ISSN 1060-992X, Optical Memory and Neural Networks, 2022, Vol. 31, No. 2, pp. 123–144. © Allerton Press, Inc., 2022.

Advanced Techniques for Perception and Localization in Autonomous Driving Systems: A Survey

Qusay Sellat[a, *] **and Kanagachidambaresan Ramasubramanian**[a, **]

[a] *Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, 600062 India*

**e-mail: qusaysellat123@gmail.com*

***e-mail: kanagachidambaresan@gmail.com*

Received December 4, 2021; revised March 30, 2022; accepted April 11, 2022

Abstract—Autonomous driving research has progressed significantly in recent years. In order to travel safely, comfortably, and effectively, an autonomous car must completely comprehend the driving scenario at all times. The key criteria for a comprehensive understanding of the driving situations are proper perception and localization. The research on perception and localization of autonomous cars has increased substantially as a result of recent breakthroughs in AI, such as the wide range of deep learning approaches. However, owing to environmental uncertainties, sensor noise, and the complex interaction between the parts of the driving environment, additional study is required to achieve totally trustworthy perception and localization systems. In this survey, we demonstrate the advanced perception and localization processes in the field of autonomous driving. We show how cutting-edge approaches and practices have brought today's autonomous cars closer than ever to completely comprehending the driving environment.

Keywords: machine learning, deep neural networks, autonomous driving, object detection, semantic segmentation

DOI: 10.3103/S1060992X22020084

1. INTRODUCTION

Technology developments, particularly incredible improvements in computing technologies, artificial intelligence, and hardware capabilities, have allowed us to broaden our horizons in terms of what we can create or enhance. As a result, one application that was once thought to be a dream, autonomous driving, now appears to be on the verge of becoming a reality. Being able to travel quickly, safely, and comfortably under various types of situations and weather conditions using driver-less cars has become a main research subject in recent years due to the anticipated benefits, such as ease of use, reduced number of accidents and traffic jams, and more efficient energy management [1].

In general, each autonomous driving system comprises three main stages (Fig. 1). The first one is the perception and localization stage that derives values that describe the driving situation so that the autonomous vehicle understands the surrounding environment. In this stage, data from different sensors and digital maps are combined, processed, and then the driving situation is interpreted. The second and third stages are responsible for decision making. The second stage is motion planning, which involves deciding on a reference trajectory that represents a driving behavior and sending it to the trajectory controller. The third stage is the trajectory controller, which seeks to keep as close as possible to the reference trajectory chosen by the previous stage while taking into account comfort, safety, and efficiency criteria. Acceleration and steering commands are computed by the trajectory controller and sent to the actuators [2].

This paper provides a survey of the advanced methods for perception and localization of autonomous driving:

- Sensors used for autonomous driving.
- Current trends in the research of perception and related methodologies.
- State-of-the-art methods used for localization.

In Section 2 we list the sensors and other sources used for autonomous driving research. In Section 3, perception of autonomous driving is discussed in detail with focus on object detection and semantic seg-

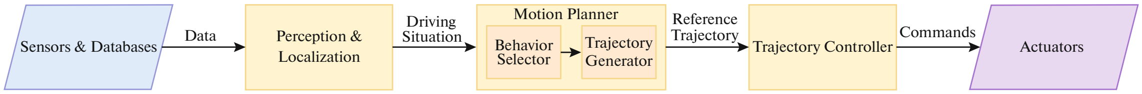

Fig. 1. Autonomous driving system general architecture (adopted from [2]).

mentation. In Section 4 we discuss localization and current research on simultaneous localization and mapping. Finally, we conclude.

2. SENSORS AND OTHER RESOURCES USED FOR AUTONOMOUS DRIVING

A summary of the various types of sensor that are used in modern autonomous driving research is shown in Table 1.

Each sort of sensor is utilized in a different circumstance than the others or in combination with others to execute a task as efficiently as possible. In the context of autonomous driving, cameras are critical. On the ego vehicle, one or more cameras are mounted to take images of the surrounding environment, which are subsequently forwarded to the perception system, which is in charge of situation interpretation. As a result, researchers employ a variety of cameras to ensure that their self-driving cars can visually comprehend their surroundings. A stereo camera is a camera that has two or more lenses, each with its own image sensor. It is used to take 3D pictures of the surrounding area. Because 3D recognition is critical for autonomous driving, a growing number of academics are working to make the ego car perceive its environment in 3D using this technology.

However, sensors such as cameras have a lot of difficulty capturing excellent images at night and in adverse weather situations such as rain, fog, and snow. The employment of a radar can ensure that the surrounding environment is monitored at all times. Radars send out pulses of radio waves and measure the locations and speeds of nearby objects based on the reflections. LiDARs transmit laser pulses rather than radio pulses. LiDAR sensors generate a detailed 3D map from reflected signals by firing invisible lasers. These signals produce “point clouds”, which depict the vehicle’s surroundings. In the dark, LiDARs can do significantly better than cameras. Ultrasonic sensors, on the other hand, emit ultrasonic pulses that are reflected by the surrounding objects. After that, the echo signals are received and analyzed. They are mostly used in tasks that need very short-range perception such as parking.

An inertial measurement unit (IMU) is an electronic device that uses a combination of accelerometers, gyroscopes, and magnetometers to detect a vehicle’s specific force, angular rate, and orientation. Recent advancements have made it possible to manufacture IMU-enabled GPS devices that can be used in autonomous cars in combination with other types of sensor (such as wheel odometry, LiDAR, and camera) to measure the ego vehicle’s position and estimate its dynamic velocity.

The main usage conditions of various types of sensor are summarized in Table 2.

3. PERCEPTION IN AUTONOMOUS DRIVING

The accuracy and efficiency of nearly all perception tasks have increased as a result of recent research breakthroughs, particularly those linked to machine learning.

3.1. Detection, Classification, and Recognition

Object detection is critical for autonomous driving in general. Deep learning technologies [17] such as convolutional neural networks (CNNs) are increasingly used in modern research to produce cutting-edge findings [18, 19]. CNNs are increasingly being deployed in object detection applications when combined with other deep learning techniques such as autoencoders, recurrent neural networks (RNNs), and Long Short-Term Memory network (LSTM). For autonomous driving perception, 3D object detection is very important. 3D scene information can be acquired via stereo imaging. A CNN is supplied with context and depth information and produces the 3D bounding boxes coordinates and poses of the detected objects using the acquired high-quality 3D object proposals. The input of the CNN consists of a variety of proposals, including object size priors, object location on the ground plane, and numerous depth characteristics that reflect free space, point cloud densities, and ground distance [20]. Object detection for autonomous driving is increasingly using hybrid deep learning architectures. A hybrid of CNN and RNN architecture can be used to create a model that understands the 3D traffic scene in general, rather than just

Table 1. A summary of sensor usage in autonomous driving research

Paper	Camera	LiDAR	Radar	Ultra-sonic	GNSS/GPS/IMU	Odometry	Notes
[3]	√	√					Two SICK LiDARs and one IBEO LiDAR are mounted on the vehicle's front bumper guard, and three SICK LiDARs are mounted on the custom-made roof rack. Three cameras are mounted behind the vehicle's windscreen. Bumps and pedestrian crossings in front of the car are detected by two cameras. To recognise lanes, a single colour camera is utilised
[4]	√	√					There are eight laser scanners and seven cameras. Three SICK-LMS291 single-line scanners and two analogue cameras with a frame rate of 25 frames/s and an image resolution of 640480 pixels were utilised for road detection from these
[5]	√	√	√		√		To characterise the surroundings around the vehicle, one Velodyne HDL-64E S2 laser was used. To receive GPS data and measure velocity, acceleration, and rotation angles, an Applonix POS LV 220 linked GNSS/IMU is used. For image processing and computer vision applications, a camera system is used. The vehicle's body is covered in 12 omnidirectional radars that monitor distant regions over a 180 m range
[6]		√					For placement optimization, two 16-channel LiDARs (Velodyne VLP-16) were utilised, and one 32-channel LiDAR (Velodyne HDL-32E) was used for comparison
[7]	√	√			√	√	Hall Effect sensors have been installed on each wheel of the vehicle to enable basic offline localization inside a low-level system. The IMU utilised is an Xsens MTi-G-710 with a global navigation satellite system (GNSS) running at 50 Hz. LiDARS (two) (SICK LMS111-1010 and ibeo LUX 4). To conduct visual navigation tasks, a pair of FLIR Blackfly GigE cameras are placed on the vehicle's frame
[8]		√			√		A Velodyne HDL-32E is utilised, as well as a GPS/IMU suite
[9]	√	√			√	√	A variety of onboard sensors, including wheel-speed sensors, a steering-wheel-angle sensor, and a yaw-rate sensor. To identify distant objects, two multilayer laser scanners (Ibeo LUX) are placed on the front bumper. Four single-layer laser scanners (LMS 151) are mounted on each corner of the ego vehicle to identify nearby objects. To identify obstacles, two single-layer laser scanners (LMS 291) have been placed on the roof. On the interior of the windshield, one colour camera and three monocameras are placed to identify and categorise mission objects. A Real-Time Kinematics GPS (RTK-GPS) and a Differential GPS receiver are both utilised (DGPS). The vehicle's dynamic motion is estimated using IMU
[10]	√	√			√	√	To detect the car's dynamic motion, the ESC system uses various types of vehicle motion sensors (a steering angle sensor, wheel speed sensors, and a yaw rate sensor). For initialization and GPS measurement updates, positioning data from a low-cost GPS (Ublox EVK-6T) was used. Three types of cameras were installed in the test vehicle: a front camera, a rear camera, and an AVM system
[11]	√				√	√	A Camera to capture front images, a MEM gyroscope for yaw-rate measurement, a GNSS receiver for initial position estimation. Odometer and IMU are used for mapping
[12]	√	√	√	√	√	√	To compare three sensor plans, all sorts of sensors are employed

Table 1. (Contd.)

Paper	Camera	LiDAR	Radar	Ultra-sonic	GNSS/GPS/IMU	Odometry	Notes
[13]	√		√		√		There is a millimeter-wave radar, a GPS system, a video surveillance system, and a vehicle gyro
[14]	√	√			√		At the front of the car, a stereo camera and a four-layer LiDAR are fitted. An RTK-DGPS receiver is used to locate the car used in the testing
[15]	√	√	√		√		Multiple cameras, a 360 deg 64-beam high-definition LiDAR (Velodyne HDL64e), millimeter-wave radars, and an Applanix POS LV 220 (fused GPS + IMU positioning)
[16]	√	√	√		√		To estimate the vehicle's pose, a NovAtel virtual reference station-aided GPS is mounted on the roof. The GPS is used in conjunction with the IMar IMU to improve real-time kinematic performance. To detect and track nearby cars, a 77 GHz long- and midrange Delphi multimode ESR radar, two multilayered LD-MRS LiDARs connected to the front edge, two two-dimensional (2D) SICK LMS LiDARs attached to the rear edge, and a 3D Velodyne HDL-64 LiDAR mounted to the roof are utilised. Traffic signals are detected using a vision sensor (camera) mounted in the middle of the windshield.

certain elements of it, and makes driving strategy recommendations based on multimodal data [21]. Here, a CNN is employed with an LSTM (Fig. 2). In terms of general object detection, feature fusion, and low-level semantic understanding, the proposed model is said to outperform the current state-of-the-art.

It might be difficult to interpret and comprehend collected data in particular traffic scenarios. Some academics have developed a framework to address these traffic-related issues [22]. The framework can identify and distinguish items in a driving scenario, as well as anticipate the intentions of pedestrians in that scene. They employed YOLOv4 [23] to recognize ten different item types. They also used Part Affinity Fields [24] to determine the pedestrians' poses. They leverage Explainable Artificial Intelligence (XAI) technology [25] to explain and help with the outcomes of the risk assessment task. Finally, they built an end-to-end system that allowed them to combine multiple models with the greatest precision. [26] introduced YOLOv4-5D, a deep learning-based one-stage object detector (no region proposal stage required) geared for autonomous driving. Their model is a modified version of YOLOv4. It has an accuracy comparable to those highly accurate, two-stage models [27, 28], while also has the ability to run in real-time. The backbone and feature fusion phases of YOLOv4 are modified in YOLOv4-5D to meet the precision and real-time demands. Even for greater image resolutions, the suggested technique now runs properly and in

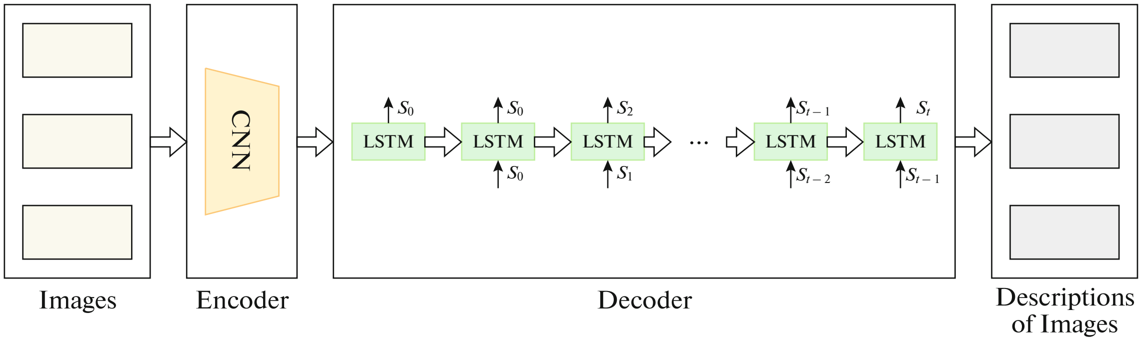

Fig. 2. The overall system architecture in [21].

Table 2. A summary of sensor usage in autonomous driving research

Sensor	Usage	Best for	Not for
Camera	Object detection (cars, road signs, traffic signs, etc.) [3, 9, 10, 15, 16], semantic segmentation [4, 7], visual odometry [7], localization (positioning) [7, 10, 11]	Perception in normal weather conditions in daytimes. Can be used to detect detailed and small objects. Can be used to detect colors. Best for object detection	Adverse weather conditions or nighttimes
LiDAR	Object and obstacle detection [3, 7, 9, 15, 16], semantic segmentation [4, 8], localization [7, 10]	Perception and localization in Normal weather conditions at all times. Usually combined with other sensors	Adverse weather conditions Color detection Long range Low price requirements
Radar	Scanning of distant areas [5], perception [15], object detection [16]	Perception in all weather conditions and at all times. Cheap choice	Detailed and small object detection
Ultrasonic	Low speed parking [12]	Perception in very short range (mainly used for parking). Cheap choice	Medium and long ranges
IMU/GPS/GNSS	Localization [5, 7, 9–11, 14], vehicle dynamic motion [9], vehicle pose [16]	Localization	High accurate localization (needs to be combined with other types of sensor such as LiDAR and camera)
Odometry	Linear and rotational velocities [7], wheel speed and yaw rate [9, 10]		

real-time thanks to the changes. For instance, the feature fusion module of the new model is able to handle five scales of detection, and the proposed network pruning method based on sparse scaling factor reduces the redundant convolutional channels, leading the complexity of the model to drop significantly. In order to determine the location and orientation of objects, costly components such as 3D LiDARs and stereo vision are frequently employed. Recent works, such as [29], proposed a less expensive method for object recognition that relies only on images from a single RGB camera. The depth information acquired from contact points of the respective objects and the ground plane is used to estimate the 3D position of obstacles on the road. Deep neural network approaches for monocular 3D object detection and monocular depth prediction leverage the calculated positions.

When there is a large object scale variation, CNNs usually struggle. To overcome this problem, some study suggests that deconvolutional and fusion layers be added to retrieve context and depth information. Furthermore, some difficulties, such as object occlusion and poor lighting, can be overcome by employing non-maximal suppression (NMS), which can be applied across object proposals at several feature sizes [30]. In order to conduct both single-class and multi-class classification on imbalanced categories, researchers recommended employing a multi-level cost function (Fig. 3). A deep data integration technique can also be utilized for hard or rare samples [31]. The atrous spatial pyramid pooling (ASPP) [32, 33] was used by some researchers to solve the problem of poor performance of object detection systems when there are small objects, low illumination, or blurred outline in the received camera images. This approach extracts features from multiple scales so that object detection accuracy is higher with the same number of parameters [34]. CNNs can further be used for a variety of different applications, including pedestrian movement detection with pedestrian movement direction recognition [35].

Lane detection and recognition, as well as the detection and recognition of road surface markings, are essential tasks for autonomous driving. They should be performed in real-time while taking into account the various angles and sizes of the discovered objects. Unsupervised learning techniques based on a live feed from a camera mounted on the dashboard of a moving car are proposed to satisfy these constraints. A spatio-temporal incremental clustering algorithm is used with curve-fitting to detect lanes and road surface markings at the same time. This method of learning is appropriate for a variety of object kinds and

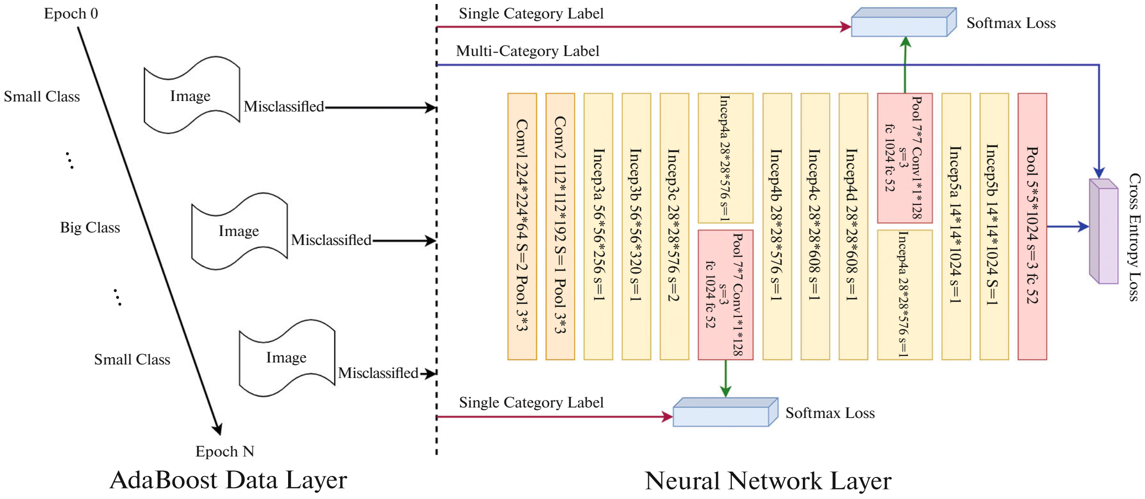

Fig. 3. Proposed network architecture in [31].

scales [36]. Others [19] created a hybrid approach that integrates CNN and RNN principles and is said to outperform previous approaches in lane detection, particularly in difficult scenarios.

As illustrated in Fig. 4, the proposed technique extracts information from many image frames (instead of using a single frame as most previous research used to do) using a fully-connected CNN that produces low-dimension feature maps, and then feeds the derived features from the successive frames to an RNN (An (LSTM) model is used), and finally, the resulted features are fed into another CNN that is used for lane prediction. This way, if a frame is unable to give sufficient information for lanes to be accurately detected, for example owing to heavy shadows or severe road mark degradation, the information gathered from prior frames may be applied to estimate lane information. Deep learning was also applied for lane detection and fitting in more recent works such as [37]. However, as long as learning relies on input data, the data should be diverse enough to allow for the appearance of target objects in a variety of sizes and shapes.

Other detection issues, such as curb detection, require further investigation. Curb detection is a challenging task, especially at intersections where the curbs become fragmented, and the computational cost of the detection is significant, making it difficult to perform in real-time. A technique for detecting curb areas has been proposed in [38]. The point cloud data that describes the driving scene is initially collected using a 3D LiDAR sensor. The point cloud data is then analyzed to distinguish between on-road and off-road regions. Following that, a sliding beam [39] method is applied to choose the road area using off-road data. Finally, a curb detection algorithm is used to pick curb locations based on the chosen road area. Instead of performing online curb detection by taking advantage of video cameras or 3D LiDARs that may yield erroneous information in poor lighting or bad weather, [40] developed an offline curb detection approach that employs aerial images to fulfil these tasks. Their approach is built on cutting-edge learning techniques including deep learning and imitation learning.

The ability to interpret a situation can be improved by combining vision and LiDAR data. To identify the drivable area more accurately, several researchers proposed combining data fusion with an optimal selection approach [4].

Then, depending on the automated classification of the optimum drivable area, lane detection is achieved selectively (Fig. 5). It is stated that the suggested system is effective and dependable in both structured and unstructured traffic situations, and that no human switching is required. A higher level of data fusion is suggested this time among vision, LiDAR, and radar data to conduct perception in all-weather situations, including hard illumination conditions like dim light or total darkness. The fused data is uniformly aligned and projected onto the image plane, and the output is fed into a probabilistic driving model for motion planning model (Fig. 6). The proposed approach is said to operate well in high-traffic areas and in adverse weather conditions [41]. In [42] the Multi-Sensor Fusion Perception technique (MSFP) is proposed. MSFP is based on AVOD [43], a 3D object proposal system that fuses multimodal information derived from the aggregation of RGB and point cloud data. To produce respected features, the MSFP architecture first analyses camera and LiDAR data. The characteristics are then fed into an introduced Region Proposal Network (RPN), which creates 3D region proposals.

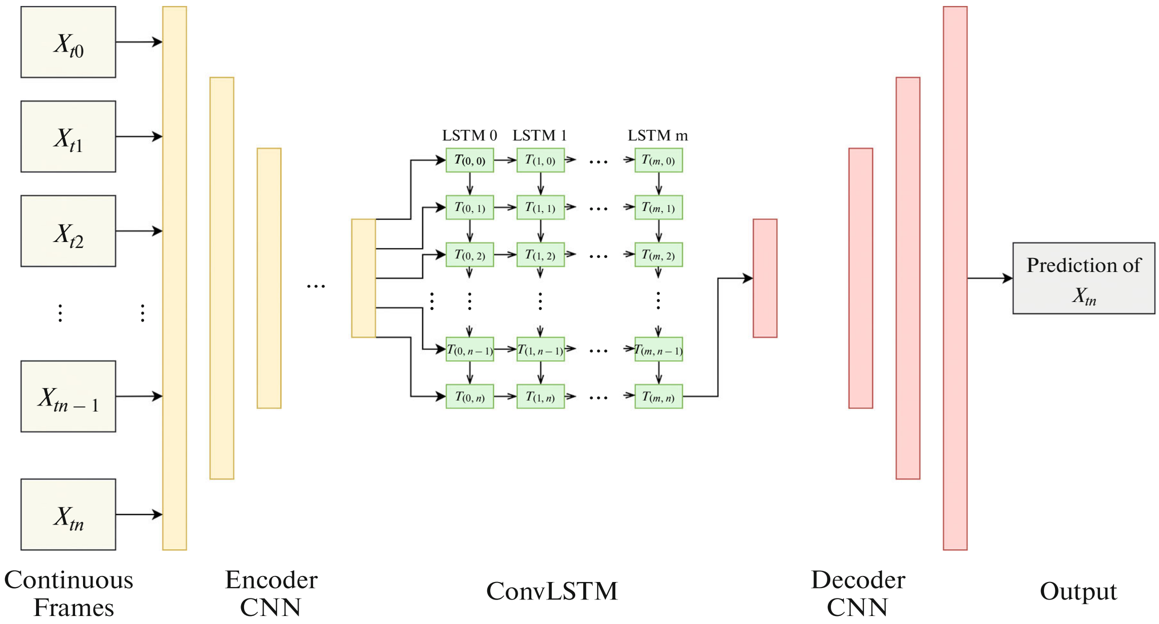

Fig. 4. Architecture of the proposed network in [19].

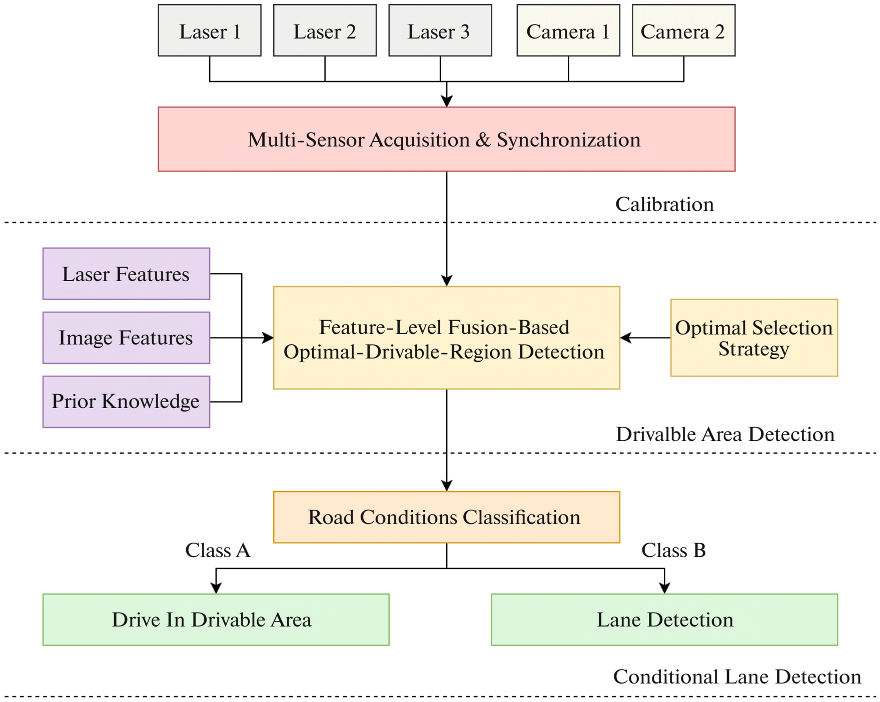
Laser 1
Laser 2
Laser 3
Camera 1
Camera 2
Multi-Sensor Acquisition & Synchronization
Calibration
Laser Features
Image Features
Prior Knowledge
Feature-Level Fusion-Based Optimal-Drivable-Region Detection
Optimal Selection Strategy
Drivalble Area Detection
Road Conditions Classification
Class A
Class B
Drive In Drivable Area
Lane Detection
Conditional Lane Detection

Fig. 5. System architecture in [4].

Despite the fact that images from video streams include a wealth of spatial and temporal information on object poses and scales, 3D multi-object tracking (MOT) in complicated driving settings with significant occlusions among numerous tracked objects is difficult. To address this issue, a variety of approaches have been developed. Some of these solutions [44, 45] assume that all of the frames (past, current, and future ones) are accessible and process all of them to generate object tracking information (global tracking). Those methods are accurate, but they don't meet real-time requirements. Other techniques [46, 47] solely incorporate information from current and previous frames, with no knowledge of future frames

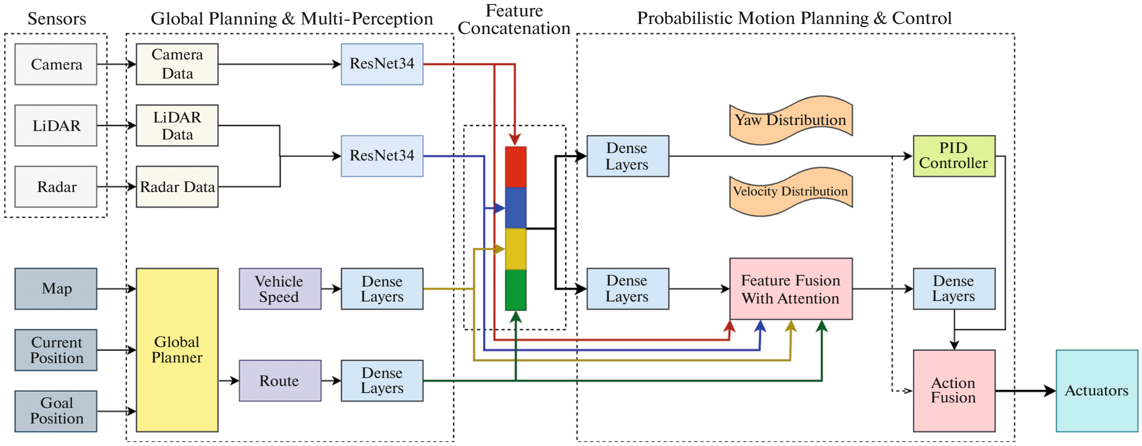

Fig. 6. The architecture of our probabilistic motion planning network (PMP-net) in [41].

(online tracking). Real-time tracking is possible using online tracking methods. However, further effort is required to provide very precise tracking in real-time. For 3D MOT, [48] presented a real-time online technique.

They employed deep learning to jointly perform both object tracking and detection by taking into account the spatial-temporal characteristics of the driving scene, resulting in a more accurate MOT than traditional detection approaches that rely on bounding boxes and association algorithms [49, 50]. The driving behavior awareness is derived from learned object characteristics adopting Deep Layer Aggregation (DLA) [51] as the network backbone and considering 2D center point, depth, rotation, and translation in parallel.

Road sign recognition is a crucial element of the perception of autonomous driving systems. It enables the ego car to gain a comprehensive understanding of the driving environment and situations. Many recent studies have presented effective ways to deal with difficulties connected to road sign recognition as a result of the rapid development of deep learning techniques related to object detection. A CNN design was developed by several academics to detect speed limit road signs, which are among the most challenging things to recognize in the US traffic sign set. They claim that their method improved the area under the precision–recall curve (AUC) for detecting speed limit signs by more than 5% [52]. Others applied advanced CNN techniques like depthwise separable convolution (DSC) [53] to reduce the complexity of the recognition process and speed it up. They increased accuracy even more by breaking down the problem of traffic sign identification into two parts: traffic sign classification (TSC) and traffic sign detection (TSD), and proposing an efficient network for each (ENet for TSC and EmdNet for TSD) [54]. Others developed a framework for both symbol and text-based road sign classification, refinement, and detection. They used the Mask R-CNN architecture [55] for detection and refinement and the CNN architecture for classification to obtain high accuracy [56]. For classification of degraded traffic signs, a multiscale CNN with dimensionality reduction is also implemented [57]. Furthermore, some studies claim to get higher accuracy outcomes by focusing on certain regions of interest. Every traffic image is preprocessed to get these regions, and then a CNN is applied to recognize the signs [58]. Many advanced road traffic sign recognition research ideas are listed in Table 3.

3.2. Semantic and Motion Segmentation

In order to accomplish a scene understanding task, a self-driving car is required to know the segment label under which each point of the received sensor signal is classified (e.g. road, car, building, pedestrian, etc.). This problem is known as semantic segmentation. Many tasks that were formerly difficult or impossible to accomplish have become achievable or even easy in the era of deep learning. As a result of these advancements, study in a variety of fields has flourished. Semantic segmentation was one of these areas that had limited success until recently, when a lot of research has been done and significant results have been obtained. The development of efficient point-wise classification is predicted to assist several industries that need precise semantic segmentation, such as autonomous driving [67, 68].

CNNs and autoencoders [17] are the most common deep learning architectures used for semantic segmentation. Convolutional autoencoders are hybrid models that combine CNNs and autoencoders to per-

Table 3. Advanced road traffic sign recognition research ideas

Paper	Topic	Proposed approach	Dataset	Result
[52]	Speed limit road signs recognition	Preprocessing technique of Contrast-limited Adaptive Histogram Equalization (CLAHE) [59]. CNN for classification. Region Proposal Generation [60], R-CNN [61], and Hard Negative Mining [62] for detection	The US traffic sign set – the LISA-TS extension [63]	The area under precision–recall curve (AUC) of the speed limit sign detection is improved by more than 5%
[54]	Efficient traffic sign recognition	CNN with advanced techniques such as depthwise separable convolution (DSC) to increase the speed of the algorithm and data mining and multi-scale operation to increase the accuracy	German Traffic Sign Recognition benchmark (GTSRB) [64]	Accuracy of 98.6% with a small number of parameters
[56]	Classification, refinement, and detection of both symbol and text based road signs	Mask R-CNN architecture for detection and refinement and CNN architecture for classification	German Traffic Sign Recognition benchmark (GTSRB)	High accuracy achieving comparable performance with the state of the art
[57]	Classification of degraded traffic signs	Multi-scale CNN with dimensionality reduction	A subset of the German Traffic Sign Recognition benchmark (GTSRB).	Good performance comparable to the state of the art
[58]	Efficient traffic sign detection and recognition	Regions of interest are determined with efficient and parallelized preprocessing of every traffic image, after which a CNN model is applied for traffic sign recognition. YOLO [65] architecture is used as a baseline detector	DFG traffic sign dataset [66]	Real-time detection in high-definition images with high recognition accuracy

form research with far more accuracy and efficiency than before. A convolutional autoencoder is an autoencoder with convolutional and deconvolutional layers as encoder and decoder, respectively. To put it another way, CNN models serve as the “backbone” structures for the autoencoders that are employed. To construct their model, the majority of contemporary studies have employed some form of convolutional autoencoder architecture. One of the first attempts in this direction was FCN [69]. It employed a fully convolutional autoencoder architecture and accomplished semantic segmentation with a huge number of parameters, making it unsuitable for real-time usage. It was also one of the first tries to eliminate fully connected layers. SegNet [70] and SegNet-Basic [71] used VGG architecture [72] as a backbone for the encoder and the decoder. They used the pooling indices of the encoder part to upsample data in the decoder part. Other designs, such as UNet [73], improved segmentation accuracy by employing skip connections between the encoder and decoder, as well as other approaches like data augmentation. Although the above-mentioned models and other architectures such as PSPNet [74], Dilated [75], and DeepLab [76] increased the accuracy of semantic segmentation models, there was still a need for a less complex architecture to perform real-time semantic segmentation because some fields of application (e.g. autonomous driving and robotics) require very accurate semantic segmentation with a minimum amount of processing time. This was not a simple task because capturing the complexity of the received images and point clouds necessitated a large amount of trainable parameters, and developing light segmentation models without reducing segmentation accuracy was a challenging mission. Some models were designed with a smaller number of parameters. FPN [77] and LinkNet [78] and other super lite models such as ApesNet [79], Enet [80], ESPNet [81], ESCNet [82] and EDANet [83] tried to minimize the number of parameters so that the semantic segmentation can be done in real-time or implemented for embedded systems. Despite the fact that these models provided practical solutions to satisfy the real-time condition, crucial applications such as road scene understanding in autonomous vehicles need much more segmentation accuracy.

Road detection and segmentation is a very important task in autonomous driving. The live feed from the camera can be processed in order to extract valuable information. From the driving scene, a saliency image map [84] can be obtained to represent visual perception by extracting spatial, spectral, and temporal information from the live stream and then applying entropy driven image-context-feature data fusion. Both for still objects and non-station objects, the fusion output of the last step includes high-level descriptors. After segmentation and object detection, road surface regions are selected by an adaptive maximum likelihood classifier [85]. To allow the ego vehicle to recognize the road area, some researchers employ supervised learning approaches. However, several recent studies recommend using semisupervised learning methods to solve the difficulty of a huge quantity of data necessary for training purposes. Advanced deep learning approaches like generative adversarial networks (GANs) [86] and conditional GANs [87] are said to produce excellent outcomes [88]. In dynamic situations, neural networks may also be utilized to gather real-time road static information. For this purpose, several studies suggest light neural networks. These kinds of efficient multi-task networks can perform occlusion-free road segmentation, dense road height estimation, and road topology recognition all at the same time. In order for these networks to perform successfully, the loss function must be chosen carefully [89].

Semi-supervised learning approach using CNN architecture can also be used for semantic segmentation of 3D LiDAR data. A CNN model can be trained using both supervised samples that are manually labeled and pairwise constraints [8]. However, recent research tends to benefit from the significant improvement of deep learning technology. For real-time RGB-D semantic segmentation, RFNet [90] is a designed fusion network with a CNN and an autoencoder. Two distinct branches are employed in the encoder component of the proposed model to extract features for RGB and depth images independently. The semantic segmentation function is completed using a simple decoder with upsampling modules and skip connections. [91] developed a deep learning-based end-to-end model that applies a sensor fusion approach. In order for semantic segmentation to be completed effectively, visual information from the camera is fused with depth information from the LiDAR. This allows for accurate scene interpretation and vehicle control. FSFnet [92] is a deep neural network that was proposed as an accurate and efficient semantic segmentation model for autonomous driving. The general architecture is made of convolutional and deconvolutional layers with the lightweight ResNet-18 [93] as a backbone (Fig. 7). They created the feature selective fusion module (FSFM) for multiscale and multilevel feature fusion, which collects valuable information from respected feature maps both spatially and channelwise, employing adaptive pooling layers to make a balance between accuracy and efficiency. To build a comprehensive understanding of the driving scene while also considering objects of various scales and sizes, context aggregation module (CAM) was introduced. CAM combines both global and multiscale context information efficiently using the lightweight ANNN [94]. PointMoSeg [95] is another advanced segmentation model designed to distinguish moving obstacles in the driving scene. PointMoSeg is an end-to-end model in which a convolutional autoencoder inspired by ResUNet [96] is adopted to online acquire point cloud data from a 3D LiDAR and process it to produce a segmented current frame with segments representing moving obstacles. The design of the proposed architecture benefits from modern deep learning technologies such as sparse tensors [97] and sparse convolutions [98], and introduces two novel modules - a temporal module and a spatial module – that are used to efficiently and effectively extract the features and context information of the moving obstacles. There is no need for flat road assumption or ego-motion estimation since the suggested end-to-end architecture assures that high performance is generalized to scenarios when the driving road has some slope or the driving environment is urban. Some other recent works such as OmniDet [99] proposed to train the deep learning model jointly on a variety of perception tasks including semantic and motion segmentation, object detection, and depth estimation. OmniDet is claimed to perform multi-task surround-view fisheye camera perception in real-time.

Advanced semantic segmentation research ideas are listed in Table 4.

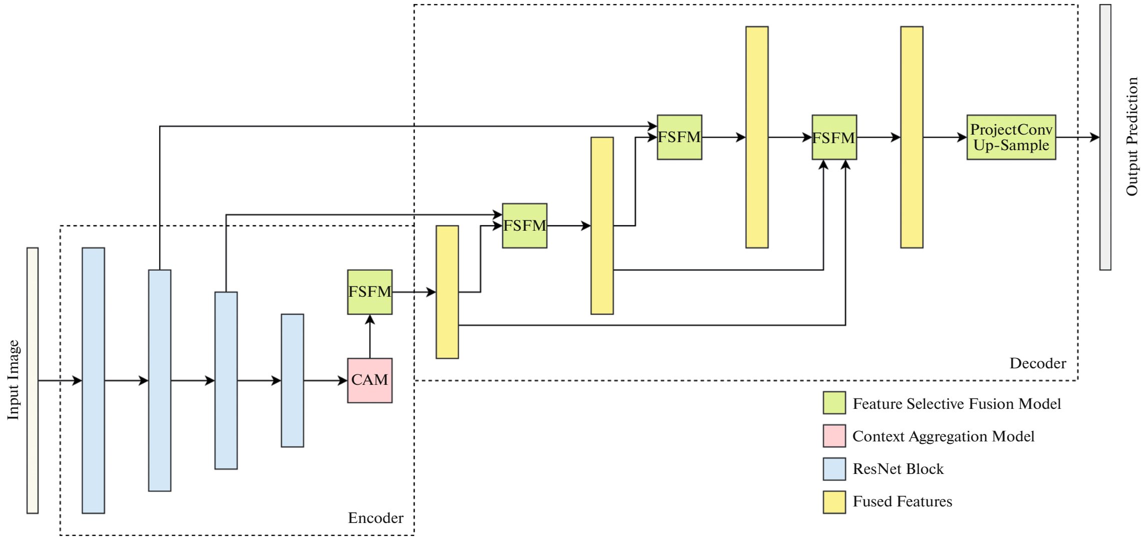

Fig. 7. The architecture of FSFnet [92].

3.3. Other Perception Related Research

Many aspects of perception have improved as a result of the recently created technology. Deep learning algorithms, for example, can be used in data acquisition to overcome the problem of high transmission bandwidth required for delivering high quality image frames in autonomous driving applications. Using deep learning algorithms, only the most essential information bits can be picked, and the bandwidth required to send this information is considerably lower than when the entire data is broadcasted. These deep learning algorithms are incorporated in the camera, allowing it to record only the task-critical data required for the autonomous car's correct operation. It is difficult to design a smart camera system that also works in real-time. To do this task properly, several researchers [110] recommended using a 3D-stacked sensor architecture with a Digital Pixel Sensor (DPS) [111, 112]. So, sensor data from an autonomous car can be extremely huge. Compressing, storing, and transmitting LiDAR point cloud data in real-time is challenging. Noise is also a major problem. Deep learning provides a solution to the previously mentioned data volume issues. For point cloud data compression, a real-time deep learning technique based on the UNet architecture [73] has been suggested. UNet reduces the point cloud data stream's temporal redundancy. The data from LiDAR point clouds is converted into a sequence of video frames. After that, certain frames are utilised as reference frames, and the other frames are interpolated using a UNet architecture. Temporal redundancy may be significantly reduced in this approach. A padding technique is also employed to mitigate the harmful effects of noise [113]. ReViewNet [114] is another example of deep learning in action. ReViewNet is described as an efficient deep learning-based solution for dehazing driving environment images, which is a critical prerequisite for perception. Their algorithm uses 12 channels representing a quadruple-color space (RGB, HSV, YCrCb, and Lab). It also has a multi-look architecture that provides the ability to further control the dehazing process and add more learning structs such as multi-output (two MSE losses are used) and spatial pooling block (to extract features in shallow layers without going deeper). This end-to-end architecture outperforms other state-of-the-art works, even the models that use advanced deep learning architectures such as GANs [115–117], allowing the ego car to remove haze from images so they are more realistic and detailed.

An autonomous system's sensors can be subjected to a variety of assaults. Because reliability is so important in autonomous driving, any autonomous vehicle design must be able to withstand many types of assaults. Some academics recommend utilizing a 3D data hiding approach based on 3D-QIM (quantization index modulation) [118–121] on point cloud data collected by 3D LiDARs, for example. Data integrity is guaranteed at the decision-making level by introducing a binary watermark into the point cloud data at the sensor level. Advanced methods can also aid in the analysis of some LiDAR parameters that have previously been unavailable for analysis. Spot pattern, waist, and divergence, in particular, may be investigated using data obtained from spot reflection on flat surfaces. This is useful for analysing the LiDAR's performance [122].

Table 4. Advanced semantic segmentation research ideas

Paper	Topic	Proposed approach	Dataset	Result
[88]	Road detection	A semisupervised learning (SSL) road detection method based on generative adversarial networks (GANs) and a weakly supervised learning (WSL) method based on conditional GANs	KITTI ROAD benchmark [100]	The Max F1-measures of SSL method and the WSL method are 95.53 and 95.70%
[90]	RGB-D real-time semantic segmentation	CNN and autoencoder	Cityscapes [101]	MIoU 72.22%
[8]	Semantic segmentation of 3D LiDAR data in dynamic scenes	Semi-supervised learning approach using a CNN architecture	A dataset is generated using 3D LiDAR data from a dynamic campus scene	The combination of a few annotations and large amount of constraint data significantly enhances the effectiveness and scene adaptability, resulting in greater than 10% improvement
[89]	Real-time occlusion-free road segmentation, dense road height estimation and road topology recognition in dynamic environments	CNN and autoencoder	A multi-task dataset, named MultiRoad, is built based on SemanticKITTI dataset [102]	Acc 97.4%, MIoU 84.1%
[103]	Real-time and accurate semantic segmentation of 3D LiDAR data	Graph, clustering, and Gaussian process regression [104]	Dataset is generated from a campus road at Seoul National University. For more experiments, two KITTI datasets are tested [105]	tracking accuracy increased by 11.4%
[67]	Efficient semantic segmentation	CNN and autoencoder: asymmetric encoder–decoder structure	CamVid [106], CityScapes, Gatech [107], KITTI Road Detection, and KITTI Semantic Segmentation	Comparable performance to the state-of-the-art
[68]	Real-time semantic segmentation	CNN and autoencoder	Cityscapes	71.8% mIoU on validation set, 69.3% on test set, 100+ frames per second (FPS) at resolution 640 × 360 on NVIDIA 1080Ti
[78]	Semantic segmentation	CNN and autoencoder	Camvid	IoU 68.3%, mIoU 55.8%.
			Cityscapes	IoU 76.4%, mIoU 58.6%.
[80]	Semantic segmentation	Lite CNN and autoencoder	Camvid	IoU 68.3%, mIoU 51.3%

Table 4. (Contd.)

Paper	Topic	Proposed approach	Dataset	Result
[92]	Efficient and accurate semantic segmentation	Deep neural network with a novel feature selective fusion module (FSFM) and a novel context aggregation module (CAM)	Cityscapes	MIoU 77.1%
			Camvid [108]	MIoU 75.1%
[95]	Moving-obstacle segmentation	CNN and autoencoder with novel temporal and spatial modules	nuScenes dataset [109]	Acc 67.81%, IoU 47.89%

4. LOCALIZATION IN AUTONOMOUS DRIVING

Localization is the process of determining the precise location of an autonomous vehicle by employing systems such as GPS, dead reckoning, and road maps. The autonomous vehicle's ability to perform well in the localization process is critical since it keeps the vehicle moving along the correct planned path. In the vast majority of situations, centimeter-level precision is necessary. Localization is challenging, particularly in urban areas where the signal is constantly disrupted by nearby buildings and moving objects. Mapping is the process of producing accurate maps of driving regions that may be used for navigation. In most cases, conventional map resources are insufficient to function securely. As a result, high-resolution HD maps of the environment are necessary. The simultaneous localization and mapping (SLAM) principle has recently become more popular and it is now used to derive most practical localization methods as it combines signals from various sensors to build a complete understanding of the object's location inside the surrounding environment, and therefore, it is more accurate than traditional localization methods that depends only on GPS and fixed anchors in the external environment [12, 123]. SLAM refers to the method of constructing and updating a map of the driving environment while also identifying the precise location of the autonomous vehicle inside that map. Conventional methods of SLAM exploit onboard sensors to accomplish localization. Visual SLAM and LiDAR SLAM are the most common ways [124].

Traditional localization methods use a GNSS receiver in collaboration with other on-board sensors such as IMU. The received information from those sources is then processed by methods such as Kalman Filter [125, 126], Hidden Markov Model [127, 128], and Bayesian Network [129]. In [9], they use an interacting multiple model (IMM)-filter-based information system [130, 131] for localization. They combined GPS data with data from other sources (wheel speed sensor, steering angle sensor, and yaw rate sensor), as per current study trends. This integration is necessary since GPS data is influenced by a variety of factors, causing localization to drift. The localization algorithm can adjust to changing driving circumstances thanks to the usage of the IMM filter. The GPS-bias technique is also included in order to increase the accuracy and reliability of the localization process. In [11], they propose a low-cost localization method. The localization approach is said to be more accurate than earlier efforts since it uses lane detection, dead reckoning, map-matching, and data fusion techniques. A lane marking detector with a short range is used to detect the lanes. The precise vehicle location is established by creating a buffer called Back Lane Markings Registry (BLMR) online, and integrating this registry with information received from gyroscopes, odometry, and a prior road map.

Buildings and other artifacts, as previously noted, can create severe perturbations in the received signals required for localization, affecting conventional localization systems such as GPS and dead reckoning. In general, these systems are very accurate. However, owing to the noise of the dead reckoning sensors and the inaccuracy in the dead reckoning integration, reliable localization cannot be accomplished in metropolitan areas during long periods of GPS outage. To address these issues, some researchers proposed a precise Monte-Carlo localization technique based on a precise digital map and perceptual data fusion of motion sensors, low GPS, and cameras [10]. Particle filtering [132] is used for combining several types of sensor probabilistic distributions (as they are non-Gaussian distributions). A method for modelling sensor noise is also provided in order to improve localization performance.

Despite the fact that numerous visual SLAM algorithms have achieved accurate and robust localization [133, 134], real-time restrictions must yet be investigated. [135] demonstrated an effective visual SLAM capability that eliminates the problem of the frame losses associated with real-time performance.

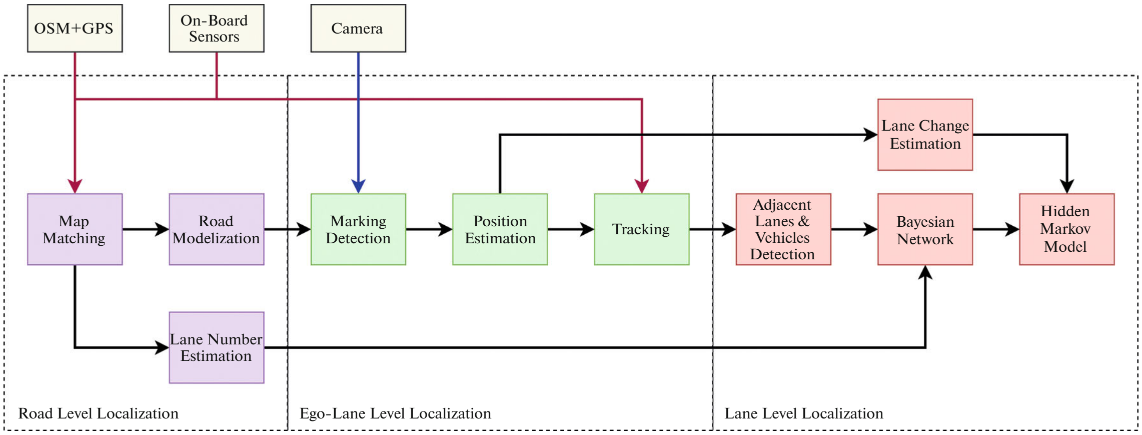

Fig. 8. The localization system architecture proposed in [140].

A four-threads architecture was implemented. Frontend thread is implemented to recognize the pose of the camera in real-time. The mapping thread then analyses additional keyframes in order to construct 3D map points while keeping drift to a minimum. The keyframe poses and 3D map point placements are refined by the state optimization thread. An online bag-of-words based loop closer thread [136, 138] is used specifically to allow long-term localization in large-scale maps. [139] provided a map-saving mechanism to extend ORB-SLAM 2 that was presented in [134], leading localization error to drop. [140] proposed an innovative end-to-end framework that takes input from a GPS receiver, OpenStreetMap (OSM) datasets [141], IMU and a front camera, and then, takes full advantage of probabilistic modeling (specifically Hidden Markov Model (HMM) and Bayesian Network (BN)) in order to conduct all aspects of localization (road level, lane level, and ego-lane level localization) (Fig. 8).

Many approaches were proposed for LiDAR SLAM. For example, LOL [142], and HDL-Graph-SLAM [143] used Iterative Closest Point (ICP) [144] that tries to minimize the Euclidean distance between point pairs in order for the transformation matrix to converge. The computing cost of these approaches is significant, and they are sensitive to environmental factors. Some other models such as LOAM [145], and LeGO-LOAM [146] used the extracted geometric features to accomplish convergence.

Other conditions such as snowy weather and wet road surfaces reduce the quality of the received signals, lowering localization accuracy. For example, in case of using LiDAR sensors, LiDAR reflectivity drops when the road surface is wet. Also, the probable snow lines may deform the expected road context in LiDAR images. These issues can be overcome by employing advanced techniques such as principal component analysis (PCA) [17] to gain a better understanding of the surrounding environment, and then using the leading PCA components to support an accumulation approach aimed at increasing the density of LiDAR data. In addition, for snow situations, an edge matching technique that matches previously recorded map pictures with online LiDAR images is developed to ensure correct localization [5].

Occupancy grid is an important way to represent the driving environment, where each cell of the grid contains information characteristics of the objects inside it [147]. The ground reflectivity inference grid is a variant of occupancy grid used in autonomous driving, in which the ground's reflection is linked with a position in the grid proportional to the projection measurement of its range into the grid's reference frame [148]. The authors of [149] recommended adopting a novel grid representation that incorporates reflectivity edges generated by combining distinct reflectivity gradient grids from fixed laser views. This representation is claimed to enable laser-perspective and vehicle motion invariance, negating the requirement for further post-factory laser reflectivity calibrations and facilitating many processes including localization.

Particle filtering was also used in [150] where they proved that accurate localization in large-scale environments can be achieved by simple input components including light-weight satellite and road maps from the internet, car odometry and a 3D LiDAR. [151] created a method to deal with the challenging 3D localization of LiDAR point cloud data. Data accumulation is used in their solution. Due to the high complexity of the problem, the processing step of 3D point cloud data of successive frame steps may need a large number of computations. They kept various data structures among those frames by employing data accumulation, which allowed the time steps to save a lot of computations. The information about recently occupied voxels is collected by filtering a stream of 3D point cloud data via a proposed Dynamic Voxel

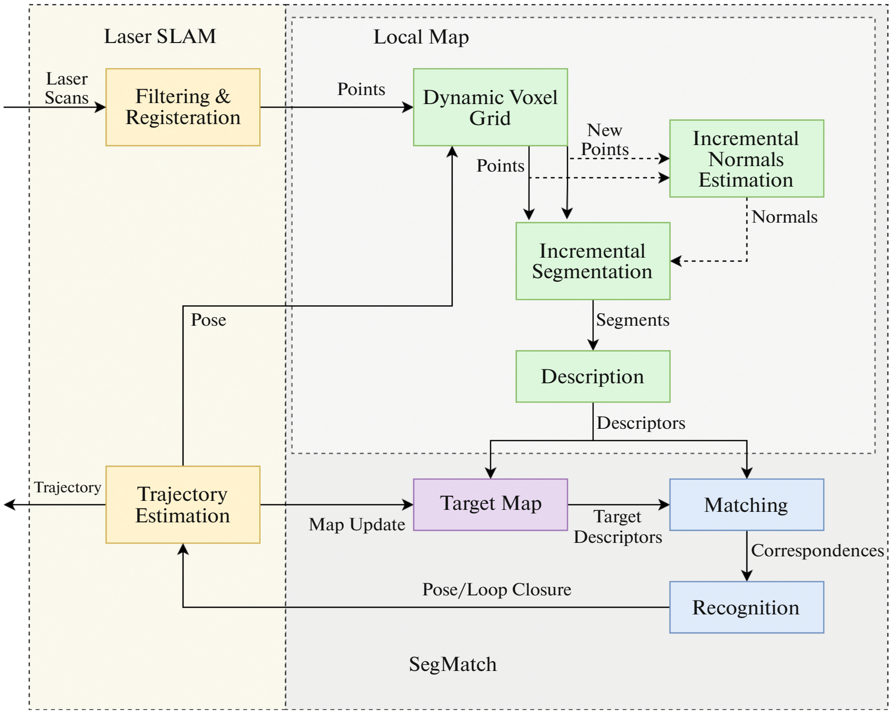

Fig. 9. System architecture in [151].

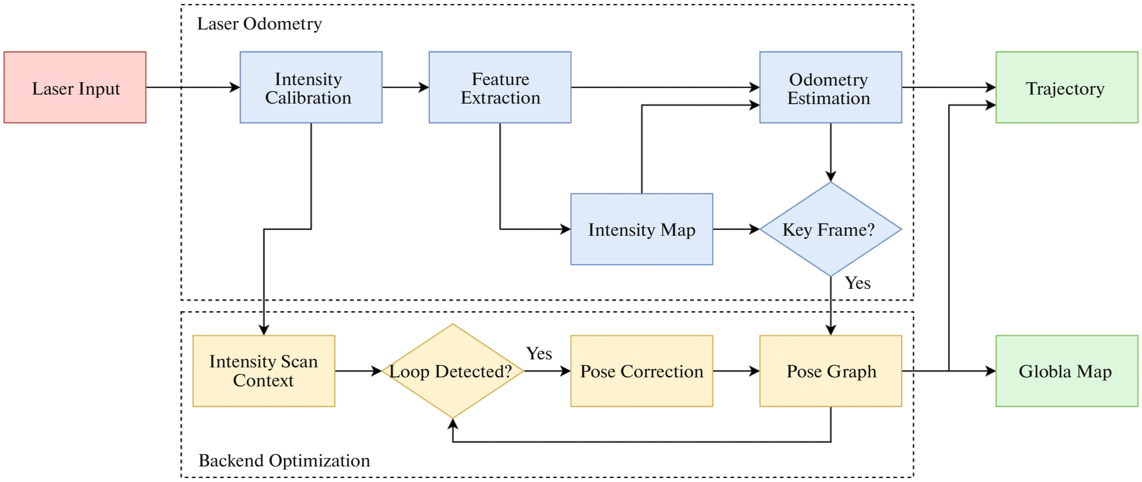

Fig. 10. System architecture in [124].

Grid (DVG). The first phase's output is then sent into an incremental region growing segmentation module [152], which combines it with the already clustered data (Fig. 9).

Even though many research works took advantage of intensity measures to extract intensity features and used them for accurate SLAM [153, 154], most of these proposals require huge data collection effort, are not consistent when the surrounding environment changes, or are applied only for 2D localization. [124] combined intensity features with geometric features in order to perform LiDAR SLAM accurately and reliably. An intensity-based loop closure detection (frontend odometry estimation) in addition to factor graph optimization (backend optimization) are used together to localize the autonomous car (Fig. 10). By using both geometric and intensity features, the performance of the SLAM system is claimed to improve.

Occupancy grid map (OGM) based 2D LiDAR SLAM [155, 156] was implemented to perform well in indoor environments. However, they cannot be generalized to work sufficiently in outdoor environments. In addition, localization methods that employ 3D LiDAR data in the form of point cloud [5], voxels [151], or extracted features [124] are very computationally expensive and, in many scenarios, they cannot be implemented for real-time LiDAR SLAM. Therefore, many researchers [157, 158] suggest using 2.5D heightmaps for SLAM, claiming improved performance compared to conventional 2D and 3D methods that overcomes the problems of local maximum, drifting, and complexity.

For autonomous driving, precise SLAM is required. However, due to a variety of reasons, SLAM remains a difficult problem to address. For example, SLAM (along with other localization methods) has a tendency to drift over a lengthy driving distance. As a result, the SLAM algorithm gets less accurate as the driving distance increases. Another issue is that while maps are necessary for SLAM, they are not always feasible under all driving situations. In other words, when an autonomous vehicle is given a true trajectory to follow, employing a common SLAM approach causes the followed trajectory to deviate from the true one over time. As a result, very accurate maps under a variety of weather situations are needed to help with the localization process (which are not always available).

5. CONCLUSIONS

Autonomous driving is a rapidly evolving area with a plethora of anticipated advantages. For self-driving cars to make the correct decision at the appropriate moment, they need accurate and quick perception and localization. A well-chosen sensor type and arrangement assists the perception and localization system in generating a complete picture of the surrounding region, allowing for a flawless comprehension of the driving scenario. The autonomous driving area has achieved outstanding achievements in all aspects, including object identification, semantic segmentation, and simultaneous localization and mapping, thanks to the recently developed technologies such as the modern deep learning methods. These accomplishments enabled the ego vehicle to attain extremely high accuracy in perception and localization tasks while also taking into account real-time concerns. However, inclement weather and congested metropolitan areas cause even the state-of-the-art systems to fail. As a result, further study is needed to increase the performance of the perception and localization systems in all driving environments and weather conditions.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Hoel, C.J., Driggs-Campbell, K., Wolff, K., Laine, L., and Kochenderfer, M.J., Combining planning and deep reinforcement learning in tactical decision making for autonomous driving, *IEEE Trans. Intell. Veh.*, 2019, vol. 5, no. 2, pp. 294-305.
2. Lefevre, S., Carvalho, A., and Borrelli, F., A learning-based framework for velocity control in autonomous driving, *IEEE Trans. Autom. Sci. Eng.*, 2015, vol. 13, no. 1, pp. 32–42.
3. Choi, J., Lee, J., Kim, D., Soprani, G., Cerri, P., Broggi, A., and Yi, K., Environment-detection-and-mapping algorithm for autonomous driving in rural or off-road environment, *IEEE Trans. Intell. Transp. Syst.*, 2012, vol. 13, no. 2, pp. 974–982.
4. Li, Q., Chen, L., Li, M., Shaw, S.L., and Nüchter, A., A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios, *IEEE Trans. Veh. Technol.*, 2013, vol. 63, no. 2, pp. 540–555.
5. Aldibaja, M., Suganuma, N., and Yoneda, K., Robust intensity-based localization method for autonomous driving on snow–wet road surface, *IEEE Trans. Ind. Inf.*, 2017, vol. 13, no. 5, pp. 2369–2378.
6. Kim, T.H. and Park, T.H., Placement optimization of multiple LIDAR sensors for autonomous vehicles, *IEEE Trans. Intell. Transp. Syst.*, 2019, vol. 21, no. 5, pp. 2139–2145.
7. Lim, K.L., Drage, T., Zhang, C., Brogle, C., Lai, W.W., Kelliher, T., Adina-Zada, M., and Bräunl, T., Evolution of a reliable and extensible high-level control system for an autonomous car, *IEEE Trans. Intell. Veh.*, 2019, vol. 4, no. 3, pp. 396–405.
8. Mei, J., Gao, B., Xu, D., Yao, W., Zhao, X., and Zhao, H., Semantic segmentation of 3d lidar data in dynamic scene using semi-supervised learning, *IEEE Trans. Intell. Transp. Syst.*, 2019, vol. 21, no. 6, pp. 2496–2509.
9. Jo, K., Kim, J., Kim, D., Jang, C., and Sunwoo, M., Development of autonomous car – Part II: A case study on the implementation of an autonomous driving system based on distributed architecture, *IEEE Trans. Ind. Electron.*, 2015, vol. 62, no. 8, pp. 5119–5132.

10. Jo, K., Jo, Y., Suhr, J.K., Jung, H.G., and Sunwoo, M., Precise localization of an autonomous car based on probabilistic noise models of road surface marker features using multiple cameras, *IEEE Trans. Intell. Transp. Syst.*, 2015, vol. 16, no. 6, pp. 3377–3392.
11. Vivacqua, R.P.D., Bertozzi, M., Cerri, P., Martins, F.N., and Vassallo, R.F., Self-localization based on visual lane marking maps: An accurate low-cost approach for autonomous driving, *IEEE Trans. Intell. Transp. Syst.*, 2017, vol. 19, no. 2, pp. 582–597.
12. Zong, W., Zhang, C., Wang, Z., Zhu, J., and Chen, Q., Architecture design and implementation of an autonomous vehicle, *IEEE Access*, 2018, vol. 6, pp. 21956–21970.
13. Wang, C., Sun, Q., Li, Z., Zhang, H., and Ruan, K., Cognitive competence improvement for autonomous vehicles: A lane change identification model for distant preceding vehicles, *IEEE Access*, 2019, vol. 7, pp. 83229–83242.
14. Artunedo, A., Villagra, J., Godoy, J., and del Castillo, M.D., Motion planning approach considering localization uncertainty, *IEEE Trans. Veh. Technol.*, 2020, vol. 69, no. 6, pp. 5983–5994.
15. Okumura, B., James, M.R., Kanzawa, Y., Derry, M., Sakai, K., Nishi, T., and Prokhorov, D., Challenges in perception and decision making for intelligent automotive vehicles: A case study, *IEEE Trans. Intell. Veh.*, 2016, vol. 1, no, 1, pp. 20–32.
16. Noh, S., Decision-making framework for autonomous driving at road intersections: Safeguarding against collision, overly conservative behavior, and violation vehicles, *IEEE Trans. Ind. Electron.*, 2018, vol. 66, no. 4, pp. 3275–3286.
17. Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.
18. Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., Wiesbeck, W., and Dietmayer, K., Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges, *IEEE Trans. Intell. Transp. Syst.*, 2020, vol. 22, no. 3, pp. 1341–1360.
19. Zou, Q., Jiang, H., Dai, Q., Yue, Y., Chen, L., and Wang, Q., Robust lane detection from continuous driving scenes using deep neural networks, *IEEE Trans. Veh. Technol.*, 2019, vol. 69, no. 1, pp. 41–54.
20. Chen, X., Kundu, K., Zhu, Y., Ma, H., Fidler, S., and Urtasun, R., 3d object proposals using stereo imagery for accurate object class detection, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, vol. 40, no. 5, pp. 1259–1272.
21. Li, W., Qu, Z., Song, H., Wang, P., and Xue, B., The traffic scene understanding and prediction based on image captioning, *IEEE Access*, 2020, vol. 9, pp. 1420–1427.
22. Li, Y., Wang, H., Dang, L.M., Nguyen, T.N., Han, D., Lee, A., Jang, I., and Moon, H., A deep learning-based hybrid framework for object detection and recognition in autonomous driving, *IEEE Access*, 2020, vol. 8, pp. 194228–194239.
23. Bochkovskiy, A., Wang, C.Y., and Liao, H.Y.M., Yolov4: Optimal speed and accuracy of object detection, *arXiv preprint arXiv:2004.10934*, 2020.
24. Cao, Z., Simon, T., Wei, S.E., and Sheikh, Y., Realtime multi-person 2d pose estimation using part affinity fields, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 7291–7299.
25. Adadi, A. and Berrada, M., Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), *IEEE Access*, 2018, vol. 6, pp. 52138–52160.
26. Cai, Y., Luan, T., Gao, H., Wang, H., Chen, L., Li, Y., Sotelo, M.A., and Li, Z., YOLOv4-5D: An effective and efficient object detector for autonomous driving, *IEEE Trans. Instrum. Meas.*, 2021, vol. 70, pp. 1–13.
27. Ren, S., He, K., Girshick, R., and Sun, J., Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, vol. 39, no. 6, pp. 1137–1149.
28. Cai, Z. and Vasconcelos, N., Cascade r-cnn: Delving into high quality object detection, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2018, pp. 6154–6162.
29. Liu, Y., Yixuan, Y., and Liu, M., Ground-aware monocular 3d object detection for autonomous driving, *IEEE Rob. Autom. Lett.*, 2021, vol. 6, no. 2, pp. 919–926.
30. Wei, J., He, J., Zhou, Y., Chen, K., Tang, Z., and Xiong, Z., Enhanced object detection with deep convolutional neural networks for advanced driving assistance, *IEEE Trans. Intell. Transp. Syst.*, 2019, vol. 21, no. 4, pp. 1572–1583.
31. Chen, L., Zhan, W., Tian, W., He, Y., and Zou, Q., Deep integration: A multi-label architecture for road scene recognition, *IEEE Trans. Image Process.*, 2019, vol. 28, no. 10, pp. 4883–4898.
32. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A.L., Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, vol. 40, no. 4, pp. 834–848.
33. Chen, L.C., Papandreou, G., Schroff, F., and Adam, H., Rethinking atrous convolution for semantic image segmentation, *arXiv preprint arXiv:1706.05587*, 2017.
34. Li, G., Xie, H., Yan, W., Chang, Y., and Qu, X., Detection of road objects with small appearance in images for autonomous driving in various traffic situations using a deep learning based approach, *IEEE Access*, 2020, vol. 8, pp. 211164–211172.

35. Dominguez-Sanchez, A., Cazorla, M., and Orts-Escolano, S., Pedestrian movement direction recognition using convolutional neural networks, *IEEE Trans. Intell. Transp. Syst.*, 2017, vol. 18, no. 12, pp. 3540–3548.
36. Gupta, A. and Choudhary, A., A framework for camera-based real-time lane and road surface marking detection and recognition, *IEEE Trans. Intell. Veh.*, 2018, vol. 3, no. 4, pp. 476–485.
37. Seo, E., Lee, S., Shin, G., Yeo, H., Lim, Y., and Choi, G., Hybrid tracker based optimal path tracking system of autonomous driving for complex road environments, *IEEE Access*, 2021, vol. 9, pp. 71763–71777.
38. Zhang, Y., Wang, J., Wang, X., and Dolan, J.M., Road-segmentation-based curb detection method for self-driving via a 3D-LiDAR sensor, *IEEE Trans. Intell. Transp. Syst.*, 2018, vol. 19, no. 12, pp. 3981–3991.
39. Thrun, S., Burgard, W., and Fox, D., *Probabilistic Robotics,* USA: Massachusetts Inst. of Technology, 2005.
40. Xu, Z., Sun, Y., and Liu, M., iCurb: Imitation learning-based detection of road curbs using aerial images for autonomous driving, *IEEE Rob. Autom. Lett.*, 2021, vol. 6, no. 2, pp. 1097–1104.
41. Cai, P., Wang, S., Sun, Y., and Liu, M., Probabilistic end-to-end vehicle navigation in complex dynamic environments with multimodal sensor fusion, *IEEE Rob. Autom. Lett.* 2020, vol. 5, no. 3, pp. 4218–4224.
42. Lin, C., Tian, D., Duan, X., and Zhou, J., 3D Environmental perception modeling in the simulated autonomous-driving systems, *Complex Syst. Model. Simul.*, 2021, vol. 1, no. 1, pp. 45–54.
43. Ku, J., Mozifian, M., Lee, J., Harakeh, A., and Waslander, S.L., Joint 3d proposal generation and object detection from view aggregation, in *2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (*IROS*), IEEE, 2018, pp. 1–8.
44. Zhang, L., Li, Y., and Nevatia, R., Global data association for multi-object tracking using network flows, in *2008 IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, 2008, pp. 1–8.
45. Yoon, K., Kim, D.Y., Yoon, Y.C., and Jeon, M., Data association for multi-object tracking via deep neural networks, *Sensors*, 2019, vol. 19, no. 3, p. 559.
46. Hu, H.N., Cai, Q.Z., Wang, D., Lin, J., Sun, M., Krahenbuhl, P., Darrell, T., and Yu, F., Joint monocular 3D vehicle detection and tracking, in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, 2019, pp. 5390–5399.
47. Zhou, X., Koltun, V., and Krähenbühl, P., Tracking objects as points, in *European Conf. on Computer Vision*, Cham: Springer, 2020, pp. 474–490.
48. Li, Q., Hu, R., Wang, Z., and Ding, Z., Driving behavior-aware network for 3D object tracking in complex traffic scenes, *IEEE Access*, 2021, vol. 9, pp. 51550–51560.
49. Leal-Taixé, L., Canton-Ferrer, C., and Schindler, K., Learning by tracking: Siamese CNN for robust target association, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 33–40.
50. Kong, X., Xin, B., Wang, Y., and Hua, G., Collaborative deep reinforcement learning for joint object search, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 1695–1704.
51. Yu, F., Wang, D., Shelhamer, E., and Darrell, T., Deep layer aggregation, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2018, pp. 2403–2412.
52. Li, Y., Møgelmose, A., and Trivedi, M.M., Pushing the “Speed Limit”: high-accuracy US traffic sign recognition with convolutional neural networks, *IEEE Trans. Intell. Veh.*, 2016, vol. 1, no. 2, pp. 167–176.
53. Chollet, F., Xception: Deep learning with depthwise separable convolutions, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
54. Bangquan, X. and Xiong, W.X., Real-time embedded traffic sign recognition using efficient convolutional neural network, *IEEE Access*, 2019, vol. 7, pp. 53330–53346.
55. He, K., Gkioxari, G., Dollár, P., and Girshick, R., Mask r-cnn, in *Proc. of the IEEE Int. Conf. on Computer Vision*, 2017, pp. 2961–2969.
56. Serna, C.G. and Ruichek, Y., Traffic signs detection and classification for European urban environments, *IEEE Trans. Intell. Transp. Syst.*, 2019, vol. 21, no. 10, pp. 4388–4399.
57. Mannan, A., Javed, K., Rehman, A.U., Babri, H.A., and Noon, S.K., Classification of degraded traffic signs using flexible mixture model and transfer learning, *IEEE Access*, 2019, vol. 7, pp. 148800–148813.
58. Avramović, A., Sluga, D., Tabernik, D., Skočaj, D., Stojnić, V., and Ilc, N., Neural-network-based traffic sign detection and recognition in high-definition images using region focusing and parallelization, *IEEE Access*, 2020, vol. 8, pp. 189855–189868.
59. Reza, A.M., Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, 2004, vol. 38, no. 1, pp. 35–44.
60. Uijlings, J.R., Van de Sande, K.E., Gevers, T., and Smeulders, A.W., Selective search for object recognition, *Int. J. Comput. Vision*, 2013, vol. 104, no. 2, pp. 154–171.
61. Girshick, R., Donahue, J., Darrell, T., and Malik, J., Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.
62. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., and Ramanan, D., Object detection with discriminatively trained part-based models, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, vol. 32, no. 9, pp. 1627–1645.

63. Møgelmose, A., Liu, D., and Trivedi, M.M., Detection of US traffic signs, *IEEE Trans. Intell. Transp. Syst.*, 2015, vol. 16, no. 6, pp. 3116–3125.

64. Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., and Igel, C., Detection of traffic signs in real-world images: The German traffic sign detection benchmark, in *2013 Int. Joint Conf. on Neural Networks* (*IJCNN*), IEEE, 2013, pp. 1–8.

65. Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., You only look once: Unified, real-time object detection, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

66. Tabernik, D. and Skočaj, D., Deep learning for large-scale traffic-sign detection and recognition, *IEEE Trans. Intell. Transp. Syst.*, 2019, vol. 21, no. 4, pp. 1427–1440.

67. Zhang, X., Chen, Z., Wu, Q.J., Cai, L., Lu, D., and Li, X., Fast semantic segmentation for scene perception, *IEEE Trans. Ind. Inform.*, 2018, vol. 15, no. 2, pp. 1183–1192.

68. Wang, W., Fu, Y., Pan, Z., Li, X., and Zhuang, Y., Real-time driving scene semantic segmentation, *IEEE Access*, 2020, vol. 8, pp. 36776–36788.

69. Long, J., Shelhamer, E., and Darrell, T., Fully convolutional networks for semantic segmentation, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

70. Badrinarayanan, V., Kendall, A., and Cipolla, R., Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, vol. 39, no. 12, pp. 2481–2495.

71. Badrinarayanan, V., Handa, A., and Cipolla, R., Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling, arXiv preprint arXiv:1505.07293, 2015.

72. Simonyan, K. and Zisserman, A., Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.

73. Ronneberger, O., Fischer, P., and Brox, T., U-net: Convolutional networks for biomedical image segmentation, in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer, 2015, pp. 234–241.

74. Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J., Pyramid scene parsing network, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 2881–2890.

75. Yu, F. and Koltun, V., Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122, 2015.

76. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A.L., Semantic image segmentation with deep convolutional nets and fully connected crfs, arXiv preprint arXiv:1412.7062, 2014.

77. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S., Feature pyramid networks for object detection, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.

78. Chaurasia, A. and Culurciello, E., *December. Linknet: Exploiting encoder representations for efficient semantic segmentation, in 2017 IEEE Visual Communications and Image Processing (VCIP)*, IEEE, 2017, pp. 1–4.

79. Wu, C., Cheng, H.P., Li, S., Li, H., and Chen, Y., ApesNet: a pixel-wise efficient segmentation network for embedded devices, *IET Cyber.-Phys. Syst.: Theory Appl.*, 2016, vo.1, no. 1, pp. 78-85.

80. Paszke, A., Chaurasia, A., Kim, S., and Culurciello, E., Enet: A deep neural network architecture for real-time semantic segmentation, arXiv preprint arXiv:1606.02147, 2016.

81. Mehta, S., Rastegari, M., Caspi, A., Shapiro, L., and Hajishirzi, H., Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation, in *Proc. of the European Conf. on Computer Vision* (*ECCV*), 2018, pp. 552–568.

82. Kim, J. and Heo, Y.S., Efficient semantic segmentation using spatio-channel dilated convolutions, *IEEE Access*, 2019, vol. 7, pp. 154239–154252.

83. Lo, S.Y., Hang, H.M., Chan, S.W., and Lin, J.J., Efficient dense modules of asymmetric convolution for real-time semantic segmentation, in *Proc. of the ACM Multimedia Asia*, 2019, pp. 1–6.

84. Watanabe, S., *Pattern Recognition: Human and Mechanical*, Wiley, 1985.

85. Altun, M. and Celenk, M., Road scene content analysis for driver assistance and autonomous driving, *IEEE Trans. Intell. Transp. Syst.*, 2017, vol. 18, no. 12, pp. 3398–3407.

86. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., Generative adversarial nets, in *Proc. of the 27th Int. Conf. on Neural Information Processing Systems*, Cambridge, MA, USAL MIT Press, 2014, vol. 2, pp. 2672–2680.

87. Mirza, M. and Osindero, S., Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784, 2014.

88. Han, X., Lu, J., Zhao, C., You, S., and Li, H., Semisupervised and weakly supervised road detection based on generative adversarial networks, *IEEE Signal Process. Lett.*, 2018, vol. 25, no. 4, pp. 551–555.

89. Yan, F., Wang, K., Zou, B., Tang, L., Li, W., and Lv, C., LiDAR-based multi-task road perception network for autonomous vehicles, *IEEE Access*, 2020, vol. 8, pp. 86753–86764.

90. Sun, L., Yang, K., Hu, X., Hu, W., and Wang, K., Real-time fusion network for RGB-D semantic segmentation incorporating unexpected obstacle detection for road-driving images, *IEEE Rob. Autom. Lett.*, 2020, vol. 5, no. 4, pp. 5558–5565.

91. Huang, Z., Lv, C., Xing, Y., and Wu, J., Multi-modal sensor fusion-based deep neural network for end-to-end autonomous driving with scene understanding, *IEEE Sens. J.*, 2020, vol. 21, no. 10, pp. 11781–11790.

92. Pei, Y., Sun, B., and Li, S., Multifeature selective fusion network for real-time driving scene parsing, *IEEE Trans. Instrum. Meas.t*, 2021, vol. 70, pp. 1–12.

93. He, K., Zhang, X., Ren, S., and Sun, J., Deep residual learning for image recognition, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

94. Zhu, Z., Xu, M., Bai, S., Huang, T., and Bai, X., Asymmetric non-local neural networks for semantic segmentation, in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, 2019, pp. 593–602.

95. Sun, Y., Zuo, W., Huang, H., Cai, P., and Liu, M., PointMoSeg: Sparse tensor-based end-to-end moving-obstacle segmentation in 3D lidar point clouds for autonomous driving, *IEEE Rob. Autom. Lett.*, 2020, vol. 6, no. 2, pp. 510–517.

96. Choy, C., Park, J., and Koltun, V., Fully convolutional geometric features, in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, 2019, pp. 8958–8966.

97. Graham, B., Engelcke, M., and Van der Maaten, L., 3d semantic segmentation with submanifold sparse convolutional networks, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2018, pp. 9224–9232.

98. Choy, C., Gwak, J., and Savarese, S., 4d spatio-temporal convnets: Minkowski convolutional neural networks, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2019, pp. 3075–3084.

99. Kumar, V.R., Yogamani, S., Rashed, H., Sitsu, G., Witt, C., Leang, I., Milz, S., and Mäder, P., Omnidet: Surround view cameras based multi-task visual perception network for autonomous driving, *IEEE Rob. Autom. Lett.*, 2021, vol. 6, no. 2, pp. 2830–2837.

100. Fritsch, J., Kuehnl, T., and Geiger, A., A new performance measure and evaluation benchmark for road detection algorithms, in *16th Int. IEEE Conf. on Intelligent Transportation Systems* (*ITSC 2013*), IEEE, 2013, pp. 1693–1700.

101. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B., The cityscapes dataset for semantic urban scene understanding, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.

102. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., and Gall, J., Semantickitti: A dataset for semantic scene understanding of lidar sequences, in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, 2019, pp. 9297–9307.

103. Shin, M.O., Oh, G.M., Kim, S.W., and Seo, S.W., Real-time and accurate segmentation of 3-D point clouds based on Gaussian process regression, *IEEE Trans. Intell. Transp. Syst.*, 2017, vol. 18, no. 12, pp. 3363–3377.

104. Williams, C.K., and Rasmussen, C.E., *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT press, 2006.

105. Geiger, A., Lenz, P., Stiller, C., and Urtasun, R., Vision meets robotics: The kitti dataset, *Int. J. Rob. Res.*, 2013, vol. 32, no. 11, pp. 1231–1237.

106. Brostow, G.J., Shotton, J., Fauqueur, J., and Cipolla, R., Segmentation and recognition using structure from motion point clouds, in *European Conf. on computer vision*, Berlin, Heidelberg: Springer, 2008, pp. 44–57.

107. Hussain Raza, S., Grundmann, M., and Essa, I., Geometric context from videos, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, pp. 3081–3088.

108. Brostow, G.J., Fauqueur, J., and Cipolla, R., Semantic object classes in video: A high-definition ground truth database, *Pattern Recognit. Lett.*, 2009, vol. 30, no. 2, pp. 88–97.

109. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., and Beijbom, O., Nuscenes: A multimodal dataset for autonomous driving, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2020, pp. 11621–11631.

110. Mudassar, B.A., Saha, P., Long, Y., Amir, M.F., Gebhardt, E., Na, T., Ko, J.H., Wolf, M., and Mukhopadhyay, S., Camel: An adaptive camera with embedded machine learning-based sensor parameter control, *IEEE J. Emerging Selected Top. Circuits Syst.*, 2019, vol. 9, no. 3, pp. 498–508.

111. Kleinfelder, S., Lim, S., Liu, X., and El Gamal, A., A 10000 frames/s CMOS digital pixel sensor, *IEEE J. Solid-State Circuits*, 2001, vol. 36, no. 12, pp. 2049–2059.

112. Skorka, O. and Joseph, D., CMOS digital pixel sensors: Technology and applications, in *Nanosensors, Biosensors, and Info-Tech Sensors and Systems 2014*, International Society for Optics and Photonics, 2014, vol. 9060, p. 90600G.

113. Tu, C., Takeuchi, E., Carballo, A., and Takeda, K., May. Point cloud compression for 3d lidar sensor using recurrent neural network with residual blocks, in *2019 Int. Conf. on Robotics and Automation* (*ICRA*), IEEE, 2019, pp. 3274–3280.

114. Mehra, A., Mandal, M., Narang, P., and Chamola, V., ReViewNet: A fast and resource optimized network for enabling safe autonomous driving in hazy weather conditions, *IEEE Trans. Intell. Transp. Syst.*, 2020.

115. Qu, Y., Chen, Y., Huang, J., and Xie, Y., Enhanced pix2pix dehazing network, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2019, pp. 8160–8168.

116. Engin, D., Genç, A., and Kemal Ekenel, H., Cycle-dehaze: Enhanced cyclegan for single image dehazing, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 825–833.

117. Mehta, A., Sinha, H., Narang, P., and Mandal, M., Hidegan: A hyperspectral-guided image dehazing gan, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 212–213.

118. Chen, B. and Wornell, G.W., Quantization index modulation: A class of provably good methods for digital watermarking and information embedding, *IEEE Trans. Inform. Theory*, 2001, vol. 47, no. 4, pp. 1423–1443.

119. Chen, B. and Wornell, G.W., Quantization index modulation methods for digital watermarking and information embedding of multimedia, *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, 2001, vol. 27, no. 1, pp. 7–33.

120. Malik, H., Subbalakshmi, K.P., and Chandramouli, R., Nonparametric steganalysis of QIM data hiding using approximate entropy, in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, International Society for Optics and Photonics, 2008, vol. 6819, p. 681914.

121. Malik, H., Chandramouli, R., and Subbalakshmi, K.P., Steganalysis: Trends and challenges, in *Multimedia Forensics and Security*, IGI Global, 2009, pp. 245–265.

122. Changalvala, R. and Malik, H., LiDAR data integrity verification for autonomous vehicle, *IEEE Access*, 2019, vol. 7, pp. 138018–138031.

123. Bresson, G., Alsayed, Z., Yu, L., and Glaser, S., Simultaneous localization and mapping: A survey of current trends in autonomous driving, *IEEE Trans. Intell. Veh.*, 2017, vol. 2, no. 3, pp. 194–220.

124. Wang, H., Wang, C., and Xie, L., Intensity-SLAM: Intensity Assisted localization and mapping for large scale environment, *IEEE Rob. Autom. Lett.*, 2021, vol. 6, no. 2, pp. 1715–1721.

125. White, C.E., Bernstein, D., and Kornhauser, A.L., Some map matching algorithms for personal navigation assistants, *Transp. Res., Part C: Emerging Technol.*, 2000, vol. 8, no. 1–6, pp. 91–108.

126. El Najjar, M.E. and Bonnifait, P., A road-matching method for precise vehicle localization using belief theory and kalman filtering, *Autonom. Rob.*, 2005, vol. 19, no. 2, pp. 173–191.

127. Newson, P. and Krumm, J., Hidden Markov map matching through noise and sparseness, in *Proc. of the 17th ACM SIGSPATIAL Int. Conf. on advances in geographic information systems*, 2009, pp. 336–343.

128. Ballardini, A.L., Cattaneo, D., Izquierdo, R., Parra, I., Sotelo, M.A., and Sorrenti, D.G., Ego-lane estimation by modeling lanes and sensor failures, in *2017 IEEE 20th Int. Conf. on Intelligent Transportation Systems* (*ITSC*), IEEE, 2017, pp. 1–7.

129. Popescu, V., Danescu, R., and Nedevschi, S., On-road position estimation by probabilistic integration of visual cues, in *2012 IEEE Intelligent Vehicles Symposium*, IEEE, 2012, pp. 583–589.

130. Jo, K., Chu, K., and Sunwoo, M., Interacting multiple model filter-based sensor fusion of GPS with in-vehicle sensors for real-time vehicle positioning, *IEEE Trans. Intell. Transp. Syst.*, 2011, vol. 13, no. 1, pp. 329–343.

131. Gwak, M., Jo, K., and Sunwoo, M., Neural-network multiple models filter (NMM)-based position estimation system for autonomous vehicles, *Int. J. Autom. Technol.*, 2013, vol. 14, no. 2, pp. 265–274.

132. Gustafsson, F., Particle filter theory and practice with positioning applications, *IEEE Aerosp. Electron. Syst. Mag.*, 2010, vol. 25, no. 7, pp. 53–82.

133. Klein, G. and Murray, D., Parallel tracking and mapping for small AR workspaces, in *2007 6th IEEE and ACM Int. Symposium on Mixed and Augmented Reality*, IEEE, 2007, pp. 225–234.

134. Mur-Artal, R. and Tardós, J.D., Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras, *IEEE Trans. Rob.*, 2017, vol. 33, no. 5, pp. 1255–1262.

135. Ferrera, M., Eudes, A., Moras, J., Sanfourche, M., and Le Besnerais, G., OV $^{2} $ SLAM: A fully online and versatile visual SLAM for real-time applications, *IEEE Rob. Autom. Lett.*, 2021, vol. 6, no. 2, pp. 1399–1406.

136. Angeli, A., Filliat, D., Doncieux, S., and Meyer, J.A., Fast and incremental method for loop-closure detection using bags of visual words, *IEEE Trans. Rob.*, 2008, vol. 24, no. 5, pp. 1027–1037.

137. Nicosevici, T. and Garcia, R., Automatic visual bag-of-words for online robot navigation and mapping, *IEEE Trans. Rob.*, 2012, vol. 28, no. 4, pp. 886–898.

138. Garcia-Fidalgo, E. and Ortiz, A., ibow-lcd: An appearance-based loop-closure detection approach using incremental bags of binary words, *IEEE Rob. Autom. Lett.*, 2018, vol. 3, no. 4, pp. 3051–3057.

139. Nobis, F., Papanikolaou, O., Betz, J., and Lienkamp, M., Persistent map saving for visual localization for autonomous vehicles: An ORB-SLAM 2 extension, in *2020 Fifteenth Int. Conf. on Ecological Vehicles and Renewable Energies* (*EVER*), IEEE, 2020, pp. 1–9.

140. Kasmi, A., Laconte, J., Aufrère, R., Denis, D., and Chapuis, R., End-to-end probabilistic ego-vehicle localization framework, *IEEE Trans. Intell. Veh.*, 2020, vol. 6, no. 1, pp. 146–158.

141. Ramm, F., Topf, J., and Chilton, S., OpenStreetMap: using and enhancing the free map of the world, *SoC Bulletin*, Cambridge: *UIT Cambridge*, 2011, vol. 45, p. 55.

142. Rozenberszki, D. and Majdik, A.L., *May. LOL: Lidar-only odometry and localization in 3D point cloud maps, in 2020 IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 4379–4385.

143. Koide, K., Miura, J., and Menegatti, E., A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement, *Int. J. Adv. Rob. Syst.* 2019, vol. 16, no. 2, p. 1729881419841532.

144. Chetverikov, D., Svirko, D., Stepanov, D., and Krsek, P., The trimmed iterative closest point algorithm, in *Object Recognition Supported by User Interaction for Service Robots*, IEEE, 2002, vol. 3, pp. 545–548.

145. Zhang, J. and Singh, S., LOAM: Lidar odometry and mapping in real-time, in *Robotics: Science and Systems*, 2014, vol. 2, no. 9.

146. Shan, T. and Englot, B., Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain, in *2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (*IROS*), IEEE, 2018, pp. 4758–4765.

147. Elfes, A., Using occupancy grids for mobile robot perception and navigation, *Computer*, 1989, vol. 22, no. 6, pp. 46–57.

148. Levinson, J., Montemerlo, M., and Thrun, S., Map-based precision vehicle localization in urban environments, in *Robotics: Science and Systems*, 2007.

149. Castorena, J. and Agarwal, S., Ground-edge-based LIDAR localization without a reflectivity calibration for autonomous driving, *IEEE Rob. Autom. Lett.*, 2017, vol. 3, no. 1, pp. 344–351.

150. de Paula Veronese, L., Auat-Cheein, F., Mutz, F., Oliveira-Santos, T., Guivant, J.E., de Aguiar, E., Badue, C., and De Souza, A.F., Evaluating the limits of a LiDAR for an autonomous driving localization, *IEEE Trans. Intell. Transp. Syst.*, 2020, vol. 22, no. 3, pp. 1449–1458.

151. Dubé, R., Gollub, M.G., Sommer, H., Gilitschenski, I., Siegwart, R., Cadena, C., and Nieto, J., Incremental-segment-based localization in 3-d point clouds, *IEEE Rob. Autom. Lett.*, 2018, vol. 3, no. 3, pp. 1832–1839.

152. Whelan, T., Ma, L., Bondarev, E., de With, P.H.N., and McDonald, J., Incremental and batch planar simplification of dense point cloud maps, *Rob. Autonom. Syst.*, 2015, vol. 69, pp. 3–14.

153. Lu, W., Wan, G., Zhou, Y., Fu, X., Yuan, P., and Song, S., DeepICP: An end-to-end deep neural network for 3D point cloud registration, *IEEE Int. Conf. on Computer Vision (ICCV)*, 2019, pp. 12–21.

154. Khan, S., Wollherr, D., and Buss, M., Modeling laser intensities for simultaneous localization and mapping, *IEEE Rob. Autom. Lett.*, 2016, vol. 1, no. 2, pp. 692–699.

155. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., and Leonard, J.J., Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age, *IEEE Trans. Rob.*, 2016, vol. 32, no. 6, pp. 1309–1332.

156. Kohlbrecher, S., Von Stryk, O., Meyer, J., and Klingauf, U., A flexible and scalable SLAM system with full 3D motion estimation, in *2011 IEEE Int. Symposium on Safety, Security, and Rescue Robotics*, IEEE, 2011, pp. 155–160.

157. Sun, L., Zhao, J., He, X., and Ye, C., June. Dlo: Direct lidar odometry for 2.5 d outdoor environment, in *2018 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2018, pp. 1–5.

158. Li, J., Zhao, J., Kang, Y., He, X., Ye, C., and Sun, L., DL-SLAM: Direct 2.5 D LiDAR SLAM for Autonomous Driving, in *2019 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2019, pp. 1205–1210.