Sensing and Machine Learning for Automotive Perception: A Review

Pandharipande, Ashish; Cheng, Chih-Hong; Dauwels, Justin; Gurbuz, Sevgi; Ibanez-Guzman, Javier; Li, Guofa; Piazzoni, Andrea; Wang, Pu; Santra, Avik

TR2023-089 July 15, 2023

Abstract

Automotive perception involves understanding the external driving environment as well as the internal state of the vehicle cabin and occupants using sensor data. It is critical to achieving high levels of safety and autonomy in driving. This paper provides an overview of different sensor modalities like cameras, radars, and LiDARs used commonly for perception, along with the associated data processing techniques. Critical aspects in perception are considered, like architectures for processing data from single or multiple sensor modalities, sensor data processing algorithms and the role of machine learning techniques, methodologies for validating the performance of perception systems, and safety. The technical challenges for each aspect are analyzed, emphasizing machine learning approaches given their potential impact on improving perception. Finally, future research opportunities in automotive perception for their wider deployment are outlined.

IEEE Sensors Journal 2023

Sensing and Machine Learning for Automotive Perception: A Review

Ashish Pandharipande, Chih-Hong Cheng, Justin Dauwels, Sevgi Z. Gurbuz, Javier Ibanez-Guzman, Guofa Li, Andrea Piazzoni, Pu Wang, Avik Santra

Abstract—**Automotive perception involves understanding the external driving environment as well as the internal state of the vehicle cabin and occupants using sensor data. It is critical to achieving high levels of safety and autonomy in driving. This paper provides an overview of different sensor modalities like cameras, radars, and LiDARs used commonly for perception, along with the associated data processing techniques. Critical aspects in perception are considered, like architectures for processing data from single or multiple sensor modalities, sensor data processing algorithms and the role of machine learning techniques, methodologies for validating the performance of perception systems, and safety. The technical challenges for each aspect are analyzed, emphasizing machine learning approaches given their potential impact on improving perception. Finally, future research opportunities in automotive perception for their wider deployment are outlined.**

Index Terms—**Automotive perception, radars, cameras, LiDAR, sensor data processing, advanced driver assistance system, autonomous driving, safety.**

I. Introduction

Different levels of automation are being included in modern vehicles, from an Advanced Driver Assistance System (ADAS) to a fully automated driving system (ADS). These systems use sensor and control technologies to improve driving safety and comfort. Automotive perception is a core module in such systems. Perception information relies on using one or more sensor modalities like camera, radar, and LiDAR. By suitably processing raw sensor data, information on the environment around the vehicle (external perception) and the state of the vehicle cabin (internal perception) is derived. Each sensor has its strengths and limitations, and its signal response will vary according to the driving environment [1]. The processing of sensor data is thus key to deriving reliable environment information for safe vehicle driving. This includes different aspects: (i) processing architectures considering what data to combine from sensors at which level of the processing chain, (ii) processing algorithms for external perception capabilities like object detection and classification, range and velocity estimation, or for internal perception capabilities like occupancy detection, occupant alertness, (iii) incorporation of physics/model/data-driven methods into data processing, (iv) application-level performance metrics and validation approaches, and (v) safety in perception-driven vehicle functions. In this paper, our objectives are to identify the key challenges in sensor processing for automotive perception, to review the advances towards addressing them, and to identify the gaps that exist to attain higher levels of perception.

High-quality, robust automotive perception is needed to reduce the number of traffic accidents and fatalities resulting from human driving errors. Automotive perception has seen significant progress in the past years due to the emergence of advanced sensors, computing power, and the successful application of machine learning techniques. This has led to the deployment of multiple driving functions with increased levels of autonomy in commercial vehicles. Automotive perception, however, is a challenging problem for several reasons. First, the operational design domain (ODD) is complex, and perception needs to be reliable across different environmental and driving conditions. Second, the interaction between perception and driving controls may lead to propagation errors when the human is no longer part of the control loop. Third, the design and deployment of perception-based autonomous vehicles involves new technological as well as social challenges.

The remainder of the paper is organised as follows. Section II outlines the architecture of ADAS/ADS systems to provide the context within which automotive perception systems are used. Section III describes the different applications of automotive perception systems to infer information about the exterior and interior of the vehicle. Section IV provides details of state-of-the-art methods used to infer information from data acquired from major classes of sensors, namely radar, camera and LiDAR. It includes the challenges involved in deriving reliable and robust perception. Section V addresses the emergent validation domain, discussing the norms, safety metrics, and the monitoring of abnormal situations especially considering machine learning. Finally, Section VI concludes the paper by presenting future opportunities for automotive perception.

II. Generic ADAS/ADS architecture

An ADAS/ADS as depicted in Fig. 1, has multiple components [2]: a sensor system with sensors, data processing involv-

The co-authors contributed equally to this article.

A. Pandharipande is with NXP Semiconductors, The Netherlands (Email: ashish.pandharipande@nxp.com). C.-H. Cheng is with Fraunhofer IKS, Germany (Email: chih-hong.cheng@iks.fraunhofer.de). J. Dauwels is with TU Delft, The Netherlands (Email: J.H.G.Dauwels@tudelft.nl). S. Z. Gurbuz is with the Dept. of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487 USA (Email: szgurbuz@ua.edu). J. Ibanez-Guzman is with Renault, France (Email: javier.ibanez-guzman@renault.com). G. Li is with the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China (e-mail: hanshan198@gmail.com), A. Piazzoni is with Nanyang Technological University, Singapore (Email: andrea006@ntu.edu.sg), P. Wang is with Mitsubishi Electric Research Laboratories, Cambridge 02139 USA (Email: pwang@merl.com). A. Santra is with Infineon Technologies, Germany (Email: Avik.Santra@infineon.com).

ing sensing and perception, a decision and planning system, and advanced driving controls and automotive services.

The perception system covers awareness information both external and internal to the automotive. External perception covers information on static and dynamic objects on the street, traffic and street signs, and is obtained using sensors like cameras, radars, and LiDARs. External perception provides a real-time picture of the dynamic environment around the vehicle, either by advanced processing of data from a sensor modality [3], [4], [5], or by fusing data from multiple modalities [6], [7]. Another aspect of external perception is localization to determine the location of the vehicle. Global Positioning System (GPS) is commonly used for localization to provide a global position of the vehicle as well as a velocity estimate. A limitation of GPS is that the received signals suffer from blockage and heavy multipath conditions typical in dense urban environments. To compensate for resulting localization vulnerabilities in GPS, an inertial measurement unit (IMU) with sensors like gyroscopes and accelerometers is employed. In such a fusion-based localization system, GPS position errors are corrected using IMU data along with additional constraints on vehicle motion, orientation and its position on a map. Note that external perception sensors can also be used to derive relative localization information using cameras [8] and radars [9]. Internal perception provides information on the occupants and objects in an automotive. Using sensor technologies like camera and radar, presence, activity and attention levels of the driver and passengers may be monitored to support various levels of autonomous driving. Understanding the state of occupants is crucial for effective human-vehicle interaction in ADAS/ADS.

The decision and planning system determines the maneuvers for a vehicle. Driving decisions are based on previously acquired knowledge about the environment, such as the drivable area and traffic rules, and also real-time information such as objects in the vicinity and traffic patterns. Decision and planning can be divided into three stages: global routing, behavior inference, and local motion planning. Global routing determines vehicle routes from source point A to destination point B, according to some criteria like shortest travel time or least number of traffic signs encountered. This determination is done using graph routing algorithms, a digital map and a traffic management system. After a global route has been determined, the automotive must be able to navigate the selected route and interact with other traffic participants according to driving conventions and rules. Given a sequence of road segments specifying the selected route, the behavioral inference stage is responsible for selecting an appropriate driving behavior at any point of time based on the perceived behavior of other traffic participants, road conditions, and other available signals from the infrastructure. The local motion planning stage translates the behavioral inference stage decisions into a feasible local path plan. It determines a path that is dynamically feasible for the automotive, comfortable for the passenger, and avoids collisions with obstacles determined by the perception system.

Vehicle driving control executes the reference path defined by the decision and planning system by selecting appropriate actuator inputs to carry out the planned motion path. Controls need to be accurate for safe automotive driving and robust under various driving conditions. As such, the control system should also be able to deal with diverse physical vehicle characteristics and dynamics. The vehicle control system is intimately linked to advanced driving functions like adaptive cruise control, emergency braking, and lane keeping assistance; automotive services like assisted parking, traffic alerts and diagnostics further enhance the experience and safety of a user.

III. Sensors for Automotive Perception applications

A. *Automotive-external perception*

Different sensor-based vehicle applications, as depicted in Fig. 2, rely on information from the perceived environment for situational understanding and decision making. These exteroceptive sensors acquire data from the vehicle's environment, which is then transformed into meaningful information such as the occupancy grid map, the 3D position of different traffic agents, and road characteristics (e.g., lane markings). There are two types of sensors: Passive sensors, like video cameras and infrared/thermal imaging sensors, which measure ambient environmental energy entering the sensor; and active sensors which emit energy into the environment and then measure the environmental response. These sensors can manage more controlled interactions with the environment; however, the emitted energy is limited by safety constraints or from interference between its signal and those of other active sensors which may impact sensing performance (refer to [10], [11] for automotive radar transmission power limits). Examples of active sensors include ultrasonic, LiDARs and Radar.

Exteroceptive sensors are sensitive to the operating outdoor environment conditions, and their performance can vary considerably according to their deployment location and weather conditions. The performance of perception algorithms depends on the ODD where the automotive operates [12], with various factors coming into play like: Dynamic range (e.g., the ratio of the largest to the smallest measurable signal for a radar), Range (e.g., how far or how near a radar or LiDAR can detect), Resolution (e.g., the number of pixels in an image), and frame rate (e.g., the rate at which data is acquired or the frames per second of a camera). In addition, the performance of perception sensors is also impacted by its layout. For example, while radars are commonly placed around vehicle bumpers or brand emblems, the same placement may not be suitable for other modalities like cameras due to an impeded field-of-view. Environment conditions have a role on sensor lifetime performance - very low or high temperatures, dust, humidity are factors that affect sensor performance and means to weather-proof sensors or to service them must be accounted for.

B. *Automotive-internal perception*

Automotive-internal perception, also termed in-cabin monitoring, systems are an indispensable feature of vehicle safety systems. They are a part of cyber-physical human systems (CPHS) capable of responding or taking actions based on

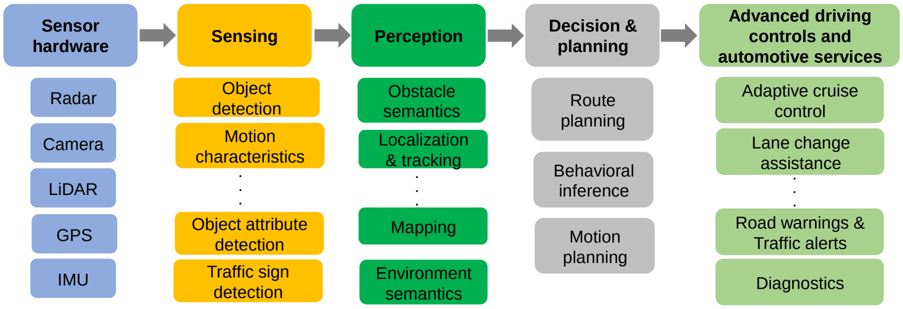


Fig. 1. Architecture of an ADAS/ADS equipped automotive.

perception of the human condition in the vehicle, as shown in Figure 3. Four principle tasks have been the focus of development: 1) occupancy detection/characterization, 2) driver monitoring, 3) passenger monitoring, and 4) human-vehicle interaction. Occupancy detection/characterization refers to the ability to automatically detect where people are located in the vehicle and especially the presence of children and infants. Occupancy sensing can provide valuable input on the proper use of child seats and seat belts, while also enabling the optimization of airbag function according to the height of the person sitting in the seat. This function is also an important safety feature for preventing the death of children and pets left in vehicles on hot days, as a vehicle perceiving this situation could take preventive measures, such as notifying the owner of the vehicle, turning on the air conditioning or opening a window. New U.S. federal regulations require all new cars to be equipped with a back-seat alert system [13], while the European New Car Assessment Program also prescribes new Child Presence Detection (CPD) protocols [14] to prevent in-vehicle heat-related child deaths.

Driving monitoring systems are primarily targeted towards ensuring safety by monitoring a driver's ability to effectively drive the vehicle. This can be indicated by a variety of measurable variables, such as driver vital signs (heart rate and respiration), driver fatigue, drowsiness and attention. Examples of attention monitoring include tracking the direction that the driver is gazing, head movements, and eye blinking, especially blink duration and frequency. Because there is a correlation between fatigue and heart rate, vital sign monitoring can be used for detection of critical health events, such as a heart attack, but also for drowsiness detection. Other health related indicators that have been considered include blood pressure measurement and blood glucose level monitoring. Passenger monitoring includes features of vital signs and health monitoring, as well as general activity within the vehicle. Especially when children are present, monitoring seat belt usage and whether potentially dangerous passenger activity is occurring (such as children changing seats) can be important for safety.

Human-vehicle interaction (HVI) has been predominantly considered within the context of non-contact gesture recognition for control of user interfaces and vehicle sub-systems, such as the radio, infotainment systems, or air conditioning, among others. However, the prospect of autonomous vehicles raises new dimensions in HVI, whereby two-way human communications with the vehicle must also be considered. Examples motivating such functionality include not just in-cabin HVI, but interactions that might occur if a passengerless autonomous vehicle were pulled over by a police officer. HVI also arises in environmental scene understanding outside the car, as an autonomous vehicle must also be able to navigate based on directions from a police officer directing traffic with the additional consideration that some detours may not be well marked with driving lanes. Moreover, while current collision avoidance systems simply try to detect and steer away from obstacles, HVI systems of the future could also include pedestrian injury mitigation features in the event of collision.

Currently, automotive-internal perception systems imple-

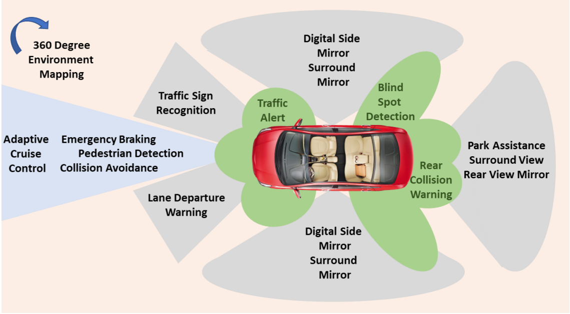


Fig. 2. Driving features based on automotive-external perception.

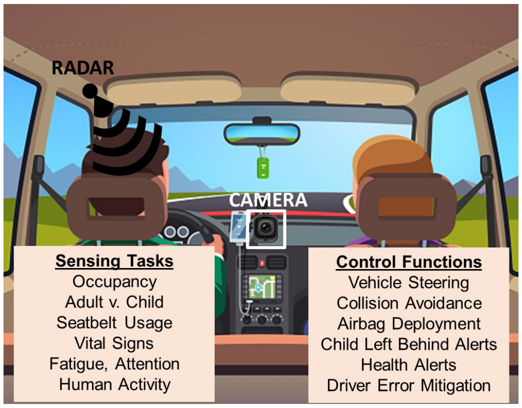


Fig. 3. Driving features based on automotive-internal perception.

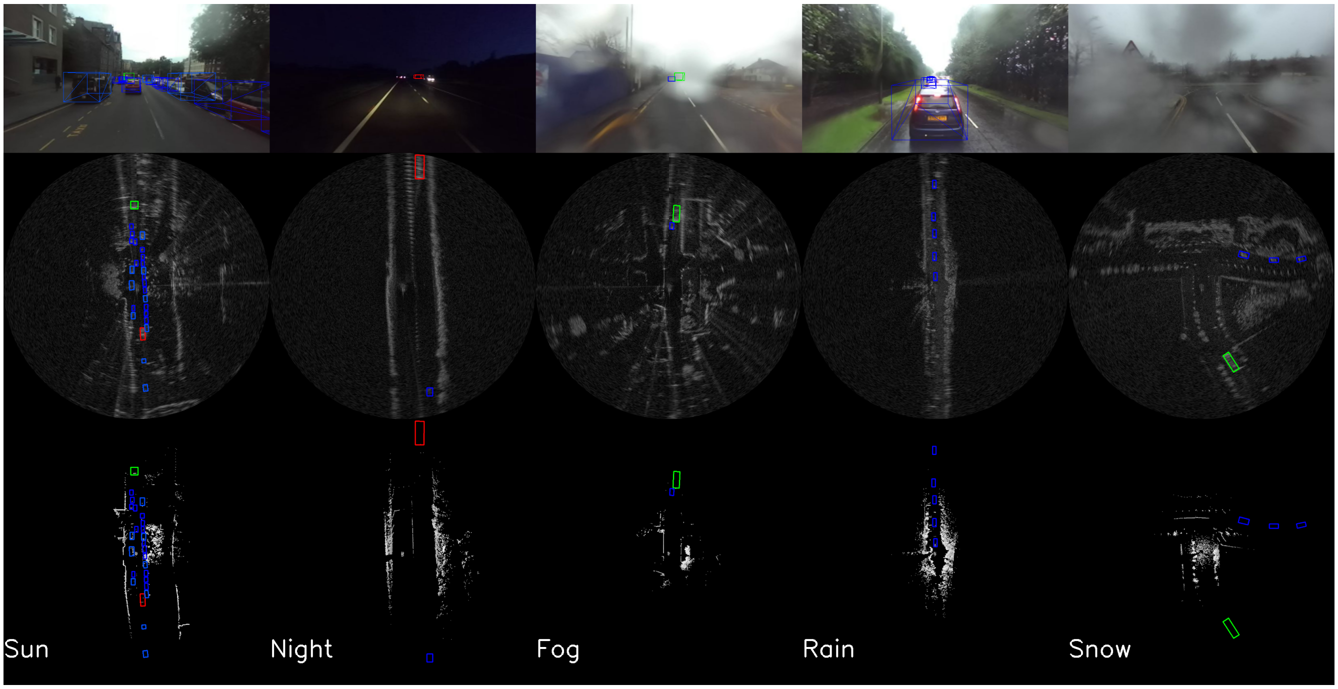


Fig. 4. Automotive external perception frames of stereo camera (top row), Radar (middle row) and LiDAR (bottom row) under different weather and light conditions (sun, fog, rain, snow, and night in five columns). Figures are plotted from data in the open RADIATE dataset [15].

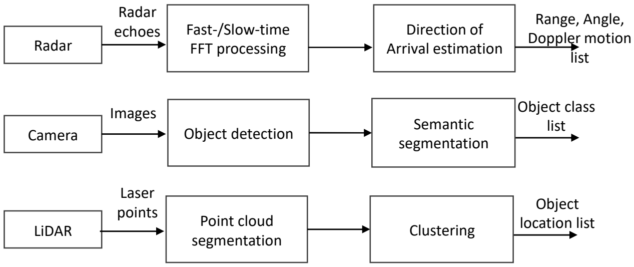


Fig. 5. Illustrative processing for radar, camera and LiDAR external-perception.

mented in commercially-available vehicles rely on video-based technologies [16] for extracting passenger information such as occupancy, posture, whether part of the body is outside the vehicle, seat belt usage, use of a car seat, whether a child has been left unattended, whether there are objects or belongings in the vehicle, and detection of pets, animals, or miscellaneous irregular situations. In addition to these functions for safety, video can also be used for security purposes to detect incidents of vandalism, abuse, presence of dangerous weapons, theft or other illegal activities. Video has also been used to enable non-contact user interface control via gesture recognition. In this case, a small camera is positioned such that it can monitor the area immediately in front of the touch screen next to the driver.

Although cameras have enabled some in-cabin monitoring features, their performance is adversely impacted by changing and low ambient light conditions. There has been an increased interest in radars recently for in-cabin monitoring, as millimeter wave radars can be used to remotely measure vital signs, track eye blinking, recognize gestures and detect vehicle occupancy. Radar offers sensing solutions that are less invasive of privacy in comparison to video, while not being dependent upon ambient lighting conditions. Over the last few years, several studies examining radar-based occupancy detection have been published. While some studies have considered continuous wave [17], pulsed [18], and impulse radio ultra-wide band (IR-UWB) radar [19] systems, most works have focused on utilization of high-range resolution FMCW radars [20], especially multi-channel FMCW [21]–[23] due to its greater angular resolution.

Driver attention monitoring studies have focused on radar-based vital sign recognition [24]–[26], monitoring of driver head movements [27], eye blinking [28]–[31], fatigue, concentration and drowsiness [32]–[35], and health indicators, such as blood pressure [36] and blood glucose levels [37]. The presence of body movements and respiration effects both the measurement of heart rate as well as that of eye blinking frequency and duration. Thus, methods for jointly estimating heartbeat and blink rate have also been proposed [38].

Gesture recognition using radar was postulated early in [39], but was made practically possible with the development of integrated, millimeter wave RF transceivers. Radar-based gesture recognition gained significant attention due to the Google SOLI project [40]. While gesture recognition research has predominantly focused on the design of deep neural networks to improve classification accuracy of common, ubiquitous hand gestures, such as virtual knob, slider, and push button, the efficacy of radar has also been demonstrated for sign language recognition [41]. In automotive environments, studies have considered radar-based intelligent driver assistance [42], the utilization of LiDAR, camera and radar for traffic signalling gesture recognition [43], and radar-enabled HVI for in-car infotainment control [44]. In-vehicle behavior and gesture recognition has also been proposed using Wi-Fi signals [45].

IV. Processing and Learning Algorithms

As discussed in the previous section, exteroceptive sensors exhibit considerably varying performance and sensitivity to operating outdoor environment conditions. This is manifested in Fig. 4 where stereo camera, radar, and LiDAR frames from the RADIATE dataset [15] are plotted under various light/weather conditions (e.g., sun, night, fog, rain, and snow) in different driving scenarios (e.g., urban, suburban, highway, etc.). Specifically, Fig. 4 clearly shows the adverse impact (e.g., blur) of fog, rain, and snow on the stereo camera, while LiDAR frames show dense false point clouds due to snowflakes in snow conditions. On the other hand, radar frame quality is limited by its resolution, particularly at long distances. In the following, we overview the processing and learning algorithms for each sensor modality and the need for sensor fusion.

A. Radar

Automotive radar has been traditionally a part of ADAS for safety features such as emergency braking, adaptive cruise control, and self-parking systems. These features have been enabled using traditional frequency-modulated continuous wave (FMCW) chirp signals along with signal processing techniques such as pulse compression, low-pass filtering, analog-to-digital conversion (ADC), fast Fourier transform (FFT), Doppler processing, clutter removal, and constant false alarm rate (CFAR) detection. For each receiver RF chain, by applying FFTs on the fast-time and slow-time samples of the baseband signal as depicted in Fig. 5, the range-Doppler (RD) heatmap can be constructed with options to apply a window function to suppress sidelobes. These RD heatmaps from multiple receiver RF chains can be combined to increase signal-to-interference-and-noise ratio (SINR). Then radar detection is performed by comparing the output power of a receiving filter with a threshold. If the measured power exceeds the threshold, an object is detected and the associated signals are then processed to estimate object parameters such as range, radial velocity (or range rate), and angles. In this context, a CFAR detection or its variant is commonly used to maximize detection probability while maintaining a fixed probability of false alarm for a given SINR. One example of CFAR detection is the cell-averaging CFAR which computes the threshold from the average power of neighboring range-Doppler cells that are separated by several guard cells to the cell of interest (COI) to avoid possible object contamination and determine whether the COI contains an object (H_1 hypothesis) or not (H_0 hypothesis) by comparing the detection statistic with the threshold. Comprehensive overview of basic FMCW-based automotive radar signal processing techniques can be found in [46] and [47]. Conventionally, these automotive radar systems were designed to achieve desired resolution and maximum unambiguous limits in the range and velocity domain by optimizing waveform parameters (e.g. bandwidth, chirp period).

Beyond the safety features, current automotive radar is limited in terms of its pixel information (i.e., limited azimuth and elevation angular resolutions) to deliver high-quality perception information [46]. To meet the challenge, a new wave of chip developments eye on improving the angular resolution of automotive radar [59] to deliver LiDAR-like radar perception with enriched semantic features. Particularly, high-resolution 3D (range-velocity-azimuth)/4D (3D+elevation) radar images enable the ability to correctly detect and classify objects in the environment. For example, it is crucial to differentiate overhead objects such as bridges and traffic signs while being aware of low-lying objects such as manholes, road debris, and curbs.

Since the angular resolution is determined by the beam width that is inversely proportional to the aperture size, chip vendors take various approaches to form the beam and synthesize a large aperture. Mechanically scanned FMCW radars, e.g., Navtech CTS350-X, have been used to collect 360° bird's-eye view (BEV) radar images in the range-azimuth domain but without the Doppler velocity [15]. Assuming that the ego vehicle's motion is known, synthetic aperture radar (SAR) techniques can coherently combine returned radar waveforms to create a high-resolution two-dimensional image of the scene [60]. For instance, a 0.1° azimuth resolution was achieved for imaging static objects and can deliver $\sim 1,000,000$ points for a typical scene [61].

To achieve high angular resolution, another popular approach is to use multiple-input multiple-output (MIMO) radar [62] where multiple N_t transmitting (Tx) and N_r receiving (Rx) antennas are used to form a virtual array with $N_t N_r$ elements. The combined MIMO-FMCW automotive radar only employs $N_t + N_r$ RF chains to reduce the hardware cost. The shape of the virtual array is determined as the convolution of the transmitter array and the receiver array. To achieve this, one needs to separate corresponding waveform to each transmitter at each receiver, provided that the transmitting waveforms from different Tx antennas can be separable or orthogonal. Several orthogonal MIMO signaling schemes can be realized in time-division multiplexing (TDM), frequency-division multiplexing (FDM), and Doppler-division multiplexing (DDM) (also referred to as slow-time MIMO) modes [63]–[66]. Once the waveforms are separated, the received MIMO radar waveforms can be arranged along the fast-time (range/distance), slow-time (Doppler/velocity), azimuth (horizontal array orientation) and elevation (vertical array orientation) dimensions. Depending on the computing resources and power consumption budget, the MIMO-FMCW automotive radar may coherently process the separated MIMO waveforms in one or more dimensions at once or in a cascading fashion (e.g., range-Doppler domain first and then angular domains) with standard FFT operations or more advanced super-resolution spectrum estimation methods such as MUSIC, ESPRIT, and compressed sensing. The TDM-MIMO mode has been commercialized by various chip vendors due to relatively easy implementation and less computational requirements on the waveform separation to achieve more than 100 or even

TABLE I
OPEN AUTOMOTIVE RADAR DATASETS

Datasets	Year	Size	Annotation	Data Format	Adverse Weather
nuScenes [48]	2020	large	bounding box, Track ID	sparse points	Yes
RadarScenes [49]	2021	large	Point Annotation	sparse points	Yes
RADIATE [15]	2020	medium	bounding box, Track ID	dense points (scanning)	Yes
Oxford robotcar [50]	2020	large	object pose	dense points (scanning)	Yes
MulRan [51]	2020	large	object pose	dense points (scanning)	No
TJ4DRadSet [52]	2022	medium	bounding box, Track ID	dense points	No
VoD [53]	2022	small	bounding box, Track ID	dense points	No
CARRADA [54]	2020	medium	bounding box, Track ID	RAD heatmap	No
CRUW [55]	2021	medium	object class, position	RAD heatmap	No
RaDICal [56]	2021	large	bounding box	Raw data	No
RADDet [57]	2021	medium	bounding box	RAD heatmap	No
RADIal [58]	2021	medium	object, segmentation	RAD heatmap+points	No

$\sim 2,000$ virtual channels in the azimuth and elevation domains [63], [67], [68].

Open radar-included automotive perception datasets have begun to emerge by using commercial radar chips; see Table I. Earlier efforts focused on the collection of radar detection points for perception. nuScenes [48] is one of the first large-scale (1 million annotated frames) automotive perception datasets with commercial automotive radar included. However, due to poor angular resolution, the radar point clouds per vehicle are very sparse (e.g., less than 10). RadarScenes [49] combines detection points from 4 automotive radar sensors operating at 79 GHz to boost the number of detection points per object. For high angular-resolution automotive radar datasets, the Oxford radar robotcar [50], MulRan [51], and RADIATE [15] (see the middle row of Fig 4) datasets used mechanically scanning FMCW radar to get sub-1 degree angular resolution for BEV radar images but without the Doppler velocity information. More recent efforts focus on the radar heatmap in range-Doppler (RD), range-angle (RA), and range-angle-Doppler (RAD) domains [54]–[58]. For instance, the CARRADA dataset provides three (RA, RD and RAD) heatmaps on a scenario of test tracks [54].

Driven by the availability of high-resolution automotive radar hardware platforms and open datasets, advanced model-based signal processing and learning-based pipelines [69] have shown great potential to achieve state-of-the-art performance in radar-assisted object detection, segmentation, multi-object tracking, simultaneous localization and mapping, trajectory and behavior prediction, multi-modal sensor fusion, and scene understanding.

1) Radar Detection Point: For sparse radar detection points, model-based object detection and tracking algorithms have been considered in the context of extended object tracking (EOT) [70]. One of the key challenges for EOT is to model the spatial distribution of radar detection points over the extent of an automotive vehicle and the subsequent prediction and update of expanded state (e.g., vehicle position, orientation, speed, turn rate, length, width, etc.) using Bayesian filtering [71]–[77]. These sparse radar detection points can also be processed by heuristic distance-based clustering algorithms such as DBSCAN or PointNet-like learning algorithms. For instance, PointNet [78] and PointNet++ [79], [80] were applied to segment radar detection points and estimate 2D bounding boxes from those segmented radar points by taking into account unique radar features such as the Doppler velocity and radar cross-section (RCS) [81]. A two-branch radar point segmentation network was considered in [82] with a convolution network branch performing semantic segmentation over radar grid maps in the static environment and the other recurrent segmentation network on radar point clouds of moving objects. Then a merging step takes the output class probabilities of each cell in the grid map from the two classifiers to form point clouds.

Compared with sparse points, dense radar points from high-resolution automotive radar platforms (e.g., scanning-based, SAR or MIMO-based) may enable radar feature extraction at a level closer to LiDAR perception networks. Particularly, PointPillars was applied to the 4D radar data in a new VoD dataset [53] and a reduced performance gap in terms of object detection can be achieved between a 64-line LiDAR sensor and a high-resolution 4D radar sensor with the utilization of elevation resolution and integration of successive radar frames. [83] proposed a radar transformer that uses both vector and scalar attention mechanisms to construct attention maps over 3D (spatial, Doppler, and RCS) domains. Exploiting the temporal relation of successive radar frames can further enhance feature extraction of radar point cloud. [84] proposed a cross-attention network that exploits the consistence of objects over successive radar frames in different levels (e.g., the input level by permuting the frame order and the feature level by introducing the cross-attention feature module). These selected temporally enhanced features are then used to regress oriented bounding boxes (OBB) at each of successive radar frames, similar to the CenterPoint framework [85].

2) Radar Heatmap: Radar heatmaps may have more semantic features for low-RCS and static objects than the after-CFAR radar detection points. It might be arguable that the neural network-based feature extraction can be more representative from the heatmap domain for complex-shape objects than the model-based CFAR detection and point-based networks. For instance, the CA-CFAR detection is optimal in the Neyman-Pearson criterion if the noise and interference amplitude is Rayleigh distributed, which may not hold in practice. Moreover, the choice of guard cells is critical to avoid object contamination. Direct heatmap-based approaches, on the other hand, skip the traditional model-based CFAR detection and can directly backpropagate the training loss into the heatmap feature extraction networks. To this end, image-based backbone networks and downstream pipelines have been applied to the radar heatmap [86]. A straightforward way is to treat the 3D RAD heatmap as an RGB image but with the number of channels the same as the number of Doppler bins. Then, image-domain object detection, segmentation and tracking pipelines can be applied to the input RGB-like radar heatmap [57], [87], [88]. In [57], a one-stage anchor-based YOLO framework with a dual detection head was considered to generate both 3D RAD and 2D Cartesian bounding boxes. Evaluation on their own RADDet dataset shows a 56.3% average precision (AP) at intersection over union (IoU) of 0.3 on 3D bounding box predictions and, respectively, 51.6% at IoU of 0.5 on 2D bounding box prediction. To reduce the input dimension of a full 4D heatmap with two angular azimuth and elevation domains, the 4D RAD heatmaps can be decomposed or projected into multiple 2D heatmaps. [87] proposed a RAMP-CNN approach to bypass the complexity of 4D convolutions and to fuse extracted features from multiple 2D heatmaps. [88] takes three projected 2D "views" of the RAD heatmaps as the input and three feature extraction networks are used for each projected view. These view-dependant features are then concatenated and fed to a decoder for the segmentation task. Evaluation on the CARRADA dataset shows a mean IoU at 58.7 on the RD heatmap and 41.3% on the RA heatmap over 4 categories of pedestrian, cyclist, car and background.

The tremendous progress in the recent years, driven by the open access of high-resolution automotive radar datasets, provides a promising future for radar-based external perception. For the external perception task, achieving higher angular resolution radar detection points or heatmaps poses further strain on the cost, computational resources, and power consumption budget. Similar to the ResNet and Vision Transformer, strong and unified radar-specific feature extraction backbone networks are needed. The need of strong backbone networks also call for diverse downstream tasks such as object detection, segmentation, and tracking, or self-supervised learning without any (or with limited accuracy) annotation labels.

B. Camera

As evident from Fig. 5 and the top row of Fig. 4, camera-based images provide distinct features to separate objects of interest (e.g., vehicles, pedestrians) from the background, and such features can be integrated into state-of-the-art deep learning frameworks for downstream tasks including object detection, classification, depth regression, object association and tracking, pixel and instance segmentation [89].

Large-scale datasets involving cameras have been collected for the development of autonomous driving technologies. The details of these datasets are listed in Table II. The most widely used ones are the KITTI and BDD100K because of its early release with various sensor signals or large number of images in various traffic situations to examine the effectiveness of developed methods.

Current learning-based methods can be generally classified into two categories including two-stage detection and one-stage detection. The two-stage detection, also called region-based detection, firstly scans the complete image to find the potential regions of interest and then focuses on these regions for deeper understanding, which generally imitates the attentional mechanism of human brain [103]. The two stages for detection are respectively responsible for generating a set of proposals and making predictions for these proposals. During the proposal generation phase, a set of proposals is generated. In the prediction phase, the feature vectors of generated proposals are encoded by deep convolutional neural networks and then classifiers are used to determine the category labels of the proposals [104]. R-CNN is a pioneering two-stage object detector proposed by Girshick et al. [105]. Xie et al. [106] proposed an improved object detection approach based on R-CNN. High-quality oriented proposals were firstly generated in an almost cost-free way, and then regression and classification technologies were used for prediction. The testing results on two datasets including DOTA and HRSC2016 showed that the mean average precisions (mAPs) were 75.87% and 96.50%, respectively. Despite the advances in learning detectors based on R-CNN networks, proposal generation still relies on traditional methods such as selective search [104]. Studies [107] [108] show that CNN has a remarkable ability to locate objects in convolution layers. Therefore, the faster-CNN method is proposed in [109] by developing a region proposal network based on CNNs. [110] proposed to use faster R-CNN for object detection in rainy weather for autonomous driving. The detection results show that faster R-CNN incorporating image translation and domain adaptation performed the best among the examined methods in rainy weather.

Different from two-stage detection algorithms which divide the detection pipeline into two parts, one-stage detection assumes that each region in the image is with a possible detected object, and each region of interest is categorized into background or target object without a separate stage to generate proposals [104]. Directly mapping from image pixels to bounding boxes with category probabilities generally saves time when comparing with the two-stage methods [103]. According to the searching methods for areas of interest, the one-stage methods can be further divided into anchor-based methods and anchor-free methods. The main idea of anchor-based methods is to predict searching anchors based on prior definition. Anchor boxes with different sizes slid over each position of an image, predicting the searching anchor box as background or object based on the ground-truth pre-defined anchors. YOLO [111], as a typical anchor-based method, considered object detection as a regression problem and spatially divided the whole image into a number of grid cells. Each cell was considered as a proposal to detect the presence of objects [103]. [112] proposed three deep learning methods for pedestrian detection in haze weather based on YOLO. The evaluation results showed that the proposed method MNPrioriBoxes-Yolo with separable depthwise convolution and bottlenecks had obvious advantages with fewer parameters for pedestrian detection in haze weather. Improved lightweight detection algorithms based on YOLO (e.g., YOLOv4 and YOLOv5) further enhance the recognition ability on small objects with limited number of parameters [113].

TABLE II
OPEN-SOURCE DATASETS FOR AUTONOMOUS DRIVING

Datasets	Year	Number of images	Viewing Angle	Resolution	Camera Type	Adverse Weather
CamVid [90]	2008	18×10^3	Dashboard	960×720	Monocular	No
KITTI [91]	2012	15×10^3	Vehicle Roof	1392×512	2 Stereo	No
Cityscapes [92]	2016	25×10^3	Windshield	–	Stereo	No
Oxford RobotCar [93]	2017	19×10^3	360 Degree	1280×960 1024×1024	Stereo & Monocular	Yes
Mapillary [94]	2017	25×10^3	–	1920×1080	–	Yes
BDD100K [95]	2017	100×10^6	Windshield	720p	Video	Yes
ApolloScape [96]	2018	144×10^3	360 Degree	3384×2710	Stereo	No
CULane [97]	2018	130×10^3	Windshield	1640×590	–	No
H3D [98]	2019	25×10^3	Vehicle Roof	1920×1200	3 Monocular	–
NuScenes [48]	2019	1.4×10^6	360 Degree	1600×900	6 Monocular	Yes
Foggy [99]	2019	14×10^3	Windshield	960×1280	Stereo	Yes
Sim 10K [100]	2017	10×10^3	Windshield	–	Gaming Engine	Yes
TuSimple [101]	2017	6.5×10^3	Windshield	1280×720	–	No
RDD2020 [102]	2020	26×10^3	Windshield	600×600 720×720	Smartphone	No

The main idea of anchor-free methods is using keypoints to describe the boxes used for detection, hence the main task is transformed into keypoint detection. The related methods have two branches, namely corner-based methods and center-based methods [104]. For corner-based methods, also called multiple keypoints estimation methods [114], the confidence scores of bounding boxes are predicted through joint corner information in the feature map. Compared to RepPoints [115] using 9 keypoints, [116] used a large number of adaptive points to model objects, which achieves the state-of-the-art performance on instance segmentation tasks. Center-based methods simplify object detection to a central point detection task by estimating the probability of a pixel as the central point. Inspired by region proposal network in the anchor-based method, [114] proposed the FII-CenterNet (foreground information introduction CenterNet) method for traffic object detection based on CenterNet [117].

For improvement based on these two-stage or one-stage methods, many advanced data augmentation and deep learning technologies have been developed to help learn effective features for better prediction. Data augmentation changes characteristics of images by cropping, flipping, rotating, scaling, translating, color perturbations, and adding noise to enrich the diversity of data samples for training to learn stable features [108]. The mainly incorporated deep learning modules for performance improvement include attention mechanism, pyramid pooling, linear bottleneck and inverted residuals, depthwise separable convolution, atrous convolution, knowledge distillation, domain adaptation, SEBlock, ResBlock, mask mechanism, network pruning and quantification [110] [118] [119] [120].

Anther advanced technology used to improve the performance of learning-based algorithms in cameras is a transformer proposed in [121]. Inspired by the success of transformer in neuro-linguistic programming, transformer has also been widely used in computer vision tasks. These methods mainly include pure transformer, transformer with convolution, and self-supervised representation learning with transformer. Transformer with convolution combines transformer modules with convolutional network modules, and self-supervised representation learning uses transformer self-supervised mechanism for training. As a typical pure transformer, vision transformer [122] divides the 2D image data into image blocks as the input to the standard transformer for supervised training. [123] adopted the transformer encoder structure and the convolution module for lane detection, and the verification results showed that optimal performance was achieved in both efficiency and accuracy. Here a novel decoder with dense queries and rectified attention field was proposed, which alleviates the deficiency in pedestrian detection by using the transformer decoder DETR (DEtection TRansformer).

Most existing approaches are supervised and rely on large-scale datasets with reliable annotated labels, which is difficult to obtain especially for extreme weather and driving situations. Developing unsupervised and weakly supervised learning algorithms is a promising solution. In [124], an image-level multi-label classifier was integrated on the detection backbone to obtain sparse but critical image regions corresponding to the classification information to bridge the gaps between source and target domains. In [125], a cross-domain adaptive clustering approach was proposed by pulling into distances between peers while simultaneously pulling away distances from different categories, which achieves the state-of-the-art performance in semi-supervised domain adaptation. Although approaches have been developed for unsupervised solutions [126], [127], [128], more efforts are still needed for further improvement in this research area.

The other remaining challenges in camera-based processing and learning algorithms include: (1) Deep learning based methods are end-to-end with insufficient model interpretability. More attention to provide model insights is required in the future. (2) Most of the learning based methods are with complex networks, with high computation requirements on hardware. This makes these ML methods infeasible to implement in the current generation of ADAS. Lightweight ML technologies need to be developed for wide deployment in automotives. (3) Detection of small objects is still not satisfactory. Fusing the signals from multiple sources for multi-modal fusion may be a solution to this challenging issue.

C. LiDAR

As seen in Fig. 4, LiDAR provides a direct depth profile over the angular (azimuth and elevation) domains with a resolution even finer than the automotive radar. Most LiDAR type sensors are based on different types of scanning mechanisms that allow for the laser beams to be projected over a large field of view following specific patterns and according to the technology used. Scanning can be through mechanical spinning or solid state. The former often includes a bulky rotating module which will spin a mirror around a vertical axis and tilt it along the pitch orientation. The latter refers to a scanning system though micro-mirrors based on MEMS technology [129]. A LiDAR generates streams of 3D-points, with intensity data associated to each point (proportional to the reflected signal). Unlike video cameras, LiDARs measure the distance applying mostly the time of flight method. Photonics principles are part of the technology used to operate with light sources that need to convert signals rapidly and process them to generate the desired measurements, whilst at the same time mapping the direction and position of the beams as they scan to attain the sensor field of view. It must be considered that there is substantial processing prior to the generation of the sensor data, thus purpose-built processors are often used that need to comply with automotive operating standards (e.g. temperature, vibration, etc.).

The data streamed out from the LiDAR is interpreted through perception algorithms into hierarchical object descriptions. This process can be divided into ground segmentation, object detection, tracking, recognition and motion prediction. Previously, these phases were addressed separately using geometric and early machine learning techniques (e.g., Support Vector Machines, based on statistical learning frameworks). However, the success encountered on the use of deep learning in machine vision is also reflected on its effective use on 3D-point clouds. Deep Learning technologies can automatically extract features from the raw input in a single phase. Convolutional neural networks (CNN) and recurrent neural networks (RNN), such as long short-term memory (LSTM), are the most frequently used models. Ground segmentation can be achieved by applying CNN to LiDAR points represented by multi-channel range images [130]. Deep neural network (DNN) based solutions achieve object detection by recognition, keeping to the paradigm of supervised learning. For example, vehicles can be detected by CNN based neural networks on a bird's eye view (BEV) representation of LiDAR 3D-points [131]. A major constraint is the low density of LiDAR 3D-points at long distances, methods using CNN on the range image and BEV representation like in [132], detect mainly vehicles and no pedestrians. A compact representation of a LiDAR point cloud as a graph was proposed as a Point-GNN (graph neural network) method in [133]. A semi-supervised using temporal GNNs to leverage the rich spatio-temporal information in 3D LiDAR point cloud videos for object detection was considered in [134]. A novel approach is to integrate evidential theory into a deep learning architecture for LiDAR based road segmentation and mapping [135]. Currently, object tracking is implemented mainly using deep learning, replacing, the conventional tracking algorithm based on estimation filters [136]. A detection net will process first a sequence of LiDAR 3D-points and images to generate detection proposals. Then, tracks are estimated by finding the best detection associations. This is achieved by a marching net and scoring net.

Point-wise semantic segmentation, which was previously difficult to attain using model-based methods, is now possible using deep learning models. One of the most popular networks is PointNet which provides a unified architecture for applications ranging from object classification, part segmentation, to scene semantic parsing, directly from LiDAR 3D-point clouds [78]. As the point cloud density is increased, together with more annotated datasets, LiDAR performance for semantic segmentation should improve providing not only classes of objects but also spatial information. A major constraint for the application of these methods is the need for large datasets. However, different annotated datasets have recently emerged including the SemanticKITTI dataset [137]. It is based on the KITTI dataset and is considered one of the largest pointwise annotated dataset. Synthethic datasets like the PRESIL dataset [138] that provide labelled scenarios for particular situations are also available.

D. Fusion

Table III shows a comparison of different sensor modalities in terms of sensing and operational features. Sensing features depict sensing performance characteristics, while operational features capture robustness and system integration aspects. Typically, radars can provide range, velocity and angular information with high resolution, in comparison to visible light and infrared cameras. Cameras on their own are unreliable in situations of abrupt change in illumination, such as when entering/exiting a tunnel or extreme weather conditions. LiDARs also suffer from performance degradation in extreme weather conditions, while radars are robust under adverse weather, environmental and illumination conditions. Compared to radars or LiDARs, cameras can capture contour, texture and color information of the scene enabling excellent recognition capabilities under non-extreme environments. Although LiDARs are superior to radars in ranging accuracy and denser point cloud, their cost is much higher and have a larger form-factor making it difficult for integration in a flexible and aesthetic way. In summary, visible light cameras are superior in determining object features and hence find use in traffic scene/sign understanding, radars have the better performance in determining object motion characteristics with high resolution and low cost, while LiDARs have superior ranging performance with a wide detection coverage. However each sensor modality also has limitations, with no single modality providing the needed sensing and perception functionalities.

The aim of sensor fusion is the collectively processing of inputs from various modalities to perceive and derive interpretations with defined level of certainty about the environment around the vehicle. Based on the discussion in earlier sections and depicted in Fig. 4 and Table III, it is clear that each individual sensor cannot work independently under all scenarios and deliver accurate information with precision

		Visible Light Camera	Infrared Camera	Radar	LiDAR
Sensing Features	Distance measurement	Medium	Medium	Very High	High
	Velocity measurement	Low	Low	High	Low
	Angle measurement	Low	Low	Medium	High
	Measurement resolution	High	Low	High	High
	Object Features	Color & contour	—	Intensity	Intensity
	Field of View	Medium	Medium	Medium	360°
	Sampling Rate	High	Medium	Medium	Low
Operational Features	Weather (rain, snow, fog)	Vulnerable	Vulnerable	Robust	Vulnerable
	Visibility (dust, smoke)	Vulnerable	Vulnerable	Robust	Vulnerable
	Illumination (low-light, glare)	Vulnerable	Robust	Robust	Robust
	Sensor Interference	None	None	Yes	Yes
	Processing Requirements	Medium	Medium	Medium	Medium
	Sensor Layout & Aesthetics	Good	Medium	Good	Low
	Costs	Low-Medium	High	Low	High

TABLE III
COMPARISON OF SENSORS FOR EXTERNAL PERCEPTION.

required to operate an autonomous vehicle with the highest degree of safety. Sensor fusion allows information from all sensors to be fused meaningfully to extract the best of all sensors while offsetting the disadvantages of an individual modality.

For ADAS/ADS, it is important that the sensor fusion architecture combines data or processed data at different meaningful stages of the pipeline. To enable perception functions, there are three fundamental sensor fusion approaches to associate and integrate data across modalities to enable an informed decision [139], [140], [141].

- Late fusion: Each sensor is operating individually and then the processed data, i.e. likelihood function, gets fused at the end to make a collective decision for the system. In [142], a multi-modal vehicle detection system employing late fusion strategy was proposed combining optical image and 3-D LiDAR detections. Individual modalities have their own detection pipeline, and then the detection information is fused via a joint re-scoring and non-maximum suppression, and demonstrated improved detection performance over single modality detection.
- Early fusion: Sensor fusion happens at the initial data stage with no to minimal data pre-processing to align and normalize the raw data. The fused data is collectively used to improve detection, classification, segmentation and monitoring of the objects. In [143], early fusion of LiDAR and camera data into a multi-dimensional occupation grid representation as input to fully convolutional networks for lane detection was proposed.
- Mid-level (or cross) fusion: This fusion approach combines the early and late fusion approaches. Targeted information derived from different sensors are fused at an initial data stage given that a certain predefined criterion is fulfilled, while other target information are fused at higher levels under other pre-defined criteria, such as low signal-to-noise ratio conditions. In [144], radar detections were associated to preliminary detection results obtained from a camera image, and then generates radar feature maps in addition to image features to estimate 3D object bounding boxes. Camera and LiDAR features are fused in a shared bird's eye view space in [145] showing improved mAP for 3D object detection in comparison to [144] and individual sensors in adverse weather and illumination conditions.

V. VALIDATION METHODOLOGIES AND SAFETY CONSIDERATIONS

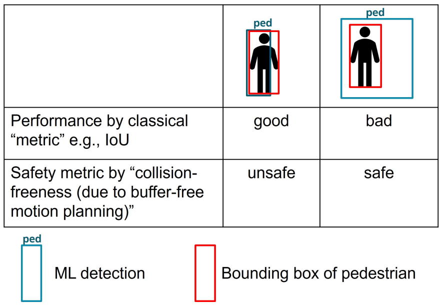
Evaluating the performance of a perception system is a complex task. It involves regulation concerning software to ensure safety, defining system performance metrics and designing methodologies for robust ML perception systems.

A. *Safety standards and guidelines for perception*

Currently, ISO 26262 [146] defines processes and measures for the functional safety of systems including one or more electrical and/or electronic systems. The requirements in ISO 26262 are considered to be sufficient to deal with risks due to random hardware faults or classic systematic software faults (e.g., array-out-of-bounds) for sensors.

While a system involving ML components may (ideally) be free of hardware or systematic software errors as governed by ISO 26262 (functional safety), the performance limitations of ML (functional insufficiencies) within the Operational Design Domain (ODD) can still lead to risks. ISO 21448 [147] focuses on processes and measures to ensure the absence of unreasonable risk due to a hazard caused by functional insufficiencies, where the performance limitation of sensors is also explicitly mentioned as one of the sources of functional insufficiencies. The basic safety specification considers the occurrence of an error pattern [148] within the ODD being sufficiently low, commonly reflected as a probability term. While the appendix of the ISO 21448 covers some high-level aspects and some of the process-oriented results aim at offering a general argumentation framework [148]–[152], we consider the key technical challenge to be the "implementation aspects" of such a process.

ANSI/UL 4600 [153] is a standard to promote a proper consideration of safety issues for generic autonomous systems, and specifically adopt autonomous vehicles as a concrete case. Specifically for perception, the standard describes how an



	ped	ped
Performance by classical "metric" e.g., IoU	good	bad
Safety metric by "collision-freeness (due to buffer-free motion planning)"	unsafe	safe

Fig. 6. Illustration where standard performance metrics may not be directly used in safety-critical driving contexts.

acceptable perception system can only be achieved after a proper ODD definition. The ODD determines a perception ontology (objects and events for perception functionality) and the challenges encountered (e.g., seasonal effects). Thus, a perception module's performance is defined by its ability to map sensor data to the ontology.

B. Performance evaluation and safety-aware metrics

State-of-the-art perception algorithms are often compared on benchmark datasets such as [48], [91], [154]. Leaderboards in categories such as 3D Camera-Only Detection, LiDAR segmentation are maintained by developers and researchers employing a set of metrics, often derived from metrics in computer vision literature. Standard metrics are based on Intersection over Union (IoU), precision, recall (both at pixel/point level or object level) with some aggregation function applied (e.g., mAP on averaging over all considered prediction classes). Moreover, tracking is a crucial step for perception, so tracking metrics such as Multiple Object Tracking Accuracy/Multiple Object Tracking Precision are also employed [155]. These standard metrics are designed to measure the average difference between specific features of the ground truth and the perception output. However, the resulting safety is much more relevant than perception accuracy. In fact, we can observe the seemingly subtle fact that optimizing a DNN following standard performance metrics does not necessarily imply that the DNN produces a safe prediction. One intuition can be observed from the fact that for autonomous driving, pedestrians that are distant should not be equally weighted compared to pedestrians being close by. Another visual example is shown in [156] where the standard IoU metric for bounding box detection will indicate that within Figure 6, the prediction in the left image is better than the right. However, when safety is defined at the ML-level to be "completely covering the object" as any region outside the bounding box is considered to be an empty space, the right image, although "worse" in terms of IoU, is safe. To be used in safety-critical ADAS, the developed metrics need to be connected to concrete performance limitations and driving applications, and the acceptance threshold should be justified. Despite recent developments [157]–[160] in developing safety-aware metrics by including various factors such as reaction time or imperfection of the labeling, these metrics are not direct reflectors of safety. To be used in a concrete safety case, one needs to fine-tune these metrics by matching the definition of acceptably safe defined in a concrete application. A similar phenomenon also occurs in associating the degree of robustness related to safety. Within the field of machine learning, researchers formulate robustness using concepts such as L_∞ norms, characterizing the minimum amount of input change that maintains prediction consistency. Nevertheless, deciding the required minimum robustness bound of an ML model in a given application with convincing rationale (so that the noise to the ML model will not be the source of harm) is far from trivial.

Moreover, detection, tracking, and scene segmentation are only sub-tasks of the overall perception algorithm required in ADAS. This makes the connection between safety and perception evaluation challenging to be tackled with summarized metrics. Vehicles are deployed in uncontrolled traffic, and the same perception error (e.g., misclassification), may be irrelevant in some situations while crucial in others.

Another reason for this limitation is the decision and planning module (refer to Fig. 1), which is crucial in determining how the ADAS reacts to the perception output. The same perception error can be a safety issue or not depending on the driving style of the ADAS-equipped vehicle. For example, a vehicle driving at a higher speed may demand better perception at longer distances. At the same time, a slower vehicle could achieve adequate safety even with sensors with a shorter operating range. Thus, an end-to-end testing step is necessary to evaluate the perception quality while accounting for both traffic scenarios and the decision and planning module. On this topic, Piazzoni et al. propose perception error models to test the impact of specific perception errors of safety (e.g., detection accuracy over time, tracking-loss probability) via virtual testing and scripted scenarios [161]. This approach requires accurate modeling of the perception errors that affect the perception algorithm under test [162], [163].

C. Data considerations and ML systems

A public dataset for automotive perception is usually a combination of sensor data collected by a vehicle on the road with annotated ground truth, typically 2D/3D bounding boxes of obstacles and road users. Each data collection campaign generates a dataset with unique limitations and features, as shown in Table I. Moreover, each dataset may include a different set of labels and detected objects. Perception datasets are expensive, with inaccurate ground truth, and limited in nature. The high cost of data collection results from the need to drive on the road with sensor-equipped vehicles, which is either expensive in terms of time or amount of vehicles. Alongside the cost of collecting data, there is also a constant need for new data collection campaigns that employ more recent sensor hardware or firmware releases.

As part of data collection, data needs to be labeled with ground truth. Common solutions are manual labeling that suffer from high cost and possible interpretation errors, or the application of offline algorithms which offer higher reliability than online algorithms by exploiting causal information. A data collection campaign can only collect and label a limited amount of data. Besides planned limitations (e.g., type of sensor used, location, and time of the day), features such as weather conditions and traffic events cannot be controlled. Since the traffic environment is not controlled, the accuracy of automated labeling techniques may be limited. Furthermore, data reflecting life-threatening events (e.g., close-to-collision) may not be easily obtained.

D. Virtual testing and sensor simulation

To overcome issues with real-world testing, synthetic data can be obtained in virtual environments [100], [164]–[166], e.g. using photo-realistic simulation or emulating humans and motion behaviors. This approach has a few key advantages. Firstly, it is much less time-consuming and resource intensive, as no vehicle is driven on the road in uncontrolled traffic. Additionally, virtual environments can offer perfect ground truth values, and every simulation aspect can be controlled (e.g., weather conditions or traffic events). Thus, this approach can provide a huge amount of labeled data, which ML algorithms could exploit.

However, the use of synthetic data unavoidably raises the problem of domain gap, where one can not ensure the performance demonstrated on synthetic data can be faithfully transferred to the real world. Model training for domain adaptation is under active research (e.g., [167]–[169]). For demonstrating diversity, it is also related to ODD, where one should develop methods to have a systematic understanding regarding how data is collected. Evidence based on combinatorial testing is based on characterizing the ODD, followed by ensuring that the collected data set can have a reasonable amount of data for any arbitrary pair of criteria. The idea has been applied in the ML setup [150], [170], [171] for highway and urban autonomous driving.

Moreover, the standard ANSI/UL 4600 [153] recommends using virtual environments for autonomous vehicle testing, for both Hardware-in-the-Loop and Software-in-the-loop modalities. The employment of virtual simulators is common practice [165], [166], [172]–[174]). Figure 7 illustrates a typical architecture for a co-simulation loop. The simulator handles the objects in the scene (e.g., traffic vehicles and pedestrians) and employs sensor models to generate synthetic data. The vehicle stack process the synthetic data and determine the response, which is then sent to the simulator. Most simulators also include complex vehicle dynamics, road maps, and tools to script traffic scenarios.

Along with individual strengths and weaknesses, most simulators share the advantage of offering safe, scalable, controlled, and scriptable test solutions. However, a major challenge is their fidelity, i.e., their ability to provide results that are representative of real-life situations. This aspect is particularly relevant for perception, as virtual environments have to generate synthetic signals (e.g., images and point clouds) to feed the autonomous vehicle stack. Thus, a high-fidelity sensor model requires accurate modeling of materials, physical properties, and effects of weather conditions [175], [176].

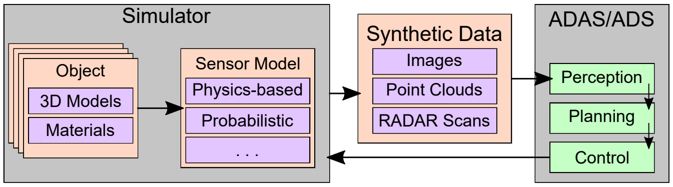


Fig. 7. Typical architecture of virtual testing involving sensor simulation.

E. Monitoring against abnormal situations

Finally, as the ML model is only expected to function in the ODD, the ML system should have the mechanism to notify if the current input is “outside the ODD”. Practically, the specified operational domain can be ambiguous (e.g., autonomous highway driving in Germany), but the ML function can still produce incorrect outputs due to errors in generalization. This problem is mediated by building up a monitor to “detect unknowns in runtime” by various means. As research in this field is still under active development, we only discuss some representative ideas.

The simplest method, similar to the work of Hendrycks et al. [177], is to use softmax in the output for proxying the likelihood of being a specific class. Then a warning is raised when the prediction is not particularly strong for every class. The ODIN approach utilizes the softmax classification but uses the temperature scaling [178] to perform uncertainty calibration. Apart from interpreting the output values, another direction considers the inspection of features within intermediate layers. The work of Lee et al. [179] assumes that feature vectors of intermediate layers produced by the training data are approximately Gaussian-distributed. The authors use the Mahanalobis distance as a confidence score for adversarial or out-of-distribution (OoD) detection. Extending this line of thoughts, recent work also aims at training the DNN to allow directly outputting uncertainty (the DUQ network) [180], where for classification, the network is trained in a way such that within the feature space, each class has a representative vector. In operation, the training input is translated into the feature vector; the classification and the uncertainty are based on the distance to all class-representative vectors. The application to object detection is reflected in CertainNet [181], where the classification is done similarly to DUQ. However, for regression, the estimation is based on computing the variance of all nearby grids having overlapping predictions over the same object. Yet another possibility is to consider OoD detectors built using abstraction-based approaches [182]–[184], where DNN-generated feature vectors from the training dataset are clustered and enclosed using hyperrectangles. Note that input outside the training data distribution may not imply that it is not in the operational design domain.

Uncertainty can also be measured using redundancy and majority voting. This leads to Bayesian approaches such as drop out at runtime [185] or ensemble learning. Deep Ensembles [186] achieve state-of-the-art uncertainty estimation but at a large computational overhead (since one needs to train many models by taking different random seeds), thus recent work attempts to mitigate this with various ideas [187], [188] such as parameter sharing in earlier layers for different models. Beyond direct measurement, the recent work on evidential learning [189] aims to learn parameters in a higher-order setup, where the learned parameters are used to uncover the distribution of the uncertainty model.

Apart from technical challenges, a crucial implementation-related challenge for monitors lies in the imbalanced data for "unknowns" or input outside the ODD. Many of the presented techniques here use in-distribution data in a self-supervised fashion, and use out-of-distribution data for calibrating the boundary. However, to perform such a calibration and as well as to test the applicability, one only has "known unknowns" collected before the development.

VI. Future outlook

The selection of sensor modalities, sensing and perception algorithms, data processing and multi-sensor fusion architectures, to deliver robust, high-quality perception for ADAS/ADS vehicles will remain a topic of interest. Perception systems need to provide high levels of functional performance for safe driving under diverse ODDs during the entire operational lifecycle. It is expected that perception can support high levels of autonomous driving functions in restricted ODDs, e.g., on well-mapped highways in good weather conditions. Validation of sensing and perception methods, especially based on machine learning methods, can thus be done in different ODDs to ensure safety and incremental adoption of autonomous features.

Although sensor fusion strategies leveraging supervised algorithms are able to mitigate the shortcomings in an individual sensing modality, these are still not perfect. Utilizing reinforcement learning paradigms in conjunction with supervised learning algorithms within the sensor fusion context could assist in scenario-based learning. Furthermore, reinforcement learning algorithms can be used to assess the risk of failure of the sensor fusion solution early on and facilitate human intervention.

Besides local fusion discussed in Section IV-D, cooperative perception is another approach to enhance the capabilities of local perception sensors by sharing information among vehicles, or by communicating with the infrastructure. Vehicle connectivity is a means to enable such information sharing. The role of connectivity in extending automotive perception capabilities is however outside the scope of this paper - the reader is referred to [190], [191]. Connectivity also enables updates and thus improvements in ML models whilst offering new perception-driven services [192], [193]. However it also brings additional challenges like transformation errors between the different reference frames, delays, uncertainty with respect to the shared information, security and trust, that need to be addressed.

It is expected that ML-driven automotive sensing and perception will lead to "better than human" capabilities like having obstacle information over a 360° field-of-view. However, the response of ML components cannot be guaranteed by traditional system engineering and software validation approaches. Aspects like explainability and reproducibility of ML processing become critical in ADAS/ADS and, given their safety-critical nature, remain a challenge [194], [195].

One of the challenges in reliable automotive sensing and perception information across diverse ODDs is the limited availability of data in difficult driving and weather conditions, diverse in-cabin conditions, and along the sensor operational lifecycle. The availability of quality sensor datasets is crucial to avoid issues like class imbalance in training ML models. As discussed in the earlier sections, the validation of automotive perception systems is a substantial challenge. Traditional computer vision metrics used in automotive perception are not context-aware; there is a need to consider safety-awareness and ADAS/ADS applications. Due to the lack of large real-world datasets, there is a need to use synthetic data. For this purpose, the development of virtual simulators could lead to more effective and efficient ways of end-to-end perception system and ADAS/ADS testing. Improving their fidelity via more realistic sensor models can provide better synthetic data for ML training.

With greater driving autonomy, there is also a concern that over-reliance on machine-based decisions may lead to bad driving habits and increased driving distractions. This brings about the need for human-vehicle interaction mechanisms that enable humans to take over driving operations on time by overriding autonomous systems when necessary. New decision-making and control designs that leverage both automotive-external and automotive-internal perception information whilst taking into account deployment differences are needed.

Research on automotive perception technologies to support the holy grail of fully autonomous driving should address synergy between technology and the fields of ethical, legal and social sciences. Greater collaboration among diverse disciplines like hardware and software reliability, algorithm designs, safety and quality engineering, security and privacy, human-machine designs, insurance and legal, is required for perception based ADAS/ADS designs to be successfully deployed at scale to provide enhanced safety and driving comfort.

References

[1] E. Marti, M. A. de Miguel, F. Garcia, and J. Perez, "A review of sensor technologies for perception in automated driving," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 4, pp. 94–108, 2019.

[2] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 33–55, 2016.

[3] Y. Chen, Y. Wang, F. Qu, and W. Li, "A graph-based track-before-detect algorithm for automotive radar target detection," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6587–6599, 2021.

[4] O. Bialer, A. Jonas, and T. Tirer, "Super resolution wide aperture automotive radar," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17 846–17 858, 2021.

[5] Y. Sun, T. Fei, and N. Pohl, "A high-resolution framework for range-doppler frequency estimation in automotive radar systems," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11 346–11 358, 2019.

[6] H. Lian, X. Pei, and X. Guo, "A local environment model based on multi-sensor perception for intelligent vehicles," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15 427–15 436, 2021.

[7] R. Ravindran, M. J. Santora, and M. M. Jamali, "Camera, lidar, and radar sensor fusion based on bayesian neural network (clr-bnn)," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 6964–6974, 2022.

[8] D. Kang and D. Kum, "Camera and radar sensor fusion for robust vehicle localization via vehicle part localization," *IEEE Access*, vol. 8, pp. 75 223–75 236, 2020.

[9] A. Venon, Y. Dupuis, P. Vasseur, and P. Merriaux, "Millimeter wave fmcw radars for perception, recognition and localization in automotive applications: A survey," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 533–555, 2022.

[10] Federal Communications Commission, "Radar Services in the 76-81 GHz Band; Report and Order – ET Docket No. 15-26," *FCC-CIRC1707-07*, 2017.

[11] International Telecommunications Union - Radiocommunication Sector, "Systems characteristics of automotive radars operating in the frequency band 76-81 GHz for intelligent transport systems applications," *Recommendation ITU-R M.2057-1*, 2018.

[12] H. Cho, "Operational design domain (odd) framework for driver-automation integrated systems," *Massachusetts Institute of Technology, Ph.D. Thesis*, 2020.

[13] National Highway Traffic Safety Administration, "New car assessment program," *Docket No. NHTSA–2021–0002*, 2022. [Online]. Available: https://www.govinfo.gov/content/pkg/FR-2022-03-09/pdf/2022-04894.pdf

[14] European New Car Assessment Programme, "Child presence detection test and assessment protocol v1.1," 2023. [Online]. Available: https://cdn.euroncap.com/media/75474/euro-ncap-cpd-test-and-assessment-protocol-v11.pdf

[15] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A radar dataset for automotive perception," *arXiv preprint arXiv:2010.09076*, 2020.

[16] A. Mishra, S. Lee, D. Kim, and S. Kim, "In-cabin monitoring system for autonomous vehicles," *Sensors*, vol. 22, no. 12, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/12/4360

[17] E. Yavari, P. Nuti, and O. Boric-Lubecke, "Occupancy detection using radar noise floor," in *IEEE/ACES International Conference on Wireless Information Technology and Systems and Applied Computational Electromagnetics*, 2016, pp. 1–3.

[18] A. Lazaro, M. Lazaro, R. Villarino, and D. Girbau, "Seat-occupancy detection system and breathing rate monitoring based on a low-cost mm-wave radar at 60 ghz," *IEEE Access*, vol. 9, pp. 115 403–115 414, 2021.

[19] J. h. Huh and S. h. Cho, "Seat belt reminder system in vehicle using ir-uwb radar," in *International Conference on Network Infrastructure and Digital Content*. IEEE, 2018, pp. 256–259.

[20] M. Hoffmann, D. Tatarinov, J. Landwehr, and A. R. Diewald, "A four-channel radar system for rear seat occupancy detection in the 24 ghz ism band," in *2018 11th German Microwave Conference*. IEEE, 2018, pp. 95–98.

[21] A. R. Diewald, J. Landwehr, D. Tatarinov, P. Di Mario Cola, C. Watgen, C. Mica, M. Lu-Dac, P. Larsen, O. Gomez, and T. Goniva, "Rf-based child occupation detection in the vehicle interior," in *International Radar Symposium*. IEEE, 2016, pp. 1–4.

[22] A. Caddemi and E. Cardillo, "Automotive anti-abandon systems: a millimeter-wave radar sensor for the detection of child presence," in *14th International Conference on Advanced Technologies, Systems and Services in Telecommunications*. IEEE, 2019, pp. 94–97.

[23] H. Abedi, S. Luo, V. Mazumdar, M. M. Y. R. Riad, and G. Shaker, "Ai-powered in-vehicle passenger monitoring using low-cost mm-wave radar," *IEEE Access*, vol. 10, pp. 18 998–19 012, 2022.

[24] S. Pisa, E. Pittella, and E. Piuzzi, "A survey of radar systems for medical applications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 11, pp. 64–81, 2016.

[25] S. M. Islam, O. Boric-Lubecke, V. M. Lubecke, A.-K. Moadi, and A. E. Fathy, "Contactless radar-based sensors: Recent advances in vital-signs monitoring of multiple subjects," *IEEE Microwave Magazine*, vol. 23, no. 7, pp. 47–60, 2022.

[26] G. Paterniani, D. Sgreccia, A. Davoli, G. Guerzoni, P. Di Viesti, A. Valenti, M. Vitolo, G. Vitetta, and B. Giuseppe, "Radar-based monitoring of vital signs: A tutorial overview," 03 2022.

[27] R. Chae, A. Wang, and C. Li, "Fmcw radar driver head motion monitoring based on doppler spectrogram and range-doppler evolution," in *IEEE Topical Conference on Wireless Sensors and Sensor Networks*, 2019, pp. 1–4.

[28] K. Staszek, K. Wincza, and S. Gruszczynski, "Driver's drowsiness monitoring system utilizing microwave doppler sensor," in *2012 19th International Conference on Microwaves, Radar & Wireless Communications*, vol. 2, 2012, pp. 623–626.

[29] Y. Kim, "Detection of eye blinking using doppler sensor with principal component analysis," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 123–126, 2015.

[30] K. Yamamoto, K. Toyoda, and T. Ohtsuki, "Doppler sensor-based blink duration estimation by analysis of eyelids closing and opening behavior on spectrogram," *IEEE Access*, vol. 7, pp. 42 726–42 734, 2019.

[31] E. Cardillo, G. Sapienza, C. Li, and A. Caddemi, "Head motion and eyes blinking detection: a mm-wave radar for assisting people with neurodegenerative disorders," in *European Microwave Conference*, 2021, pp. 925–928.

[32] C.-H. Tseng, J.-R. Lin, C.-L. Lin, Y.-C. Wu, and L.-T. Huang, "The prototype of a driver attention level monitoring system: The sanbao radar," in *IEEE International Conference on Consumer Electronics-Taiwan*, 2018, pp. 1–2.

[33] X. Gu, L. Zhang, Y. Xiao, H. Zhang, H. Hong, and X. Zhu, "Non-contact fatigue driving detection using cw doppler radar," in *IEEE MTT-S International Wireless Symposium*, 2018, pp. 1–3.

[34] G. Ciattaglia, S. Spinsante, and E. Gambi, "Slow-time mmwave radar vibrometry for drowsiness detection," in *IEEE International Workshop on Metrology for Automotive*, 2021, pp. 141–146.

[35] Z. Dong, M. Zhang, J. Sun, T. Cao, R. Liu, Q. Wang, and Danliu, "A fatigue driving detection method based on frequency modulated continuous wave radar," in *IEEE International Conference on Consumer Electronics and Computer Engineering*, 2021, pp. 670–675.

[36] S. Ishizaka, K. Yamamoto, and T. Ohtsuki, "Non-contact blood pressure measurement using doppler radar based on waveform analysis by lstm," in *IEEE International Conference on Communications*, 2021, pp. 1–6.

[37] A. E. Omer, S. Safavi-Naeini, R. Hughson, and G. Shaker, "Blood glucose level monitoring using an fmcw millimeter-wave radar sensor," *Remote Sensing*, vol. 12, no. 3, 2020.

[38] K. Yamamoto, K. Toyoda, and T. Ohtsuki, "Spectrogram-based simultaneous heartbeat and blink detection using doppler sensor," in *IEEE International Conference on Communications*, 2019, pp. 1–6.

[39] J. Holzrichter and L. Ng, "Speech articulator and user gesture measurements using micropower, interferometric em-sensors," in *IEEE Instrumentation and Measurement Technology Conference*, vol. 3, 2001, pp. 1942–1946 vol.3.

[40] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum," in *29th Annual Symposium on User Interface Software and Technology*. ACM, 2016, pp. 851–860.

[41] M. M. Rahman, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, and S. Z. Gurbuz, "Effect of kinematics and fluency in adversarial synthetic data generation for asl recognition with rf sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 4, pp. 2732–2745, 2022.

[42] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range fmcw monopulse radar for hand-gesture sensing," in *IEEE Radar Conference*, 2015, pp. 1491–1496.

[43] B. B. James, "Recognizing traffic signalling gestures through automotive sensors." M.S. Thesis, Dept. Electrical and Computer Engineering, Mississippi State University, 2022.

[44] K. A. Smith, C. Csech, D. Murdoch, and G. Shaker, "Gesture recognition using mm-wave sensor for human-car interface," *IEEE Sensors Letters*, vol. 2, no. 2, pp. 1–4, 2018.

[45] M. Raja, V. Ghaderi, and S. Sigg, "Wibot! in-vehicle behaviour and gesture recognition using wireless network edge," in *IEEE International Conference on Distributed Computing Systems*, 2018, pp. 376–387.

[46] G. L. Charvat, *Small and Short-Range Radar Systems*, 1st ed. USA: CRC Press, Inc., 2014.

[47] H. Rohling and M.-M. Meinecke, "Waveform design principles for automotive radar systems," in *2001 CIE International Conference on Radar Proceedings (Cat No.01TH8559)*, 2001, pp. 1–4.

[48] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 621–11 631.

[49] O. Schumann, M. Hahn, N. Scheiner, F. Weishaupt, J. F. Tilly, J. Dickmann, and C. Wöhler, "Radarscenes: A real-world radar point cloud data set for automotive applications," *arXiv preprint arXiv:2104.02493*, 2021.

[50] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *International Conference on Robotics and Automation*, 2020, pp. 6433–6438.

[51] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, "Mulran: Multimodal range dataset for urban place recognition," in *International Conference on Robotics and Automation*, 2020, pp. 6246–6253.

[52] L. Zheng, Z. Ma, X. Zhu, B. Tan, S. Li, K. Long, W. Sun, S. Chen, L. Zhang, M. Wan, L. Huang, and J. Bai, "TJ4DRadSet: A 4D radar dataset for autonomous driving," *ArXiv*, vol. abs/2204.13483, 2022.

[53] A. Palffy, E. Pool, S. Baratam, J. F. P. Kooij, and D. M. Gavrila, "Multi-class road user detection with 3+1D radar in the view-of-delft dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4961–4968, 2022.

[54] A. Ouaknine, A. Newson, J. Rebut, F. Tupin, and P. Pérez, "Carrada dataset: camera and automotive radar with range-angle-doppler annotations," in *25th International Conference on Pattern Recognition*, 2021, pp. 5068–5075.

[55] Y. Wang, Z. Jiang, Y. Li, J.-N. Hwang, G. Xing, and H. Liu, "Rodnet: A real-time radar object detection network cross-supervised by camera-radar fused object 3d localization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 954–967, 2021.

[56] T.-Y. Lim, S. A. Markowitz, and M. N. Do, "RaDICaL: A synchronized FMCW radar, depth, IMU and RGB camera data dataset with low-level FMCW radar signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 941–953, 2021.

[57] A. Zhang, F. E. Nowruzi, and R. Laganiere, "RADDet: Range-azimuth-doppler based radar object detection for dynamic road users," in *18th Conference on Robots and Vision*, 2021, pp. 95–102.

[58] J. Rebut, A. Ouaknine, W. Malik, and P. P'erez, "Raw high-definition radar for multi-task learning," *ArXiv*, vol. abs/2112.10646, 2021.

[59] K. Doris, A. Filippi, and F. Jansen, "Reframing fast-chirp fmcw transceivers for future automotive radar: The pathway to higher resolution," *IEEE Solid-State Circuits Magazine*, vol. 14, no. 2, pp. 44–55, 2022.

[60] A. Laribi, M. Hahn, J. Dickmann, and C. Waldschmidt, "Performance investigation of automotive SAR imaging," in *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility*, 2018, pp. 1–4.

[61] M. Mostajabi, C. M. Wang, D. Ranjan, and G. Hsyu, "High resolution radar dataset for semi-supervised learning of dynamic objects," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (Workshop Contribution)*, 2020, pp. 450–457.

[62] J. Li and P. Stoica, *MIMO Radar Signal Processing*. John Wiley & Sons, 2008.

[63] S. Rao, "White paper: MIMO radar," *Texas Instruments (TI) Technical Report SWRA554A*, 2017.

[64] P. Wang, P. Boufounos, H. Mansour, and P. V. Orlik, "Slow-time MIMO-FMCW automotive radar detection with imperfect waveform separation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 8634–8638.

[65] S. Sun, A. P. Petropulu, and H. V. Poor, "MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98–117, 2020.

[66] F. Jansen, "Automotive radar doppler division mimo with velocity ambiguity resolving capabilities," in *2019 16th European Radar Conference (EuRAD)*, 2019, pp. 245–248.

[67] A. Och, C. Pfeffer, J. Schrattenecker, S. Schuster, and R. Weigel, "A scalable 77 ghz massive MIMO FMCW radar by cascading fully-integrated transceivers," in *2018 Asia-Pacific Microwave Conference (APMC)*, 2018, pp. 1235–1237.

[68] "4D Imaging Radar: The World's First $2K$ Ultra-High Resolution Radar Platform," https://arberobotics.com/wp-content/uploads/2021/05/4D-Imaging-radar-product-overview.pdf, 2021, [Online; accessed September-27-2022].

[69] A. Santra and S. Hazra, *Deep learning applications of short-range radars*. Artech House, 2020.

[70] K. Granström, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview, and applications," *Journal of Advances in Information Fusion*, vol. 12, no. 2, 2017.

[71] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, 2008.

[72] U. Orguner, "A variational measurement update for extended target tracking with random matrices," *IEEE Trans. on Signal Processing*, vol. 60, no. 7, pp. 3827–3834, 2012.

[73] M. Baum and U. D. Hanebeck, "Extended object tracking with random hypersurface models," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 149–159, 2014.

[74] N. Wahlström and E. Özkan, "Extended target tracking using Gaussian processes," *IEEE Trans. on Signal Processing*, vol. 63, no. 16, pp. 4165–4178, 2015.

[75] P. Broßeit, B. Duraisamy, and J. Dickmann, "The volcanormal density for radar-based extended target tracking," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–6.

[76] K. Granström, M. Fatemi, and L. Svensson, "Poisson multi-Bernoulli mixture conjugate prior for multiple extended target filtering," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 56, pp. 208–225, 2019.

[77] Y. Xia, P. Wang, K. Berntorp, L. Svensson, K. Granström, H. Mansour, P. Boufounos, and P. V. Orlik, "Learning-based extended object tracking using hierarchical truncation measurement model with automotive radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 1013–1029, 2021.

[78] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 652–660.

[79] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *31st International Conference on Neural Information Processing Systems*, 2017, p. 5105–5114.

[80] J. Liu, W. Xiong, L. Bai, Y. Xia, T. Huang, W. Ouyang, and B. Zhu, "Deep instance segmentation with automotive radar detection points," *IEEE Transactions on Intelligent Vehicles*, pp. 1–1, 2022.

[81] O. Schumann, M. Hahn, J. Dickmann, and C. Wöhler, "Semantic segmentation on radar point clouds," in *International Conference on Information Fusion*, 2018, pp. 2179–2186.

[82] O. Schumann, J. Lombacher, M. Hahn, C. Wöhler, and J. Dickmann, "Scene understanding with automotive radar," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 188–203, 2020.

[83] J. Bai, L. Zheng, S. Li, B. Tan, S. Chen, and L. Huang, "Radar transformer: An object classification network based on 4d MMW imaging radar," *Sensors*, vol. 21, no. 11, 2021.

[84] P. Li, P. Wang, K. Berntorp, and H. Liu, "Exploiting temporal relations on radar perception for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 071–17 080.

[85] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *CoRR*, vol. abs/1904.07850, 2019. [Online]. Available: http://arxiv.org/abs/1904.07850

[86] S. Gurbuz, Ed., *Deep Neural Network Design for Radar Applications*. IET/Sci Tech Publishers.

[87] X. Gao, G. Xing, S. Roy, and H. Liu, "RAMP-CNN: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5119–5132, 2021.

[88] A. Ouaknine, A. Newson, P. Pérez, F. Tupin, and J. Rebut, "Multi-view radar semantic segmentation," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 671–15 680.

[89] R. Qian, X. Lai, and X. Li, "3D Object Detection for Autonomous Driving: A Survey," *Pattern Recognition*, vol. 130, p. 108796, 2022.

[90] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *European Conference on Computer Vision*. Springer, 2008, pp. 44–57.

[91] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[92] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.

[93] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[94] G. Neuhold, T. Ollmann, S. Rota Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *IEEE/CVF International Conference on Computer Vision*, 2017, pp. 4990–4999.

[95] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving video database with scalable annotation tooling," *arXiv preprint arXiv:1805.04687*, vol. 2, no. 5, p. 6, 2018.

[96] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, "The apolloscape open dataset for autonomous driving and its application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2702–2719, 2019.

[97] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in *AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[98] A. Patil, S. Malla, H. Gang, and Y.-T. Chen, "The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes," in *IEEE International Conference on Robotics and Automation*, 2019, pp. 9552–9557.

[99] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, no. 9, pp. 973–992, 2018.

[100] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, "Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?" in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Press, 2017, p. 746–753.

[101] TuSimple, "Tusimple lane detection benchmark," https://github.com/TuSimple/tusimple-benchmark, 2017.

[102] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "Rdd2020: An annotated image dataset for automatic road damage detection using deep learning," *Data in brief*, vol. 36, p. 107133, 2021.

[103] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[104] X. Wu, D. Sahoo, and S. C. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.

[105] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[106] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented r-cnn for object detection," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3520–3529.

[107] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2921–2929.

[108] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *International journal of computer vision*, vol. 128, no. 2, pp. 261–318, 2020.

[109] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[110] M. Hnewa and H. Radha, "Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 53–67, 2020.

[111] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[112] G. Li, Y. Yang, and X. Qu, "Deep learning approaches on pedestrian detection in hazy weather," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8889–8899, 2019.

[113] Y. Cai, T. Luan, H. Gao, H. Wang, L. Chen, Y. Li, M. A. Sotelo, and Z. Li, "YOLOv4-5D: An effective and efficient object detector for autonomous driving," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[114] S. Fan, F. Zhu, S. Chen, H. Zhang, B. Tian, Y. Lv, and F.-Y. Wang, "Fii-centernet: an anchor-free detector with foreground attention for traffic object detection," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 121–132, 2021.

[115] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, "Reppoints: Point set representation for object detection," in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9657–9666.

[116] Z. Yang, Y. Xu, H. Xue, Z. Zhang, R. Urtasun, L. Wang, S. Lin, and H. Hu, "Dense reppoints: Representing visual objects with dense point sets," in *European Conference on Computer Vision*. Springer, 2020, pp. 227–244.

[117] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "Centernet: Keypoint triplets for object detection," in *IEEE/CVF international conference on computer vision*, 2019, pp. 6569–6578.

[118] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, pp. 1–20, 2023.

[119] G. Li, Z. Ji, and X. Qu, "Stepwise domain adaptation (sda) for object detection in autonomous vehicles using an adaptive centernet," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[120] G. Li, Z. Ji, X. Qu, R. Zhou, and D. Cao, "Cross-domain object detection for autonomous driving: A stepwise domain adaptative yolo approach," *IEEE Transactions on Intelligent Vehicles*, 2022.

[121] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[122] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[123] L. Liu, X. Chen, S. Zhu, and P. Tan, "Condlanenet: a top-to-down lane detection framework based on conditional convolution," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3773–3782.

[124] C.-D. Xu, X.-R. Zhao, X. Jin, and X.-S. Wei, "Exploring categorical regularization for domain adaptive object detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 724–11 733.

[125] J. Li, G. Li, Y. Shi, and Y. Yu, "Cross-domain adaptive clustering for semi-supervised domain adaptation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2505–2514.

[126] G. Wang, C. Zhang, H. Wang, J. Wang, Y. Wang, and X. Wang, "Unsupervised learning of depth, optical flow and pose with occlusion from 3d geometry," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 308–320, 2020.

[127] W. Kim, A. Kanezaki, and M. Tanaka, "Unsupervised learning of image segmentation based on differentiable feature clustering," *IEEE Transactions on Image Processing*, vol. 29, pp. 8055–8068, 2020.

[128] H. Fan, P. Liu, M. Xu, and Y. Yang, "Unsupervised visual representation learning via dual-level progressive similar instance selection," *Ieee transactions on cybernetics*, 2021.

[129] Y. Li and J. Ibanez-Guzman, "LiDAR for autonomous driving: The principles, challenges, and trends for automotive LiDAR and perception systems," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, 2020.

[130] M. Velas, S. Michal, H. Michal, and H. Adam, "CNN for Very Fast Ground Segmentation in Velodyne LiDAR Data," in *IEEE International Conference on Autonomous Robot Systems and Competitions*, 2018, pp. 97–103.

[131] D. Nachiket and M. M. Trivedi, "Convolutional Social Pooling for Vehicle Trajectory prediction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (Workshop Contribution)*, 2018, pp. 1468–1476.

[132] X. Chen, M. Huimin, W. Ji, L. Bo, and X. Tian, "Multi-view 3d object detection network for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1907–1915.

[133] W. Shi and R. Rajkumar, "Point-gnn: Graph neural network for 3d object detection in a point cloud," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[134] J. Wang, H. Gang, S. Ancha, Y.-T. Chen, and D. Held, "Semi-supervised 3d object detection via temporal graph neural networks," in *2021 International Conference on 3D Vision (3DV)*, 2021, pp. 413–422.

[135] E. Capellier, F. Davoine, V. Cherfaoui, and Y. Li, "Transformation-adversarial network for road detection in LiDAR rings, and model-free evidential road grid mapping," in *11th Workshop on Planning, Perception, Navigation for Intelligent Vehicles*, 2019, pp. 47–52.

[136] D. Frossard and R. Urtasun, "End-to-end learning of multi-sensor 3d tracking by detection," in *International Conference on Robotics and Automation*, 2018, pp. 635–642.

[137] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "Semantickitti: A dataset for semantic scene understanding of LiDAR sequences," in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9297–9307.

[138] B. Hurl, K. Czarnecki, and S. Waslander, "Precise synthetic image and lidar (presil) dataset for autonomous vehicle perception," in *IEEE Intelligent Vehicles Symposium*, 2019, pp. 2522–2529.

[139] M. Haberjahn and K. Kozempel, "Multi level fusion of competitive sensors for automotive environment perception," in *Proceedings of the 16th International Conference on Information Fusion*, 2013, pp. 397–403.

[140] M. Aeberhard and T. Bertram, "Object classification in a high-level sensor data fusion architecture for advanced driver assistance systems," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 416–422.

[141] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object

detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *Trans. Intell. Transport. Syst.*, vol. 22, no. 3, p. 1341–1360, mar 2021.

[142] A. Asvadi, L. Garrote, C. Premebida, P. Peixoto, and U. J. Nunes, "Multimodal vehicle detection: fusing 3d-lidar and color camera data," *Pattern Recognition Letters*, vol. 115, pp. 20–29, 2018.

[143] F. Wulff, B. Schäufele, O. Sawade, D. Becker, B. Henke, and I. Radusch, "Early fusion of camera and lidar for robust road detection based on u-net fcn," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1426–1431.

[144] R. Nabati and H. Qi, "Centerfusion: Center-based radar and camera fusion for 3d object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 1527–1536.

[145] Z. Liu, H. Tang, A. Amini, X. Yang, H. Mao, D. Rus, and S. Han, "BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[146] "ISO 26262:2018 road vehicles — functional safety," https://www.iso.org/standard/68383.html, 2018.

[147] "ISO 21448:2022 road vehicles - safety of the intended functionality," https://www.iso.org/standard/77490.html, 2022.

[148] R. Salay, K. Czarnecki, H. Kuwajima, H. Yasuoka, T. Nakae, V. Abdelzad, C. Huang, M. Kahn, and V. D. Nguyen, "The missing link: Developing a safety case for perception components in automated driving," *CoRR*, vol. abs/2108.13294, 2021.

[149] S. Burton, L. Gauerhof, and C. Heinzemann, "Making the case for safety of machine learning in highly automated driving," in *International Workshop on Assurance Cases for Software-Intensive Systems*. Springer, 2017, pp. 5–16.

[150] C.-H. Cheng, C.-H. Huang, and G. Nührenberg, "nn-dependability-kit: Engineering neural networks for safety-critical autonomous driving systems," in *IEEE/ACM International Conference on Computer-Aided Design*, 2019, pp. 1–6.

[151] X. Zhao, A. Banks, J. Sharp, V. Robu, D. Flynn, M. Fisher, and X. Huang, "A safety framework for critical systems utilising deep neural networks," in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2020, pp. 244–259.

[152] Y. Jia, T. Lawton, J. McDermid, E. Rojas, and I. Habli, "A framework for assurance of medication safety using machine learning," *arXiv preprint arXiv:2101.05620*, 2021.

[153] ANSI/UL 4600, Standard for Safety for Evaluation of Autonomous Products. [Online]. Available: https://ulse.org/UL4600

[154] Waymo Open Dataset: An autonomous driving dataset. [Online]. Available: https://www.waymo.com/open

[155] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.

[156] C.-H. Cheng, T. Schuster, and S. Burton, "Logically sound arguments for the effectiveness of ml safety measures," in *International Conference on Computer Safety, Reliability, and Security (Workshop Contribution)*. Springer, 2022, pp. 343–350.

[157] C.-H. Cheng, C.-H. Huang, H. Ruess, H. Yasuoka *et al.*, "Towards dependability metrics for neural networks," in *ACM/IEEE International Conference on Formal Methods and Models for System Design*, 2018, pp. 1–4.

[158] M. Lyssenko, C. Gladisch, C. Heinzemann, M. Woehrle, and R. Triebel, "From evaluation to verification: Towards task-oriented relevance metrics for pedestrian detection in safety-critical domains," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (Workshop contribution)*, 2021, pp. 38–45.

[159] G. Volk, J. Gamerdinger, A. von Bernuth, and O. Bringmann, "A comprehensive safety metric to evaluate perception in autonomous systems," in *IEEE International Conference on Intelligent Transportation Systems*, 2020, pp. 1–8.

[160] C.-H. Cheng, A. Knoll, and H.-C. Liao, "Safety metrics for semantic segmentation in autonomous driving," in *IEEE International Conference on Artificial Intelligence Testing*, 2021, pp. 57–64.

[161] A. Piazzoni, J. Cherian, M. Slavik, and J. Dauwels, "Modeling perception errors towards robust decision making in autonomous vehicles," in *International Joint Conference on Artificial Intelligence*, 7 2020, pp. 3494–3500, main track.

[162] M. Hoss, M. Scholtes, and L. Eckstein, "A review of testing object-based environment perception for safe automated driving," *Automotive Innovation*, pp. 1–28, 2022.

[163] J. Sadeghi, B. Rogers, J. Gunn, T. Saunders, S. Samangooei, P. K. Dokania, and J. Redford, "A step towards efficient evaluation of complex perception tasks in simulation," *arXiv preprint arXiv:2110.02739*, 2021.

[164] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *European Conference on Computer Vision*. Springer, 2016, pp. 102–118.

[165] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[166] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta, E. Agafonov, T. H. Kim, E. Sterner, K. Ushiroda, M. Reyes, D. Zelenkovsky, and S. Kim, "LGSVL Simulator: A High Fidelity Simulator for Autonomous Driving," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE Press, 2020, p. 1–6.

[167] L. Yang, X. Liang, T. Wang, and E. Xing, "Real-to-virtual domain unification for end-to-end autonomous driving," in *European Conference on Computer Vision*. Springer, 2018, pp. 530–545.

[168] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "Dada: Depth-aware domain adaptation in semantic segmentation," in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7364–7373.

[169] A. Robey, G. J. Pappas, and H. Hassani, "Model-based domain generalization," *Advances in Neural Information Processing Systems*, vol. 34, pp. 20 210–20 229, 2021.

[170] C.-H. Cheng, C.-H. Huang, and H. Yasuoka, "Quantitative projection coverage for testing ML-enabled autonomous systems," in *International Symposium on Automated Technology for Verification and Analysis*. Springer, 2018, pp. 126–142.

[171] C. Gladisch, C. Heinzemann, M. Herrmann, and M. Woehrle, "Leveraging combinatorial testing for safety-critical computer vision datasets," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (Workshop Contribution)*, 2020, pp. 324–325.

[172] MSC, "Vires VTD, https://vires.mscsoftware.com," 2022.

[173] IPG, "IPG CarMaker, https://ipg-automotive.com/en/products-solutions/software/carmaker/," 2022.

[174] Siemens, "Siemens Simcenter Prescan, https://plm.automation.siemens.com/global/en/products/simcenter/prescan.html," 2022.

[175] J. Zhao, Y. Li, B. Zhu, W. Deng, and B. Sun, "Method and applications of lidar modeling for virtual testing of intelligent vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 2990–3000, 2020.

[176] C. Linnhoff, P. Rosenberger, and H. Winner, "Refining object-based lidar sensor modeling — challenging ray tracing as the magic bullet," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 238–24 245, 2021.

[177] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *International Conference on Learning Representations*, 2017. [Online]. Available: https://openreview.net/forum?id=Hkg4TI9xl

[178] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=H1VGkIxRZ

[179] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Conference on Neural Information Processing Systems*, vol. 31, 2018.

[180] J. Van Amersfoort, L. Smith, Y. W. Teh, and Y. Gal, "Uncertainty estimation using a single deep deterministic neural network," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9690–9700.

[181] S. Gasperini, J. Haug, M.-A. N. Mahani, A. Marcos-Ramiro, N. Navab, B. Busam, and F. Tombari, "Certainnet: Sampling-free uncertainty estimation for object detection," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 698–705, 2021.

[182] T. A. Henzinger, A. Lukina, and C. Schilling, "Outside the box: Abstraction-based monitoring of neural networks," in *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, ser. Frontiers in Artificial Intelligence and Applications, G. D. Giacomo, A. Catalá, B. Dilkina, M. Milano, S. Barro, A. Bugarín, and J. Lang, Eds., vol. 325. IOS Press, 2020, pp. 2433–2440.

[183] C.-H. Cheng, C.-H. Huang, T. Brunner, and V. Hashemi, "Towards safety verification of direct perception neural networks," in *Design, Automation & Test in Europe Conference & Exhibition*. IEEE, 2020, pp. 1640–1643.

[184] C. Wu, Y. Falcone, and S. Bensalem, “Customizable reference runtime monitoring of neural networks using resolution boxes,” *arXiv preprint arXiv:2104.14435*, 2021.

[185] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in *International Conference on Machine Learning*. PMLR, 2016, pp. 1050–1059.

[186] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS’17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6405–6416.

[187] M. Dusenberry, G. Jerfel, Y. Wen, Y. Ma, J. Snoek, K. Heller, B. Lakshminarayanan, and D. Tran, “Efficient and scalable bayesian neural nets with rank-1 factors,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 2782–2792.

[188] M. Havasi, R. Jenatton, S. Fort, J. Z. Liu, J. Snoek, B. Lakshminarayanan, A. M. Dai, and D. Tran, “Training independent subnetworks for robust prediction,” in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=OGg9XnKxFAH

[189] M. Sensoy, L. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” *Advances in neural information processing systems*, vol. 31, 2018.

[190] A. Caillot, S. Ouerghi, P. Vasseur, R. Boutteau, and Y. Dupuis, “Survey on cooperative perception in an automotive context,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 204–14 223, 2022.

[191] B. Häfner, V. Bajpai, J. Ott, and G. A. Schmitt, “A survey on cooperative architectures and maneuvers for connected and automated vehicles,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 380–403, 2022.

[192] S. Wang, C. Li, Q. Hao, C. Xu, D. W. K. Ng, Y. C. Eldar, and H. V. Poor, “Federated deep learning meets autonomous vehicle perception: Design and verification,” *arXiv preprint arXiv:2206.01748*, 2022.

[193] A. C. Marosi, R. Lovas, A. Kisari, and E. Simonyi, “A novel iot platform for the era of connected cars,” in *IEEE International Conference on Future IoT Technologies*, 2018, pp. 1–11.

[194] A.-M. Leventi-Peetz and T. Östreich, “Deep learning reproducibility and explainable ai (xai),” *arXiv preprint arXiv:2202.11452*, 2022.

[195] B. Li, P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi, and B. Zhou, “Trustworthy AI: From Principles to Practices,” *ACM Comput. Surv.*, vol. 55, no. 9, jan 2023.