Array 10 (2021) 100057

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/journals/array/2590-0056/open-access-journal

Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues

Abhishek Gupta, Alagan Anpalagan *, Ling Guan, Ahmed Shaharyar Khwaja

Ryerson University, 350 Victoria Street, Toronto, M5B2K3, Ontario, Canada

A R T I C L E I N F O

Keywords:
Self-driving cars
Levels of automation
Machine learning
Deep learning
Convolutional neural networks
Scene perception
Object detection
Multimodal sensor fusion
LiDAR
Computer vision
Autonomous driving initiatives

A B S T R A C T

This article presents a comprehensive survey of deep learning applications for object detection and scene perception in autonomous vehicles. Unlike existing review papers, we examine the theory underlying self-driving vehicles from deep learning perspective and current implementations, followed by their critical evaluations. Deep learning is one potential solution for object detection and scene perception problems, which can enable algorithm-driven and data-driven cars. In this article, we aim to bridge the gap between deep learning and self-driving cars through a comprehensive survey. We begin with an introduction to self-driving cars, deep learning, and computer vision followed by an overview of artificial general intelligence. Then, we classify existing powerful deep learning libraries and their role and significance in the growth of deep learning. Finally, we discuss several techniques that address the image perception issues in real-time driving, and critically evaluate recent implementations and tests conducted on self-driving cars. The findings and practices at various stages are summarized to correlate prevalent and futuristic techniques, and the applicability, scalability and feasibility of deep learning to self-driving cars for achieving safe driving without human intervention. Based on the current survey, several recommendations for further research are discussed at the end of this article.

1. Introduction

With recent advances in artificial intelligence (AI), machine learning (ML) and deep learning (DL), various applications of these techniques have gained prominence and come to fore. One such application is self-driving cars, which is anticipated to have a profound and revolutionary impact on society and the way people commute [1]. Although, the acceptance and domestication of technology can face initial or prolonged reluctance, yet these cars will mark the first far reaching integration of personal robots into the human society [2]. The last decade has witnessed growing research interest in applying AI to drive cars [3]. Due to rapid advances in AI and associated technologies, cars are eventually poised to evolve into autonomous robots entrusted with human lives, and bring about a diverse socio-economic impact [4]. However, for these cars to become a functional reality, they need to be equipped with perception and cognition to tackle high-pressure real-life scenarios, arrive at suitable decisions, and take appropriate and safest action at all times [5].

Embedded in the self-driving vehicles' AI are visual recognition systems (VRS) that encompass image classification, object detection, segmentation, and localization for basic ocular performance [6]. Object detection is emerging as a subdomain of computer vision (CV) that benefits from DL, especially convolutional neural networks (CNNs) [7]. This article discusses the self-driving cars' vision systems, role of DL to interpret complex vision, enhance perception, and actuate kinematic manoeuvres in self-driving cars [8]. This article surveys methods that tailor DL to perform object detection and scene perception in self-driving cars. In the survey, we also answer the following questions while appreciating the contribution of DL in these areas [9,10]:

1. What are the mutually reinforcing and fundamental operational requirements for fully functional self-driving cars?
2. What landmarks and developments have been achieved in self-driving cars in the last 20 years and what are some promising research directions for the next decade?
3. What is DL and how does DL create artificial perception? With the arrival of DL, is it eventually feasible to attain human level cognition and perception in self-driving cars?
4. Why is DL a promising technique for solving object detection and scene perception in self-driving cars? What are the cutting-edge DL models used for object detection and scene perception in self-driving cars?

* Corresponding author.
E-mail address: alagan@ee.ryerson.ca (A. Anpalagan).

https://doi.org/10.1016/j.array.2021.100057
Received 2 September 2020; Received in revised form 8 December 2020; Accepted 26 January 2021
Available online 23 February 2021

5. With deployment of 5G mobile communication and conceptualization of ultra-fast 6G technology, how is multi-sensor data fusion and 3D point cloud analysis realized and impacted in autonomous vehicles?
6. What are the most recent and successful object detection techniques applied to autonomous vehicles and promising directions for further research?

The rest of the article is structured as follows: Section II provides a brief introduction to the evolution of self-driving cars, and discusses the levels of automation to gradually and progressively achieve fully autonomous vehicles. A list of abbreviations used in the paper is also presented in Table 1 and the Table 2 presents a summary of existing literature related to deep learning and self-driving cars. Section III introduces big data, the role of big data in autonomous vehicles and collecting driving data using LiDAR cameras. Processing driving data captured using various sensors in real-time is a significant challenge, and some promising solutions such as multimodal sensor fusion, road scene analysis in adversarial weather conditions, and polarimetric image analysis for object detection in autonomous driving scenarios are discussed in this section. Section IV introduces deep learning (DL) and the factors that make DL a powerful technique in computer vision. Section IV delves deeper into CNNs, RNNs, DBNs, and other widely used DL techniques in CV. In section V, we investigate the role of deep reinforcement learning (DRL) to enable vision in self-driving cars. We discuss unsupervised learning and explore the possibilities of scene perception in self-driving cars without being explicitly trained on data, leading to artificial general intelligence (AGI) in self-driving cars. Self-driving cars and their ability to achieve human level driving is jointly reviewed from a deep learning perspective with a focus on scene perception and object detection to complement advanced vision. We also provide insights into current applications of DL to achieve AGI that could enable self-driving cars to perceive their environment and take appropriate actions without the need of human intervention. Lastly, we enlist some promising future directions to achieve next generation autonomous vehicles based on the survey and conclude the paper.

2. Evolution of self-driving cars

2.1. Brief history of self-driving vehicles

The concept of self-driving cars has been around for almost 80 years, first reported in 1939 World's Fair in New York by general motor's (GM) Futurama [51]. Contemporary developments in communication networks and wireless connectivity, arrival of accurate and robust sensors that continuously miniaturize in size and cost, coupled with AI have been the cornerstone for autonomous driving systems [52]. Embedded in these self-driving systems are human-machine interface applications, network enabled controls, multiple-sensor data fusion, 3D drive scene analysis, and software-defined signal processing to transport materials, payloads, goods, and people [53]. The AI based self-driving machines must be able to navigate successfully in all situations at all times [54]. The accuracy of autonomous navigation depends significantly on attaining precise localization, unobtrusive data collection, fused data-set generation, and uninterrupted high-level communication with other vehicles and surrounding smart infrastructure [55]. In the longer run, self-driving technology can also expected to be extended to tractor-trailers, cargo trucks, mining trucks, and buses [56]. In the last decade, Carnegie Mellon University and the Defense Advanced Research Projects Agency (DARPA) self-driving cars have contributed to autonomous vehicles advancement [57]. Tesla Motors implemented an autopilot technology to its electric vehicles where the cameras and sensors predicted collisions with up to 76% accuracy leading to collision prevention rate of over 90%. Google, Tesla Motors, General Motors, Waymo, Uber, nuTonomy and other automobile companies envision a future with autonomous vehicles in approximately 15–20 years time [57]. Several infrastructure upgrades such as automated highway system, robotic vehicle cruising management

Table 1
List of abbreviations in alphabetical order.

Acronym	Explanation	Acronym	Explanation
2D	Two Dimensional	IVS	Intelligent Vision Systems
3D	Three Dimensional	IVSS	Intelligent Visual Surveillance System
5G	Fifth Generation Mobile Networks	KITTI	Karlsruhe Institute of Technology Dataset
ADAS	Advanced Driver Assistance Systems	L1 & L2, L0-L5	NA (regularization techniques), levels of automation defined by SAE
AE	Auto Encoder	LiDAR	Light Detection and Ranging
AGI	Artificial General Intelligence	LSTM	Long Short-Term Memory
AI	Artificial Intelligence	MAP	Map Attention Decision
ANN	Artificial Neural Network	MCP	McCulloch & Pitts neural network
AR	Average Recall	MIT-AVT	Massachusetts Institute of Technology-Advanced Vehicle Technology
ATRI	American Transport Research Institute	ML	Machine Learning
AV	Autonomous Vehicles	MLP	Multilayer Perceptron
AWS	Amazon Web Services	NHTSA	National Highway Traffic Safety Administration
BP	Back Propagation	NN	Neural Network
CNN	Convolutional Neural Network	OBU	On-board Unit
CoreML	Core Machine Learning	openCV	Open Source Computer Vision
CPU	Central Processing Unit	PB	Petabytes
CUDA	Compute Unified Device Architecture	PD-DBM	Partially Directed DBM
cuDNN	CUDA Deep Neural Network Library	RADAR	Radio Detection And Ranging
CV	Computer Vision	RBM	Restricted Boltzmann Machines
DAE	Denoising Autoencoder	rCDN	Reverse Content Distribution Network
DBN	Deep Belief Network	R–CNN	Region-CNN
DBM	Deep Boltzmann Machine	ResNet	Residual Network
DIP	Digital Image Processing	RGB	Red, Green, & Blue
DIVS	Deep Intelligent Visual Surveillance	RNN	Recurrent Neural Network
DL	Deep Learning	RoI	Region of Interest
DLib	Deep Library	RPN	Region Proposal Network
DLR	Docklands Light Railway	S3C	Spike & Slab Sparse Coding
DMV	Department of Motor Vehicles	SAE	Society of Automotive Engineers, Stacked Autoencoder
DNN	Deep Neural Networks	SciPy	Scientific Python
DPM	Deformable Parts Model	SSD	Singleshot Multibox Detection
DRL	Deep Reinforcement Learning	STR	Smart Transportation Robots
DSRC	Dedicated Short-Range Communication	SVM	Support Vector Machines
DVS	Deep Vision Systems	SSVM	Structured Support Vector Machines
FCNN	Fully Connected Neural Network	TB	Terabytes
FPS	Frames Per Second	TL	Transfer Learning
GAN	Generative Adversarial Network	TLI	Traffic Light Information
GLAD	GoogLeNet for Autonomous Driving	TPU	Tensor Processing Unit
GM	General Motors	UAV	Unmanned Aerial Vehicle
GPS	Global Positioning System	V2V	Vehicle to Vehicle
GPU	Graphical Processing Unit	V2I	Vehicle to Infrastructure
HD	High Definition	V2X	Vehicle to Everything
		VANET	Vehicular ad hoc Network

(*continued on next page*)

Table 1 (*continued*)

Acronym	Explanation	Acronym	Explanation
HoG/HOG	Histograms of Oriented Gradients		
HRPN	Hyper Region Proposal Network	VaaS	Vehicle-as-a-Service
iOS	iPhone Operating System	VAE	Variational Autoencoder
ICT	Information and Communication Technology	VGG	Visual Geometry Group
IoT	Internet of Things	VOC	Visual Object Classes
IoU	Intersection over Union	VRS	Visual Recognition Systems
IoV	Internet of Vehicles	XOR	Exclusive-or
ITS	Intelligent Transportation Systems	YOLOv3	You Look Only Once version 3

systems, 6G cell-free mobile communication systems with real-time video processing and near-zero latency are parallel research areas that would contribute to realizing full-fledged autonomous vehicles kick-starting a greener future through autonomous electric vehicles [57].

Self-driving cars, also known as autonomous vehicles, driver-less cars, smart transportation robots (STR) or robocars are one of the most speculated scientific invention with a potential to change the world [54]. The recent and broader implications of self-driving cars incorporate integration with novel infrastructure, smart cities, urban planning with provisions for advanced cyber-security, privacy, and insurance [58]. It is worth mentioning that while the self-driving cars have gained intense attention in the last decade, driver-less transportation has been in existence for over a decade [59]. Trains are a prominent example of widespread use of self-driving technology [60]. Some of such train examples include the:

- SkyTrain in Vancouver, Canada [61].
- Docklands Light Railway (DLR) in London, United Kingdom [59].
- Yurikamome in Tokyo, Japan [60].
- London Heathrow airport's ultra-pods [59].

These autonomous rail systems transport thousands of passengers on a daily basis. Authors in Ref. [59] note that the majority of passengers commuting through self-driving trains were not worried about using those trains. However, the aforementioned trains and autonomous pods operate on enclosed tracks, isolated from the public roads, and bypass the need to interact with other vehicles or pedestrians [61]. In contrast, self-driving cars are set to encounter various users, thereby resulting in complex interactions and the possibility of collision [12]. Whether people will be as accepting of self-driving cars as they appear to be of existing autonomous transport is an active area of research [61].

2.2. Advantages of self-driving cars

The advances in wireless networking, software-defined networking, and information and communication technology (ICT) have found applications in intelligent transportation systems (ITS) to reduce collisions, reduce pollution, ameliorate mobility issues, provide newer ways of public transportation, and share resources, materials and space [8]. According to studies, there are 1.3 million deaths every year due to drunk, drugged, distracted and drowsy driving, which can potentially be saved with the help of autonomous AI systems by eliminating some of the human follies [62]. The following advantages motivate the current research in self-driving cars:

- For users, the advantages may be reduced stress, faster commutes, reduced travel times, enhanced user productivity, optimum fuel consumption, reduced carbon emissions. These cars can be programmed to drive defensively, stay clear of blind spots, and follow speed limits [63].

Table. 2
Summary of existing literature related to deep learning and self-driving cars.

Publication	Brief summary
[1]	Self-driving cars: human perception, factors affecting their acceptance, psycho-social impacts
[11]	SAE levels of automation in vehicles
[12]	Transition-time requirements and driver behaviour in automated driving scenarios
[13]	Studies on driver behaviour changes in automated and autonomous driving scenarios
[14]	Studies on lane-change behaviours, traffic flow optimization, vehicular trajectory planning, and bottlenecks in intelligent vehicle assistance systems
[15]	Visual, manual, and cognitive aspects in deciding optimum takeover time
[16–18]	Deep reinforcement learning and deep Q-learning for self-driving cars
[19]	Current state-of-the-art in autonomous driving; social, legal, and technological challenges
[20,21]	Multimodal sensor fusion and LiDAR camera technology coupled with DL for self-driving cars
[22]	V2V, V2X, V2I, cloud platform, fog, and edge computing in self-driving cars
[23]	Challenges in scene understanding, object detection, artificial perception in self-driving cars
[24,25]	Evolution of DL and significant milestones
[26]	DL for object detection, object localization, object categorization
[27]	Computer vision, scene perception, object detection, and localization in self-driving cars
[28]	Advances in AI techniques for object detection, inference in self-driving utilizing geometrical features and event reasoning
[23]	Object detection for self-driving cars using DL: techniques to detect other vehicles, road markings, curb, bicyclist, pedestrian, traffic light, road constructions and signs
[29]	3D object detection in self-driving cars from ground truth labelled data
[30]	Crowdsourcing, social media analytics, surveys using questionnaires discussing public acceptance, attitude, and fears towards self-driving cars
[31]	Tests conducted on self-driving cars to navigate in real-world: industrial, startups, and University initiatives; study of miles driven by different self-driving cars and role of DL for scene perception and object detection
[32]	Crashes involving self-driving cars and their impact on researchers
[33]	Deep neural networks (DNN) for object cetegorization such as pedestrian detection and traffic scene understanding
[6]	Surveys on DL and CV for visual understanding, pedestrian detection and tracking
[34]	A survey on DL for CV and scene understanding
[7]	Surveys on recent advances in 3D semantic image segmentation using NN, DL, and CNN
[35]	A survey on advances in deep vision systems (DVS) and their adoption in self-driving cars
[36]	Learning in DNN: feature extraction and optimization, CNN for image detection
[37]	Unsupervised DL for object detection in self-driving cars
[38]	CNN for driver distraction, pedestrian, and video object detection in self-driving scenarios
[39]	A survey of class imbalance problem in CNN, impact on self-driving cars' perception ability
[40]	CNN as feature learners and feature extractors in self-driving scenarios
[41]	Helmholtz machines, gradient descent, convex optimization, and batch normalization for visual odometry and instance-level segmentation in self-driving cars
[42]	Object position, orientation detection, vehicle counting, and parking occupancy using DNN & CNN
[43]	GPU and DL libraries and application in self-driving
[44]	State-of-the-art CNN architectures
[45]	Visual perception in self-driving cars using DL
[46]	Ethics, AI and self-driving cars as smart transportation robots (STR)
[47]	Advances in unsupervised DL, AlphaGo, AlphaGoZero, and AGI
[48]	Transfer learning: accelerating self-driving research
[49]	Data Science, data collection, and publicly available datasets for autonomous driving applications
[50]	A survey on DL for big data

- For Governments, self-driving cars would assist in traffic enforcement, enhance roadway capacity, reduce road casualties and the number of on-road driving related accidents, and lead to better observance of speed limits [12].
- Self-driving cars are envisioned to eliminate drunk driving issues, eliminate issues related to distracted driving, texting and other cell phone use, less braking and accelerating, and less gridlock on highways [64]. Reduced accidents are expected to be beneficial for children and the elderly, encouraging people to feel comfortable and amiable towards self-driving cars [65].
- Autonomous electric vehicles would introduce a greener mode of transport, leading to less greenhouse and noise pollution, along with increased mobility for the elderly and disabled people [65]. In current driving landscape, cars are parked for a long time. With self-driving cars, parking lots can be converted to parks and other green infrastructure [65].
- Self-driving cars would be equipped to improve scheduling and routing, and provide best routes to improve travel times, while also lowering the travel cost [5].
- Although self-driving cars would reduce or even eliminate car ownership, they would expand shared access, keep transportation personalized, efficient and reliable [65].

2.3. Probable disadvantages and drawbacks of self-driving cars

Cars are one of the most widespread and readily available modes of transportation and while technology has developed safer cars, driving is still a dangerous activity [66]. Self-driving cars formulate a scenario where a few lines of source codes, coupled with AI get to decide the life of a human beings [67]. Some disadvantages of self-driving cars are outlined as follows:

- The foremost catastrophic consequence of self-driving cars would be elimination of jobs in the transportation industry [62].
- Although the role of AI in our society is consistently evolving, an AI system making critical decisions need to respect societal values and conform to social norms to gain acceptance [68]. The acceptance of self-driving technology at philosophical, ethical and technological levels is a fundamental research problem in psychology and cognitive science. It is argued that in case autonomous vehicles and AI systems malfunction, a person would not die or suffer injuries if they themselves were in control of the system [66].
- Driving at intersections without traffic lights, malfunctioning traffic lights, uncontrolled intersections, busy intersections, regions with humans in close proximity are a challenge for self-driving cars [69]. As self-driving cars use global positioning system (GPS) for localization, they are deemed unsuitable to drive in non-mapped areas [70].
- The scope of car's connectivity, the car being online at all times, makes it susceptible to hacking. The safety and convenience offered by self-driving cars might compromise privacy of passengers as their moves will be tracked and logged [22].

2.4. Communication between different entities in self-driving cars

Two well-known collisions mentioned below, involving vehicles operating with a certain degree of autonomous technology emphasize the benefits of vehicle to vehicle (V2V), vehicle to infrastructure (V2I), and vehicle to everything (V2X) communication in self-driving cars:

- A fatal accident involving a semi tractor-trailer that turned in front of a Tesla car operating on its autopilot program in Florida, caused due to sensors failing to detect a turning vehicle [32].
- A fatal Uber crash in Arizona [32].

Investigation and analysis of these accidents indicates that these accidents could have been avoided if the involved vehicles were communicating with each other [32]. The V2V and V2X broadcast a vehicle's current location to nearby traffic, alert traffic to upcoming maneuvers, traffic jams, accidents and road constructions. A crash, three cars ahead, is too far to be detected by sensors but can easily be communicated over longer distances using V2V [71]. The V2I technology consists of sending traffic light information (TLI) to self-driving cars' acceleration and braking systems, which can assist in plannning routes based on the frequency of traffic light changes [72]. The V2V can provide 360-degree road-situation awareness to enhance safety. Although the user of these techniques requires all vehicles to operate on a standard mode of communication such as dedicated short range communication (DSRC) to relay critical information, a formal policy to mandate DSRC in vehicles is still under development [73]. Scientists, researchers, and experts have historically viewed the lack of computational infrastructure as a major bottleneck that prevents achieving reliable V2V, V2I, and V2X communication. Deploying roadside infrastructure partly mitigated the problem by providing uninterrupted wireless coverage while also improving handover and coverage [72]. Authors in Ref. [73] proposed vehicle-as-a-service (VaaS) by leveraging vehicular cloud, vehicular fog and internet of vehicles (IoV) to provide the necessary real-time computing platform that will define self-driving vehicular environment [74,75].

2.5. Levels of automation: semi automated, automated and self-driving cars

Autonomy in self-driving cars is based on progression from human centered autonomy to complete autonomy where all the driving tasks are governed and controlled by the vehicle's AI system, and human interaction is summoned only when necessary [11]. To investigate the capabilities of the present AI systems [76], Fig. 1 briefly outlines the six levels of vehicular automation defined by the society of automotive engineers (SAE). The Fig. 1 also highlights the differences between fully automated (level-5) and partially automated vehicles (levels-3 & 4). The functioning of involved AI systems at an accuracy less than 100% is dangerous to human life [10]. Even if the AI algorithms function exceptionally well on the engineering side, their performance remains disputable from an ethical point of view [77]. As an example, if the driver is intoxicated or incapable to drive, is it safe for AI system to demand human takeover? Offering technology to humans can make them trust the system all the time, rendering them lazy and not able to re-engage themselves [78]. A test on level-2 cars revealed that drivers fell asleep when the vehicle was set on auto-pilot, questioning the ways in which humans interact with AI [46]. The prevailing deep learning architectures for scene perception and object detection in autonomous vehicles are depicted in Fig. 2.

Driving is an activity used by many for recreation, relaxation, and for exercising control [79]. Imparting it into the hands of an AI system tests their ability to perceive environment and interact and build trust with the driver and passengers. Upon encountering situations when they cannot handle control themselves, these systems need to efficiently and voluntarily transfer control to humans [80]. The transfer of control to humans represents a fascinating opportunity for testing the performance of AI [81]. The authors in Ref. [13] studied the effect of vehicle automation on driver behavior and concluded that secondary task engagement by the driver cannot be avoided entirely. Whereas listening to music, lectures and discourses are prevalent in today's driving environment, a significant degree of autonomy could encourage occupants in self-driving cars to engage in activities such as playing board games, getting actively engaged in smartphones, engrossed in conversation without attention on road, thereby totally disconnected with the driving environment [13]. It is feared that when AI asks for help, it is not guaranteed as the human might be distracted or not paying enough attention to the AI system. Recapturing human attention, exercising vigilance by alerting a sleepy human, relaying multi-sensory, localized, and meaningful warning signals to human to assume control continues to be a central AI research problem [46]. Autonomous vehicles incorporate inherent human

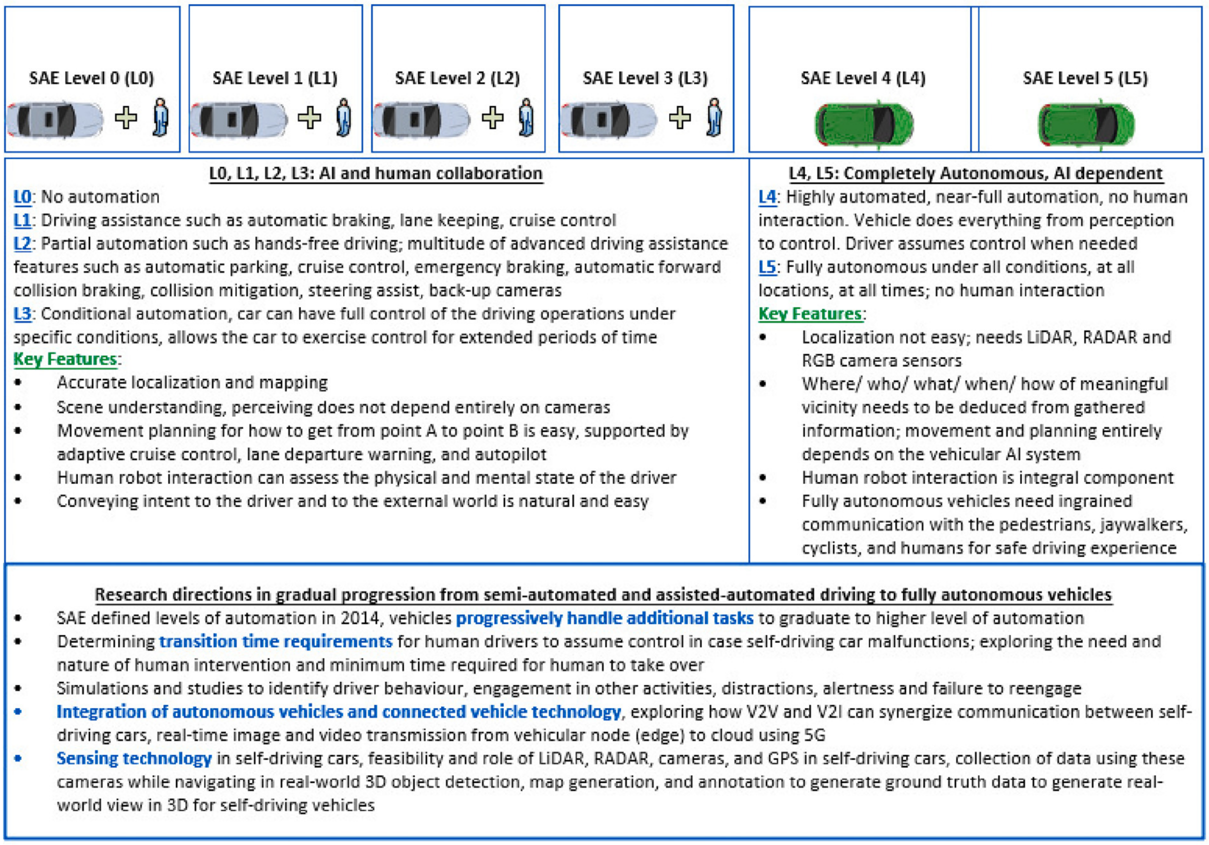

Fig. 1. SAE Levels of automation.

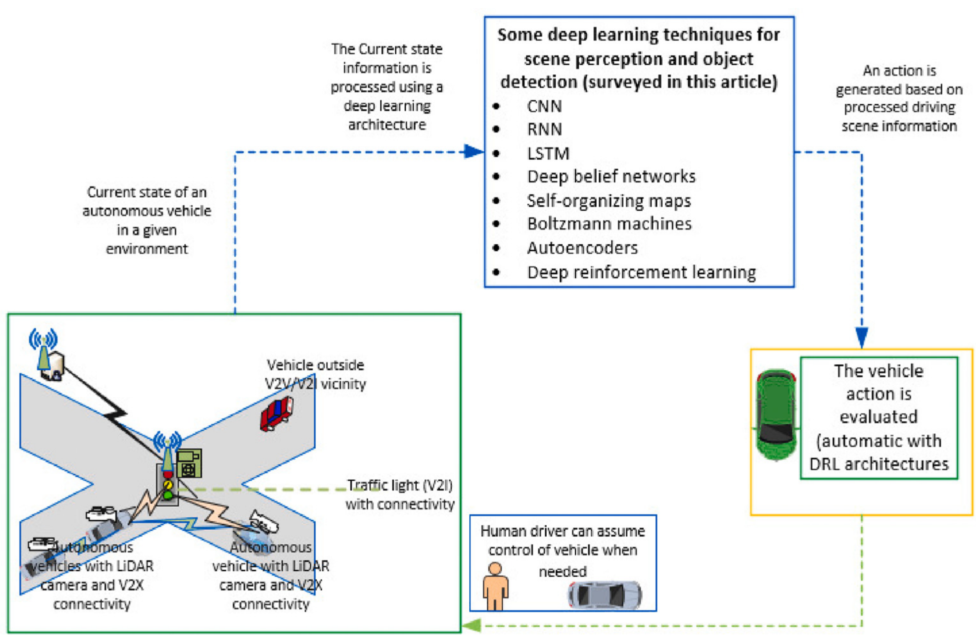

Fig. 2. Deep learning architectures for scene perception and object detection in autonomous vehicles.

element in driving as humans must develop the algorithms and write the code that control them. The Department of Motor Vehicles (DMV) in the USA has issued a preliminary draft that addresses the requirements of steering wheel, brakes, registration, certifications, licensing, and cyber-security and privacy [82]. The DMV mandates that a licensed driver be present inside the self-driving cars at all times, capable of assuming control in case of a technology failure or emergency, vehicle operator having the ability to override the autonomous features [83].

SAE level-3 combines several automated functions and is the first layer where rather than needing to instantly seize control, the driver can be fairly oblivious to road conditions and driving operations. The transition between self-driving and driver-based control can be more relaxed, allowing up to 10 s for the driver to take over [84]. An active avenue for discussion is the number of seconds ideally needed for a driver to take over and assume control. Level-0 to level-2 semi-autonomous vehicles are primarily human-assisted vehicles, where AI systems can make decisions with the driver's permission, such as forward-collision and automatic braking [85]. The National Highway Traffic Safety Administration (NHTSA) defines level-4 automation as fully self driving under certain conditions, where the vehicle can operate on its own with the driver only providing the destination information. A human driver will be necessary to assume control outside the ideal condition zones [11]. SAE level-5 self-driving cars are fully equipped to monitor their surroundings and react safely, without a human in the driver's seat [11]. Waymo, nuTonomy and other leading manufacturers have such vehicles operating on customized driving terrains [31]. Whereas level-3 automated cars follow orders about destination and route, lane-keeping or car-following, a level-5 autonomous car would be able to decide on destination, route, as well as control within the lanes.

Some impediments towards achieving level-4 and level-5 include federal and provincial regulations, insurance issues, and comprehensive street mapping [82]. Waymo tested a level-5 car without a driver, engineer or staff member inside the car. It is recorded as a fully autonomous trip, with transportation from point A to point B without a driver or human assistance [86]. Auto supplier Delphi's Roadrunner drove from San Francisco to New York City traversing 15 states, spanning approximately 3400 miles in 9 days. While the car had a driver to assume control if needed, the car reportedly completed 99% of the trip without needing any driver intervention [87]. The drive successfully faced varying conditions such as construction zones, tunnels, bridges, traffic circles, other aggressive drivers, different weather conditions. The trip successfully tested a number of self-driving components, as well as amassed 3Terabytes (TB) of real-world driving data for research [87].

3. Big data and big-sensed data for self-driving cars

The availability of big data related to self-driving cars facilitates the application of data-driven learning methods to autonomous vehicles. Data emphasizes the role of ML and DL as it is infeasible to craft all possible if-then-else rules that learn all possible situations a self-driving vehicle might encounter in the drive-terrain [89]. Training on data allows self-driving cars to learn by driving, develop efficient inference algorithms, identify patterns in data models and relate complex dependencies in real-time. A fundamental problem in self-driving cars is localization, which is solved using maps. Mapping is expensive and building road maps of the world is an even more expensive task [90]. Usually, mapping is done through dedicated vehicles with many sensors. Often, these sensors have limited coverage, providing a minuscule and narrow view of the driving environment. Drones, unmanned aerial vehicles (UAVs), planes, and satellites have been used to create high definition (HD) maps, capable of capturing information such as parking spots, and sidewalks. Universities and organizations across the globe have voluntarily made these data available to the community at large [91]. In addition, capturing hyper-accurate mapping information from other vehicles on the road results in a massive, but not insurmountable amount of data transmission for real-time analysis [92].

3.1. Role of big data in self-driving cars

Defining a self-driving vehicle problem, formalizing it, collecting sufficient related data on it, and to devise solutions through general purpose AI such as reinforcement and unsupervised learning usually requires raw sensor information and low-level data [93]. Deep learning, however, involves training and testing on labelled data, which can be labelled in case of self-driving cars and annotated by means of ground-truth bounding boxes. By training self-driving cars on these datasets, they are expected to respond to new input data they have never seen before [77]. The self-driving cars need information on their surroundings such as one-way streets, navigation routes, no-entry status, and speech recognition [92]. Localization, i.e. to know where the self-driving car is in the scene, and to decide if and where it has seen this place before is crucial for autonomous systems. While big-data analytics offers affordable self-localization, the scalability depends on aspects such as how the driving environment looks like at every point in time, in every season, at every time of the day, just as humans have the ability to navigate in unknown areas [46,94].

3.2. Collecting big data for self-driving cars

Data collection on public roads is valuable to aid autonomy in self-driving vehicles [95]. The vehicles used for data discovery and collection are equipped with a vast array of sensors and cameras, specifically LiDAR [96]. The data from the sensors are logged on a disk or transmitted to the nearby cloud, which are used to train and test various algorithms for self-driving cars such as vehicle detection, pedestrian detection, or motion estimation [97]. Sensors collect data from external environment, the software analyzes the data, and recreates road conditions in three dimensions. Data collection is a long and costly process, and redundancy is avoided by directly exploiting existing datasets as well as collaborating with data collected by other researchers. To facilitate the analysis of ML or DL controlled driving, these datasets vary in terms of traffic conditions, application focus, detector setup, format, size, tool support, and performance aspects [98]. Driving regularly on the public roads amidst real traffic to evaluate performance and refine technology is crucial to the future of self-driving cars [49]. As DL thrives on datasets, and powerful computational resources enable data processing on GPUs, Massachusetts Institute of Technology's advanced vehicle technology (MIT-AVT) consortium provides a large-scale naturalistic dataset collected over 275,000 miles and having 4.7 billion video frames actively annotated [99]. The dataset is based on 25 types of car pictures to represent various real-world driving scenarios covering Waymo, Cruise, Uber and many prominent self-driving car manufacturers engaged in conducting public trials and test drives in the United States [95,99].

Authors in Ref. [100] present 27 publicly existing vehicular datasets, compare them based on different criteria, and provide suggestions for choosing appropriate dataset pertaining to specific objectives. Another example of collecting big data for self-driving cars is presented in Ref. [101], the KITTI benchmark dataset. The KITTI dataset established that vehicular datasets are significantly large in size. The authors in Ref. [101] also evaluate how well can LiDAR and RGB cameras collect real-time driving data. In self-driving cars, high accuracy is usually defined as 100% accuracy and less than 100% could lead to fatality. The driving data comprises vast range of weather conditions, traffic at different times of the day, roundabouts, intersections, traffic lights, blind turns, curved roads, lane markings, lane change behaviours, and on-street parking [102]. This data spans into petabytes (PB), manually collected, annotated and deduced with accurate ground truth representations. Once the data is ready, the next step is to choose a DL model and architecture [103].

Human visual system is the benchmark for self-driving cars to classify objects, perform edge detection, track lanes and expand visibility range [104]. The AI system in self-driving cars must be able to reveal when it does not see some aspects of a scene. The main sources of raw data in self-driving cars are the automotive sensors. Whereas LiDAR is the most powerful camera, it is expensive and researchers argue that images captured using RGB cameras are sufficient for self-driving applications in certain conditions [105]. Research is underway to manufacture low-cost LiDAR. A brief comparison of images generated by the three cameras is presented below [106].

1. Camera: Cameras are image sensors that operate on RGB values. Cameras capture infrared visual data and offer high resolution information. Cameras can be used as readily available and cheap sensors to capture information that can be learned and inferred to

interpret external scene. Human brain uses similar sensor technology, with eyes acting as sensors that work under illumination, and operate in RGB space to detect and segment lane markings, traffic lights, pedestrians, etc. Cameras work well in visible light but their performance degrades in darkness and extreme weather, and are bad at depth estimation [105].
2. RADAR: Highly reliable and provide higher resolution and accuracy. They are ultrasonic, cheap, and work extremely well in extreme weather. However, under low resolution, they are mostly used as automotive sensors for object detection [106].
3. LiDAR: Although expensive, LiDAR provides extremely accurate depth information, has resolution much higher than RADAR, and provides 360 degress visibility. LiDAR has been a successful source of 3D ground truth data in driving environment [107].

In summary, the need for annotated data increases the demands on camera technology [70]. Cloud based real-time transmission of these annotated images and videos has drawn attention to multi-agent deep reinforcement learning (DRL) techniques to control multiple cars [17]. Although cameras operate in clear, well lit conditions, good visibility requirements over a long range in dark conditions, heavy rain, snow, and fog makes CV and image processing a critical and open research problem [107]. A novel technique suggests cameras, RADAR, and LiDAR be integrated together through sensor fusion, where DL methods can be used to interpret the spatial-visual characteristics to understand, interpret and track the dynamics of the environment [69].

3.3. Multimodal sensor fusion in autonomous vehicles

Sensor modalities allow reconstruction of images for regularization and feature-based reconstruction on data from multiple sources and sensors, where each modality provides significant knowledge and valuable information pertaining to the object of interest [108]. This enables driving under uncertainty through dynamic sampling, characterization and image denoising, deblurring and segmentation [108,109].

3.3.1. Multimodal data fusion

Deep multimodal representation learning incorporates meaningful information from heterogeneous multimodal driving data spanning a multitude of autonomous driving scenarios [110]. Multimodal sensor fusion in self-driving vehicles enhances the representation ability, abstraction and coordinated representation [111] that use variational auto encoders (VAE) and generative adversarial networks (GANs) to resolve multiple data modalities [112]. In autonomous driving scenarios, deep convolutional neural networks (DCNN) have been used to automate feature extraction from vehicular sensor inputs, whereas to capture and model the driving scene temporal dynamics from multimodal LiDAR sensors, recurrent neural networks (RNN) with convolutional and LSTM recurrent units have been used [21,113]. Multimodal RNNs label the driving scenes using gate recurrent cells with computation-intensive GPU clusters to model self-driving vehicle behavior. Multimodal sensor fusion is widely used to ensure robustness of data acquired from different sensors in different driving scenarios and cross-modality generalization, and to approximate missing data through correlation between different available modalities [114]. Visual driving features captured using multimodal RNNs can be enhanced using restricted Boltzmann machines (RBM), a generative graphic model that captures the probability distribution between visible units and hidden units [113,115]. Driving scene object detection captioning is usually followed by stacked autoencoder (SAE) to model time dependency of the intrinsic temporal features, where bounds-based filtering and delta product reduce the redundant dot product calculations in multi-modal computations [111,116]. Modality-specific unsupervised Boltzmann models capture the inherent representation of the driving scene transform the modality-specific representations to semantic features using probabilistic graphical networks [109].

3.3.2. Multimodal driving scene understanding

For collision avoidance in autonomous vehicles, multimodal sensor data fusion is critical for decision making [114]. The KITTI benchmark dataset comprises stereo and laser driving information collected at various driving environments spanning surveillance cameras to discover variables from multi-source data [101]. The 3D images in autonomous driving environment consists of data modalities in camera images due to image depth and semantic labels [20]. Multimodal fusion for detecting objects such as pedestrian, traffic lights, and road infrastructure uses multi spectral identification based on visual odometry to capture deep features of the object direction [117]. Interpreting the multiple object-dynamics from driving image sequences is efficiently done through unsupervised learning using a probabilistic model that estimates positions for objects through a linear state-space model and localizes the positions of the objects through a non-linear process to encompass spatial and temporal dependencies of the multimodal driving data [118]. The driving scene consists of features distributed across spatially and temporally distinct feature-space and involves finding groups of closely related data samples [116]. Multimodal data fusion and multi-instance unsupervised learning in autonomous driving helps to estimate the distributions in the driving space where data labels are learned from specific ensemble learning to improve generalization and distributed representation of objects in driving environment [119]. Autonomous driving surveillance videos contain complex visual events [120], and to generate video descriptions effectively and efficiently without human supervision needs recognizing multiple events in a given video [121].

3.4. Robust decision making in uncertainty and artificial general intelligence in autonomous vehicles

Maneuver planning, driving scene perception, and decision-making for autonomous vehicles is based on some predefined functional requirements that define an initial state of the vehicle, a route map, the obstacles in the region of interest, and a destination [122]. Randomized prior functions help to estimate the uncertainty of decisions in autonomous driving to identify trajectories and the likelihood of a self-driving vehicle following them [123]. Dynamically feasible trajectories for collision avoidance, overtaking, path planning for autonomous vehicles, and collaborative autonomy in decision-making have been tried to achieve in autonomous driving using hidden Markov model and partially observable Markov decision processes [124]. The intelligence and smartness of an autonomous vehicle are strongly related to the use of AI [125]. Whether artificially intelligent autonomous vehicles can safeguard humanity from on-road accidents, drivers frustration and other driving related catastrophes is an active field of research [126]. The ethical aspect of AI technologies, algorithms, and their current and prospective applications to achieve artificial general intelligence (AGI) in autonomous vehicles is a current research topic in terms of driving efficiency, control and planning, obstacle detection and avoidance [127,128]. AGI ultimately aims to integrate sustainability, security, safe transportation and urban infrastructure in autonomous vehicles [129].

3.5. Road scene analysis in adversarial weather conditions

Object detection is critical to autonomous driving assistance systems, especially in road scenes with poor illumination, strong reflections, or other adversarial weather conditions [130], as reflection from rain and ice over roads could lead to significant detection errors [131]. Polarization images characterize light waves and can be used to robustly capture and describe physical properties of the objects in the driving vicinity [132]. Polarimetric imaging modality significantly overcomes the limitations of classical object detection methods in adverse weather conditions, as compared to RGB images for object detection [133]. Semantic foggy scene understanding (SFSU) is an active research area where neural networks trained on RGB images are tested on images augmented by polarimetry for recognition of road-based objects [134]. As intelligent visual traffic surveillance systems depend on cameras and sensors fusion systems [135], adverse weather conditions add constraints to camera functionality impacting computer vision and scene understanding ability of autonomous vehicles [136,137].

3.6. Polarimetric images for autonomous driving scene perception

In autonomous driving scenarios, object detection involves locating and classifying instances of semantic objects [138]. Typical object detection models such as simultaneous localization and mapping (SLAM) have high accuracy on benchmark datasets and are used in applications such as mobile robotics, self-driving cars, unmanned aerial vehicles, drones, and underwater vehicles [139]. LiDAR-SLAM fusion in situations with lack of light, lack of visual features, or haphazard motion of vehicles uses different spatial resolutions to interpolate the missing data with quantifiable uncertainty [140]. Optical technology for optimizing data for driving tasks uses light scan technology, stereo/depth cameras, monocular (RGB) and time-of-flight (TOF) cameras for mapping, obstacle detection, avoidance and localization [141]. Multimodal polarimetric data streams are spatially, geometrically and temporally aligned using low-cost and compact 3D visual sensing devices that offer panoramic background subtraction for visual scene understanding, spectral signal processing to locate objects in a scene, and polarization signatures to detect glare from hazardous road conditions [134,142].

3.6.1. Data labeling to deal with multi-modalities with different signal forms and different resolutions, LiDAR pros and cons, and how to overcome range issue in LiDAR

LiDAR imaging performance for long-range, high spatial resolution in the daytime, real-time autonomous driving environment depends significantly on tolerance to solar background noise, short and long-range narrow and wide fields of view, as well as on the detection range of single-beam LiDAR [20]. LiDAR has a rotating wheel at high speed and multiple stacked single-photon avalanche diode (SPAD) detectors [140]. LiDAR uses a time-of-flight method for detecting laser light in presence of natural light [141]. The output of the SPAD-LiDAR is a monocular image data and a single beam operating at high pulse rate to record a point cloud quickly [131] by comparing a return pulse with one of the transmitted pulses to compute the correct distance [143]. Airborne sensors in LiDARs transmit a single, relatively high-power, pulse in order to maximize the detection range, noise reduction amidst multiple return pulses from each transmitted pulse [144]. Constrained to a specific range of spatial resolutions, it aims to optimize the effectiveness of sensors to solve real-world autonomous driving [145]. Whereas the eariler LiDAR sensors passively searched a scene to detect objects against fixed backgrounds in both time and space, the later solid-state LiDAR sensors enable intelligent information capture through active search by actual acquisition of classification attributes and the nuances of objects in real time [146]. The frame rate of LiDAR imaging systems include object revisit rate to capture instantaneous resolution and detection range as a measure of sensor's ability to intelligently enhance perception [147].

4. Deep learning: A subset of artificial intelligence and machine learning

Deep learning is a multi-layered computational model used for feature extraction and representation learning at various levels of abstraction [24]. DL is a branch of ML that automatically extracts features and patterns from raw data and makes predictions or takes actions based on some reward function [148]. It comprises techniques such as neural networks, hierarchical probabilistic methods, supervised and unsupervised learning models, and deep reinforcement learning (DRL). As stated in Table 4, DL builds on the earliest developments in artificial neural networks (ANN), reported in 1943, that tried to understand how human brain transmits and perceives information [24]. The basic entity of ANN was neuron, which is the fundamental computational unit in all the DL models and architectures [149]. Later in 1958, the first ANN consisting of a connection of six neurons in a single layer, termed as basic perceptron was proposed [24]. The model was critiqued on the basis that the perceptron could not solve the basic ex-or gate (XOR) problem. The

Table 3
Self-driving cars: Industrial and academic initiatives to build prototypes and test in real-world driving environments [31,88].

Industrial/ Academic initiative	Principal entity	Miles travelled	Notes
Industrial	Tesla Autopilot	One billion miles of automated driving, 33% miles driven autonomously	Remarkable adoption rate with drivers actively disengaged from driving. L2 system with autopilot equipped perception control system
Industrial	Waymo	Four million miles driven autonomously by Dec 2017	No driver although operated in restricted conditions. Marks a significant step forward from human-AI collaboration to fully autonomous vehicles
Industrial	Uber	Two million miles driven autonomously	Drivers over-trusting the system is an issue
Industrial	Google	Most autonomously driven miles accumulated on road	Huge investment in DL based self-driving technology research
Industrial	Delphi	San Francisco to New York, 99% in fully automated mode	Detailed report available in [87]
Industrial	Audi A8	To be released	Introduction to commercially available traffic jam assist, first L3 model. Driver obliged to keep hands on vehicle at all times, and to monitor the traffic and intervene immediately when required
Academic	University of Michigan	In-campus autonomous driving shuttle	Feedback available in newspapers and online blogs

Table 4
Significant milestones in artificial neural networks (ANN) leading to the current deep learning era.

Milestone	Principal instigator	Reference
MCP model: The foundation of ANN	McCulloch and Pitts, 1943	[24,153]
Basic (single-layer) perceptron	Rosenblatt, 1958	[24]
Multilayer perceptron	Minsky and Papert, 1969	[24,154]
Backpropagation of errors	Werbos, 1974	[149]
Neocognitron: a hierarchical multilayer ANN, first used for pattern recognition and serves as a foundation for CNN	Kunihiko Fukushima, 1980	[149]
Boltzmann Machines	G. Hinton et al., 1985	[155]
Restricted Boltzmann Machines	Smolensky, 1986	[26]
Recurrent neural networks	Jordan, 1986	[26]
Autoencoders	Rumelhart, Hinton, and Williams, 1986	[156, 157]
LeNet	LeCun, 1990	[150]
Long short-term memory (LSTM)	Hochreiter and Schmidhuber, 1997	[26]
Deep belief networks: Beginning of deep learning	G. Hinton, 2006	[158]
Deep Boltzmann machine	G. Hinton and Salakhutdinov, 2009	[155]
SuperVision (AlexNet): Marks the beginning of CNN revolution and transfer learning	G. Hinton and Krizhevsky, 2012	[151, 157]

concept of multi layer perceptron, proposed in 1969, led to a reinvigorated interest in ANN. A significant way to train ANN was to learn from errors, or back-propagation, that is still used today with convex optimization and gradient descent [25]. The Table 3 outlines some industrial and academic initiatives to build prototypes and test in real-world driving environments.

Later, models such as LeNet, deep belief networks (DBN), restricted Boltzmann machines (RBM) were developed in the 2000s [150]. The interest in ANN and DL has been intermittent due to lack of availability of fast computation [151]. With the advances in cloud computing, and arrival of computing platforms such as Google Cloud, and Amazon Web Services (AWS), computing large datasets has become fast, easy, and accessible [152]. The growth in internet and wireless communication technology has led to developments of computationally intensive tasks such as CV [34]. The combination of DL, Internet connectivity, 5G communication, data science and analytics, and cloud computing engines is critical to the development of self-driving cars [25]. Self-driving vehicles have garnered considerable interest and investment in recent years largely due to breakthroughs in DL, CNN, and deep neural networks (DNN) [95].

Supervised learning requires data to be annotated by human beings [158]. Indirectly, humans are at the core of DL performance. Human data annotation is replaced by augmented annotation methods [159]. Unsupervised learning in autonomous driving interprets the driving environment and surroundings with minimal input from humans [97]. Unlike support vector machines (SVM), the DL can solve complex and non-linear problems without projecting them onto a higher dimension [25]. DL uses a large number of hyper-parameters and layers to solve problems. Table 5 summarizes tools and techniques that enable deploying DL for vision and perception self-driving cars.

To achieve object-detection, cognition and scene perception, self-driving cars are expected to perceive surroundings in a way at least similar to the way human eye processes information [165] leading to cognitive AI systems that can learn, relearn, take actions [166]. To achieve human level driving from CV perspective, self-driving cars need to be able to recognize environment, interpret 3D representation of world, to discern the movement of objects, pedestrians, and other cars, and deal with human emotions [167]. The feasibility of DL is being established as some state-of-the-art results have been achieved by Google cars and Uber cars in maps-based localization, that were trained to drive with little prior knowledge of the roads [86]. These vehicles use DL for path planning, obstacle avoidance, and try to process camera based information to solve complex CV problems [168]. While DL algorithms learn effective perception control from data, LiDAR costs and the expenses involved in manually annotating the maps restrict the application of DL to autonomous driving [49]. The Table 6 summarizes a few DL architectures and deep vision systems that have been applied for object detection in self-driving cars [169]. DL and ML can be classified into three categories; supervised, unsupervised, and reinforcement learning [170]. The mathematical framework to implement these DL techniques involves the following steps:

- Backpropagation: the primary method of learning. Calculates error, computes error function and tries to minimize error function in subsequent forward passes [171].
- Use gradient descent to backpropagate the error function [25,171].
- Subtract a fraction of the gradient from the weight [171].
- Recalculate the weights responsible for making a correct or incorrect decision [171].
- The objective is to minimize the error function, by updating the weights using minibatch or stochastic gradient descent [171].
- More data and very large networks lead to too many parameters, and increased training times [171,172].

4.1. Deep learning architectures for object detection and computer vision in autonomous vehicles

4.1.1. Convolutional neural networks

Convolutional neural networks (CNN) have been extensively applied to image classification and computer vision, and have returned 100% classification rates on datasets such as ImageNet [180]. In CNN architecture, successive layers of neurons learn progressively complex features in a supervised way by back-propagating classification errors, with the last layer representing output image categories [181]. CNNs do not use a distinct feature extraction module or a classification module, i.e. CNNs do not have an unsupervised pre-training and the input representation is implicitly through supervised training but eliminate the need for manual feature description and feature extraction [182]. It extracts features from raw data based on pixel values leading to final object categories [173]. The CNNs have been used in self-driving cars to determine driver behavior and information such as where they are looking, emotional state, cognitive load, body pose estimation, and drowsiness [36]. The CNNs explore the nature of AI and the role of artificially intelligent systems in the society, as full autonomy in self-driving cars involves creating intelligence [183].

The inputs to CNN can be images, video, text, audio depending on different data with one-to-one, one-to-many, or many-to-many relation

Table 5
Summary of tools and techniques that enable deploying deep learning for vision and perception self-driving cars.

Technique	Examples	Scope	Functionality	Performance
Advanced parallel computing [160]	GPU, TPU, CUDA, cuDNN	Vehicle onboard units, smart IoT devices, mobile servers, workstations	Enable fast, real-time training and inference of DL models in mobile vehicular environments	High
Dedicated deep learning libraries [43–152]	TensorFlow, Theano, Caffe, Torch, Keras	Edge nodes, servers	High-level toolboxes that help to build novel CNN and DL architectures	High
Fog, Edge, and Cloud computing [22–161]	CoreML	Mobile devices, workstations	Facilitate edge and fog based DL computation for real-time mission critical applications, uses 5G	Medium
Advanced image processing and CV [34–162]	openCV, Tesseract, DLib, SciPy, TensorFlow	Mobile servers, workstations, GPU	Process images for scene understanding, object detection, and object localization in real-time	Medium
Regularization and prevention of overfitting in DL architectures [41–150]	Momentum, L1 & L2 regularization, dropout, early stopping	Mobile devices, workstations, IoT devices	Avoid overfitting to improve model performance	Medium
Fast optimization algorithms [163]	Nesterov, Adagrad, RMSprop, Adam	Fast training of DL architectures and novel CNNs	Accelerate model optimization process	Medium
Distributed machine learning systems [25]	MLbase, Adam, GeePS	Distributed data centers, cross-server, IoT devices, smart infrastructure	Support DL frameworks in mobile systems, cloud, and data centers	High
Big data captured using LiDAR and cameras [50–164]	LiDAR, sensors	Mounted on vehicles	Capture 2D and 3D images of driving environment	High

Table 6
Summary of Different deep learning architectures and their application for vision and perception in self-driving cars.

Model	Learning scenarios	Example architectures	Object detection problems in autonomous driving	Pros	Cons	Applications in autonomous driving object detection
MLP [24]	Supervised, unsupervised, reinforcement	ANN, AdaNet	Identifying correlations in visual data	Easy structure, design and implementation	Slow convergence, less patterns	Can be used as a component of DL to model multi-attribute data
RBM [173]	Unsupervised	DBN	Entropy maximization, robust scene representation	Data points generation	Complex training, lack of self-awareness	Learning from unlabeled data, augmented data, scene/object prediction
AE [174,175]	Unsupervised	DAE, SAE, VAE	Learning representations and features from sparse data, in locations with less mapping	Effective unsupervised learning	Time consuming to train, real-time communication not guaranteed	Weight initialization, data dimensionality reduction
CNN [38–176]	Supervised, unsupervised, reinforcement	LeNet, AlexNet, ResNet, VGG, GoogLeNet, DenseNet	Spatial data modelling	Invariant to image transformations	Finding optimal hyper-parameters needs deep structures	Spatial data analysis, object detection, object localization
RNN [177]	Supervised, unsupervised, reinforcement	LSTM	Sequential, time-series data modelling	Capture temporal dependencies	Gradient vanishing problem	Traffic flow analysis, spatio-temporal data modelling
GAN [178]	Unsupervised	GAN	Data generation in varying driving scenarios	Produces real-world data artefacts	Difficult convergence	Virtual driving scene generation, simulation, assist supervised learning
DRL [179]	Reinforcement	Deep Q-learning	Analyze high dimensional data	Models high dimensional driving environment	Slow convergence	Driving scene control, management and AI decision-making

between input data and output classes. Depth in CNNs is provided by the number of layers, and is analogous to features, taking all the features generated by filters of different sizes and using backpropagation to arrive at the best features [184]. Novel CNN architectures are designed using transfer learning where a series of predefined convolutions are followed by series of fully connected layers, without having to train CNNs from scratch [185]. In a CNN, a driving image is convolved with activating functions to obtain feature maps, which can be further scaled down to identify patterns in an image or signal [25]. CNNs are robust to translational invariance and rotational invariance, as convolution multiplies the same weight everywhere on the given input. Each layer in the CNN finds successively complex features where the first layer finds a small, simple feature anywhere on the image, the second layer finds more complex features and so on [6]. At the last layer, these feature maps are processed using fully connected neural networks (FCNN). In addition to reducing the driver's responsibilities and assist them in critical tasks, the end result envisioned is to eliminate the active need for driver engagement, which extends much beyond currently available semi-autonomous models with ADAS [186].

4.1.2. Recurrent neural networks

Recurrent neural networks (RNN) recognize sequences and patterns in structures consisting recurrent computations to sequentially process the input data [177]. Long short-term memory (LSTM) is an RNN based method which uses feedback connections for sequences and patterns recognition using input, output, and forget gates [120]. LSTM remembers the output computed from the previous time step, and provides output based on the current input [121]. The connection between units forms a directed cycle and the RNN input and output are related as the edges of RNN feed output from one timestep into the next time step [187]. RNNs have been applied for robust and accurate visual tracking in autonomous vehicles in constrained scenarios [188]. Temporal correlation of RNNs predicts the object at next time frame based on region of interest (ROI) and uses that image as the input of the next frame, resembling a prediction model for object tracking [189].

4.1.3. Deep belief networks

A deep belief network (DBN) is a hybrid multi-layered, generative graphical model used for learning robust features from high dimensional data and consists multiple layers of stochastic, latent (hidden) variables connected between the neurons of different layers, but not between the units of each layer [179]. The undirected Restricted Boltzmann Machines (RBM) where each layer trains separately to produce an expected input are the building blocks of DBN architecture [158]. Each RBM layer communicates with the previous and next layers for accuracy and computational efficiency [155]. DBN handles non-convex objective functions and local minima through multiple layers of latent variables where a hidden layer acts as the visible input layer for the adjacent layer [190].

4.1.4. Stacked autoencoders

Autoencoders are a class of unsupervised feature extractors that find a generalized transformation of the input and assist a classifier in a supervised task [191]. Stacked autoencoders (SAE) are used in autonomous vehicles vision systems to visualise high-dimensional data to find clusters and to create similarity metrics between samples [192]. SAE are sued to reduce the dimensions of the input image data captured using LiDAR sensorsin self-driving vehicles. Dimension reduction avoids learning the identity function without any explicit changes to the driving accuracy and gives smaller reconstruction errors.SAE restrict the driving scene output to be sparse, imposing a sparsity constraint [25]. SAE add random noise to the input, requiring a reconstruction of the original input. This forces the driving vision system to learn the structure of the input distribution to undo the effects of the added noise [23], which makes the system more robust to small changes of the input [25]. The learned features of the autoencoder are tolerant to the changes in the input space [26]. DBN and SAE assist in building a non-linear, distributed representation of the input, where DBN captures the representation in stochastic distribution form, while the SAE learns a direct mapping of the input to another space [25]. Unlike CNNs, other methods such as RNN, SAE, DBN, LSTM, and deep reinforcement learning (DRL) models do not exploit prior knowledge about the structure of images. If the pixels of all images in the driving environment were randomly permuted, it would not affect the performance of the models in scene analysis, environment perception and object detection [193].

4.2. Singleshot multibox detection

A previous technique known as sliding window utilizes CNNs for better image detection. In realistic self-driving scenarios, some images might have zero objects, and some might have up to 50 objects. The output of a CNN is still going to be a fixed set of numbers that the processing infrastructure or the vehicle AI system needs to interpret [194]. Single shot multibox detection (SSD) is largely viewed as a milestone in digital image processing (DIP) and CV research to enhance real-time performance requirements [195]. Self-driving cars have to recognize objects as soon as they see them, which is a CV problem as well as a security requirement in self-driving cars [89]. Singleshot multibox detection improves both speed and accuracy by looking for the presence of an object, and its location. If an image has multiple class instances such as people, cars, trucks, bicycles, people, traffic lights, traffic signs, landmarks, lane markers in a single image, in such settings, image classification has limited applications [196]. This leads to research into object localization, which tells a vehicle whether an object is present in an image and its location. This is done using bounding boxes or rectangles, and depending on applications these can be ellipses, facial key points known as landmarks, or fingerprints and retina structures. A requirement for images is to have translational invariance, rotational invariance, color invariance, and size invariance [197]. Comparable techniques are you only look once (YOLO) and region-CNN (R–CNN) and have been outperformed by SSD as it addresses the following two problems in an efficient manner [198].

4.2.1. Problem of scale

According to the problem of scale, a car is much larger than a bicycle in all three dimensions, height, width and length. Singleshot multibox detection makes an assumption that our window size is pre-known. Utilizing the general principle of CNNs that an image shrinks at each layer, and the feature becomes larger, SSD attaches mini-neural networks to intermediate layers of a pre-trained network, known as region proposals [199]. This technique inherent in SSD eliminates the need to define where to search for an object in an image and tackles the problem of scale. Cross entropy does not fit as a suitable loss function due to increased dimensions and increased classes [197]. A requirement of bounding box is that it should be normalized coordinate between 0 and 1. Then a code maps a particular windows' bounding box to the full image's box in de-normalized pixel coordinates. A drawback of this sliding window technique is that it needs $O(N^2)$ computations, where N implies the number of steps per image dimension [195]. In SSD, we only have to pass an image through the CNN once rather than $O(N^2)$ times. This is an improvement over R–CNN.

4.2.2. Problem of shape

Cars are wide and people are tall, therefore an efficient window size is based on the aspect ratio. In an autonomous driving scenario, multiple objects can occur in the same window, one occluding the other [48]. In general, humans are endowed with an inherent ability to effortlessly figure the difference through the context of surrounding image. The current CV research aims to impart such human level perception to computers, enhancing their pattern recognition abilities. Authors in Ref. [200] tackled the problem of image localization as the opposite of regression, where instead of trying to find the object, they placed the box and let the CNN decide if an object is present there or not [200]. SSD shows considerable improvement over YOLO that operates over a single scale feature map [198]. Singleshot multibox detection searches everywhere in an image through parallel computing and convolution. Instead of considering the whole CNN as a feature extractor, each subpart of the CNN is a feature extractor that takes output from multiple parts of CNN [182]. Mini-neural networks attached at various convolution or pooling layers of CNN lead to their own object detection at each layer. Thus, SSD efficiently applies to any neural network and makes use of transfer learning [201].

4.3. Object detection approaches in self-driving vehicles: pros and cons

Object detection is a mechanism to identify class instances to which an object belongs [202]. To gain a complete 3D view of the environment, self-driving cars need to classify different objects present in an image, and precise locations of these objects [203]. Object detection for semantic scene understanding is broadly divided into following three categories:

1. Region proposal/region selection: The most popopular technique for region selection in pre-DL era was to scan the whole image using a multi-scale sliding window. However, for self-driving cars, sliding window technique is computationally intensive and fails to meet real-time need to find all positions of an object exhaustively [201].
2. Feature extraction: Techniques such as Haar-transform, Haar-like features and histograms of oriented gradients (HoG) were widely used for feature extraction. However, in self-driving scenarios, these techniques do not provide robustness to fluctuating environmental conditions [204].
3. Classification: Once the objects are perceived and localized, classification is accomplished using ML algorithms such as MLP and SVM. Deformable parts model (DPM) gained widespread acceptance for object classification and has been proposed in self-driving scenarios [205].

Histogram of oriented gradients (HoG) is a feature detection technique that found application in pedestrian detection. Histogram of oriented gradients calculates gradients for the whole image at multiple scales and employs linear classifier for each scale at every position. The HoG model for self-driving cars in heavy traffic situations with multiple vehicles needs near perfect accuracy to avoid crashes. The model proved to be too slow for real-life situations involving pedestrians and other vehicles [206]. As HoGs represent edges, edges symbolize convolutions and histograms represent pooling, CNNs can be seen analogous to fast HoG. Later, a modification known as DPM replaced the HoG linear classifier with a linear template for the object [207]. Another variation used latent SVM to learn and evaluate, leading to a powerful classifier that allowed more deformability in the model. Classification was still computationally demanding, and needed to test many positions and scales that became infeasible for cluttered images [205]. An improvement was region proposals, that found regions likely to contain objects. The technique was not very accurate but was fast to run. An advancement to this approach was selective search that merged adjacent pixels having similar colour and texture. Merged regions at multiple scales led to reduced search space [208].

As the image scenario approaches more and more real-world conditions, certain constraints and added complexities arise in form of:

- Partial view [209] and side/angled view [209].
- Occluded object [210].
- Multiple objects in the vicinity [194].
- Similar objects at varying distance, hence appearing to be of distinct sizes when in reality they belong to the same class instance [23].
- Illumination variation based on time-of-day [18].
- Road conditions such as slippery road and unclear road markings [28].
- Weather conditions such as snow, fog, and rain [211].
- Faded lane markings
- Changing traffic lights [98].

Table 7
DL based object detection and scene perception for meaningful AI actions in self-driving cars.

Self-driving car behaviour type	Manoeuvre and intervention type	Scenarios emphasizing the underlying role of DL and generating sufficient input for the vehicle AI system	Achievable action
Operational	Initiate response	Vehicle sensors detect an obstacle, DL processes the image in real-time and the AI system initiates deceleration, braking, or steering. The AI system also issues alerts to the driver/human inside the vehicle to intervene to lessen the impact of a probably unavoidable crash [224].	Adaptive cruise control in modern vehicles works in a similar way, defined by SAE level-2. Fully autonomous, level-4 and level-5 vehicles would need increased real-time information processing and transmitting capabilities.
Operational	Supplement response	Sensors detect a threat, DL processes the image in real-time and infer a response from other vehicles using V2V and V2I communication. The nearby vehicles adjust trajectories and speed to accomodate the threat [61].	Elicit nearby self-driving vehicle's collaboration in real-time to an inevitable crash that exceeds a vehicle's ability to respond alone
Operational/ Tactical	Guide response	Sensors detect a threat, DL processes the information and guides the vehicle AI system to initiate appropriate response [30–183].	DL detects vechicle fllowing too close and activates forward-collision warning system. The AI system initiates braking.
Precautional	Direct attention	Sensors detect a crash or roadside hazard, DL processes the image and alerts the AI system and the vehicles behind, but does not initiate a response [225].	An alarm that indicates the presence of hazardous weather conditions, or a roadside hazard.
Tactical/ Strategic	Enhance awareness	DL processes the images and estimates deficiencies in roadside conditions, weather conditions, and driving performance of other autonomous vehicles. The AI system determines the likelihood of a lane drift or a crash [226].	The system constantly sends out alerts to nearby V2I cloud, and nearby vehicles using V2V to engage in more appropriate driving behaviour.
Tactical	Deliver information	GPS and sensor data to constantly update location information to ensure vehicle is enroute to destination. A new scene in image could alert the vehicle to having taken wrong path [227–229].	The guidance and regulatory route information based on DL scene understanding is analogous to human drivers following signs and directions. This also eliminates the human propensity to occasionally fail to obey traffic controls, but needs accurate scene perception and object localization.
Tactical/ Strategic	Enhance feedback and tune expectations	Sensors capture images, DL processes them and AI system repeatedly exposes drivers/ humans to alerts based on specific hazards. This enhances driver awareness and tunes their expectations to situations in which self-driving vehicles can demand attention [95].	Such a system motivates a human to stay alert to roadside conditions, be ready to intervene, and be cautious of specific situations when self-driving car can demand takeover. Marks a huge step towards level-5 self-driving vehicles.
Tactical/ Strategic	Calibrate capacity and demand	Sensors capture information, DL processes them and detects objects and repeated exposure of AI system to those images makes it learn the roadway occupancy pattern at specific times [230].	Such systems will be crucial to have human passengers adapt their behaviour to self-driving cars' intervention-demand cycle, based on time-of-the-day.
Strategic/ Tactical	Postdrive feedback	The DL system provides the images of same route over and over again to the AI system. The AI system compares it with decisions made during previous drives, identifies suboptimal or erroneous decisions and tries to imprive behaviour, either on its own or with human collaboration [43].	The DL system captures images, AI system generates a report of risky situations encountered and decisions taken. Human experts designing the system can assiciate consequences and retrain the system accordingly.

- Invariancy requirements such as translation-invariance, rotational invariance, colour and size invariance

Although various object detection architectures such as R–CNN, fast R–CNN, faster R–CNN, YOLO, and SSD have produced remarkable accuracy and very low error rates (less than 5%) on ImageNet and Pascal VOC datasets, the speed of these architectures to yield low error in real-time self-driving scenarios is still a concern. Authors in Ref. [212] present a thorough evaluation of different region proposal mechanisms analyzing the pros and cons of each technique. Combining region proposals and CNNs became known as R–CNN. R–CNN showed good results for AlexNet and visual geometry group-16 (VGG-16). However, it was slow as it needed to run full forward pass of CNN for each region proposal [204]. For a large number of regions, the need to evaluate CNN for each region leads to CNN weights update in response to the parts of the network pointing towards similar regions. Box annotations to train mask R–CNN were used to detect vehicles in aerial imagery. Vehicle detection methods based on sliding-window search led to powerful feature representations than faster R–CNN [212], which used region proposal networks (RPN) and had poor performance for accurately locating small-sized vehicles due to their inability to distinguish vehicles from complex backgrounds. Although CNN operation interweaved with computation of region proposals for an image is faster, yet it is not a robust mechanism to train the whole system end-to-end at once in self-driving scenarios. A solution to the weights update problem was known as fast R–CNN [213]. Fast R–CNN introduces region of interest (RoI) pooling by passing an input image through convolution and pooling layer, followed by the FCNN layer to compute low-level features using backpropagation, maxpooling, and instance segmentation. Fast R–CNN tackles object detection as a variable size regression problem, which proved to be slow for real-time object detection in self-driving cars [199]. In self-driving cars, sliding window technique for extracting regions followed by CNN leads to slow R–CNN at test-time. Faster R–CNN has found application in traffic monitoring from unmanned aerial vehicle (UAV) imagery captured over signalized intersections for vehicle detection. It is robust to illumination changes and detection insensitive to the number of cars in a frame. Faster R–CNN has also shown high accuracy for parking lot car detection, and its application to detect other transportation modes such as buses, trucks, motorcycles, and bicycles is still research domain. Processing real-time video streams to find the objects of interest is computationally intensive for self-driving scenarios [214].

Object detection with faster R–CNN was a 101 layer ResNEt with faster R–CNN as the FCNN [215]. You look only once tries to pose localization as regression and divides an image into an S x S grid. Within each grid cell, it predicts B bounding boxes with four coordinates plus the confidence function [216]. Although YOLO was faster than faster R–CNN, it still showed performance degradation on benchmark datasets. Different variants of faster R–CNN on VGG, ResNEt, and GoogLeNet (inception) were slower than YOLO but outperformed YOLO performance [217]. Whereas YOLO is orders of magnitude faster than other object detection algorithms for intelligent visual surveillance system (IVSS), it struggles to detect small objects within the image. In self-driving cars, YOLO could detect nearby cars with high confidence, but faraway cars that appear small were detected and localized with lesser confidence, attributed to the spatial constraints of the algorithm [49]. Being able to detect and localize objects in meaningful vicinity using DL will serve the following AI related tasks in self-driving vehicles, stated in Table 7.

4.3.1. 3D outdoor object detection and point cloud analysis for autonomous vehicles

Automated processing of uneven, unstructured, noisy, and massive 3D point clouds in autonomous driving environment is a challenging task [218]. Deep learning models assist LiDAR point clouds segmentation, detection, classification, and accurate distance measurement to detect objects in dynamic urban environments [219]. Processing large volumes of information in real-time, detecting objects in autonomous driving trajectory has crucial applications for driving safety and security [220]. YOLOv3 method is among the most widely used deep learning-based object detection methods [221]. It uses k-means clustering to estimate the initial width and height of the predicted bounding boxes [222]. The estimated width and height are sensitive to the initial cluster centers [223].

4.4. Deep learning libraries

Modern deep learning libraries such as Theano, PyTorch, TensorFlow, and Keras make designing neural networks easier [173]. TensorFlow allows for plug-and-play script [157]. As training neural networks takes long time, ranging from days to weeks and months, these DL libraries make use of GPUs, that speed up matrix multiplications and other mathematical operations by orders of magnitude. Table 8 presents a brief comparison of prevalent DL libraries used in neural network designs, transfer learning (TL), and object detection techniques. Utilizing these DL libraries in conjunction with GPU and vehicular cloud for real-time CV in self-driving cars and carrying out the task in Table 7 is an active area of research.

Table 8
Comparison of deep learning platforms.

Platform	Workstation hardware supported	Mobile hardware supported	Speed	Code size	Mobile compatibility	Open sourced
TensorFlow [26–173]	CPU/GPU/TPU	CPU	Slow	Medium	Medium	Yes
Caffe [26–173]	CPU/GPU/TPU	CPU	Slow	Large	Medium	Yes
Keras [26–173]	CPU/GPU/TPU	CPU	Medium	Small	Medium	Yes
ncnn [26–173]	CPU/GPU/TPUs	CPU	Medium	Small	Good	Yes
CoreML [26–173]	CPU/GPU/TPU	CPU/GPU	Fast	Small	Only iOS 11+ supported	No
DeepSense [26–173]	CPU/GPU/TPU	CPU	Medium	Not known	Medium	No

5. Deep reinforcement learning for computer vision in self-driving vehicles

With images captured using powerful LiDAR, RADAR, and RGB cameras, DL has been applied extensively for object detection and scene perception in self-driving cars [231], which is broadly concerned with following tasks:

- Self localization [23].
- Understanding driving environment and reconstructing the environment [193].
- Pixel labeling, differentiating between individual objects, 3D detection and tracking [179].
- Generalize to unseen driving scenarios and predict trajectories where car has never driven before [232].
- Semantic scene understanding, and understanding high level semantics and scene understanding in traffic patterns [34].

5.1. Is training on data always necessary: the case of AlphaGoZero

Application of AI in a variety of domains such as speech recognition, image classification, and natural language processing is based on leveraging human expertise and enormous amounts of data [94]. A drawback of supervised learning is that AI systems are trained to achieve best performance based on a reward function, and defining a good reward function is difficult. The AI system risks discovering local pockets of high reward ignoring the implied bigger picture goal. In addition, specifying a reward function for self-driving cars raises ethical questions [96]. How the actual system behaves is important for the acceptance of autonomous vehicles. However, training self-driving cars with no data and expecting them to learn driving scenarios on their own is an open avenue for research. However, human knowledge is usually too expensive, unreliable or unavailable. Consequently, AI research aims to bypass the need for human expertise and create algorithms that achieve superhuman performance with no human input, even in most challenging situations [94]. A significant step towards achieving this goal was observed in 2017, when a board game known as AlphaGoZero defeated its previous version, AlphaGo. Whereas prior versions of the game were AI systems trained on human data, AlphaGoZero was not trained on training data or human input. AlphaGoZero beat AlphaGo and many of its variants by playing itself and generating moves that surprised human experts [96]. However, authors in Ref. [233] argue that detecting patterns and outliers in unlabeled data in self-driving domain is an ill-posed problem until the spatio-temporal relation between objects is adequately represented and the inconsistencies filtered out.

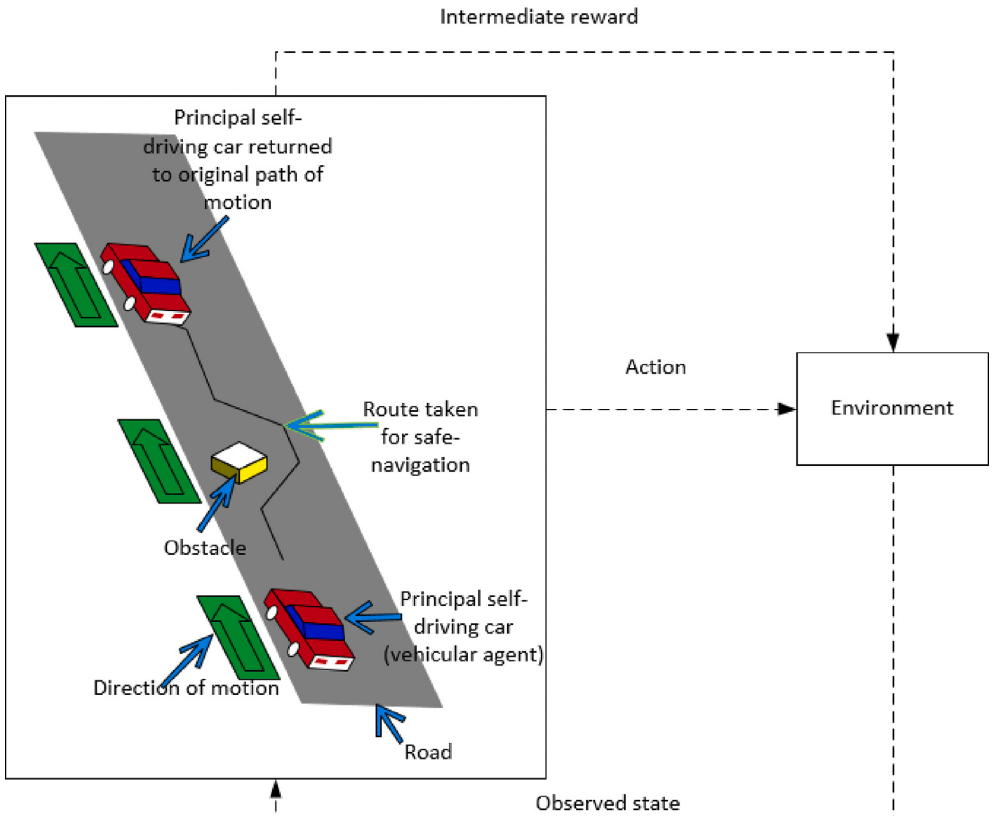

Fig. 3. Deep reinforcement learning architectures for scene perception and object detection in autonomous vehicles.

5.2. Deep reinforcement learning for computer vision and object detection in autonomous vehicles

To classify an image, to assign one or more labels from a given set to describe the contents of the image, image segmentation, object detection, and object recognition are some of the challenges in autonomous vehicle scene perception [234]. To design vision systems independent of variations in the input image, such as translation, rotation, scaling, stretching and other distortions of the image is also a significant challenge [173]. For real-time image segmentation in self-driving cars, less communication across V2V and V2I servers is desired. This led to designing smaller DNN that require low bandwidth to communicate from the cloud to a self-driving car's limited memory [176]. A direct perception approach for autonomous driving focused on feature extraction from road markings and other vehicles in the scene was known as GoogLenet for Autonomous Driving (GLAD). This architecture eliminated the need for unrealistic assumptions about a self-driving cars' surroundings using five avoidance parameters as compared to 14 used by earlier models, both on empty roads and while driving in the presence of other surrounding vehicles [185]. A DRL approach for scene perception and object detection is shown in Fig. 3.

Strong nonlinear mapping capability, and increasing number of layers and neurons in a given layer enhanced their representation ability [235]. However, these models suffered from overfitting as the vehicle datasets lacked diversity in captured scenes [159]. To prevent overfitting, dropout was used as a simple and efficient way, where some nodes in the network were randomly removed, along the incoming and outgoing edges. A stacked denoising autoencoder combined with dropout, achieved better performance than singular dropout [174]. Existing CNN implementations were improved using recurrent neural networks (RNNs) [236]. Spike-and-slab models for object recognition in self-driving cars was proposed using spike-and-slab sparse coding (S3C), and a variant of S3C, partially directed deep Boltzmann machine (PD-DBM) [237]. PD-DBM generated better results than deep Boltzmann machine (DBM) and deep belief networks (DBN) and was trained successfully without greedy layerwise training. Simultaneously training all layers of the DBM led to single generative models and RNNs that maximized approximation likelihood and improved accuracy based on shared network parameters [200]. In self-driving scenarios, inference approximation and classification in case of missing inputs is inherently dangerous. To avoid approximation, a structured support vector machine (SSVM) in the last two layers of a CNN trained a DPM by backpropagating the error of the DPM to the CNN [205]. Authors in Ref. [238] proposed a deep vision technique by formulating the DPM as a CNN. The technique introduced detection proposals to search objects, avoiding exhaustive sliding window search across images. Each step of the DPM was mapped to an equivalent CNN layer, replacing the DPM image features with a CNN feature extractor. The model was named DeepPyramid DPM [238].

Model-based RL in autonomous driving uses driving experience to construct a model of the driving state transitions and outcomes in a driving environment [239]. Based on a reward function, actions are chosen by the vehicular agent to plan next set of actions in driving environment [231]. Model-free RL uses driving experience to learn state-action values and based on an optimal driving policy, achieve an optimal behavior [240]. Deep reinforcement learning (DRL) combines deep learning and reinforcement learning to solve sequential decision-making problems such as driving trajectory planning [16]. A few actor-critic and proximal policy optimization (PPO) methods, and their applications in autonomous vehicle object detection are briefly described below [17]:

- In policy optimization methods, a vehicular agent learns a stochastic policy function that maps a state to an action, that outputs a probability distribution over actions in a given state [239]. This process is called Partially Observable Markov Decision Process (POMDP) [231].

- In policy gradient methods, a driving policy based on a driving parameter outputs a probability distribution of probable actions [239], and aim to find the best parameter that improves the policy [239]. Gradient descent optimization in policy gradient methods uses a random search to provide a function approximation for arbitrary policies [25].
- In asynchronous advantage actor-aritic (A3C) methods, several agents are trained in an environment the experience of each agent is independent of the experience of the others [240]. The overall experience becomes more diverse and combines the benefits of both approaches from policy-iteration method and value-iteration method [240].
- Trust region policy optimization (TRPO) is an on-policy algorithm used or environments with both discrete or continuous action spaces [16]. In autonomous vehicles, TRPO updates policies by taking the largest step possible to improve performance based on the driving environment constraints [17].
- Proximal policy optimization (PPO) is an on-policy algorithm applicable to both discrete as well as continuous action spaces [16]. In autonomous drivnig scenarios, PPO aims to increase policy improvement by limiting the policy update at each training step [16].
- Q-learning is a value-iteration method that learns the action-value function to determine how good will be a particular driving action when the autonomous vehicle agent is in a given state [17].
- Distributional reinforcement learning, for each state-action pair, a distribution of value-functions is learned to improve the overall driving policy [17].

5.3. Critical evaluation of recent implementations and tests conducted on self-driving cars

This section summarizes the applications of deep machine learning to autonomous vehicles and the methods are briefly compared. The above mentioned object detection mechanisms provide repeatability, ground truth annotation recall, and impact DPM, R–CNN, and Fast R–CNN detection performance [241]. Smaller bounding boxes on high-resolution layers with a smaller stride and larger bounding boxes on low-resolution layers with a larger stride inspired the conv/deconv structure [238]. While this fully leveraged the low-level local details and high-level regional semantics, it was slow to identify the objects in an image [213]. Instead of using images, video frames have also been used in CV. A reverse content distribution network (rCDN) is a collection of related video streams emerging from multiple content sources, leading to highly dynamic video data [117]. The rCDN uses fog computing to serve connected and autonomous vehicles, where video content from vehicles cameras and street cameras are distributed across rCDN nodes [117]. This allows communication from downstream devices (cameras) to upstream devices (e.g., IoT gateway, road side unit, vehicular edge node) to stream the content directly to the cloud or proccess data in a distributed fashion [161]. To optimize the overall training loss, map attention decision (MAD) bridged the gap between DL and conventional object detection frameworks. MAD aggressively searches for neuron activations to gather dense local features from discriminatively trained CNNs [177]. However, this model is highly sensitive to initialization, and even if a single neuron is initialized poorly, it might not fire for the entire training dataset [242].

In autonomous vehicles' vision systems, deep learning and deep reinforcement learning (DRL) aim to learn data-representation through consecutive non-linear transformations [243]. Representation refers to a vector of values comprising compressed and encoded statistical information about the driving-environment input space [244]. For object detection, obstacle avoidance, and scene perception, the images containing similar objects are usually represented by similar values in the vector, even in different backgrounds, enhancing the vehicle's ability to capture meaningful information in various types of input scenarios [245]. The representation learning methods need an exponential number of parameters to approximate the function that maps the input to the output [246]. The mean square expected error for a given number of driving images rises exponentially with the number of the dimensions of the input space [247]. Hyper-parameter tuning in autonomous vehicle computer vision is described by functions that have underlying patterns with high variance that form linear combination of the most dominant features [248]. However, due to real-time communication requirements and computational resource constraints, the number of parameters based on prior knowledge about the driving task reduce the dimensions of the input [249]. Moreover, this approach is object specific, and a system designed to recognize traffic lights might not be able to recognize pedestrians [250].

In autonomous driving scenarios, such as one depicted in Fig. 3, the relation between the input and the output is highly non-linear with underlying patterns [247]. Kernel-based methods such as SVM need a large number of parameters to reduce training error and fail to generalize to previously unseen inputs [246]. The representations of the driving environment input learned by DL architectures are Gaussian distributed and have a mutually exclusive feature representation [17,248]. The DL architectures need $O(N)$ parameters to distinguish between $O(N)$ features and using multiple parameters to represent an input space can distinguish up to $O(2^N)$ features, as the number of features and object combinations is exponential to the number of parameters [16,17,250] in highly non-linear spaces, such as complex drive terrains.

6. Conclusion and future directions

In this article, we reviewed and studied the recent trends and developments in deep learning for computer vision, specifically vision, object detection, and scene perception for self-driving cars. The analysis of prevailing deep learning architectures, frameworks and models revealed that CNN and a combination of RNN and CNN is currently the most applied technique for object detection due to remarkable ability of CNNs to function as feature extractors. The CNNs can learn subtle patterns in an image, and are robust to translational and rotational variations. We outlined the ongoing initiatives taken by researchers to test self-driving cars and emphasized the role of DL in real-time object detection. With GPU and cloud based fast computation, DL could process captured information in real-time and communicate it to nearby cloud and other vehicles in the meaningful vicinity. The study also revealed that in order to improve performance metrics such as accuracy, precision, recall, and F1 scores, transfer learning is used to enhance accuracy of object detection. In this survey, we focused on the recent advancements in CNNs that are principally used for images. In self-driving cars, CNN dependent strategies still need to be fine-tuned so as to achieve the precision level of human eye. The findings reported that although DL is a key catalyst to realize object detection and scene understanding in self-driving cars, there is a huge scope for additional advancements. It is yet to be investigated that when and under what conditions CNNs cease to perform well and can pose a threat to human life in self-driving scenarios.

The artificial driving intelligence is still incapable to annotate and categorize driving environment on its own, without need for human assistance. Also, much of the earlier tests conducted on autonomous driving were predominantly on open roads and good weather, but more recent tests include weather conditions such as driving in fog, adverse weather events, or snow. Limited exposure of the self-driving LiDAR cameras has been enhanced using multimodal sensor fusion and point cloud analysis for object classification. The findings of the survey summarize that self-driving cars are no longer a question of if but more of when and how. The penetration rate of these autonomous robots into human society depends on their ability to drive safely. This puts forth a critical need for reliable object detection techniques, mathematical models and simulations to mimic reality and arrive at best parameters and configurations that can adapt with changes in surroundings. Nevertheless, with big data, DL and CNNs, we have tools at our disposal that can achieve high levels of arbitrary accuracy to solve perception problems in self-driving cars. These tools have provided researchers with the ability to break complex problems into easier ones and previously impossible problems into solvable but slightly expensive ones such as capturing and annotating data to create ground truth. A number of questions surrounding self-driving technology have emerged from this survey, given as follows:

1. How do people respond to self-driving technology ?
2. What evolutionary advances can be anticipated in LiDAR technology to capture high quality 3D data ?
3. What evolutionary advances can be expected with advent of 5G to enhance safety and communication in self-driving cars ?
4. At present, how does deep learning based image perception compare with that of the human eye ?
5. How are self-driving vehicles expected to increase in accuracy, and approach 100% accuracy over the next five years ?
6. How will the human driver respond to various system alerts and other electronic stimuli ?
7. What is the range of driver response times to assume control of the vehicle upon receiving alert, and is the response always guaranteed ?

One aspect to answer these questions is the advancement of self-driving car's ability for scene perception and object detection to navigate safely. To achieve these objectives, some future works are outlined below:

- An immediate future work is to collect data in hazardous weather conditions such as rain, hail, snow and study self-driving cars' navigation without human intervention.
- One of the current challenges involves transfer learning, as DL models are unable to transfer representations to unrelated domains. A research avenue is to apply transfer learning between domains that could lead to novel ways to interpret scenes, leading to real-time object detection.
- Current SSD techniques may be modified to work on videos [214]. A recently released dataset contains vehicular videos taken at 40 frames per second (fps) and can be used to test current cloud based DL in real-time [159].
- Self-driving cars are expected to be rolled out in phases, merging with existing cars. One application of DL is to process sounds of emergency vehicles up to a considerable distance and gauge the direction the sound is coming from. If self-driving cars are deployed with existing transportation networks, a mixture of human driven and autonomous vehicles will lead to highly complex driving scenarios. The beneficial effects of autonomous driving will depend on how well these cars differentiate between human driven cars and other self-driving cars. Scene inference and subsequent control actions could play a crucial role to integrate new technologies with legacy, conventional and existing systems in the least disruptive manner.
- Driving scene understanding and segmentation can be integrated with temporal propagation of information to understand both space and time. Different DL architectures such as RNNs can be used to generate automatic captioning of images, localizations and detection.
- The AI systems take decisions based on a cost/reward function, and execute a maneuver associated with lowest cost/highest reward. In cases where an unsafe/dangerous maneuver is the lowest cost option, the self-driving car would nonetheless execute that maneuver. This calls for reviewing cost/reward strategy in vehicular AI.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Pakusch C, Stevens G, Boden A, Bossauer P. Unintended effects of autonomous driving: a study on mobility preferences in the future. Sustainability 2018;10:2404.
[2] Kellett J, Barreto R, Hengel A, Vogiatzis N. How might autonomous vehicles impact the city? the case of commuting to central adelaide. Urban Pol Res 2019;37:442–57.
[3] Tian Y, Gelernter J, Wang X, Chen W, Gao J, Zhang Y, Li X. Lane marking detection via deep convolutional neural network. Neurocomputing 2018;280:46–55.
[4] Wang P, Huang X, Cheng X, Zhou D, Geng Q, Yang R. The apolloscape open dataset for autonomous driving and its application. IEEE Trans Pattern Anal Mach Intell 2019;42(10):2702–19.
[5] Calvert S, Schakel W, van Lint J. Will automated vehicles negatively impact traffic flow?'. J Adv Transport 2017;17:1–17.
[6] Brunetti A, Buongiorno D, Trotta G, Bevilacqua V. Computer vision and deep learning techniques for pedestrian detection and tracking: a survey. Neurocomputing 2018;300:17–33.
[7] Vodrahalli K, Bhowmik A. 3d computer vision based on machine learning with deep neural networks: a review. J Soc Inf Disp 2017;25:676–94.
[8] Divakarla K, Emadi A, Razavi S, Habibi S, Yan F. A review of autonomous vehicle technology landscape. Int J Electr Hybrid Veh (IJEHV) 2019;11:320–45.
[9] Bishop R. Intelligent vehicle technology and trends. Computer 2005;50:18–23.
[10] Herrmann A, Brenner W, Stadler R. Autonomous Driving: How the Driverless Revolution Will Change the World. 2018.
[11] Inagaki T, Sheridan T. A critique of the sae conditional driving automation definition, and analyses of options for improvement. Cognit Technol Work 2019;21:569–78.
[12] Zhang B, Wilschut E, Willemsen D, Martens M. Transitions to manual control from highly automated driving in non-critical truck platooning scenarios. Transport Res Part F: Psychology and Behaviour 2019;64:84–97.
[13] Price M, Lee J, Dinparastdjadid A, Toyoda H, Domeyer J. Effect of automation instructions and vehicle control algorithms on eye behavior in highly automated vehicles. International Journal of Automotive Engineering 2019;10:73–9.
[14] Roncoli c, Papageorgiou M, Papamichail I. Traffic flow optimisation in presence of vehicle automation and communication systems part i: a first-order multi-lane model for motorway traffic. Transport Res Part C 2015;57:241–59.
[15] Xu Q. Division of area of fixation interest for real vehicle driving tests. Math Probl Eng 2017:1–10.
[16] Ravichandiran S. Deep reinforcement learning with Python: master classic. RL, Deep RL, Distributional RL, Inverse RL, and More with OpenAI Gym and TensorFlow. second ed. Birmingham: Packt Publishing, Limited; 2020.
[17] Lapan M. Deep Reinforcement Learning Hands-On: Apply Modern RL Methods, with Deep Q-Networks, Value Iteration, Policy Gradients, TRPO, AlphaGo Zero and More. Limited, Birmingham: Packt Publishing; 2018.
[18] Masuda S, Nakamura H, Kajitani K. Rule-based searching for collision test cases of autonomous vehicles simulation. IET Intell Transp Syst 2018;12:1088–95.
[19] Montanaro U. Towards connected autonomous driving: review of use-cases. Veh Syst Dyn 2018:1–36.
[20] Royo S, Ballesta-Garcia M. An overview of lidar imaging systems for autonomous vehicles. Appl Sci 2019;9(19):4093.
[21] Bae K, Park J, Lee J, Lee Y, Lim C. Flower classification with modified multimodal convolutional neural networks. Expert Syst Appl 2020;159:113455.
[22] Borraz R. Cloud incubator car: a reliable platform for autonomous driving. Appl Sci 2018;8:303.
[23] Uã§ar A, Demir Y, Gã¼zelis C. Object recognition and detection with deep learning for autonomous driving applications. Simulation 2017;93:759–69.
[24] Yakovlev S, Borisov A. A synergy of the rosenblatt perceptron and the Jordan recurrence principle. Automat Contr Comput Sci 2009;43:31–9.
[25] Goodfellow I, Heaton J, Bengio Y, Courville A. Deep learning. first ed. The MIT Press; 2016, ISBN 0262035618800pp.
[26] Abbas Q, Ibrahim M, Jaffar M. A comprehensive review of recent advances on deep vision systems. Artif Intell Rev 2018:1–38.
[27] Lefevre S, Carvalho A, Borrelli F. A learning-based framework for velocity control in autonomous driving. IEEE Trans Autom Sci Eng 2016;13:33–42.
[28] Geng X. A scenario-adaptive driving behavior prediction approach to urban autonomous driving. Appl Sci 2017;7:426.
[29] Xue J, Fang J, Zhang P. A survey of scene understanding by event reasoning in autonomous driving. Int J Autom Comput 2018;15:249–66.
[30] Lin R, Ma L, Zhang W. An interview study exploring tesla drivers' behavioural adaptation. Appl Ergon 2018;72:37–47.
[31] Anonymous, Autobloggreen: Nutonomy will test a self-driving car in boston later this year, Newstex Trade and Industry Blogs.
[32] FavarÃ F, Nader N, Eurich S, Tripp M, Varadaraju N. Examining accident reports involving autonomous vehicles in California. PloS One 2017;12:e0184952.
[33] Tahboub K, Guera D, Reibman A, Delp E. Quality-adaptive deep learning for pedestrian detection. IEEE International Conference on Image Processing 2017:4187–91.
[34] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018;300. 7068349–13.
[35] Abbas Q, Ibrahim M, Jaffar M. A comprehensive review of recent advances on deep vision systems. Artif Intell Rev 2019;52:39–76.
[36] Kim J, Hong H, Park K, K J. Convolutional neural network-based human detection in nighttime images using visible light camera sensors. Sensors 2017;17:1065.

[37] Dairi A, Harrou F, Senouci M, Sun Y. Unsupervised obstacle detection in driving environments using deep-learning-based stereovision. Robot Autonom Syst 2018; 100:287–301.
[38] Zhang W, Wang Z, Liu X, Sun H, Zhou J, Liu Y, Gong W. Deep learning-based real-time fine-grained pedestrian recognition using stream processing. IET Intell Transp Syst 2018;12:602–9.
[39] Buda M, Maki A, Mazurowski M. A systematic study of the class imbalance problem in convolutional neural networks. Neural Network 2018;106:249–59.
[40] Meng Z, Fan X, Chen X, Chen Y, Tong. Detecting small signs from large images. Mach Vis Appl 2017;28:793–802.
[41] Lin W, Ren X, Zhou T, Cheng X, Tong M. A novel robust algorithm for position and orientation detection based on cascaded deep neural network. Neurocomputing 2018;308:138–46.
[42] Zhu Y, Zhang C, Zhou D, Wang X, Bai X, Liu W. Deep learning for decentralized parking lot occupancy detection. Expert Syst Appl 2017;72:327–34.
[43] Ramos S. Detecting unexpected obstacles for self-driving cars: fusing deep learning and geometric modeling. arXiv.org, https://arxiv.org/pdf/1612.06573.pdf.
[44] Diari A. Unsupervised obstacle detection in driving environments using deep-learning-based stereovision. Robot Autonom Syst 2018;100:287–301.
[45] Schuster M, Paliwal K. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997;45:2673–81.
[46] Batin M, Turchin A, Markov S, Zhila A, Denkenberger D. Artificial intelligence in life extension: from deep learning to superintelligence. Informatica 2017;41: 401–17.
[47] Shevlin H, Vold K, Crosby M, Halina M. The limits of machine intelligence: despite progress in machine intelligence, artificial general intelligence is still a major challenge. EMBO Rep 2019;20:e49177.
[48] Uijlings J. Selective search for object recognition. Int J Comput Vis 2013;104: 154–71.
[49] Li Y, Wang J, Xing T, Liu T, Li C, Su K. Tad16k: an enhanced benchmark for autonomous driving. IEEE International Conference on Image Processing 2017: 2344–8.
[50] F. Kurz, D. Waigand, P. Pekezou-Fouopi, E. Vig, C. Henry, N. Merkle, D. Rosenbaum, V. Gstaiger, S. Azimi, S. Auer, P. Reinartz, S. Knake-Langhorst, Dlrad a first look on the new vision and mapping benchmark dataset for autonomous driving, the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1.
[51] Wasson H. The other small screen: moving images at New York's world fair, 1939. Can J Film Stud 2012;21:81–103.
[52] Santos V, A D S, Miguel O. Special issue on autonomous driving and driver assistance systems. Robot Autonom Syst 2017;91:208–9.
[53] Cho E, Jung Y. Consumers understanding of autonomous driving. Inf Technol People 2018;31:1035.
[54] Brell T, P R, Martina Z. Suspicious minds? - users' perceptions of autonomous and connected driving. Theor Issues Ergon Sci 2019;20:301.
[55] Laes E, Gorissen L, Nevens F. A comparison of energy transition governance in Germany, The Netherlands and the United Kingdom. Sustainability 2014;6: 1129–52.
[56] Johannesson L, Murgovski N, Jonasson E, Hellgren J, Egardt B. Predictive energy management of hybrid long-haul trucks. Contr Eng Pract 2015;41:83–97.
[57] Ryan M. The future of transportation: ethical, legal, social and economic impacts of self-driving vehicles in the year 2025. Sci Eng Ethics 2019;26(3):1185–208.
[58] M. Hill, Cyber security in intelligent mobility transport systems catapult, DOI: 10.1049/ic.2016.0020.
[59] Nordhoff S, de Winter J, Kyriakidis M, van Arem B, Happee R. Acceptance of driverless vehicles: results from a large cross-national questionnaire study. J Adv Transport 2018:1–22.
[60] Deb S, Strawderman L, Carruth D, DuBien J, Smith B, Garrison T. Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. Transport Res Part C 2017;84:178–95.
[61] Robertson R, Meister S, Vanlaar W, Hing M. Automated vehicles and behavioural adaptation in Canada. Transport Res Part A 2017;104:50–7.
[62] de Winter J, Happee R, Martens M, Stanton N. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. Transport Res Part F: Psychology and Behaviour 2014;27: 196–217.
[63] Casey A, Niblett A. Self-driving contracts. J Corp Law 2017;43:1–33.
[64] Louw T, Markkula G, Boer E, Madigan R, Carsten O, Merat N. Coming back into the loop: drivers perceptual-motor performance in critical events after automated driving. Accid Anal Prev 2017;108:9–18.
[65] Gish J, Grenier A, Vrkljan B, Van Miltenburg B. Older people driving a high-tech automobile: emergent driving routines and new relationships with driving. Can J Commun 2017;42:235.
[66] Tivesten E, Victor T, Gustavsson P, Johansson J. Out-of-the-loop crash prediction: the automation expectation mismatch (aem) algorithm. IET Intell Transp Syst 2019;13:1231–40.
[67] Naujoks F, Purucker C, Neukum A, Wolter S, Steiger R. Controllability of partially automated driving functions does it matter whether drivers are allowed to take their hands off the steering wheel? Transport Res Part F: Psychology and Behaviour 2015;35:185–98.
[68] Anonymous. Intelligent sensing, planning and control for autonomous driving vehicles. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2017;47: 1759–60.
[69] Feng D, Rosenbaum L, Dietmayer K. Towards safe autonomous driving: capture uncertainty in the deep neural network for lidar 3d vehicle detection. arXiv.org, https://arxiv.org/pdf/1804.05132.pdf.
[70] Wolcott R, Eustice R. Robust lidar localization using multiresolution Gaussian mixture maps for autonomous driving. Int J Robot Res 2017;36:292–319.
[71] Vinel A. Vehicular networking for autonomous driving: guest editorial. IEEE Commun Mag 2015;53:62.
[72] Hobert L. Enhancements of v2x communication in support of cooperative autonomous driving. IEEE Commun Mag 2015;53:64–70.
[73] Liu S. A unified cloud platform for autonomous driving. Computer 2017;50:42–9.
[74] Noy I, Shinar D, Horrey W. Automated driving: safety blind spots. Saf Sci 2018; 102:68–78.
[75] Bansal P, Kockelman K. Are we ready to embrace connected and self-driving vehicles? a case study of texans. Transportation 2018;45:641–75.
[76] Petropoulos, G. (2017). Machines that learn to do, and do to learn: What is artificial intelligence?. Brussels: Bruegel. Retrieved from http://ezproxy.lib.ryerson.ca/login?url=https://www-proquest-com.ezproxy.lib.ryerson.ca/blogs,-podcasts,-websites/machines-that-learn-do-what-is-artificial/docview/1888642452/se-2?accountid=13631.
[77] Ajoudani A, Zanchettin A, Ivaldi S, Albu-SchÃ¤ffer A, Kosuge K, Khatib O. Progress and prospects of the humanrobot collaboration. Aut Robots 2018;42: 957–75.
[78] Lu Z, Coster X, de Winter J. How much time do drivers need to obtain situation awareness? a laboratory-based study of automated driving. Appl Ergon 2017;60: 293–304.
[79] Kindelsberger J. Designing toward minimalism in vehicle hmi. arXiv.org, https://arxiv.org/pdf/1805.02787.pdf.
[80] Banks V, Stanton N. Driver-centred vehicle automation: using network analysis for agent-based modelling of the driver in highly automated driving systems. Ergonomics 2016;59:1442–52.
[81] Sentouh C. Driver-automation cooperation oriented approach for shared control of lane keeping assist systems. IEEE Trans Contr Syst Technol 2018:1–17.
[82] Maurer M. Autonomous Driving: Technical, Legal and Social Aspects. 2016.
[83] Li S, Sui P, Xiao J, Chahine R. Policy formulation for highly automated vehicles: emerging importance, research frontiers and insights. Transport Res Part A 2019; 124:573–86.
[84] Zeeb K, Buchner A, Schrauf M. What determines the take-over time? an integrated model approach of driver take-over after automated driving. Accid Anal Prev 2015;78:212–21.
[85] Etzioni A, Etzioni O. Incorporating ethics into artificial intelligence. J Ethics 2017; 21:403–18.
[86] Birdsall M, , Google, ite. The road ahead for self-driving cars, Institute of Transportation Engineers. ITEA J 2014;84:36.
[87] Uhlemann E, H A M. Automakers are teaming up with cellular providers [connected vehicles]. IEEE Veh Technol Mag 2015;10:23–6.
[88] D. Etherington, Apple could use machine learning to shore up lidar limitations in self-driving, Techcrunch.
[89] D. L T., G. Rad, K. Choo, Driverless vehicle security: Challenges and future research opportunities, Future Generation Computer Systems.
[90] Ramanagopal M, Anderson C, Vasudevan R, Johnson-Roberson M. Failing to learn: autonomously identifying perception failures for self-driving cars. IEEE Robotics and Automation Letters 2018;3:3860–7.
[91] Hof R. Deep learning: with massive amounts of computational power, machines can now recognize objects and translate speech in real time. artificial intelligence is finally getting smart. MIT Technology Review 2013;116:32.
[92] Li L, Lin Y, Zheng N, Wang F, Liu Y, Cao D, Wang K, Huang W. Artificial intelligence test: a case study of intelligent vehicles. Artif Intell Rev 2018;50: 441–65.
[93] Goertzel B. Artificial general intelligence: concept, state of the art, and future prospects. Journal of Artificial General Intelligence 2014;5:1–48.
[94] Chen J. The evolution of computing: Alphago. Comput Sci Eng 2016;18:4–7.
[95] Van Brummelen J, OBrien M, Gruyer D, Najjaran H. Autonomous vehicle perception: the technology of today and tomorrow. Transport Res Part C 2018;89: 384–406.
[96] Granter S, Beck A, Papke J, David J. Alphago, deep learning, and the future of the human microscopist. Arch Pathol Lab Med 2017;141:619–21.
[97] Brownsword R. From erewhon to alphago: for the sake of human dignity, should we destroy the machines? Law, Innovation and Technology 2017;9:117–53.
[98] Madani A, Madani A, Yusof R. Traffic sign recognition based on color, shape, and pictogram classification using support vector machines. Neural Comput Appl 2010;30:2807–17.
[99] Kalra N, Paddock S. Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability? Transport Res Part A 2016;94: 182–93.
[100] H. Yin, C. Berger, When to use what data set for your self-driving car algorithm: an overview of publicly available driving datasets, IEEE 20th International Conference on Intelligent Transportation Systems (ITSC) DOI: 10.1109/ITSC.2017.8317828.
[101] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving the kitti vision benchmark suite, Computer Vision and Pattern recognition DOI: 10.1109/CVPR.2012.6248074.
[102] Endsley M. Autonomous driving systems: a preliminary naturalistic study of the tesla model s. Journal of Cognitive Engineering and Decision Making 2017;11: 225–38.
[103] Bhat A, Aoki S, Rajkumar R. Tools and methodologies for autonomous driving systems. Proc IEEE 2018;50:1–17.
[104] H. Yoo, Mems-based lidar for autonomous driving, E and i Elektrotechnik Und Informationstechnik.

[105] Meinel H. Radarsensors and autonomous drivingyesterday, today and tomorrow. E I Elektrotechnik Inf 2018;135:370–7.
[106] Kim J, Kang H. A new 3d object pose detection method using lidar shape set. Sensors 2018;18:882.
[107] H. Daraei, Tightly-coupled lidar and camera for autonomous vehicles, ProQuest Dissertations Publishing.
[108] Ramachandram D, Taylor GW. Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process Mag 2017;34(6):96–108.
[109] Ying Yang M, Rosenhahn B, Murino V. Multimodal scene understanding: Algorithms, applications and deep learning. first ed. San Diego: Elsevier Science and Technology; 2019.
[110] Lathuilire S, Massé P, Mesejo B, Horaud R. Neural network based reinforcement learning for audiovisual gaze control in humanrobot interaction. Pattern Recogn Lett 2019;118:61–71.
[111] Tian H, Tao Y, Pouyanfar S, Chen S, Shyu M. Multimodal deep representation learning for video classification. World Wide Web 2019;22(3):1325–41.
[112] Usama M, Qadir J, Raza A, Arif H, Yau K, Elkhatib Y, Hussain A, Al-Fuqaha A. Unsupervised machine learning for networking: techniques, applications and research challenges. IEEE access 2019;7:65579–615.
[113] Choi J-H, Lee J-S, Embracenet. A robust deep learning architecture for multimodal classification. Inf Fusion 2019;51:259–70.
[114] Gao M, Jiang J, Zou G, John V, Liu Z. Rgb-d-based object recognition using multimodal convolutional neural networks: a survey. IEEE access 2019;7: 43110–36.
[115] Wang W, Ramesh A, Zhu J, Li J, Zhao D. Clustering of driving encounter scenarios using connected vehicle trajectories. IEEE transactions on intelligent vehicles 2020;5(3):485–96.
[116] de Morais G, Marcos L, Bueno J, de Resende N, Terra M, Grassi Jr V. Vision-based robust control framework based on deep reinforcement learning applied to autonomous ground vehicles. Contr Eng Pract 2020;104:104630.
[117] Zhao B, Tang L, Han Y, Wang W. Deep spatial-temporal joint feature representation for video object detection. Sensors 2019;18:774.
[118] Li Q, Yu G, Wang J, Liu Y. A deep multimodal generative and fusion framework for class-imbalanced multimodal data. Multimed Tool Appl 2020;79:25023–50.
[119] Boutaba R, Salahuddin M, Limam N, Ayoubi S, Shahriar N, Estrada-Solano F, Caicedo O. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. Journal of internet services and applications 2018;9(1):1–99.
[120] Han M, Chen W, Moges A. Fast image captioning using lstm. Cluster Comput 2019; 22(S3):6143–55.
[121] Liu A, Shao Z, Wong Y, Li J, Su Y, Kankanhalli M. Lstm-based multi-label video event detection. Multimed Tool Appl 2019;78(1):677–95.
[122] Wu J, Shang S. Managing uncertainty in ai-enabled decision making and achieving sustainability. Sustainability 2020;12:8758.
[123] Aradi S. Survey of deep reinforcement learning for motion planning of autonomous vehicles. IEEE Trans Intell Transport Syst 2020:1–20.
[124] Qian L, Xu X, Zeng Y, Huang J. Deep, consistent behavioral decision making with planning features for autonomous vehicles. Electronics (Basel, Switzerland) 2019; 8(12):1492.
[125] Lu Y, Xu X, Zhang X, Qian L, Zhou X. Hierarchical reinforcement learning for autonomous decision making and motion planning of intelligent vehicles. IEEE Access 2020;8:209776–89.
[126] Bennajeh A, Bechikh S, Said L, Aknine S. Bi-level decision-making modeling for an autonomous driver agent: application in the car-following driving behavior. Appl Artif Intell 2019;33(13):1157–78.
[127] Berntorp K, Hoang T, Di Cairano S. Motion planning of autonomous road vehicles by particle filtering. IEEE transactions on intelligent vehicles 2019;4(2):197–210.
[128] Hoel C, Driggs-Campbell K, Wolff K, Laine L, Kochenderfer M. Combining planning and deep reinforcement learning in tactical decision making for autonomous driving. IEEE transactions on intelligent vehicles 2020;5(2).
[129] Hoel C, Driggs-Campbell K, Wolff K, Laine L, Kochenderfer M. Combining planning and deep reinforcement learning in tactical decision making for autonomous driving. IEEE transactions on intelligent vehicles 2020;5(2).
[130] Feng D, Haase-Schutz C, Rosenbaum L, Hertlein H, Glaser C, Timm F, Wiesbeck W, Dietmayer K. Deep multi-modal object detection and semantic segmentation for autonomous driving: datasets, methods, and challenges. IEEE Trans Intell Transport Syst 2020:1–20. https://doi.org/10.1109/TITS.2020.2972974.
[131] Chapel M, Bouwmans T. Moving objects detection with a moving camera: a comprehensive review. Computer science review 2020;38:100310.
[132] Xu LY, Wu G, Luo J. Multi-modal deep feature learning for rgb-d object detection. Pattern Recogn 2017;72:300–13.
[133] Farahnakian F, Heikkonen J. Deep learning based multi-modal fusion architectures for maritime vessel detection. Rem Sens 2020;12(16):2509.
[134] Sakaridis C, Dai D, Van Gool L. Semantic foggy scene understanding with synthetic data. Int J Comput Vis 2018;126(9):973–92.
[135] Naseer M, Khan S, Porikli F. Indoor scene understanding in 2.5/3d for autonomous agents: a survey. IEEE access 2019;7:1859–87.
[136] Jiao L, Zhang F, Liu F, Yang S, Li L, Feng Z, Qu R. A survey of deep learning-based object detection. IEEE access 2019;7:128837–68.
[137] R. Blin, S. Ainouz, S. Canu, F. Meriaudeau, Road scenes analysis in adverse weather conditions by polarization-encoded images and adapted deep learning.
[138] Abdi L, Meddeb A. In-vehicle augmented reality tsr to improve driving safety and enhance the drivers experience. Signal, image and video processing 2018;12(1): 75–82.
[139] Liu K, He L, Ma S, Gao S, Bi D. A sensor image dehazing algorithm based on feature learning. Sensors 2018;18(8):2606.
[140] Ito S, Soga M, Hiratsuka S, Matsubara H, Ogawa M. Quality index of supervised data for convolutional neural network-based localization. Appl Sci 2019;9(10): 1983.
[141] Ito S, Hiratsuka S, Ohta M, Matsubara H, Ogawa M. Small imaging depth lidar and dcnn-based localization for automated guided vehicle. Sensors 2018;18(1):177.
[142] Mahmoud Y, Okuyama Y, Fukuchi T, Kosuke T, Ando I. Optimizing deep-neural-network-driven autonomous race car using image scaling. SHS web of conferences 2020;77:4002.
[143] Debeunne C, Vivet D. A review of visual-lidar fusion based simultaneous localization and mapping. Sensors 2020;20(7):2068.
[144] Marti E, de Miguel M, Garcia F, Perez J. A review of sensor technologies for perception in automated driving. IEEE intelligent transportation systems magazine 2019;11(4):94–108.
[145] Rosique F, Navarro P, FernÃ¡ndez C, Padilla A. A systematic review of perception system and simulators for autonomous vehicles research. Sensors 2019;19(3):648.
[146] Fayyad J, Jaradat M, Gruyer D, Najjaran H. Deep learning sensor fusion for autonomous vehicle perception and localization: a review. Sensors 2020;20(15): 4220.
[147] Kolar P, Benavidez P, Jamshidi M. Survey of datafusion techniques for laser and vision based sensor integration for autonomous navigation. Sensors 2020;20(8): 2180.
[148] S. Yun, D. Kum, The multilayer perceptron approach to lateral motion prediction of surrounding vehicles for autonomous vehicles, IEEE Intelligent Vehicles Symposium (IV).
[149] Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 1980;36:193–202.
[150] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:2278–324.
[151] Webster C. Alan turings unorganized machines and artificial neural networks: his remarkable early work and future possibilities. Evolutionary Intelligence 2012;5: 35–43.
[152] Campmany V. Gpu-based pedestrian detection for autonomous driving. Procedia Computer Science 2016;80:2377–81.
[153] McCulloch W, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5:115–33.
[154] Diggavi S, Shynk J, Bershad N. Convergence models for rosenblatt's perceptron learning algorithm. IEEE Trans Signal Process 1995;43:1696–702.
[155] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[156] Andreopoulos A, Tsotsos J. 50 years of object recognition: directions forward. Comput Vis Image Understand 2013;117:827–91.
[157] Alom M. The history began from alexnet: a comprehensive survey on deep learning approaches. arXiv.org, https://arxiv.org/ftp/arxiv/papers/1803/1803.01164.pdf.
[158] Mitchell I, Lumsden S. Reinforcement learning. Account Irel 2018;50:84–5.
[159] Anonymous, Berkeley deepdrive releases 36,000 nexar videos to research community: The data fueling autonomous vehicle driving perception research amongst bdd members in the past year will now be made public, PR Newswire.
[160] Liu S. Computer architectures for autonomous driving. Computer 2017;50:18–25.
[161] Lee E. Internet of vehicles: from intelligent grid to autonomous cars and vehicular fogs. Int J Distributed Sens Netw 2016;12. 155014771666550.
[162] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew M. Deep learning for visual understanding: a review. Neurocomputing 2016;187:27–48.
[163] SoilÃ¡n M, Riveiro B, MartÃnez-SÃ¡nchez J, Arias P. Traffic sign detection in mls acquired point clouds for geometric and image-based semantic inventory. ISPRS J Photogrammetry Remote Sens 2016;114:92–101.
[164] Khalif A, Mignotte M. Segmentation data visualizing and clustering. Multimed Tool Appl 2017;76:1531–52.
[165] Vogelpohl T. Transitioning to manual driving requires additional time after automation deactivation. Transport Res Part F: Psychology and Behaviour 2018; 55:464–82.
[166] H. D. D.. Effects of platooning on signal-detection performance, workload, and stress: a driving simulator study. Appl Ergon 2017;60:116–27.
[167] Sugano Y, Matsushita Y, Sato Y. Appearance-based gaze estimation using visual saliency. IEEE Trans Pattern Anal Mach Intell 2013;35:329–41.
[168] Fridman L. Owl and lizard: patterns of head pose and eye pose in driver gaze classification. IET Comput Vis 2016;10:308–13.
[169] Dairi A, Harrou F, Senouci M, Sun Y. Technology - cybernetics; new cybernetics findings from sejong university described (efficient deep cnn-based fire detection and localization in video surveillance applications). J Technol 2019:1207.
[170] Lopez A. Training my car to see using virtual worlds. Image Vis Comput 2017;68: 102–18.
[171] Zhang Y, Chen H, Waslander S, Yang T, Zhang S, Xiong G, Liu K. Toward a more complete, flexible, and safer speed planning for autonomous driving via convex optimization. Sensors 2018;18:2185.
[172] Balado J, MartÃnez-SÃ¡nchez J, Arias P, Novo A. Road environment semantic segmentation with deep learning from mls point cloud data. Sensors 2019;19: 3466.
[173] Voulodimos W. Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018:1–13.
[174] Song X, Rui T, Zhang S, Fei J, Wang X. A road segmentation method based on the deep auto-encoder with supervised learning. Comput Electr Eng 2018;68:381–8.
[175] Ding C, Hu Z, Karmoshi S, Zhu M. A novel two-stage learning pipeline for deep neural networks. Neural Process Lett 2017;46:159–69.

[176] Han X, Lu J, Zhao C, You S, Li H. Semisupervised and weakly supervised road detection based on generative adversarial networks. IEEE Signal Process Lett 2018;25:551–5.
[177] Chung J. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv.org, https://arxiv.org/pdf/1412.3555.pdf.
[178] Lin G. Deep unsupervised learning for image super-resolution with generative adversarial network. Signal Process Image Commun 2018;68:88–100.
[179] You C, Lu J, Filev D, Tsiotras P. Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning. Robot Autonom Syst 2019;114:1–18.
[180] Li Y, Cui F, Xue X, Cheung-Wai Chan J. Coarse-to-fine salient object detection based on deep convolutional neural networks. Signal Process Image Commun 2018;64:21–32.
[181] Zhang K. Beyond a Gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans Image Process 2017;26:3142–55.
[182] Mahmoudi N, Ahadi S, Rahmati M. Multi-target tracking using cnn-based features: Cnnmtt. Multimedia Tools and Applications; 2018. p. 1–20.
[183] Liu S. Creating Autonomous Vehicle Systems. third ed. Morgan Claypool Publishers; 2017.
[184] Qian Y, Dong J, Wang W, Tan T. Feature learning for steganalysis using convolutional neural networks. Multimed Tool Appl 2018;77:19633–57.
[185] Al-Qizwini M. Deep learning algorithm for autonomous driving using googlenet. IEEE Intelligent Vehicles Symposium 2017;(IV):89–96. https://doi.org/10.1109/IVS.2017.7995703.
[186] Nguyen A, Sentouh C, Popieul J. Sensor reduction for driver-automation shared steering control via an adaptive authority allocation strategy. IEEE ASME Trans Mechatron 2018;23:5–16.
[187] Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW. Action recognition in video sequences using deep bi-directional lstm with cnn features. IEEE access 2018;6:1155–66.
[188] Abdellaoui M, Douik A. Human action recognition in video sequences using deep belief networks. Trait Du Signal 2020;37(1):37–44.
[189] Lahan GS, Talukdar AK, Sarma KK. Action recognition from depth video sequences using microsoft kinect. IEEE; 2019. p. 35–40.
[190] Nie S, Zheng M, Ji Q. The deep regression bayesian network and its applications: probabilistic deep learning for computer vision. IEEE Signal Process Mag 2018;35:101–11.
[191] Ji Z, Zhao Y, Pang Y, Li X. Cross-modal guidance based auto-encoder for multi-video summarization. Pattern Recogn Lett 2020;135:131–7.
[192] Xie D, Zhang L, Bai L. Deep learning in visual computing and signal processing. Applied Computational Intelligence and Soft Computing 2017:1–13.
[193] Asvadi A, Garrote L, Premebida C, Peixoto P, Nunes U. Multimodal vehicle detection: fusing 3d-lidar and color camera data. Pattern Recogn Lett 2018;115:20–9.
[194] TomÃ D, Monti F, Baroffio L, Bondi L, Tagliasacchi M, Tubaro S. Deep convolutional neural networks for pedestrian detection. Signal Process Image Commun 2016;47:482–9.
[195] W. Xiang, Context-aware single-shot detector, IEEE Winter Conference on Applications of Computer Vision.
[196] Carse B, Oreland J. Evolution and learning in neural networks: dynamic correlation, relearning and thresholding. Adapt Behav 2000;8:297–311.
[197] K. He, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition DOI: 10.1109/CVPR.2016.90.
[198] Zhang J. A real-time Chinese traffic sign detection algorithm based on modified yolov2. Algorithms 2017;10:127.
[199] Ren S. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137–49.
[200] Wang X, Hu Z, Tao Q, Huang G, Mu M. Bayesian place recognition based on bag of objects for intelligent vehicle localisation. IET Intell Transp Syst 2019;13:1736–44.
[201] Li H. Zoom out-and-in network with map attention decision for region proposal and object detection. Int J Comput Vis 2018:1–14.
[202] X. Chen, Monocular 3d object detection for autonomous driving, DOI: 10.1109/CVPR.2016.236.
[203] Hane C. 3d visual perception for self-driving cars using a multi-camera system: calibration, mapping, localization, and obstacle detection. Image Vis Comput 2017;68:14–27.
[204] Wang X. Regionlets for generic object detection. IEEE Trans Pattern Anal Mach Intell 2015;37:2071–84.
[205] Witoonchart P, Chongstitvatana P. Application of structured support vector machine backpropagation to a convolutional neural network for human pose estimation. Neural Network 2017;92:39–46.
[206] Watanabe T, Ito S, Yokoi K. Image feature descriptor using co-occurrence histograms of oriented gradients for human detection. J Inst Image Inf Televis Eng 2017;71:J28–34.
[207] R. Girshick, Deformable part models are convolutional neural networks, IEEE Computer Society Conference on Computer Vision and Pattern Recognition DOI: 10.1109/CVPR.2015.7298641.
[208] Hosang J. What makes for effective detection proposals? IEEE Trans Pattern Anal Mach Intell 2016;38:814–30.
[209] Shim I. An autonomous driving system for unknown environments using a unified map. IEEE Trans Intell Transport Syst 2015;16:1999–2013.
[210] Feng Y. Improved object proposals with geometrical features for autonomous driving. Mobile Inf Syst 2017:1–11.
[211] Lee U, Jung J, Jung S, Shim D. Development of a self-driving car that can handle the adverse weather. Int J Automot Technol 2018;19:191–7.
[212] R. Girshick, Rich feature hierarchies for accurate object detection and semantic segmentation, Computer Vision and Pattern Recognition DOI: 10.1109/CVPR.2014.81.
[213] He K. Mask r-cnn. IEEE Transactions on Pattern Analysis and Machine Intelligence; 2018.
[214] Yaseen M, Anjum A, Farid M, Antonopoulos N. Cloudbased video analytics using convolutional neural networks. Software Pract Ex 2019;49:565–83.
[215] Chollet F. Xception: deep learning with depthwise separable convolutions. IEEE Conference on Computer Vision and Pattern Recognition 2017:1800–7.
[216] Everingham M. The pascal visual object classes (voc) challenge. Int J Comput Vis 2010;88:303–38.
[217] Ma C. Data-driven state-increment statistical model and its application in autonomous driving. IEEE Trans Intell Transport Syst 2018:1–11.
[218] Jung Y, Seo S-W, Kim S-W. Curb detection and tracking in low-resolution 3d point clouds based on optimization framework. IEEE Trans Intell Transport Syst 2019;21(9):1–16.
[219] Song X, Zhan W, Che X, Jiang H, Yang B. Scale-aware attention-based pillarsnet (sapn) based 3d object detection for point cloud, Mathematical problems in engineering. 2020.
[220] Arnold E, Al-Jarrah OY, Dianati M, Fallah S, Oxtoby D, Mouzakitis A. A survey on 3d object detection methods for autonomous driving applications. IEEE Trans Intell Transport Syst 2019;20(10):3782–95.
[221] Shi S, Wang Z, Shi J, Wang X, Li H. From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. IEEE Trans Pattern Anal Mach Intell 2020:1. https://doi.org/10.1109/TPAMI.2020.2977026.
[222] Shen S, Wang L, Duan S, He X. Car plate detection based on yolov3. Conference series J Phys 2020;1544:12039.
[223] Luo Z, Yu H, Zhang Y. Pine cone detection using boundary equilibrium generative adversarial networks and improved yolov3 model. Sensors 2020;20(16):4430.
[224] Kohl C, Knigge M, Baader G, Bhm M, Krcmar H. Anticipating acceptance of emerging technologies using twitter: the case of self-driving cars. J Bus Econ 2018;88:617–42.
[225] Volosov V, Gubarev V, Kuntsevich V. Problems of creating intelligent autonomous robotic underwater vehicles and their application. Cybern Syst Anal 2007;43:696–703.
[226] Broggi A, Buzzoni M, Debattisti S, Grisleri P, Laghi M, Medici P, Versari P. Extensive tests of autonomous driving technologies. IEEE Trans Intell Transport Syst 2013;14:1403–15.
[227] D. Lumb, Navya puts its self-driving shuttle tech in an autonomous taxi, Endgadget.
[228] Anonymous, Transport systems catapult conducts first public autonomous driving trials, European Union News.
[229] Anonymous. Chery-chery self-driving car developed with baidu appears at ces 2018 12/1/2018. 2018. p. 5. China Automotive.
[230] Kapoor A, Grauman K, Urtasun R, Darrell T. Gaussian processes for object categorization. Int J Comput Vis 2010;88:169–88.
[231] Prianto E, Kim M, Park J-H, Bae J-H, Kim J-S. Path planning for multi-arm manipulators using deep reinforcement learning: soft actor-critic with hindsight experience replay. Sensors 2020;20(20):5911.
[232] TomÃ D, Monti F, Baroffio L, Bondi L, Tagliasacchi M, Tubaro S. Deep convolutional neural networks for pedestrian detection. Signal Process Image Commun 2016;47:482–9.
[233] Ramanagopal M, Anderson C, Vasudevan R, Johnson-Roberson M. Failing to learn: autonomously identifying perception failures for self-driving cars. IEEE Robotics and Automation Letters 2018;3:3860–7.
[234] Voannidou A, Chatzilari E, Nikolopoulos S, Kompatsiaris I. Deep learning advances in computer vision with 3d data: a survey. ACM Comput Surv 2018;50:1–38.
[235] Shafiee M. Squishednets: squishing squeezenet further for edge device scenarios via deep evolutionary synthesis. arXiv.org, https://arxiv.org/pdf/1711.07459.pdf.
[236] Chen T, Lu S, Fan J, S-cnn. Subcategory-aware convolutional networks for object detection. IEEE Trans Pattern Anal Mach Intell 2017;39.
[237] Chowdhuri S, Pankaj T, Zipser K. Multinet: multi-modal multi-task learning for autonomous driving. IEEE Winter Conference on Applications of Computer Vision 2019;13:1496–504.
[238] Lin T. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence; 2018.
[239] Wang S, Zhu B, Li C, Wu M, Zhang J, Chu W, Qi Y. Riemannian proximal policy optimization. Comput Inf Sci 2020;13(3):93.
[240] Graesser L, Keng W. Foundations of Deep Reinforcement Learning: Theory and Practice in Python. first ed. Addison-Wesley Professional; 2019.
[241] Zhang C, Kim J. Multi-scale pedestrian detection using skip pooling and recurrent convolution. Multimed Tool Appl 2018;10:1–18.
[242] Gers F, Schmidhuber J, Cummins F. Learning to forget: continual prediction with lstm. Neural Comput 2000;12:2451–71.
[243] Cunneen M, Mullins M, Murphy F. Autonomous vehicles and embedded artificial intelligence: the challenges of framing machine driving decisions. Appl Artif Intell 2019;33(8):706–31.
[244] Cunneen M, Mullins M, Murphy F. Artificial intelligence assistants and risk: framing a connectivity risk narrative. AI Soc 2020;35(3):625–34.
[245] Verganti R, Vendraminelli L, Iansiti M. Innovation and design in the age of artificial intelligence. J Prod Innovat Manag 2020;37(3):212–27.
[246] Grigorescu S, Trasnea B, Cocias T, Macesanu G. A survey of deep learning techniques for autonomous driving. J Field Robot 2019;37(3):362–86.
[247] Yurtsever E, Lambert J, Carballo A, Takeda K. A survey of autonomous driving: common practices and emerging technologies. IEEE access 2020;8:58443–69.

[248] Devi S, Malarvezhi P, Dayana R, Vadivukkarasi K. A comprehensive survey on autonomous driving cars: a perspective view. Wireless Pers Commun 2020;114(3): 2121–33.
[249] Qayyum A, Usama M, Qadir J, Al-Fuqaha A. Securing connected & autonomous vehicles: challenges posed by adversarial machine learning and the way forward. IEEE Communications surveys and tutorials 2020;22(2):998–1026.
[250] Ni J, Chen Y, Chen Y, Zhu J, Ali D, Cao W. A survey on theories and applications for self-driving cars based on deep learning methods. Appl Sci 2020;10(8):2749.