comparing the two ultrasonic sensors, LiDAR tends to be more accurate for shorter distances (<200m), while radar can detect objects at further distances away at a lower resolution [35]. Radar sensors are also cheaper and more compact than LiDAR sensors. Thus far, incorporating sensory perceptive abilities to artificial intelligence systems that come naturally to humans remains technologically challenging. Given the complementary nature of cameras, LiDAR, and radar sensors, future research directions can focus on combining these methods to leverage the unique advantages provided by each of the sensor modalities.

While a collaborative system incorporating multiple modalities is promising, a common shortcoming among most of the aforementioned technologies is that they are significantly affected by inclement weather that compromise the illumination and/or clarity of the surrounding environment. The performance of systems relying on camera-based perception (both monocular and omnidirectional) can fluctuate with changes in lighting. The operation of camera-based ADSs in low light conditions is an ongoing field of research [36] [37]. Changes in illumination due to shifting shadows or changing weather conditions can thus cause algorithms to fail, especially supervised learning methods. While LiDAR and radar sensors are not as adversely affected by changes in lighting, the performance of LiDAR sensors also suffers in weather conditions such as fog or snow [38]. While radars are more robust in poor weather, as mentioned above, the resolution of radar sensors is currently too low for short-range perception tasks. Thus, successfully performing driving tasks in inclement weather remains a central challenge in the perception module of ADS.

3.1.2.2 Localization and Mapping

Localization involves determining the ADS's position with respect to other objects in the driving environment, while mapping involves creating a map with information of the environment and context (including traffic rules). Referring to the components of Modular

Systems in Section 2.2.1.2, the Route Planning component is included in this category. Traditionally, satellite-based systems (e.g. GPS) can be used to estimate global position at a relatively low cost. GPS is often integrated with Inertial Measurement Unit (IMU) systems, and the GPS-IMU fusion is used for low-cost vehicle localization. The IMU tracks changes in position and orientation, thus estimating vehicle position relative to initial position using a dead reckoning algorithm. The primary shortcoming of dead reckoning is that it is subject to cumulative errors, which requires correction at regular intervals using GPS readings. However, while GPS readings are often accurate, they are not always available as they require a clear view of the sky. Especially in densely populated, metropolis areas with tunnels and tall urban infrastructure, the accuracy is further compromised [39]. GPS-IMU fusion systems are therefore not sufficiently accurate for microlevel localization and can at best assist with route planning.

One alternative to the GPS-IMU fusion system is using a-priori map-based localization. This involves constructing a highly detailed map of the environment a-priori. Localization then occurs when comparing the ADS's position readings (usually using a GPS reading at its initial position) with the map's information to estimate the ADS's position on the map. Map construction usually falls into two broad categories: planar map construction using layers or planes from a Geographic Information System, and construction of maps using point-cloud matching [40]. In either case, creating and maintaining accurate maps are computationally expensive and difficult to scale for potential large-scale implementations of ADS systems. A-priori map-based localization is also a relatively fragile system that is sensitive to changes in the environment. Thus, it may not be a feasible approach in the long run due to inevitable changes in the environment and the vast size of road networks.

Studies have recently emerged to tackle this challenge by investigating simultaneous localization and mapping (SLAM). SLAM involves the construction of maps during online interactions with the environment, while simultaneously localizing the vehicle within the map. Starting from an unknown location within an unknown environment, the ADS locates its own position through repeated observation of environmental features as it moves, creating an incremental map of the surrounding environment according to its own position. SLAM has a significant edge over a-priori map-based localization due to their inherent flexibility. SLAM systems can operate in any environment, and do not suffer from the same sensitivities to environmental changes as a-priori map-based localization.

However, SLAM remains computationally challenging, especially in outdoor environments. Advances in image-based computer vision algorithms, like feature extraction and matching, and 3-D reconstruction via triangulation, still need to be made for sufficiently accurate visual SLAM to be employed in ADSs. As a result. it is currently less accurate than performing localization with a detailed pre-built map. Despite these challenges, given the fragility and low scalability of a-priori map-based localization, SLAM remains a promising alternative, thus investigating methods to make SLAM more efficient is an active area of research [41] [42].

3.1.2.3 Vehicle Control

Vehicle control refers to the selection of appropriate actuator inputs and execution of the planned actions. Referring to the components of Modular Systems in Section 2.2.1.2, the components Motion Planning and Vehicle Control are included in this category. There has been a great interest in exploring the benefits that deep learning provides for vehicle control. Deep learning allows for self-optimization of ADS behaviour from data, allowing it to adapt to new scenarios. Deep learning is therefore suitable for vehicle control in complex and dynamic environments. Instead of manually tuning each parameter iteratively to optimize performance, deep learning enables developers to teach the system to perform the desired behavior and generalize to new environments through learning. Learning methods are typically classified as reinforcement learning and supervised learning, and control styles classified as lateral control, longitudinal control and joint control.

The main challenge for deep learning methods is the large amount of data and time required for adequate training, especially for deep reinforcement learning. The amount of training data required to build a reliable and robust ADS can be significant. Generalizability hence becomes an issue, since it is challenging to train an ADS for all possible scenarios on the road. Besides the large quantity of data, ensuring the fairness of the dataset is also important. Increasing the quantity of data without increasing variety runs the risk of overfitting [43]. The continuous states and actions of the dynamic environment in which the ADS operates adds to the computational complexity due to the Curse of Dimensionality: the computational complexity grows exponentially with the number of dimensions [44]. Not only does this make the development and training of networks challenging, it also implies that the vehicle must have large computing and power capabilities on-board, which affects vehicle size/weight and cost.

Furthermore, it is unclear what the ideal neural network architecture is for a given task. The selection of fundamental architecture, training method, learning rate, loss function, batch size etc., is a general problem for deep learning applications, however the complex task of autonomous driving has specific challenges [45]. Currently, most end-to-end driving systems use smaller networks to avoid overfitting to the relatively small training datasets. However, as mentioned above, deeper architectures reduce both bias and variance when training on large datasets. One possibility being investigated is conditional imitation learning, which includes a different final layer for each high-level command for driving [46]. Further research can be conducted to design architectures specific for the task of autonomous driving.

For deep reinforcement learning, the desired behaviour of the agent must be accurately captured by the reward function, otherwise unexpected and undesirable behaviour might occur. Goal specification is hence a challenge, especially for a complex task such as driving. A multi-objective reward function is required, which needs to consider a variety of objectives that may even conflict with one another. Objectives may include maintaining a safe distance from other vehicles, staying close to the centre of the lane, avoiding pedestrians, not changing lanes too often, maintaining desired velocity, and avoiding harsh accelerations/braking. These can easily conflict with each other in an emergency situation. Hence, the reward function must also take into account the weight of these factors [47].

3.1.2.4 Human-Machine Interfaces

A key component of communication between humans and vehicles is a human-machine interaction (HMI) module. Currently, the most significant challenge with regards to HMI is the handover issue during emergency situations. Studies have shown that humans generally exhibit a lower cognitive load when monitoring automated driving than other tasks. Hence, in an emergency situation, the transition from manual to automated driving and vice versa is currently prone to failure, and often increases collision risk with surrounding vehicles [48]. Investigations into methods of maintaining human engagement are being conducted, so that the occupants are at least aware of the ADS's intentions and basic decision-making processes.

The general consensus is that visual HMI modules are not ideal, as they are affected by alert response time and the time taken to read, understand the alert, and act, which takes too long. It is also difficult to simultaneously monitor a visual HMI module while driving in a takeover situation. The alternative, auditory HMIs, are premised on highly accurate automatic speech recognition (ASR). Though natural language processing (NLP) has advanced significantly in recent years, ASR is still impacted by environmental noise. In an emergency situation, there is limited room for error, thus the NLP system cannot fail to interpret or misinterpret the human driver. Furthermore, moving beyond a limited set of simple voice commands to conversational NLP also remains challenging. As a result, a real world system with an efficient handover interaction module does not exist yet [49].

3.1.2.5 Assessment and Decision-Making

ADSs should be constantly reevaluating the overall risk level of the road situation, as well as predict the intentions of surrounding vehicles and pedestrians. Most of these approaches employ quantitative methods (such as Bayesian methods or Hidden Markov Models). However, there is a short observation window for predicting the behavior of surrounding vehicles or humans, but complex computational models often require longer observation periods [50]. Therefore, the primary challenge with regards to assessment is reducing the time required for situational assessment to a few seconds or less. Especially for human drivers and pedestrians, the unpredictability of human behavior is also a significant challenge.

With regards to decision-making, a robust ADS should be able to sequence multiple decisions and organize the complex driving task. As mentioned in Section 3.1.2.3, there are several objectives that the ADS strives to achieve: staying within the lane at a safe distance from other vehicles, identifying and responding to traffic signs and signals, avoiding pedestrians, changing speed and lange where appropriate, etc. Organizing this complex task and prioritizing the actions to be taken is a significant challenge that still requires further research. This task is also difficult to generalize, as it is impossible to train a vehicle to sequence decisions to every possible driving situation.

3.2 Cybersecurity and Exploitation

A significant practical hurdle that ADSs must overcome is ensuring comprehensive security protections. Most software systems are subject to security pitfalls with zero-day vulnerabilities and various other exploits. However, unlike typical software systems, the remote access of ADSs puts into danger the physical safety of passengers. A remote attacker accessing an AV has the potential to control the vehicle with the potential to cause harm. Thus, significant security research in ADSs is necessary before wide-scale adoption.

Remote attacks on ADSs are both feasible and effective. For example, split-second "phantom" images like in Figure 4 can be used to trick the driver-assistance systems of a Tesla Model X and Mobileye 630 into thinking a depth-less projection is a physical obstacle [51]. Researchers at Israel's Ben Gurion University of the Negev demonstrated that an attacker can embed a split-second image of a road sign into an advertisement on a digital billboard causing the Tesla Autopilot system to suddenly stop in the middle of a road. The researchers found that they could reliably trick the Tesla by showing an image for just 0.42 seconds [51].

The security researchers at Ben Gurion University of the Negev propose a countermeasure consisting of four deep convolutional neural networks that can determine the authenticity of an object based on the object's light, context, surface, and depth. The system successfully achieved a true positive rate (TPR) of 0.994 [51].

Internet-connected vehicles may also have a significant impact on traffic flows. Re-

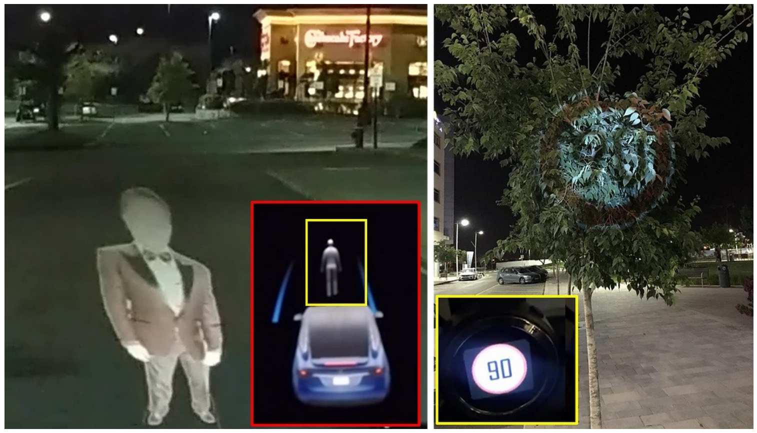

Figure 4: Projections can successfully trick the Tesla semi-autonomous system.

searchers at the Georgia Institute of Technology found that if 10–20% of cars in Manhattan were somehow disabled during rush hour, half of traffic would come to a halt, preventing emergency services from getting through [52]. Even more, they find that "with just 5% of cars hacked, a five-by-five grid could be gridlocked within 15 minutes" [52]. The researchers of this study warn that as more ADSs take the road, the greater the potential for a hack that disrupts critical emergency services in cities.

In perhaps the most concerning hack of connected vehicles, Charlie Miller and Chris Valasek, funded by a DARPA grant, demonstrated that they were able to take full control of a Jeep Cherokee on a highway, while they were in their living room [53]. Through a cellular connection in the Cherokee's entertainment system, Miller and Valasek were able to gain remote access to its steering, transmission, and brakes [53]. Although this demonstration was a proof-of-concept, Miller and Valasek echo the concerns about security in increasingly connected vehicles.

Security must be a priority for future ADSs, and a system of quickly deploying security updates must be established.

3.3 Coexistence with Humans

The final challenge is ensuring ADSs can safely coexist with humans. Many of the aforementioned developmental challenges stem from the fact that ADSs have to anticipate and respond to human behavior (for both drivers and pedestrians). The inherent randomness and unpredictability of human behavior is a significant challenge due to human free will. This is especially relevant for moving object tracking where the system has to anticipate what humans will do next. For example, Google's self-driving car collided with an oncoming bus in 2016 [54]. The ADS decided to perform a lane change under the incorrect assumption that the bus driver was going to yield. Instead, the bus driver accelerated, resulting in the collision. Further investigations conducted by the National Transportation Safety Board into the Uber Autonomous car crash in 2018 revealed that the pedestrian was jaywalking, a factor that the system failed to anticipate, resulting in the collision [55]. Therefore, while efforts have been made to develop models that more accurately predict human behavior [56], there is an inherent challenge in the randomness associated with human behavior.

As a result of this challenge, some are of the opinion that large-scale implementation of ADSs can only succeed if there is a complete overhaul of the system, meaning no human drivers on the road [57] [58] [59]. If the algorithms used for ADSs can be somewhat standardized, ADSs can then mutually predict each other's behavior facilitating the development of large-scale connected systems as mentioned in Section 2.2.2. However, whether this is necessary or even possible remains debated. Furthermore, vehicle-vehicle interaction is just one part of the problem. Elements of randomness still remain due to pedestrians, inclement weather, etc.. Robust ADS systems must be equipped to handle such situations.

4 Ethical Discussion and Policy Implications

While this paper is primarily focused on analyzing the technical development of ADSs, it is difficult to get a holistic view of the field without taking into consideration the ethical and policy implications of developments in ADSs, which we will briefly review in this section.

4.1 The Trolley Problem

An ethical dilemma relevant to the discussion about ADSs is the infamous Trolley Problem. The Trolley Problem is a series of thought experiments, with the central question revolving around whether it is justified to sacrifice one person to save five, as initially outlined in Thomson's 1976 paper [60]. Philosophical theory suggests that consequentialists will believe pulling the lever is justified, while deontologists do not. Nonetheless, the answer to the question is ambiguous. It is true that the Trolley Problem is merely a thought experiment, and developers are primarily grappling with elementary problems preventing ADSs from being fully realized. However, these ethical dilemmas persist as a reminder about how ADSs should respond in inevitable accident situations. It also begs the question of who should decide how the ADSs are programmed to respond – engineers, the government, or some third-party agency – since there is no definitive answer to the dilemma, and responses differ widely around the world due to social, cultural, and other psychological factors.

4.2 Liability and Blackbox Nature

In 2018, an Uber self-driving car in Arizona struck and killed a pedestrian as they were crossing the road [55]. An important question for ADSs is who is ultimately morally and legally liable for accidents. In the case of Uber accident, the safety driver was held responsible as evidence showed that she was distracted and could have otherwise stopped the vehicle [61]. However, in future ADSs with higher levels of automation, there may not be a human driver in charge of monitoring the vehicle. In such a case, it is unclear who will be held liable: the occupant(s), the company developing the ADS, or someone else. In light of this, AI law specialists are investigating ways of giving AI a legal personality [62].

Furthermore, studies have shown that the highly publicized accidents during the testing phases of ADS systems in recent years have not only cost lives, but have also undermined public trust in and receptiveness to ADSs. Nearly three quarters of American drivers would not trust an autonomous car, up significantly from 63% in 2017 [63]. Overcoming the inevitable initial pushback to ADSs will be an additional challenge to overcome as companies try to implement wider scale ADS systems.

4.3 Regulatory Measures

The primary goal of regulatory measures for ADSs should be to establish a safety certification method. Lawmakers and players in the industry have yet to establish a procedure to simulate and test ADSs. Since there are an almost infinite number of possible driving scenarios, simulations will likely be required to comprehensively test ADSs. Consequently, research into how to create simulations that certify low risk and reliability of ADSs is critical.

Regulation has the opportunity to influence the design of ADSs to promote safety. The secondary goal of regulation should be encourage and ensure that the design of ADSs is highly redundant to minimize the chance of a catastrophic failure. However, a balance must be achieved because regulation should not stifle innovation in the ADS space.

The National Highway Traffic Safety Administration estimates that there are 1.25 deaths per 100 million non-autonomous vehicle miles driven [64]. Although current ADSs have a higher death per 100 million miles, this may be attributed to failures in early ADSs and not enough miles driven by state-of-the-art ADSs. Overall accident severity for ADSs is lower than for human-caused accidents. Additionally, the minor accident rate among autonomous vehicles is significantly lower than that of humans [65].

Regulators will eventually have to answer an important open question: how much safer or better do ADSs need to be before they are considered "good enough" to be implemented at a wide scale? Identifying a quantitative benchmark to ascertain the safety of ADSs will be of great importance in the coming years.

5 Conclusion

5.1 Future Projections

Consuming Mobility as a Service (MaaS)

If ADSs are successful in overcoming the current challenges, there may be a large shift from personal vehicle-ownership to shared ownership. 40% of car drivers say that they would be willing to replace some of their private car trips with a car-sharing service [66]. Furthermore, 20% of drivers indicate that they may not buy a planned car or sell a current car if carsharing becomes available [66]. While this shift may already be occuring, ADSs will only decrease the cost of shared ownership and further push this trend forward.

Shared car ownership will, importantly, decrease the environmental impacts of car ownership. A shared electric, self-driving car fleet is preferred over a fossil-fuel fleet and will have significant positive impacts on the environment [66].

Logistics Revolution

When ADS trucks take the road, a logistics revolution may be imminent. Amazon's acquisition of Zoox foreshadows this revolution. While the Zoox acquisition was primarily to help the startup "bring their vision of autonomous ride-hailing to reality," according to Amazon, Zoox's technology includes developing zero-emission ADS vehicles, which could easily translate to advancing Amazon's delivery operations [67]. ADS trucks may potentially make logistics faster, cheaper, and more efficient. Additionally, ADS trucks will reduce truck accidents caused by drowsy driving.

This logistics revolution will also bring with it a displacement of trucking-driving jobs. Policy must be in place to ensure that a smooth transition for this workforce occurs. The bulk of truck-driving jobs are in short-haul and last mile trucking which are much more difficult to automate [68]. Long-haul trucking, which constitutes significantly fewer jobs, will be the first to be automated. Thus, there is time to institute proper policies to ensure that this workforce has a safety net.

Rise of Connected Systems

Though ego-only systems are currently the most common implementation for ADSs, connected systems show great promise for the future of ADSs. Connected ADSs greatly reduce traffic accidents, improve the efficiency of transportation networks, and increase quality of life [69]. A connected architecture would allow the distance between autonomous vehicles to decrease while still maintaining a high speed. As a result, the overall capacity of vehicles in the network would increase, making the entire network more efficient. This requires a greater percentage (or even a majority) of vehicles on the road to be ADSs, as connected systems are premised on the ability of ADSs to communicate with one another. Since much of the cause of inefficiency or accidents on the road are due to human error, with a connected system where most vehicles are rule-based, traffic could potentially move more quickly and efficiently when vehicles are part of a connected network. Hence, we can expect future research and development to be increasingly focused on moving towards connected systems.

Significant progress has been made in the development of ADSs in recent years. While there are a number of technological, ethical, and policy challenges to overcome through further research, many advances have been made thus far, and will continue to be made. The growing demand for ADSs will likely propel breakthroughs in AI, deep learning, and sensor modalities. This brings us closer to robust and secure ADSs, and potentially a wider-scale implementation of connected systems in the future.

References

[1] Society of Automotive Engineers. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, 2016.

[2] S. Singh. Critical reasons for crashes investigated in the national motor vehicle crash causation survey, 2015.

[3] Department for Transport. Research on the impacts of connected and autonomous vehicles (cavs) on traffic flow: Summary report, 2017.

[4] T. J. Crayton and B. M. Meier. Autonomous vehicles: Developing a public health research agenda to frame the future of transportation policy. *Journal of Transport & Health*, 6:245–252, 2017.

[5] W. D. Montgomery, E. L. Groshen R. Mudge, S. Helper, J. P. MacDuffie, and C. Carson. America's workforce and the self-driving future: Realizing productivity gains and spurring economic growth, 2018.

[6] S. Thrun. Winning the darpa grand challenge. In Johannes Fürnkranz, Tobias Scheffer, and Myra Spiliopoulou, editors, *Machine Learning: ECML 2006*, pages 4–4, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[7] H. Reese. Waymo (formerly google's self-driving car project): The smart person's guide, 2017.

[8] S. Gautam. 10 major milestones in the history of self-driving cars.

[9] WIRED Brand Lab. A brief history of autonomous vehicle technology, 2016.

[10] R. Glon and S. Edelstein. The history of self-driving cars, 2020.

[11] J. Hecht. Lidar for self-driving cars. *Optics and Photonics News*, 29(1):26–33, 2018.

[12] P. E. Ross. Velodyne will sell a lidar for $100, Jan 2020.

[13] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access*, 8:58443–58469, 2020.

[14] C. Fries and H. Wuensche. Autonomous convoy driving by night: The vehicle tracking system. In *2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, pages 1–6, 2015.

[15] A. I. Maqueda, A. Loquercio, G. Gallego, N. García, and D. Scaramuzza. Event-based vision meets deep learning on steering prediction for self-driving cars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5419–5427, 2018.

[16] D. A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In *Advances in neural information processing systems*, pages 305–313, 1989.

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[18] D. Floreano, P. Dürr, and C. Mattiussi. Neuroevolution: from architectures to learning. *Evolutionary intelligence*, 1(1):47–62, 2008.

[19] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. Paixão, F. Mutz, L. Veronese, T. Oliveira-Santos, and A. F. De Souza. Self-driving cars: A survey, 2019.

[20] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng. An empirical evaluation of deep learning on highway driving, 2015.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection, 2016.

[22] A. Rosebrock. Intersection over union (iou) for object detection, Nov 2016.

[23] X. Du, R. Vasudevan, and M. Johnson-Roberson. Bio-lstm: A biomechanically inspired recurrent neural network for 3-d pedestrian pose and gait prediction. *IEEE Robotics and Automation Letters*, 4(2):1501–1508, 2019.

[24] C. Jang, C. Kim, D. Kim, M. Lee, and M. Sunwoo. Multiple exposure images based traffic light recognition. In *2014 IEEE Intelligent Vehicles Symposium Proceedings*, pages 1313–1318, 2014.

[25] E. W. Dijkstra et al. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[26] J. Levinson and S. Thrun. Robust vehicle localization in urban environments using probabilistic maps. In *2010 IEEE International Conference on Robotics and Automation*, pages 4372–4378. IEEE, 2010.

[27] J. Bongard. Probabilistic robotics. sebastian thrun, wolfram burgard, and dieter fox.(2005, mit press.) 647 pages, 2008.

[28] L. Chi and Y. Mu. Deep steering: Learning end-to-end driving model from spatial and temporal visual cues. *CoRR*, abs/1708.03798, 2017.

[29] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013.

[30] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.

[31] S. Baluja. Evolution of an artificial neural network based autonomous land vehicle controller. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(3):450–463, 1996.

[32] C. Chen, A. Seff, A. L. Kornhauser, and J. Xiao. Deepdriving: Learning affordance for direct perception in autonomous driving. *CoRR*, abs/1505.00256, 2015.

[33] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In David Vernon, editor, *Computer Vision — ECCV 2000*, pages 445–461, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

[34] A. Saxena, S. Chung, and A. Ng. Learning depth from single monocular images. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18, pages 1161–1168. MIT Press, 2006.

[35] J. Wei, J. M. Snider, J. Kim, J. M. Dolan, R. Rajkumar, and B. Litkouhi. Towards a viable autonomous driving research platform. In *2013 IEEE Intelligent Vehicles Symposium (IV)*, pages 763–770, 2013.

[36] N. Carlevaris-Bianco and R. M. Eustice. Learning visual feature descriptors for dynamic lighting conditions. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2769–2776, 2014.

[37] V. Peretroukhin, W. Vega-Brown, N. Roy, and J. Kelly. Probe-gk: Predictive robust estimation using generalized kernels. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 817–824, 2016.

[38] R. H. Rasshofer and K. Gresser. Automotive radar and lidar systems for next generation driver assistance functions. *Advances in Radio Science*, 3:205–209, 2005.

[39] S. Edelkamp and S. Schrödl. Chapter 17 - vehicle navigation. In S. Edelkamp and S. Schrödl, editors, *Heuristic Search*, pages 737 – 757. Morgan Kaufmann, San Francisco, 2012.

[40] T. Wiemann, A. Nüchter, K. Lingemann, S. Stiene, and J. Hertzberg. Automatic construction of polygonal maps from point cloud data. In *2010 IEEE Safety Security and Rescue Robotics*, pages 1–6, 2010.

[41] B. Huang, J. Zhao, and J. Liu. A survey of simultaneous localization and mapping. *CoRR*, abs/1909.05214, 2019.

[42] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser. Simultaneous localization and mapping: A survey of current trends in autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 2(3):194–220, 2017.

[43] F. Codevilla, E. Santana, A. M. López, and A. Gaidon. Exploring the limitations of behavior cloning for autonomous driving. *CoRR*, abs/1904.08980, 2019.

[44] L. Chen. *Curse of Dimensionality*, pages 545–546. Springer US, Boston, MA, 2009.

[45] Y. LeCun et al. Generalization and network design strategies. *Connectionism in perspective*, 19:143–155, 1989.

[46] F. Codevilla, M. Müller, A. Dosovitskiy, A. M. López, and V. Koltun. End-to-end driving via conditional imitation learning. *CoRR*, abs/1710.02410, 2017.

[47] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah. A survey of deep learning applications to autonomous vehicle control. *CoRR*, abs/1912.10773, 2019.

[48] S. Sibi, H. Ayaz, D. P. Kuhns, D. M. Sirkin, and W. Ju. Monitoring driver cognitive load using functional near infrared spectroscopy in partially autonomous cars. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pages 419–425, 2016.

[49] C. A. Pickering, K. J. Bumnham, and M. J. Richardson. A review of automotive human machine interface technologies and techniques to reduce driver distraction. In *2007 2nd Institution of Engineering and Technology International Conference on System Safety*, pages 223–228, 2007.

[50] E. Yurtsever, S. Yamazaki, C. Miyajima, K. Takeda, M. Mori, K. Hitomi, and M. Egawa. Integrating driving behavior and traffic context through signal symbolization for data reduction and risky lane change detection. *IEEE Transactions on Intelligent Vehicles*, 3(3):242–253, 2018.

[51] B. Nassi, Y. Mirsky, D. Nassi, R. Ben-Netanel, O. Drokin, and Y. Elovici. Phantom of the adas: Securing advanced driver-assistance systems from split-second phantom attacks. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, CCS '20, page 293–308, New York, NY, USA, 2020. Association for Computing Machinery.

[52] S. Vivek, D. Yanni, P. J. Yunker, and J. L. Silverberg. Cyberphysical risks of hacked internet-connected vehicles. *Phys. Rev. E*, 100:012316, Jul 2019.

[53] C. Miller and C. Valasek. Remote exploitation of an unaltered passenger vehicle. *Black Hat USA*, 2015:91, 2015.

[54] D. Lee. Google self-driving car hits a bus, 2016.

[55] National Transportation Safety Board. Ntsb accident summary, 2019.

[56] C. Marina Martinez, M. Heucke, F. Wang, B. Gao, and D. Cao. Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 19(3):666–676, 2018.

[57] J. Webber. Driverless cars: once they're on the road, human drivers should be banned, 2019.

[58] CNBC. The problem with self-driving cars could turn out to be humans, 2017.

[59] L. B. Eliot. An inconvenient truth: Human drivers and autonomous cars mix like oil and water, 2019.

[60] J. J. Thomson. Killing, letting die, and the trolley problem. *The Monist*, 59(2):204–217, 1976.

[61] M. Harris. Ntsb investigation into deadly uber self-driving car crash reveals lax attitude toward safety, 2019.

[62] R. Dremliuga, P. Kuznetcov, and A. Mamychev. Criteria for recognition of ai as a legal person. *Journal of Politics and Law*, 12:105, 08 2019.

[63] American Automobile Association. Aaa: American trust in autonomous vehicles slips, 2018.

[64] NHTSA's National Center for Statistics and Analysis. 2019 fatality data show continued annual decline in traffic deaths, Oct 2020.

[65] T. Maddox. How autonomous vehicles could save over 350k lives in the us and millions worldwide, Feb 2018.

[66] F. Liao, E. Molin, H. Timmermans, and B. van Wee. Carsharing: the impact of system characteristics on its potential to replace private car trips and reduce car ownership. *Transportation*, 47(2):935–970, 2020.

[67] D. Etherington. Amazon to acquire autonomous driving startup zoox, 2020.

[68] M. Gittleman and K. Monaco. Truck-driving jobs: Are they headed for rapid elimination? *ILR Review*, 73(1):3–24, 2020.

[69] J. N Bajpai. Emerging vehicle technologies & the search for urban mobility solutions. *Urban, Planning and Transport Research*, 4(1):83–100, 2016.