INTEGRATION, the VLSI journal 78 (2021) 36–48

Contents lists available at ScienceDirect

Integration

journal homepage: www.elsevier.com/locate/vlsi

Vulnerable objects detection for autonomous driving: A review

Esraa Khatab [a,b,*], Ahmed Onsy [a], Martin Varley [a], Ahmed Abouelfarag [b]

[a] University of Central Lancashire, Preston, PR12HE, UK
[b] Arab Academy for Science, Technology, and Maritime Transport, Alexandria, 21923, Egypt

ARTICLE INFO

Keywords:
Autonomous driving
Objects detection
Sensor fusion
Deep learning

ABSTRACT

Object detection performed by Autonomous Vehicles (AV)s is a crucial operation that comes ahead of various autonomous driving tasks, such as object tracking, trajectories estimation, and collision avoidance. Dynamic road elements (pedestrians, cyclists, vehicles) impose a greater challenge due to their continuously changing location and behaviour. This paper presents a comprehensive review of the state-of-the-art object detection technologies focusing on both the sensory systems and algorithms used. It begins with a brief introduction on the autonomous driving operations and challenges. Then, different sensory systems employed on existing AVs are elaborated while illustrating their advantages, limitations and applications. Also, sensory systems employed by different research are reviewed. Moreover, due to the significant role Deep Neural Networks (DNN)s are playing in object detection tasks, different DNN-based networks are also highlighted. Afterwards, previous research on dynamic objects detection performed by AVs are reviewed in tabular forms. Finally, a conclusion summarizes the outcomes of the review and suggests future work towards the development of vehicles with higher automation levels.

1. Introduction

AUTONOMOUS Vehicles (AV)s, also known as self-driving or driverless vehicles, have the potential to be a real game-changer on the UK's roads. They offer mobility to a broader range of people, gives passengers extra free time during their journey (the average driver in England spends 235 h driving every year [1]), reduces emissions, eases congestions, and most importantly, helps improve road safety [1]. Since the early 1990s, autonomous driving has attracted many research fields. Thus; several highly automated driver assistance capabilities have reached mass production.

Accidents statistics show that 76% of all accidents were based solely on human error, while 94% have the human error factor involved, as reported by the National Highway [1]. Some of the leading causes of collisions on roads are: getting distracted, being in a hurry, misjudging other road users' movements [2].

The autonomous driving operation can be summarized in the following steps [3,4]:

- Environment perception; by detection, localization and tracking of road elements
- Ego-vehicle self-localization
- Trajectories estimation and path planning
- Controlling the vehicle

In this paper, we will be focusing on the first step, environment perception, due to its strong impact on the following steps.

1.1. Automation levels

Due to the different terminologies describing autonomous driving, the Society of Automotive Engineers (SAE) has established a ranking of vehicles [5]:

- *Level 0*: no automation
- *Level 1*: either longitudinal or lateral control, but not both
- *Level 2*: both longitudinal and lateral control capabilities
- *Levels 3*: Object and Event Detection and Response (OEDR), while the human override is mandatory whenever the vehicle is unable to execute a particular task
- *Level 4*: vehicles are entirely autonomous in most situations, but there is still an option to switch to manual driving in challenging Operational Design Domains (ODD)s

* Corresponding author. University of Central Lancashire, Preston, PR12HE, UK.
E-mail address: eahkhatab@uclan.ac.uk (E. Khatab).

https://doi.org/10.1016/j.vlsi.2021.01.002
Received 8 March 2020; Received in revised form 7 December 2020; Accepted 1 January 2021
Available online 8 January 2021

Table 1
Level 4 and 5 concept cars.

Level 4	Level 5
• Symbioz	• Mercedes Benz S-Class
• Yandex Taxi	• VW Sedric
• Volvo 360c	• Rinspeed Snap
• Ford Fusion	• Rinspeed Oasis
• Rolls Royce 103EX	• Rinspeed Microsnap
• Chrysler Pacificas	• Rinspeed ∑tos
• Toyota Edge	

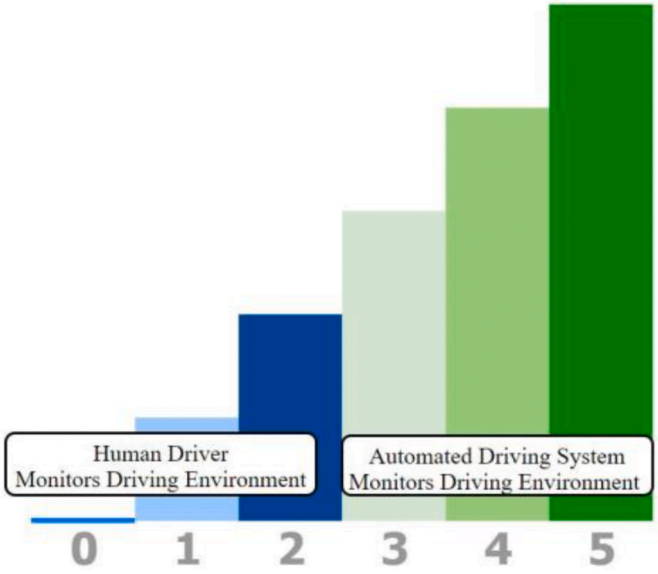

Fig. 1. SAE J3016 AUTOMOTIVE AUTOMATION STANDARD [5].

- *Level 5*: automation under any ODD, the vehicle does not feature any typical driving controls (ex: steering wheels, brake pedals, etc.)

For an easier illustration, the UK's Centre for Connected and Autonomous Vehicles described level 0–2 as "Hands-on Assisted Driving", level 3 as "Hands-off, eyes-on (the road)", and levels 4 and 5 as "Hands-off, Eyes-off" [6].

1.2. Competitions encouraging autonomous driving

Many competitions have been held in order to encourage the improvement of autonomous vehicle industry; the first was the European PROMETHEUS project, its route was from Munich to Odense in 1995 [7,8]. Also, one of the most popular challenges was organized by the Defense Advanced Research Projects Agency (DARPA) [9–11]. A summary of AVs projects and challenges was elaborated in Ref. [4]. The latest ongoing challenge is the AutoDrive Challenge sponsored by the SAE and General Motors (GM) [12], its goal is to navigate in an urban driving environment in a Level 4 automation as described by the SAE standard (J3016) [5].

1.3. Autonomous driving in industry

In addition to the academic field, car manufacturers also adopted the development of autonomous vehicles [13]. Consequently, this research field has witnessed significant technological advancements [14–17]. Many automotive manufacturers have targeted their advancements towards the autonomous capabilities of vehicles. Since 2013, Volvo has been developing the "Drive Me" project and have successfully developed Level 3 AVs. In 2017, Volvo launched a large-scale trial of autonomous driving carried out by customers on real roads. Also, in 2017, Audi released a Level 3 AV, it was the world's first production vehicle explicitly developed for conditional automated driving at level 3. The highest autonomy level AVs are developed by Waymo which also manufactures a suite of self-driving hardware. On the other hand, all new Tesla cars are manufactured with hardware capable of providing autopilot capabilities and allows for the incremental introduction of self-driving features via software updates; therefore, it is considered somewhere between level 2 and 3 autonomy. It is also worth noting that all AVs manufacturers use the same suite of sensors; except Tesla, as they do not use a Light Detection and Ranging (LiDAR) sensor (LiDAR technology will be discussed in Section 2).

Although no commercially available level 5 AVs exist, all manufacturers are working towards this goal, while making use of the same sensors (except Tesla, as they do not use a LiDAR). Table 1 shows the concept cars that manufacturers are aiming to produce in the next couple of years.

1.4. Challenges facing the spread of autonomous driving

Despite the great advancements in the autonomous driving field, many concerns are hindering its spread [18,19] (see Fig. 1). Fig. 2 unfolds a taxonomy of the different challenges affecting the widespread and operation of AVs in our streets. However, in this paper, we focus on the challenges related to the environment perception task (see Fig. 3).

Many survey papers have discussed different AVs sensor technologies [3,4,20–27], while [28–33] discussed different object detection approaches. Others focused on motion planning techniques [34]. However, in this paper we focus on the AVs' task of vulnerable road elements detection, while focusing on the closely related aspects: the sensor hardware and software techniques. This paper establishes a

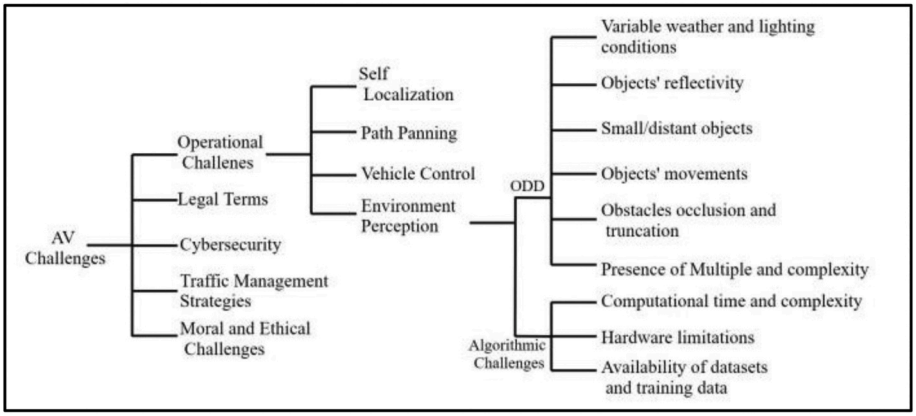

Fig. 2. Taxonomy of challenges facing the widespread and the operation of AVs.

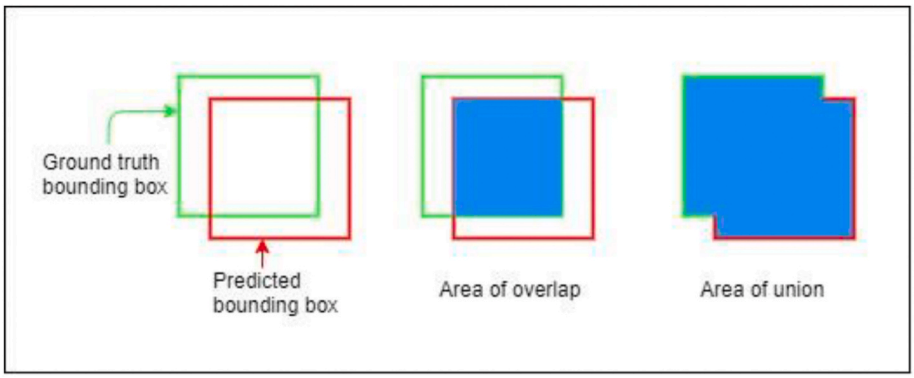

Fig. 3. IoU calculations.

Table 2
Automotive sensing categories.

Sensory system	Definition	Sensors
Self-sensing	Estimating the current state of the ego-vehicle: velocity, yaw angle, acceleration, and steering angle, using proprioceptive sensors and a Controller Area Network Bus	Proprioceptive sensors: • Odometers • IMUs • Gyroscopes
Localization	Estimating the current location of the vehicle	• External sensors: GNSS • Dead reckoning: IMUs
Environment Perception	Perceiving external data such as road markings, obstacles, traffic signs, etc.	Exteroceptive sensors: • Camera • RADAR • LiDAR • Ultrasonic sensors

baseline for further research on AVs' path planning and collision avoidance techniques. Besides, the discussion on the sensory system and detection techniques is complemented with tabular surveys on the existing research that employ these techniques on AVs.

The rest of this paper is organized as follows: Section 2 discusses the different sensory hardware used on existing AVs and different sensor fusion combinations adopted, Section 3 discusses object detection categories, available evaluation datasets, and a concise review on the state-of-the-arts object detection approaches based on Deep Neural Networks (DNN)s. Finally, the conclusion and suggestions for future work are presented in Section 0.

2. Sensory systems

The first step in autonomous driving is environment perception via the appropriate sensory system. Sensory systems fall into three main categories [35], as shown in Table 2.

Sensors can also be classified into active and passive sensors [28,36]. Active sensors emit electromagnetic waves and measure the backscatter reflected back to them, such as Ultrasonic, Radio Detection and Ranging (RADAR), and LiDAR. However, passive sensors do not have their own source of energy/signals, but they only perceive electromagnetic waves already existing in the environment, such as infrared- and light-based cameras [37–48], and acoustic systems [49–51].

2.1. Exteroceptive sensors

As mentioned in Table 2, Exteroceptive sensors are responsible for environment sensing and perception. Table 3 shows a concise overview of the most popular exteroceptive sensors used on autonomous vehicles,

Table 3
Sensors, advantages, limitations, applications.

Sensor	Advantages	Limitation	Application
Monocular camera	• Low cost • Different Fields of view • High-resolution cameras provide longer range • Provide features data	• High computational requirements • Does not provide straightforward distance calculations • Limited by weather and lighting conditions • Cannot calculate the velocity of objects	• Object detection & classification • Traffic signals recognition • Road and lane detection
Stereo-camera	In addition to the advantages of monocular cameras, stereo-cameras also provide: • Depth calculation • 3D-localization of objects • Enhanced object detection	• More expensive than monocular cameras • Higher computational requirements • Limited by weather and lighting conditions • Cannot calculate the velocity of objects	• Object detection and classification • 3D localization • 3D mapping
Short-range RADAR	• Large Field of View • Easier to develop • Resistant to bad weather	• Large package size • Shorter sensing range	• Blindspot detection • Parking aid
Long-range RADAR	• Higher accuracy • Better resolution • Smaller package size	• More data losses • Narrow Field of View at short distances	• Speed calculation of detected vehicles • Used on highways and cross-traffic alert systems
Ultrasonic	• Direct distance estimation • Can operate in harsh weather conditions • Can detect near objects (<2 m)	• Can only detect near objects • Low angular resolution	• Parking assistance • Near object detection

along with the advantages, limitations, and applications of each of them. In this section, we discuss their employment on different AVs' operation: when used solely and when integrated together.

2.1.1. Camera

Two main Camera categories can be used on outdoors AVs:

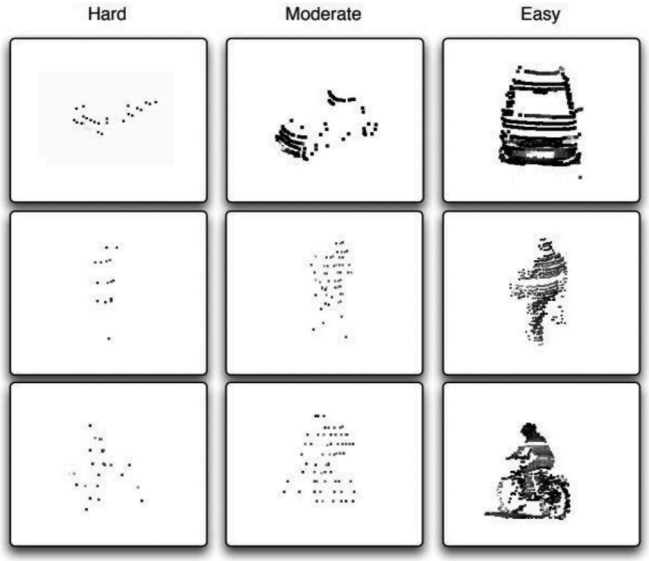

Fig. 4. Examples of labelled object instances from the training set of different difficulties. Left column: hard, instances comtaining number of measurements $m < 50$. Middle column: moderate, instances containing number of measurements $50 \leq m < 150$. Right column: easy, instances containing number of measurements $m \geq 150$.

Table 4
Different deep learning approaches for object detection.

Approach		Speed (FPS)	Testing dataset	mAP (%)
CNN	R–CNN [107]	6	VOC	53.3
	Fast R–CNN [109]	0.5	VOC	70
	Faster R–CNN (VGG16) [110]	7	VOC	73.2
SSD	SSD300 [111]	59	VOC	74.3
	SSD512 [111]	22	VOC	76.8
YOLO	YOLOv1 [112]	45	VOC	63.4
	YOLOv1-Tiny [121]	155	VOC	52.7
	YOLOv2 [120]	67	VOC	76.8
	YOLOv2 [120]	40	VOC	78.6
	YOLOv2 [121]	40	COCO	48.1
	YOLOv2-Tiny [120,121]	244	COCO	23.7
	YOLOv3 [121,135]	51	COCO	57.9
	YOLOv3-Tiny [121]	220	COCO	33.1

- *Monocular cameras*: These are the basic Camera sensors used for object detection in Ref. [47]. Also, in Ref. [52], multiple monocular cameras were used in multi-object tracking. Various algorithms have used RGB images to perform object detection and localization [53,54]. However, the results suffered from relatively low accuracy regarding depth estimation, especially at longer ranges.
- *Stereo-cameras*: They have been used in Ref. [55] to perform road surfaces condition monitoring, also as an extension to work proposed in Ref. [54], Chen et al. have proposed 3D object proposals using stereo imagery in Ref. [56]. Moreover, in Refs. [57,58], authors have achieved 3D object detection in RGB-D images demonstrating the enormous potential of 3D deep learning to perform 3D detections. Another significant boost to stereo-vision applications was the work proposed in Refs. [59,60], authors have converted data gathered from a stereo-camera into a 3D point cloud to mimic the 3D point cloud gathered by a 3D LiDAR, in this way different existing LiDAR-based detection algorithms could be applied.

2.1.2. LiDAR

It uses the Time of Flight (ToF) principle in detecting the distance between the sensor and the detected object. LiDARs can be classified into two main types: 2D and 3D; it depends on their construction and the number of laser beams projected. 2D LiDARs project a single laser beam on a rotating mirror, while 3D LiDARs use multiple laser diodes that rotate at a very high speed, the higher the number of laser diodes, the more accurate the perception becomes. 2D LiDARs can be used in 3D mapping and reconstruction by mounting them on rotating motors [61]; also multiple 2D LiDARs have been used in vehicle detection [62], pedestrian detection [63,64]. On the other hand, 3D LiDARs are used in 3D object detection [65,66]. Also, in Ref. [67] authors used a sensory system that consists of only 2D LiDARs in order to estimate both the 6D pose of the system and the surrounding's 3D structures. Examples of 2D LiDARs are LRS-1000, LMS-291 and UTM-30LX, while examples of 3D LiDARs are IBEO LUX, Velodyne, and Quanergy.

2.1.3. RADAR

It features exceptional robustness against weather and lighting conditions, multi-depth detection, long-range detection, as used in Refs. [68,69]. Also, in Ref. [70], a compact automotive RADAR has been used to evaluate its implementation for basic vehicle odometry, road structure mapping and moving object tracking.

2.1.4. Ultrasonic

It also uses the ToF principle by measuring the time of flight of sonic waves until its echo is received back. They are robust sensors that can measure distances to objects regardless of their color, material, weather conditions. However, they are only used for short distance measurements and in parking systems, therefore, they can not be used alone for objects detection.

2.2. Proprioceptive sensors

In order to achieve proper path planning of the ego vehicle, its state must be monitored using its proprioceptive sensors: Global Navigation Satellite System GNSS, Inertial Measurement Units (IMU), and wheel odometry. The GNSS is a satellite-based navigation system which provides AVs with context and online information. Most of the available GNSS sensors used in AVs have an IMU included; when integrated with other sensors, the AV gains the capability of measuring velocity, Euler angles, angular velocity, etc., which allows for ego-motion correction. A popular example of GNSS/IMU is the X-sens MTi-G. Additionally, wheel odometry tracks the wheel rotation rates and estimate the speed and acceleration of the ego vehicle [66]. Advanced Driver Assistance Systems (ADAS) developers believe that even level 3 AVs need high-precision proprioceptive sensors integrated; for example, proprioceptive sensors provide an AV with its accurate location and path to follow even if the lane view is blocked.

2.3. Sensor fusion

Using a single type of sensors has proven to be inefficient and unreliable; therefore, sensor fusion is mandatory in order to overcome the limitations of single sensors using the advantages of multiple sensors. As a result, sensor fusion enhances the reliability, accuracy, and reduces the uncertainty of measurements. Sensor fusion can be classified according to:

- *Relationship between input data sources*, as proposed by Durrant-Whyte [71]: (a) Complementary, (b) Redundant, and (c) Co-operative
- *Types of input and output data*, as proposed by Dasarathy [72]: (a) Data-In-Data-Out, (b) Data-In-Feature-Out, (c) Feature-In-Feature-Out, (d) Feature-In-Decision-Out, and (e) Decision-In-Decision-Out.

Table 5
CNN-BASED OBJECT DETECTION NETWORKS.

Network	Approach	Time Analysis	Hardware	Input	Testing Dataset	Accuracy
3DVP [123]	Translation of point clouds into voxels or image stacks using CNNs	Not noted	Not noted	RGB images	KITTI (training & testing), OutdoorScene (testing)	• 3D AP on KITTI (car: 87.46%, 75.77%, 65.38%) • AOS (86.92%, 74.59%, 64.11%) • AP on OutdoorScene (car: 90.0%, 76.5%, 62.1%)
VoxNet [124]	Volumetric occupancy grid representation with a supervised 3D CNN	Real-time, 2 ms for around 2000 points	GPU Tesla K40	• LiDAR point cloud • RGBD images • CAD data	LiDAR data: Sydney Urban Objects	Average Accuracy: • ModelNet10: 0.92 • ModelNet40: 0.83 • NYUv2: 0.71
MV3D [77]	Translation of point clouds into voxels or images stacks using CNNs	0.36s inference time for one image	TitanX GPU	• LiDAR point cloud • RGB images	KITTI	3D detection AP (56.56%–96.52%) ranging from easy 0.25 IoU to a hard 0.7 IoU
Vote3Deep [125]	Translation of point clouds into voxels or image stacks using CNN	Not noted	4-core 2.5 GHz CPU	• LiDAR point cloud • RGB images	KITTI	2D AP (easy, moderate, hard): • Car: 76.79%, 68.24%, 63.23% • Pedestrian: 68.39%, 55.37%, 52.59% • Cyclists: 79.92%,67.88%, 62.98%
PointNet [126]	Direct point cloud processing using multi-layer perceptrons (CNNs)	Not noted	GPU 1080X	• LiDAR point cloud	Stanford 3D dataset, ModelNet40	• ModelNet40: mA (86.2%), OA (89.2%) • Stanford dataset: 3D detections mAP 24.24%
Frustum [75]	Direct point cloud processing using multi-layer perceptrons (CNNs)	4 FPS	GPU NVIDIA GTX 1080i	• LiDAR point cloud • RGB images	• KITTI • SUN-RGBD	• mAP on SUN-RGBD: 54% • 3D AP on KITTI (easy, moderate, hard): o Car: 81.2%,70.39%,62.19% o Pedestrian: 51.21%, 44.89%, 40.23% o Cyclists: 71.96%, 56.77%, 50.39%
PointFusion [127]	• Image data: Faster R–CNN • Raw point data: PointNet	Not noted	Not noted	• LiDAR point cloud • RGB images	• KITTI • SUN-RGBD	• mAP on SUN-RGBD: 45.38% • 3D AP on KITTI (easy, moderate, hard): o Car: 77.92%, 63%, 53.27% o Pedestrian: 33.36%, 28.04%, 23.38% • Cyclists: 49.34%, 29.42%, 26.98%
VoxelNet [128]	Translation of point clouds into voxels or image stacks using CNNs	33 ms inference time	TitanX GPU	• LiDAR point cloud	• KITTI	3D AP (easy, moderate, hard): • Car: 77.47%, 65.11%, 57.73% • Pedestrian: 39.48%, 33.69%, 31.51% • Cyclists: 64%, 52.18%, 46.61%
AVOD [78]	Combined fusion approaches, Faster R–CNN	Real-time, 0.1s per frame	GPU TitanXP	• LiDAR point cloud • RGB camera	• KITTI	3D AP (easy, moderate, hard): • Car: 81.94%, 71.88%, 66.38% • Pedestrian: 50.8%, 42.81%, 40.88% Cyclists: 64%, 52.18%, 46.61%
PointCNN [129]	Direct point cloud processing using multi-layer perceptrons	0.031/0.012 s per batch for training/ inference of batch size 16	GPU NVIDIA Tesla P100	LiDAR point clouds	• ModelNet40, ScanNet	• ModelNet40: o Pre-aligned: mA (88.8%), OA (92.5%) o Unaligned: mA (88.1%), OA (92.2%) o ScanNet mA (55.7%), OA (79.7%)
PSMNet [130]	Spatial Pyramid pooling and 3D CNN	Real-time	Not noted	• LiDAR point cloud • RGB images	• Real driving conditions in inter-urban scenarios	92.03% positive detections, 0.59 misses per frame

- *Different Abstraction levels*, as proposed by Luo et al. [73]: (a) Signal-level, (b) Pixel-level, (c) Characteristic-level, and (d) Symbol-level
- *Architecture type* [74]: (a) Centralized, (b) Decentralized, and (c) Distributed

Different sensors data can be fused, below are examples of the most widely applied sensor fusion combinations:

2.3.1. Camera and LiDAR fusion

The fusion between a camera and a LiDAR is primarily used to achieve obstacle detection and avoidance in various approaches as presented in Refs. [54,75–85]. In Refs. [76,84,85], data from the LiDAR and camera were fused together to achieve obstacle perception; LiDAR was responsible for getting the accurate position of objects, while the camera would get its features and classification. Han et al. in Ref. [85] prompted the object detection problem by developing a framework that applies decision-level sensor fusion techniques on a Velodyne 64-beam LiDAR with an RGB camera in order to improve the detection of dim objects such as pedestrians and cyclists. Similarly, Guan et al. applied decision-level sensor fusion between a camera and a 3D LiDAR to achieve 3D vehicles detection in different illuminations levels [86]. While authors in Refs. [54,75] extracted Regions of Interest (ROI)s from images, and the LiDAR provided 3D localization. However; this limited the detection process only by the 2D imagery.

Also, a 3D object detector that processes in a Bird's Eye View (BEV)

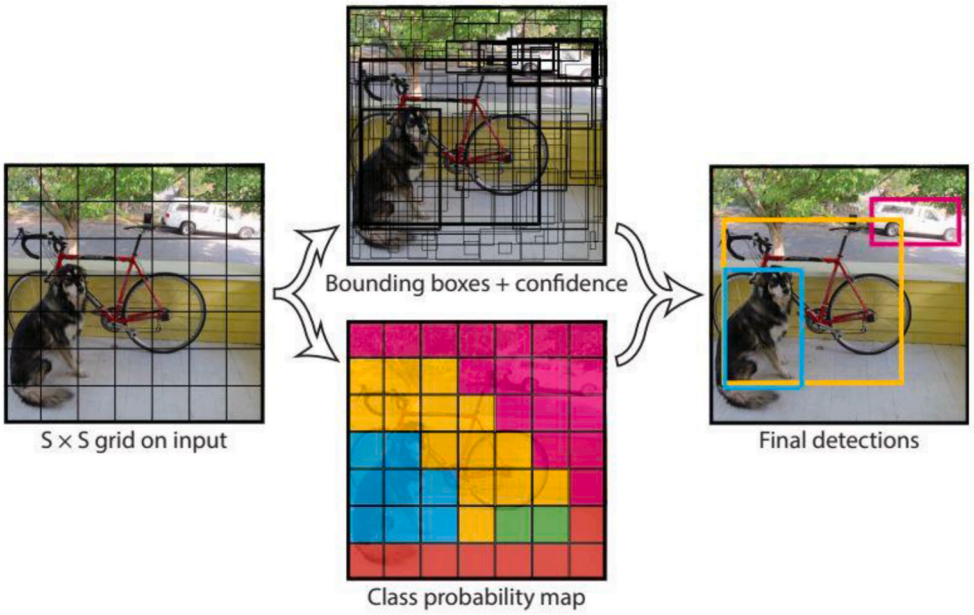

Fig. 5. YOLOv1's model. It models detection as a regression problem. It divides the image into SxS grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are encoded as an S x S x (B*5 +C) tensor [112].

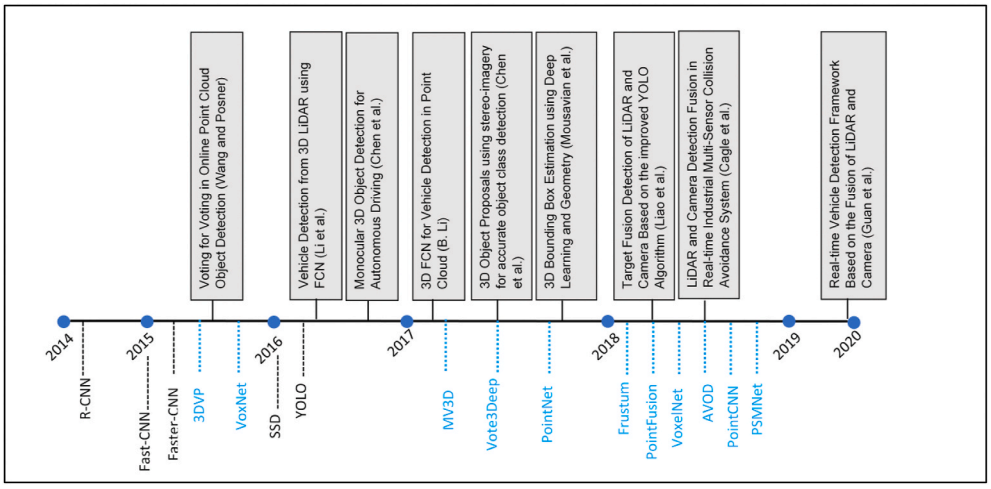

Fig. 6. Milestones of object detection for AVs, including the state-of-art DL-based approaches (R–CNN [107], Fast-CNN [109], Faster-CNN [110], SSD [111], YOLO [112]), CNN-based Networks [75,77,78,123–130], and reliable rela-time object detection approaches [53,54,56,82,85,86,132–134]. The time period prior to 2013 was dominated by handcrafted features. A transition took place in 2013 with the development of CNNs which contributed in the emergence of multiple real-time object detectors for driving scenarios.

has been presented in Ref. [81], it fuses image features by learning to project them into the BEV space. Continuous convolutions have been employed to fuse image and LiDAR feature maps at different levels of resolution. Additionally, in Ref. [77] authors designed a deep fusion framework that combines features from different views. Their 3D localization and detection tests on the KITTI benchmark outperformed the state-of-the-art at that time by 25% and 30% Average Precision (AP). Similarly, in Refs. [78,79], authors applied 2D Convolutional Neural Networks (CNN)s on both the BEV view and image feature maps and fused them at the intermediate region-wise convolutional feature map.

Some approaches have targeted the detection of specific object classes; for example, authors in Ref. [87] presented a pedestrian detector that performs pedestrians pattern matching and recognition, they also made use of context information providing a danger estimation associated with the detected pedestrians. Their experiments achieved 82.29% positive detections and 1.11% false positives (per frame). They compared these results to the cases when only camera (73.97% positive detections, 5.27% false positives) and only LiDAR (74.56% positive detections, 13.3% false positives) were used.

Also, vehicle detection problem has been addressed in Ref. [80]. The approach presented takes advantage of context and online information provided by the GPS. Fusion results achieved 92.03% positive vehicle detections, and 0.59% missed detections (per frame). These results were compared to when using only a camera (47.72% positive detections, 1.13% missed detections) and when using only a LiDAR (91.03% positive detections, 8.19% missed detections). Another example is the passive beacon detection presented in Ref. [83], authors proposed a delineation method that enhanced the performance after fusing camera and LiDAR data.

Besides object detection and localization, the fusion between LiDAR and camera sensors has enabled the operation of urban mapping and vehicle positioning, as shown in Refs. [84,88,89]. In Ref. [88] authors offered a novel approach based on the fusion between a panoramic camera and 2D LiDAR for urban mapping and localization, their results had an average error of 0.2 m in the horizontal plane over a 580 m trajectory. In Ref. [89], authors have achieved 3D mapping using the 2D Hokuyo UTM-30LX LiDAR and a monocular camera, their unique approach was mounting the LiDAR on a rotating motor which rotates at

Table 6
DL-BASED OBJECT DETECTION APPROACHES.

Approach	Base Architecture	Target Objects	Time Analysis	Hardware	Input Data	Testing Dataset	Evaluation
An efficient Pedestrian Detection Method Based on YOLOv2 [131]	YOLOv2	2D pedestrians	73 FPS	NVIDIA GPU GTX 1080Ti	RGB images	INRIA & Caltech	90.9% AP
Target Fusion Detection of LiDAR and Camera Based on the Improved YOLO Algorithm [85]	Improved YOLO	3D detection of dim objects	13 FPS	NVIDIA GPU	64-beam LiDAR & color camera	6-category dataset of 3000 frames	82.9% mAP
LiDAR and Camera Detection fusion in a Real-time Industrial Multi-Sensor Collision avoidance system [82]	YOLO and SVM	Passive beacons (traffic cones)	Real-time 5 Hz	NVIDIA Jetson TX2 GPU	8-beam LiDAR & RGB camera	Self-collected dataset	• 3–20 m range (TP rate 97.6%, FN rate: 2.4%) • 20–40 m range (TP rate: 94.8%, FN rate: 5.2%)
3D fully convolutional Network for Vehicle Detection in Point Cloud [132]	Fully Convolutional Networks	3D vehicle detection	Not noted	Not noted	64-beam LiDAR	KITTI	AP (easy, medium, hard): (84.2%, 75.3%, 68%)
Voting for Voting in Online Point Cloud Object Detection [133]	• Sliding Window Approach • CNN	Car, pedestrians, cyclists	100K- points point cloud at 8 orientations in 0.5s	Multi-core CPU (i7)	3D LiDAR	KITTI	AP (easy, medium, hard): • Car: 47.99%, 56.8%, 42.57% • Pedestrian: 35.74%, 44.48%, 33.72% • Bicyclists: 31.24%, 41.43%, 28.6%
Vehicle Detection from 3D LiDAR using Fully Convolutional Network [134]	Translation of Point clouds into voxels or image stacks using CNNs	3D vehicle detection	Not noted	Not noted	64-beam LiDAR	KITTI	• Image space AP: (60.3%, 47.5%, 42.7%) • Image space AOS (59.1%. 45.9%, 41.4%)
3D object proposals using stereo-imagery for accurate Object Class Detection [56]	CNNs	3D objects	At test time 2s to evaluate one image with 2K proposals	NVIDIA GPU TitanX	Stereo-camera	KITTI	3D AP of cars (89.49%, 81.21%, 74.32%)
3D Bounding Box Estimation using Deep Learning and Geometry [53]	Deep CNNs	3D objects	Not noted	Not noted	RGB Camera	KITTI	• AP cars: 92.98%, 89.04%, 77.17% • AOS cars: 92.9%, 88.75%, 76.76%
Monocular 3D Object Detection for autonomous Driving [54]	CNNs	3D objects	1.8s inference time on a single core (can be parallelized for real-time	Not noted	Single Monocular Camera	KITTI	AP (easy, medium, hard): • Car: 92.33%, 88.66%, 78.96% • Pedestrian: 80.35%, 66.68%, 63.44% • Cyclists: 76.04%, 66.36%, 58.87% AOS (easy, medium, hard): • Cars: 91.01%, 86.62%, 76.84% • Pedestrians: 71.15%, 58.15%, 54.94% • Cyclists: 65.56%, 54.97%, 48.77%
Real-Time Vehicle Detection Framework Based on the Fusion of LiDAR and Camera [86]	YOLOv3	3D vehicle detection	0.057s per frame which is much shorter than the response time of 0.2s for human drivers	Intel Xeon E5-2670 CPU and an NVIDIA GeForce GTX 1080Ti GPU	Monocular Camera and 3D Velodyne LiDAR	KITTI and Waymo Open Dataset	AP (vehicles): • Easy: 85.62% • Moderate: 80.16% • Hard: 70.19%

180°. In Ref. [84], Xue et al. integrated sensor data of LiDAR and camera to achieve efficient autonomous positioning and obstacle perception through geometric and semantic constraints, however, the algorithm was too complex and could not meet the real-time constraints, and also did not consider the detection of dim objects.

2.3.2. *Camera and RADAR fusion*

The integration of RADARs on AVs added the advantage of reliable and accurate obstacles detection in harsh situations; for example low visibility, challenging weather conditions, occluded and distant objects. In Ref. [90], authors addressed the problem of crossing pedestrians, especially when they are occluded by other vehicles using the ability of RADARs in detecting reflections of pedestrians even in a full occlusion

via multi-path propagation. Also, in Ref. [91], authors demonstrated the integration of RADAR along with a camera in order to boost the performance of detecting small (distant) objects. Their work was focused on two classes: vehicles and pedestrian/cyclists.

The fusion between camera and RADAR proved to not only be efficient in object detection but also classification. In Ref. [92], authors fused data from both camera and RADAR sensors in order to classify detected objects and measure the relative distance; their proposed work also works in real-time environments because of its fast detections.

Also, a novel research direction is to apply deep learning techniques on RADAR signals to detect the presence of an object in the surrounding and then estimate its 3D position. In Ref. [69], authors employed deep learning framework instead of using traditional signal processing techniques on RADAR data; their proposed system was tested on car detection in noisy environments.

2.3.3. Camera, LiDAR, and RADAR fusion

It is considered the most expensive sensor fusion approach; many research has adopted this sensors combination in different ways. For example, authors in Ref. [93] made use of already-fused measurement data and represented a novel grid-based object tracking approach. In order to achieve a precise motion estimation, RADAR Doppler velocity estimations were integrated into the input data. They evaluated their work on real sensor data with the context of autonomous driving in challenging urban scenarios.

Also, in Ref. [94] authors applied sensor fusion on a RADAR, ZED stereo camera, and an HDL-32E Velodyne LiDAR, and developed an algorithm that would track any potential object in the scene (e.g. object agnostic). Using the same LiDAR and stereo-camera as in Ref. [94] along with a 150 GHz RADAR, authors in Ref. [95] tested their approach in a scene having artificially-generated fog in order to highlight the requirement of inclusion of a RADAR sensor as part of the automotive industry.

In [11], the authors focused on the centralized sensor-independent fusion schemes in order to allow simple sensor replacement and ensure redundancy by using probabilistic and generic interfaces. The authors integrated three IBEO LUX 3D LiDARs, monochrome cameras, and several RADARs. Another approach developed by authors in Ref. [96], allowed the systems to switch between a point and a 3D-box model according to the distance of objects to the vehicle.

3. Object detection approaches

Object detection step provides a semantic understanding of the acquired data; it can be defined as determining the class and location of detected objects. In order to develop a reliable AV capable of perceiving the surrounding environment, the best fitting object detection technique must be chosen to work efficiently in the required environments. The diverse driving scenarios impose different object detection challenges; as presented in Fig. 2.

Object detection can be categorized into two main categories: 2D- and 3D-object detections. Choosing the detection technique mainly depends on the intended application, the used sensors, and the operations following the detection step (ex: additional sensor data fusion). For example, video surveillance applications can perform their task using only 2D detections. However, AVs' operation depends on the precise 3D localization of detected objects. Different sensors can be used in order to achieve object detection, but the visual system stays the primary data source as it can perform object classification as well.

In autonomous driving, different objects classes need to be detected and can be classified into static and dynamic objects. Static objects include traffic lights and signs, buildings, bridges, and curbs, while dynamic objects include pedestrians, cyclists, animals, and different vehicles. Detection of static objects is considered a straightforward task because of their definite shape and easily predicted location. In this paper, we focus on the detection of vulnerable road users (e.g. dynamic objects) due to the higher danger level they impose on the system.

There are two main approaches towards object detection: Machine Learning (ML), and Deep Learning (DL). Examples of ML-based approaches are Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), and Viola-Jones object detection framework based on Harr-like features. While examples of DL-based approaches and architectures are: Region proposals (R–CNN, Fast R–CNN, Faster R–CNN), SSD, and You Only Look Once (YOLO). Due to the diversity of object classes, lighting, and background conditions, feature extraction is not a robust approach. Therefore, DNNs are playing a significant role in object detection tasks; as they can learn and extract more complex features, also their training algorithms eliminate the need for manually designing the features that need to be extracted.

3.1. Object detection evaluation metrics

Object detection is not only concerned with classification, but also localization of objects within bounding boxes associated with its corresponding confidence value. Therefore, evaluation metrics are concerned with the closeness of the predicted bounding boxes with the ground truths.

3.1.1. Intersection over union (IoU)

It is also called the Jaccard Index; it quantifies the correspondence between the ground truth bounding boxes and the predicted bounding boxes. The IoU values range from 0 to 1. It can be formulated as follows:

$$IoU_{pred}^{truth} = \frac{truth \cap \text{pred}}{truth \cup \text{pred}}$$

3.1.2. Predictions

Predictions can be classified into True Positives (TP), False Negatives (FN), and False Positives (FP). After calculating an IoU value for each bounding box, a threshold is set to discriminate between positive detections and false detections.

3.1.3. Precision

It corresponds to the percentage of the time the detector was right, also called positive predictive value.

$$Precision = \frac{TP}{TP + FP} = \frac{true\ object\ detection}{all\ detected\ boxes}$$

3.1.4. Recall

It is also referred to as sensitivity; it measures the probability of detection of ground truth objects

$$Recall = \frac{TP}{TP + FN} = \frac{true\ object\ detection}{all\ ground\ truth\ boxes}$$

3.1.5. Precision x recall curve

This curve presents both recall and precision. It plots recall on the *x*-axis and precision on the *y*-axis, therefore, the ideal point is the closes to (1.0, 1.0).

3.1.6. Average Precision (AP)

AP represents the averages precision at a set of 11 spaced recall points (0, 0.1, 0.2,..,1). Corresponding precision values for a certain recall value (*r*) are interpolated by taking the maximum precision whose recall value is greater than r.

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, 0.2, \ldots 1\}} Pinterp\ (r)$$

Where

$$Pinterp(r) = \underset{\tilde{r} > r}{\max} p\ (\tilde{r})$$

3.1.7. Mean Average Precision (mAP)

It is the average AP over all the N-class categories present in a dataset.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

Similarily, AP_{3D} for 3D detections can be calculated using 3D data measurements.

3.1.8. Average Orientation similarity (AOS)

It was proposed by Ref. [97] in order to evaluate the performance of jointly detecting objects and estimating their 3D orientation. It is defined as:

$$AOS = \frac{1}{11}\sum_{r \in \{0,\ 0.1,\ \ldots,\ 1\}} \max_{\tilde{r}:\ \tilde{r} \geq r} s(\tilde{r})$$

Here, r is the recall while IoU threshold is 0.5. The orientation similarity s ∈[0,1] at recall r is a normalized ([0..1]) variant of the cosine similarity, defined as:

$$s(r) = \frac{1}{|D(r)|}\sum_{i \in D(r)} \frac{1 + \cos \Delta_{\theta}^{(i)}}{2}\delta_i$$

Where D(r) denotes the set of all object detections at recall rate r, and $\Delta_{\theta}^{(i)}$ is the angle difference between estimated and ground truth oritntation of detection i. δ_i is set to 1 if detection i has been assigned to a ground truth bounding box within the IoU threshold, and δ_i is set to 0 if it has not been assigned.

3.1.9. Mean IoU (mIoU)

To calculate the shape's mIoU: For each part type in category C, compute IoU between ground truth and prediction. If the union of ground truth and prediction points is empty, then count part IoU as 1. Then we average IoUs for all part types in category C to get mIoU for that shape. To cacualte mIoU for the category, we take average of mIoUs for all shapes in that category.

3.2. Object detection evaluation datasets

In order to train and validate detection networks that will be used to detect objects surrounding AVs outdoors, different datasets have been proposed with different features, including:

- The data size
- Data types (RGB images, stereo-images, LiDAR point clouds, etc.)
- Driving conditions (downtown vs suburban areas, daytime vs nighttime, weather and illumination conditions, etc.)

3.2.1. PASCAL VOC 2007/2012

It comprises 20 categories. Results are evaluated by the AP in each category and the mAP across all the 20 categories with an IoU threshold of 0.5. Additionally, it provides standard labelling and evaluation tools.

3.2.2. Microsoft COCO

It is a Microsoft-sponsored dataset which consists of more than 330K fully segmented images; each image has an average of 7 objects from a total of 91 categories, images were gathered from complex everyday scenes; therefore, it is considered more challenging than PASCAL. Results are evaluated by the Average Precision (AP) calculated under different values of IoUs (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95) and different object sizes [98].

3.1.4. ImageNet

It is organized according to the WordNet hierarchy, in which every meaningful concept is described by multiple words or phrases. ImageNet consists of 14,197,122 images organized into 21,841 subclasses; these subclasses can be considered sub-trees of 27 high-level categories [99].

3.1.5. KITTI

It was gathered by an autonomous driving platform. The complete benchmark covers different tasks, such as stereo, optical flow, visual odometry, etc. The object detection dataset has 7481 annotated training images and 7518 testing images. The annotated images have both 2D and 3D bounding boxes of cars, pedestrians, and cyclists in urban driving scenarios. The official 3D metric is the Average Orientation similarity (AOS) which is calculated by multiplying the AP of the 2D detector and the average Cosine distance similarity for the azimuth orientation [97, 100]. Moreover, there are three levels of difficulties for the Velodyne data (Fig. 4), and also RGB images:

- Easy: Minimum bounding box height is 40 pixels, fully visible objects, maximum truncation is 15%
- Moderate: Minimum bounding box height is 25 pixels, contains partly occluded objects, maximum truncation is 30%
- Hard: Minimum bounding box height is 25 pixels, contains difficult to see objects, maximum truncation is 50%.

3.1.6. SceneFlow

It is a synthetic dataset with more than 39000 stereo image pairs with a resolution of 960 × 540 pixels. Images are categorized into three classes: FlyingThings3D, Driving, and Monka. It provides dense and elaborate disparity maps as ground truth [101].

3.1.7. ModelNet

It provides a clean collection of 3D Computer-Aided Designs (CAD) models of objects acquired from online search engines. There are two ModelNet datasets, one of them only has the ten most popular object categories, while the other one has the full dataset of the 40 classes. The ModelNet40 has 12,311 3D mesh models from 40 categories, with 9843/2468 training/testing split. It also provides different objects alignment and facing-directions settings. The evaluation metrics used on the ModelNet40 are mean per-class Accuracy (mA) and Overall Accuracy (OA) [102,103].

3.1.8. OutdoorScene

It consists of 200 images; it is primarily designed in order to test (not train) detections of highly occluded and truncated objects. This dataset has 659 cars, among which 235 cars are occluded, and 135 are truncated [104].

3.1.9. Sydney urban objects

This dataset was collected by a Velodyne HDL-64E LiDAR in Sydney. It has 631 individual scans of 26 different classes of vehicles, pedestrians, signs, and trees, etc. Its main purpose is to test the detection performance in challenging viewpoints and occlusion levels [105].

3.1.10. Waymo Open Dataset

It is comprised of high-resolution sensor data collected by Waymo self-driving cars in diverse conditions. It includes different data types (1 mid-range LiDAR, 4 short-range LiDARs, 5 cameras, synchronized LiDAR and camera data, LiDAR to camera projections, sensor calibrations and vehicle poses). It has labels for 4 objects classes, vehicles, pedestrians, cyclists, and signs. This dataset also stands out for the diverse conditions it includes, as it includes diving scenes during daylight and nighttime, downtown and suburban areas, and diverse weather conditions [106].

3.2. Object detection using DNNs

Fig. 6 demonstrates a chronicle figure exploiting milestones regarding object detection approaches applied on AVs. Since the development of R–CNN networks [107], many research has turned towards the employment of DNNs in object detection [108–112]. Table 4 summarizes a comparison between the performance of the three main DL-based object detection approaches (CNNs, SSD, YOLO) comparing their speed (Frames per Second (FPS)) while running on a GPU, and accuracy measures. An extensive review of object detection with deep learning has been made in Ref. [33]. On the other hand, because of the different advantages of CNN-based approaches (ex: hierarchical feature representation, increased expressive capability, and the combination of several tasks together), many CNN-based detection networks have been developed as shown in Table 5. CNN-based detection has been used on AVs to achieve:

- Vehicle detection (ex [113])
- Pedestrian detection (ex [113–115]). Also, a survey was made to exploit deep learning techniques for pedestrians' detection and tracking in Ref. [116]. Moreover [117], is an extensive survey on camera-based pedestrian detection techniques.
- Cyclists detection (ex [118,119])

As shown in Table 4, there are many different deep learning approaches towards object detection. Therefore, choosing the right object detection is crucial and depends on the application. The performance of each of the approaches depends on many criteria, including, size of objects, number of objects, number of classes, and the detection speed.

Faster R–CNN is the best choice if only high accuracy is required. However, its very low speed does not allow it to achieve real-time performance. SSD is a better recommendation, but also if speed is the highest priority, then YOLO is the best candidate. One limitation for YOLO is that it cannot perform accurately on detecting small objects as it uses larger grids. However, this limitation does not hinder its application on AVs as vulnerable objects that are considered critical appear as large objects in the frame.

The working model of YOlOv1 is shown in Fig. 5. Due to the unparalleled speed of YOLO, different research has been conducted to extend its functionality and capabilities:

- *YOLO9000* [120]*:* Authors have jointly trained YOLO9000 on the COCO detection dataset and ImageNet classification dataset. This simultaneous training allows YOLO9000 to predict unlabeled data. The proposed YOLO9000 was validated on the ImageNet dataset with 19.7% mAP, and a mAP of 16.0% for 156 classes that were not in the COCO dataset. Therefore, it can detect more than 9000 different object categories in real-time constraints.
- *YOLO-LITE* [121]*:* IT was proposed to facilitate its usage on portable devices that lack a GPU. It was trained on the PASCAL VOC, and then COCO datasets achieving a mAP of 33.81% and 12.26% respectively. Its main advantage is the high FPS which is 21 FPS on a non-GPU platform.
- *Complex-YOLO* [122]*:* It is an extension of YOLOv2 using Euler-Region-Proposal-Network in order to make use of the 3D point cloud taken by multi-beam LiDARs (ex: 64-beam Velodyne LiDAR) in order to achieve 3D detections of all eight classes in the KITTI dataset running faster than 50 FPS on an NVIDIA TitanX GPU. The AP achieved on the three main object classes of the KITTI were:
  - o Cars: 67.72% (easy), 64.0% (medium), 63.01% (hard)
  - o Pedestrian: 41.79% (easy), 39.7% (medium), 35.92% (hard)
  - o Cyclists: 68.17% (easy), 58.32% (medium), 54.3% (hard)

Finally, Table 6 highlights the most recent approaches towards employing deep learning in the object detection task of AVs.

4. Conclusion

One of the main motivations for developing AVs was to reduce the accidents rate caused by human error. Therefore, AVs are expected to outperform human driving. In order to achieve this, the initial step: object detection, plays a significant role in the driving process. Object detection is not only bound by the sensors' efficiency and capabilities, but also the algorithms by which data is processed. According to the research reviewed in Section 2, vehicles must have multiple sensors mounted on them to acquire different properties of the surrounding. For example, monocular cameras can not perform real-time 3D detections as they can not directly acquire the third dimension. Therefore, any additional operations made on RGB images saassto convert 2D pixels into 3D pixels will disable the system to operate in real-time constraints.

On the other hand, stereo-images provide 3D data but suffer from long computation time due to the size of the acquired data. Cameras are an essential sensor in AVs as they provide feature data of the surrounding; however, they should be complemented with other sensors that acquire third dimensions, such as LiDARs, which are now considered one of the main sensors in AVs, the main drawback of multi-beam LiDARs is the high cost. Therefore, more research should focus on the evolution of low-cost and reliable LiDARs, and also the efficient employment of single-beam LiDARs.

In Section 3, we reviewed different approaches addressing real-time vulnerable objects detection task. There is no existing optimal object detection technique; however, choosing an approach depends on the available sensors, hardware capabilities, successive operations required, etc. Therefore, in this paper, we presented an up-to-date review on existing attempts along with their advantages, requirements, and limitations. It is shown that one of the main requirements for performing real-time detections is the employment of powerful GPUs.

Also, as shown in Table 6, YOLO has proved to be superior to other DNN-based approaches due to its fast performance. YOLOv3 has now achieved the best performance on AVs' real-time object detection task.

Although current perception system implemented on levels 1–4 AVs have shown to enhance vehicles operations, there is still much to improve upon before level 5 vehicles be commercially available. For example, more diverse datasets imitating the non-optimal real-life driving scenarios will enhance the performance of existing object detection networks (ex: Waymo Open Dataset). Another direction would be qualifying object detection techniques of performing under challenging ODDs (ex: rain, snow, fog, etc.). Therefore, it has been recently encouraged to develop domain-adaptive object detection techniques in order not to hinder the reliable performance of different sensors in challenging weather conditions.

Declaration of competing interest

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

References

[1] Department for Transport, The Pathway to Driverless Cars: Summary Report and Action Plan, 2015.
[2] R. Sherony, C. Zhang, Pedestrian and bicyclist crash scenarios in the US, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp. 1533–1538.
[3] H. Zhu, K. Yuen, L. Mihaylova, H. Leung, Overview of environment perception for intelligent vehicles, IEEE Trans. Intell. Transport. Syst. 18 (10) (2017) 2584–2601.
[4] J. Van Brummelen, M. O'Brien, D. Gruyer, H. Najjaran, Autonomous vehicle perception: the technology of today and tomorrow, Transport. Res. C Emerg. Technol. 89 (2018) 384–406.
[5] SAE On-Road Automated Vehicle Standards Committee and Others, in: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, SAE International: Warrendale, PA, USA, 2018.

[6] L. Jones, Driverless cars: when and where? Eng. Technol. 12 (2) (2017) 36–40, https://doi.org/10.1049/et.2017.0201. Available: https://digital-library.theiet.org/content/journals/10.1049/et.2017.0201.

[7] E.D. Dickmanns, B. Mysliwetz, T. Christians, An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles, IEEE Trans. Syst. Man Cybern. 20 (6) (1990) 1273–1284.

[8] E.D. Dickmanns, R. Behringer, D. Dickmanns, T. Hildebrandt, M. Maurer, F. Thomanek, J. Schiehlen, The seeing passenger car'VaMoRs-P, in: Proceedings of the Intelligent Vehicles' 94 Symposium, 1994, pp. 68–73.

[9] C. Rouff, M. Hinchey, Experience from the DARPA Urban Challenge, Springer Publishing Company, Incorporated, 2011.

[10] M. Walton, Robots Fail to Complete Grand Challenge, CNN News, Mar, 2004.

[11] F. Kunz, D. Nuss, J. Wiest, H. Deusch, S. Reuter, F. Gritschneder, A. Scheel, M. Stübler, M. Bach, P. Hatzelmann, Autonomous driving at ulm university: a modular, robust, and sensor-independent fusion approach, in: 2015 IEEE Intelligent Vehicles Symposium (IV), 2015, pp. 666–673.

[12] D.L. Peters, Students working towards robotic chauffeurs: the SAE/GM autodrive challenge, Mechanical Engineering Magazine Select Articles 140 (2018) S6–S11, 03.

[13] K. Bimbraw, Autonomous cars: past, present and future a review of the developments in the last century, the present scenario and the expected future of autonomous vehicle technology, in: 2015 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO), 2015, pp. 191–198.

[14] T. Luettel, M. Himmelsbach, H. Wuensche, Autonomous ground vehicles—concepts and a path to the future, Proc. IEEE 100 (2012) 1831–1839 (Special Centennial Issue).

[15] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, Stanley: The robot that won the DARPA Grand Challenge, J. Field Robot. 23 (9) (2006) 661–692.

[16] C. Urmson, J. Anhalt, D. Bagnell, C. Baker, R. Bittner, M.N. Clark, J. Dolan, D. Duggins, T. Galatali, C. Geyer, Autonomous driving in urban environments: boss and the urban challenge, J. Field Robot. 25 (8) (2008) 425–466.

[17] P. Chatterjee, Self-driving Car Pushes Sensor Technology, EDN Magazine, on-Line, 2012.

[18] I. Barabás, A. Todoruţ, N. Cordoş, A. Molea, Current challenges in autonomous driving. IOP Conference Series: Materials Science and Engineering, 2017, 012096.

[19] M. Martínez-Díaz, F. Soriguera, Autonomous vehicles: theoretical and practical challenges, Transportation Research Procedia 33 (2018) 275–282.

[20] W. Elmenreich, An Introduction to Sensor Fusion, vol. 502, Vienna University of Technology, Austria, 2002.

[21] K. Kovačić, E. Ivanjko, H. Gold, Computer Vision Systems in Road Vehicles: a Review, 2013 arXiv Preprint arXiv:1310.0315.

[22] C. Ilas, Electronic sensing technologies for autonomous ground vehicles: a review, in: 2013 8th International Symposium on Advanced Topics in Electrical Engineering, Atee), 2013, pp. 1–6.

[23] M.O. Aqel, M.H. Marhaban, M.I. Saripan, N.B. Ismail, Review of visual odometry: types, approaches, challenges, and applications, SpringerPlus 5 (1) (2016) 1897.

[24] W. Shi, M.B. Alawieh, X. Li, H. Yu, Algorithm and hardware implementation for visual perception system in autonomous vehicle: a survey," *Integration*, the VLSI Journal 59 (2017) 148–156.

[25] S. Campbell, N. O'Mahony, L. Krpalcova, D. Riordan, J. Walsh, A. Murphy, C. Ryan, Sensor technology in autonomous vehicles: a review, in: 2018 29th Irish Signals and Systems Conference, ISSC), 2018, pp. 1–4.

[26] J. Kocić, N. Jovičić, V. Drndarević, Sensors and sensor fusion in autonomous vehicles, in: 2018 26th Telecommunications Forum (TELFOR), 2018, pp. 420–425.

[27] F. Rosique, P.J. Navarro, C. Fernández, A. Padilla, A systematic review of perception system and simulators for autonomous vehicles research, Sensors 19 (3) (2019) 648.

[28] A. Mukhtar, L. Xia, T.B. Tang, Vehicle detection techniques for collision avoidance systems: a review, IEEE Trans. Intell. Transport. Syst. 16 (5) (2015) 2318–2338.

[29] K.U. Sharma, N.V. Thakur, A review and an approach for object detection in images, Int. J. Comput. Vis. Robot 7 (1–2) (2017) 196–237.

[30] M. Tiwari, R. Singhai, A review of detection and tracking of object from image and video sequences, Int. J. Comput. Intell. Res. 13 (5) (2017) 745–765.

[31] W. Zhiqiang, L. Jun, A review of object detection based on convolutional neural network, in: 2017 36th Chinese Control Conference, CCC), 2017, pp. 11104–11109.

[32] L.E. Carvalho, A. von Wangenheim, 3d object recognition and classification: a systematic literature review, in: Pattern Analysis and Applications, 2019, pp. 1–50.

[33] Z. Zhao, P. Zheng, S. Xu, X. Wu, Object detection with deep learning: a review, in: IEEE Transactions on Neural Networks and Learning Systems, 2019.

[34] D. González, J. Pérez, V. Milanés, F. Nashashibi, A review of motion planning techniques for automated vehicles, IEEE Trans. Intell. Transport. Syst. 17 (4) (2015) 1135–1145.

[35] M. Maurer, J.C. Gerdes, B. Lenz, H. Winner, Autonomous Driving, vol. 10, Springer Berlin Heidelberg, Berlin, Germany, 2016, 978-973.

[36] B. Vanholme, D. Gruyer, B. Lusetti, S. Glaser, S. Mammar, Highly automated driving on highways based on legal safety, IEEE Trans. Intell. Transport. Syst. 14 (1) (2012) 333–347.

[37] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, M. Porta, Artificial vision in road vehicles, Proc. IEEE 90 (7) (2002) 1258–1271.

[38] C.R. Wang, J.J. Lien, Automatic vehicle detection using local features–a statistical approach, IEEE Trans. Intell. Transport. Syst. 9 (1) (2008) 83–96.

[39] E.U. Haq, S.J.H. Pirzada, J. Piao, T. Yu, H. Shin, Image processing and vision techniques for smart vehicles, in: 2012 IEEE International Symposium on Circuits and Systems, 2012, pp. 1211–1214.

[40] M. Bertozzi, A. Broggi, S. Castelluccio, A real-time oriented system for vehicle detection, J. Syst. Architect. 43 (1–5) (1997) 317–325.

[41] C. Tzomakas, W. von Seelen, Vehicle detection in traffic scenes using shadows. Ir-Ini, Institut Fur Nueroinformatik, Ruhr-Universitat, 1998.

[42] M.B. Van Leeuwen, F.C. Groen, Vehicle detection with a mobile camera: spotting midrange, distant, and passing cars, IEEE Robot. Autom. Mag. 12 (1) (2005) 37–43.

[43] W. For, K. Leman, H. Eng, B. Chew, K. Wan, A multi-camera collaboration framework for real-time vehicle detection and license plate recognition on highways, in: 2008 IEEE Intelligent Vehicles Symposium, 2008, pp. 192–197.

[44] S. Ieng, J. Vrignon, D. Gruyer, D. Aubert, A new multi-lanes detection using multi-camera for robust vehicle location, in: IEEE Proceedings. Intelligent Vehicles Symposium, 2005, 2005, pp. 700–705.

[45] H.T. Niknejad, S. Mita, D. McAllester, T. Naito, Vision-based vehicle detection for nighttime with discriminately trained mixture of weighted deformable part models, in: 2011 14th International IEEE Conference on Intelligent Transportation Systems, ITSC), 2011, pp. 1560–1565.

[46] J. Kim, S. Hong, J. Baek, E. Kim, H. Lee, Autonomous vehicle detection system using visible and infrared camera, in: 2012 12th International Conference on Control, Automation and Systems, 2012, pp. 630–634.

[47] F. García, A. Prioletti, P. Cerri, A. Broggi, A. de la Escalera, J.M. Armingol, Visual feature tracking based on phd filter for vehicle detection, in: 17th International Conference on Information Fusion, FUSION), 2014, pp. 1–6.

[48] G. Saldaña González, J. Cerezo Sánchez, M.M. Bustillo Díaz, A. Ata Pérez, Vision system for the navigation of a mobile robot, Comput. Sist. 22 (1) (2018) 301–308.

[49] B. Ling, D.R. Gibson, D. Middleton, Motorcycle detection and counting using stereo camera, IR camera, and microphone array, in: Video Surveillance and Transportation Imaging Applications, 2013, p. 86630P.

[50] R. Sen, P. Siriah, B. Raman, Roadsoundsense: acoustic sensing based road congestion monitoring in developing regions, in: 2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2011, pp. 125–133.

[51] M. Mizumachi, A. Kaminuma, N. Ono, S. Ando, Robust sensing of approaching vehicles relying on acoustic cues, Sensors 14 (6) (2014) 9546–9561.

[52] K. Yoon, Y. Song, M. Jeon, Multiple hypothesis tracking algorithm for multi-target multi-camera tracking with disjoint views, IET Image Process. 12 (7) (2018) 1175–1184.

[53] A. Mousavian, D. Anguelov, J. Flynn, J. Kosecka, 3d bounding box estimation using deep learning and geometry, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7074–7082.

[54] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, R. Urtasun, Monocular 3d object detection for autonomous driving, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2147–2156.

[55] Y.L. Chen, M.R. Jahanshahi, P. Manjunatha, W. Gan, M. Abdelbarr, S.F. Masri, B. Becerik-Gerber, J.P. Caffrey, Inexpensive multimodal sensor fusion system for autonomous data acquisition of road surface conditions, IEEE Sensor. J. 16 (21) (2016) 7731–7743.

[56] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler, R. Urtasun, 3d object proposals using stereo imagery for accurate object class detection, IEEE Trans. Pattern Anal. Mach. Intell. 40 (5) (2017) 1259–1272.

[57] J. Lahoud, B. Ghanem, 2d-driven 3d object detection in rgb-d images, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4622–4630.

[58] S. Song, J. Xiao, Deep sliding shapes for amodal 3d object detection in rgb-d images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 808–816.

[59] Y. Wang, W. Chao, D. Garg, B. Hariharan, M. Campbell, K.Q. Weinberger, Pseudo-lidar from visual depth estimation: bridging the gap in 3d object detection for autonomous driving, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8445–8453.

[60] Y. You, Y. Wang, W. Chao, D. Garg, G. Pleiss, B. Hariharan, M. Campbell, K. Q. Weinberger, Pseudo-LiDAR : accurate depth for 3D object detection in autonomous driving, 2019 arXiv Preprint arXiv:1906.06310.

[61] Z. Fang, S. Zhao, S. Wen and Y. Zhang, "A real-time 3D perception and reconstruction system based on a 2D laser scanner," Journal of Sensors, vol. 2018, 2018.

[62] X. Zhang, W. Xu, C. Dong, J.M. Dolan, Efficient L-shape fitting for vehicle detection using laser scanners, in: 2017 IEEE Intelligent Vehicles Symposium (IV), 2017, pp. 54–59.

[63] T. Taipalus, J. Ahtiainen, Human detection and tracking with knee-high mobile 2D LIDAR, in: 2011 IEEE International Conference on Robotics and Biomimetics, 2011, pp. 1672–1677.

[64] X. Shao, H. Zhao, K. Nakamura, K. Katabira, R. Shibasaki, Y. Nakagawa, Detection and tracking of multiple pedestrians by using laser range scanners, in: 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007, pp. 2174–2179.

[65] Z. Rozsa, T. Sziranyi, Obstacle prediction for automated guided vehicles based on point clouds measured by a tilted LIDAR sensor, IEEE Trans. Intell. Transport. Syst. 19 (8) (2018) 2708–2720.

[66] F. García, F. Jiménez, J.E. Naranjo, J.G. Zato, F. Aparicio, J.M. Armingol, A. de la Escalera, Environment perception based on LIDAR sensors for real road applications, Robotica 30 (2) (2012) 185–193.
[67] D. Choi, Y. Bok, J. Kim, I. Shim, I. Kweon, Structure-from-motion in 3D space using 2D lidars, Sensors 17 (2) (2017) 242.
[68] C. Hung, A.T. Lin, B.C. Peng, H. Wang, J. Hsu, Y. Lu, W. Hsu, J.C. Zhan, B. Juan, C. Lok, 9.1 toward automotive surround-view radars, in: 2019 IEEE International Solid-State Circuits Conference-(ISSCC), 2019, pp. 162–164.
[69] G. Zhang, H. Li, F. Wenger, Object detection and 3D estimation via an FMCW radar using a fully convolutional network, 2019 arXiv Preprint arXiv: 1902.05394.
[70] L. Arnone, P. Vicari, Simultaneous odometry, mapping and object tracking with a compact automotive radar, in: 2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive, AEIT AUTOMOTIVE), 2019, pp. 1–6.
[71] M. Kam, X. Zhu, P. Kalata, Sensor fusion for mobile robot navigation, Proc. IEEE 85 (1) (1997) 108–119.
[72] B.V. Dasarathy, Sensor fusion potential exploitation-innovative architectures and illustrative applications, Proc. IEEE 85 (1) (1997) 24–38.
[73] R.C. Luo, C. Yih, K.L. Su, Multisensor fusion and integration: approaches, applications, and future research directions, IEEE Sensor. J. 2 (2) (2002) 107–119.
[74] F. Castanedo, A review of data fusion techniques, ScientificWorldJournal (2013) 704504. Oct 27, 2013.
[75] C.R. Qi, W. Liu, C. Wu, H. Su, L.J. Guibas, Frustum pointnets for 3d object detection from rgb-d data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 918–927.
[76] H. Moon, J. Kim, J. Kim, Obstacle detecting system for unmanned ground vehicle using laser scanner and vision, in: 2007 International Conference on Control, Automation and Systems, 2007, pp. 1758–1761.
[77] X. Chen, H. Ma, J. Wan, B. Li, T. Xia, Multi-view 3d object detection network for autonomous driving, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1907–1915.
[78] J. Ku, M. Mozifian, J. Lee, A. Harakeh, S.L. Waslander, Joint 3d proposal generation and object detection from view aggregation, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 1–8.
[79] X. Du, M.H. Ang, S. Karaman, D. Rus, A general pipeline for 3d detection of vehicles, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 3194–3200.
[80] F. Garcia, D. Martin, A. De La Escalera, J.M. Armingol, Sensor fusion methodology for vehicle detection, IEEE Intelligent Transportation Systems Magazine 9 (1) (2017) 123–133.
[81] M. Liang, B. Yang, S. Wang, R. Urtasun, Deep continuous fusion for multi-sensor 3d object detection, in: Proceedings of the European Conference on Computer Vision, ECCV), 2018, pp. 641–656.
[82] P. Wei, L. Cagle, T. Reza, J. Ball, J. Gafford, LiDAR and camera detection fusion in a real-time industrial multi-sensor collision avoidance system, Electronics 7 (6) (2018) 84.
[83] A. Rövid, V. Remeli, Towards raw sensor fusion in 3D object detection, in: 2019 IEEE 17th World Symposium on Applied Machine Intelligence and Informatics, SAMI), 2019, pp. 293–298.
[84] J. Xue, D. Wang, S. Du, D. Cui, Y. Huang, N. Zheng, A vision-centered multi-sensor fusing approach to self-localization and obstacle perception for robotic cars, Frontiers of Information Technology & Electronic Engineering 18 (1) (2017) 122–138.
[85] J. Han, Y. Liao, J. Zhang, S. Wang, S. Li, Target fusion detection of LiDAR and camera based on the improved YOLO algorithm, Mathematics 6 (10) (2018) 213.
[86] L. Guan, Y. Chen, G. Wang, X. Lei, Real-time vehicle detection framework based on the fusion of LiDAR and camera, Electronics 9 (3) (2020) 451.
[87] F. García, J. García, A. Ponz, A. De La Escalera, J.M. Armingol, Context aided pedestrian detection for danger estimation based on laser scanner and computer vision, Expert Syst. Appl. 41 (15) (2014) 6646–6661.
[88] Y. Shi, S. Ji, X. Shao, P. Yang, W. Wu, Z. Shi, R. Shibasaki, Fusion of a panoramic camera and 2D laser scanner data for constrained bundle adjustment in GPS-denied environments, Image Vis Comput. 40 (2015) 28–37.
[89] J. Zhang, S. Singh, Visual-lidar odometry and mapping: low-drift, robust, and fast, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 2174–2181.
[90] A. Palffy, J.F. Kooij, D.M. Gavrila, Occlusion aware sensor fusion for early crossing pedestrian detection, in: 2019 IEEE Intelligent Vehicles Symposium (IV), 2019, pp. 1768–1774.
[91] S. Chadwick, W. Maddern, P. Newman, Distant vehicle detection using radar and vision, in: International Conference on Robotics and Automation, ICRA), 2019.
[92] H. Jha, V. Lodhi, D. Chakravarty, Object detection and identification using vision and radar data fusion system for ground-based navigation, in: 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), 2019, pp. 590–593.
[93] S. Steyer, C. Lenk, D. Kellner, G. Tanzmeister, D. Wollherr, Grid-based object tracking with nonlinear dynamic state and shape estimation, IEEE Transactions on Intelligent Transportation Systems, 2019.
[94] A. Ahrabian, M. Emambakhsh, M. Sheeny, A. Wallace, Efficient multi-sensor extended target tracking using GM-PHD filter, in: 2019 IEEE Intelligent Vehicles Symposium (IV), 2019, pp. 1731–1738.
[95] L. Daniel, D. Phippen, E. Hoare, A. Stove, M. Cherniakov, M. Gashinova, Low-THz Radar, Lidar and Optical Imaging through Artificially Generated Fog, 2017.
[96] H. Cho, Y. Seo, B.V. Kumar, R.R. Rajkumar, A multi-sensor fusion system for moving object detection and tracking in urban driving environments, in: 2014 IEEE International Conference on Robotics and Automation, ICRA), 2014, pp. 1836–1843.
[97] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354–3361.
[98] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: common objects in context, in: European Conference on Computer Vision, 2014, pp. 740–755.
[99] J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
[100] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, The KITTI vision benchmark suite. http://Www.Cvlibs.Net/Datasets/Kitti, 2015.
[101] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, T. Brox, A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4040–4048.
[102] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3d shapenets: a deep representation for volumetric shapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1912–1920.
[103] Princeton University, ModelNet, Edu, Princeton, 2015. Available: https://modelnet.cs.princeton.edu/#.
[104] Y. Xiang, S. Savarese, Object detection by 3d aspectlets and occlusion reasoning, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 530–537.
[105] The University of Sydney, Sydney urban objects dataset, Sydney Urban Objects Dataset, 2013. Available: http://www.acfr.usyd.edu.au/papers/SydneyUrbanObjectsDataset.shtml.
[106] W.O. Dataset, No Title, An Autonomous Driving Dataset, 2019.
[107] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.
[108] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: a survey, Int. J. Comput. Vis. 128 (2) (2020) 261–318.
[109] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.
[110] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, 2015, pp. 91–99.
[111] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A.C. Berg, Ssd: single shot multibox detector, in: European Conference on Computer Vision, 2016, pp. 21–37.
[112] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[113] Y. Xiang, W. Choi, Y. Lin, S. Savarese, Subcategory-aware convolutional neural networks for object proposals and detection, in: 2017 IEEE Winter Conference on Applications of Computer Vision, WACV), 2017, pp. 924–933.
[114] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, S. Tubaro, Deep convolutional neural networks for pedestrian detection, Signal Process. Image Commun. 47 (2016) 482–489.
[115] Z. Zhao, H. Bian, D. Hu, W. Cheng, H. Glotin, Pedestrian detection based on fast R-CNN and batch normalization, in: International Conference on Intelligent Computing, 2017, pp. 735–746.
[116] A. Brunetti, D. Buongiorno, G.F. Trotta, V. Bevilacqua, Computer vision and deep learning techniques for pedestrian detection and tracking: a survey, Neurocomputing 300 (2018) 17–33.
[117] R. Benenson, M. Omran, J. Hosang, B. Schiele, Ten years of pedestrian detection, what have we learned?, in: European Conference on Computer Vision, 2014, pp. 613–627.
[118] K. Wang, W. Zhou, Pedestrian and cyclist detection based on deep neural network fast R-CNN, Int. J. Adv. Rob. Syst. 16 (2) (2019), 1729881419829651.
[119] K. Saleh, M. Hossny, A. Hossny, S. Nahavandi, Cyclist detection in LIDAR scans using faster R-CNN and synthetic depth images, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems, ITSC), 2017, pp. 1–6.
[120] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7263–7271.
[121] R. Huang, J. Pedoeem, C. Chen, YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers, in: 2018 IEEE International Conference on Big Data, Big Data), 2018, pp. 2503–2510.
[122] M. Simon, S. Milz, K. Amende, H. Gross, Complex-YOLO: an euler-region-proposal for real-time 3D object detection on point clouds. European Conference on Computer Vision, 2018, pp. 197–209.
[123] Y. Xiang, W. Choi, Y. Lin, S. Savarese, Data-driven 3d voxel patterns for object category recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1903–1911.
[124] D. Maturana, S. Scherer, Voxnet: a 3d convolutional neural network for real-time object recognition, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, pp. 922–928.
[125] M. Engelcke, D. Rao, D.Z. Wang, C.H. Tong, I. Posner, Vote3deep: fast object detection in 3d point clouds using efficient convolutional neural networks, in:

2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 1355–1361.
[126] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652–660.
[127] D. Xu, D. Anguelov, A. Jain, Pointfusion: deep sensor fusion for 3d bounding box estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 244–253.
[128] Y. Zhou, O. Tuzel, Voxelnet: end-to-end learning for point cloud based 3d object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4490–4499.
[129] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, B. Chen, Pointcnn: convolution on x-transformed points, in: Advances in Neural Information Processing Systems, 2018, pp. 820–830.
[130] J. Chang, Y. Chen, Pyramid stereo matching network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5410–5418.
[131] Z. Liu, Z. Chen, Z. Li, W. Hu, An efficient pedestrian detection method based on YOLOv2, in: Mathematical Problems in Engineering vol. 2018, 2018.
[132] B. Li, 3d fully convolutional network for vehicle detection in point cloud, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 1513–1518.
[133] D.Z. Wang, I. Posner, Voting for voting in online point cloud object detection, Robotics: Science and Systems 1 (3) (2015), 10.15607.
[134] B. Li, T. Zhang, T. Xia, Vehicle detection from 3d lidar using fully convolutional network, in: Robotics: Science and Systems, 2016.
[135] J. Redmon, A. Farhadi, Yolov3: an incremental improvement, 2018 arXiv Preprint arXiv:1804.02767.