IECE Transactions on Sensing, Communication, and Control
http://dx.doi.org/10.62762/TSCC.2024.989358

REVIEW ARTICLE

Innovations in 3D Object Detection: A Comprehensive Review of Methods, Sensor Fusion, and Future Directions

Ghulam E Mustafa Abro[1,*], **Zain Anwar Ali**[2] **and Summaiya Rajput**[3]

[1] Interdisciplinary Research Centre for Aviation and Space Exploration (IRCASE), King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Kingdom of Saudi Arabia
[2] Electronic Engineering Department, Maynooth International Engineering College (MIEC), Maynooth University, Maynooth, Co. Kildare, Ireland
[3] Department of Telecommunication Engineering, Mehran University of Engineering and Technology (MUET), Pakistan

Abstract

This review paper offers a thorough assessment of three-dimensional object recognition methods, an essential element in the perception frameworks of autonomous systems. This analysis emphasises the integration of LiDAR and camera sensors, providing a distinctive contrast with more economical alternatives like camera-only or camera-Radar combinations. This study objectively evaluates performance and practical implementation issues, such as cost and operational efficiency, thereby elucidating the limitations of existing systems and proposing avenues for further research. The insights provided render it a significant asset for enhancing 3D object recognition and autonomy in intelligent systems.

Keywords: autonomous systems, camera, fusion methods, LiDAR, object detection, radar and three-dimensional.

Academic Editor:
Quanmin Zhu

Submitted: 13 September 2024
Accepted: 22 September 2024
Published: 12 October 2024

Vol. 1, **No.** 1, 2024.
10.62762/TSCC.2024.989358

***Corresponding author:**
Ghulam E Mustafa Abro
Ghulam.abro@kfupm.edu.sa

Citation
Abro, G. E. M., Ali, Z. A., & Rajput, S. (2024). Innovations in 3D Object Detection: A Comprehensive Review of Methods, Sensor Fusion, and Future Directions. IECE Transactions on Sensing, Communication, and Control, 1(1), 3–29.

1 Introduction

Progress in 3D object identification methodologies for autonomous systems is crucial in determining the future of autonomous driving technologies. The integration of deep learning with 3D point cloud object recognition has led to substantial advancements in the perceptual capabilities of autonomous systems. These autonomous vehicles employ a variety of sensors, including mono and stereo cameras, thermal imaging, night vision, LiDAR, Radar, Inertial Navigation Systems (INS), Global Positioning Systems (GPS), and Inertial Measurement Units (IMUs), to precisely detect and classify objects in their environment, thereby facilitating safer navigation and enhanced situational awareness [1, 2]. Furthermore, the incorporation of 3D object recognition techniques has enabled the seamless integration of virtual items into the real world, transforming interactions inside augmented reality environments. The autonomous driving system depends on a perception system interconnected with multiple components, as depicted in Figure 1. It commences with sensor data collecting, followed by pre-processing for accuracy verification, and subsequently advances to feature extraction to discern pertinent information [3]. Subsequently, environmental comprehension evaluates the scene for risk assessment, followed by decision-making processes to ascertain best actions. Ultimately,

real-time control commands for vehicle maneuvering are generated, completing this iterative process crucial for safe and efficient autonomous navigation.

Deep learning techniques and LiDAR sensors are dependent on the techniques that are now considered to be the state of the art for the detection of three-dimensional objects. Additionally, certain approaches incorporate camera data in order to improve the overall efficacy of the detecting process. Despite the fact that cameras provide extensive semantic information from images and LiDARs offer exact depth measurements in the form of point clouds, this combination of sensors has gained great popularity across a variety of perception tasks [4] and [5]. This is due to the fact that their outputs are complementary to one another. Both LiDAR and camera sensors are essentially light dependent, which makes them subject to performance degradation in adverse weather circumstances such as rain or low light settings and other situations such as nighttime. Despite their usefulness, both types of sensors are sensitive to this degradation.

Radar sensors provide distinct metrics, including radial velocity and radar cross section (RCS) on the bird's eye view (BEV) plane, which are not obtainable from other sensors. These measurements offer valuable insights into the motion, shape, size, and material of objects, hence enhancing 3D object detection capabilities. In contrast to light-based sensors, radars employ radio waves, rendering them dependable in poor visibility circumstances. Moreover, radars possess an extended detection range, allowing for the effective identification of distant objects.

Nevertheless, the sparse characteristics of radar point clouds present challenges for independent 3D object detection utilizing solely radars. Conversely, cameras excel in delivering rich and dense semantic information, so facilitating accurate object recognition and scene comprehension. They record intricate texture and color, as well as contextual information, which is essential for perception tasks. However, cameras lack the capability to directly assess depth and velocity, which are required for estimating the three-dimensional position, dimensions, and motion of objects. Furthermore, given the intricate weather conditions faced by autonomous driving systems, the effectiveness of sensor data, particularly images and point clouds obtained by the perception system, may experience significant degradation due to the inherent constraints associated with each sensor. For example, visual cameras face challenges when presented with scenes exhibiting extreme brightness or blackness [6–10].

Moreover, LiDAR systems encounter difficulties in accurately identifying distant or diminutive objects due to their inherent poor resolution [11]. Furthermore, adverse weather circumstances exacerbate the challenge by introducing environmental noise and resulting in a significant reduction in perceptual distance [12] and [13]. These sensor-specific restrictions present substantial challenges for the robust and accurate perception necessary for safe autonomous driving in varied real-world environments. Radar camera fusion techniques integrate radar and camera data to identify objects, utilizing the complementary advantages of each sensor. Cameras provide extensive semantic information, complementing the limited semantic data from radar. Conversely, radar offers reliable velocity and depth measurements at extended distances, compensating for the depth and velocity constraints of cameras [13].

This complementary nature allows these methods to consistently produce precise 3D detections, even under challenging settings. Furthermore, the aggregate expense of radar and camera sensors is inferior to that of a solitary LiDAR, rendering the radar-camera amalgamation favored in advanced driving assistance systems (ADAS) [14]. Nevertheless, despite these advantages and the inherent differences between radar and cameras, challenges arise in integrating their information. Furthermore, the combination of radar and camera-based 3D object detection methods is less explored compared to LiDAR-based methods, resulting in performance lag [15]. Radar point clouds, while analogous to LiDAR point clouds, are sparser and exhibit lower accuracy and resolution, complicating the direct application of LiDAR-based techniques to radar data.

Fusion-based methodologies for integrating radar and camera data are generally categorized into three types based on the timing of the fusion process: Data level, feature level, and decision level fusion [16]. Data level fusion integrates raw data from several sensor modalities while minimizing information loss and facilitating joint feature learning. Nonetheless, data-level methodologies are inflexible and susceptible to sensor misalignment, as well as being computationally intensive. Decision-level fusion, in contrast, integrates

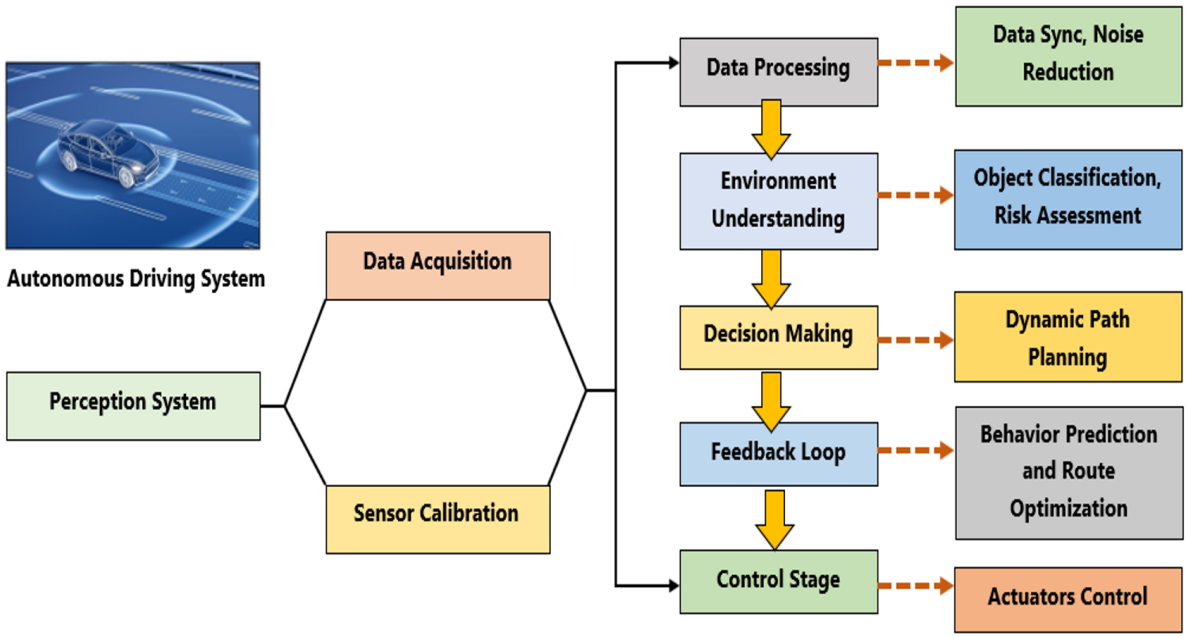

Figure 1. The components for an autonomous driving system.

detection results obtained independently from each modality, providing flexibility and robustness while minimizing computational costs. Nonetheless, decision level fusion is hindered by information loss and lacks the capacity to learn joint features. In the context of radar-camera fusion, performing data-level fusion is challenging due to disparities in sensor characteristics. Decision level fusion is unfeasible due to the inadequate performance of radar-based methods [17] and [18].

Feature level fusion techniques in radar-camera integration for autonomous driving systems offer a sophisticated method for improving object detection capabilities. These methods reconcile data level and decision level fusion by synthesizing features independently extracted from radar and camera modalities. A prevalent strategy entails projecting the radar point cloud onto surfaces, such as the bird's eye view (BEV) plane [19-27], or directly onto the camera image plane. Significant methodologies, including references [19–21], have markedly enhanced vehicle detection through the integration of radar and camera features. The integration of radar and camera data addresses critical challenges, such as detecting distant vehicles and navigating diverse environmental circumstances, as demonstrated by the frameworks in [22, 23]and the methodologies in [24] and [25]. Furthermore, references [26, 27] constitute significant additions to the field by employing cross-modal fusion to attain remarkable detection accuracy. These breakthroughs in radar camera integration are enabled by meticulous consideration of projection methods. Projecting the radar point cloud into the bird's eye view (BEV) plan preserves spatial information and facilitates feature extraction, while it necessitates intricate procedures to convert image features and integrate them with radar characteristics. Moreover, projecting the point cloud onto the image plane facilitates fusion but diminishes the depth dimension and complicates spatial feature extraction. Despite conveying semantic information, an object's details, including position, orientation, and size, along with the flattened depth dimension, complicate spatial feature extraction. This diminishes the significance of optimization feature extraction techniques for efficient radar-camera integration in autonomous driving systems.

Diverse methodologies, including Cluster Fusion and sophisticated radar camera fusion, enabling three-dimensional object detection. Extracting spatial features from radar point clouds directly within clusters, Cluster Fusion improves object recognition performance. Inspired on CenterFusion's methodology [14], it employs image-based preliminary detections to filter and cluster radar points. This method guarantees enhanced detection in autonomous driving. In our evaluation, we compared ClusterFusion with radar-monocular camera fusion methods and multi-camera setups [27, 28]. ClusterFusion notably attained the greatest nuScenes detection score (NDS) of 48.7%, surpassing other methods in mean orientation, velocity, and attribute

errors. Its competitive performance in mean attribute error underscores its effectiveness in enhancing 3D object detection for autonomous vehicles. In manufacturing, these techniques are essential for real-time defect detection, quality control, and ensuring adherence to high production standards. Furthermore, urban planners and architects utilize 3D object recognition to analyze and design efficient, sustainable, and visually appealing urban environments. Moreover, in the domain of robotics and deep learning, 3D object identification empowers robots to traverse intricate surroundings, manipulate objects, and aid in many jobs, spanning logistics to healthcare. This paper seeks to examine the accomplishments, challenges, and potential directions in 3D object recognition techniques for autonomous systems, offering valuable insights into their vital role across several industries.

This review paper thoroughly examines the progress and obstacles in 3D object identification methodologies for autonomous systems. As shown in Figure 2, section 1 delineates the importance of 3D object detection in autonomous systems. Section 2 offers a comprehensive overview of contemporary methodologies, encompassing contributions from historical to current developments. Section 3 encompasses the evaluation and examination of 3D object detection techniques utilising several sensor technologies, including LiDAR, cameras, and radar counted as trendy technologies and approaches. Furthermore, Section 4 analyses the integration of camera and LiDAR sensors, emphasizing current accomplishments and progress comparatively. Section 5 examines integration of camera, LiDAR and fusion sensor techniques. Section 6 addresses benefits of fusion techniques. Section 7 alternative approaches. Section 8 provides technical analysis whereas section 9 shares the future directions and recommendations. Finally, one may find the conclusion of the work under section 10.

2 State of the Art Approaches in 3D Object Detection

Cameras and LiDAR (Light Detection and Ranging) sensors are the most commonly employed sensors for 3D object detection [31, 32]. Although these technologies provide cost-effectiveness, their efficacy has led to extensive examinations and reviews. Nonetheless, current reviews frequently focus exclusively on particular approaches, primarily highlighting accuracy. This survey seeks to explore the essential elements of Accuracy, Latency, and Robustness for a thorough examination of existing techniques. This section emphasizes camera-based 3D object identification techniques, which have garnered considerable attention, particularly in the automotive sector for applications such as multi-view systems like BEV (bird's-eye view). Camera-based techniques can be categorized into three primary types: monocular, stereo-based, and multi-view (bird's-eye view) [33–35].

2.1 Monocular 3D Object Detection

Monocular 3D object detection substantially employs data from a single camera to identify and localize 3D objects, utilizing convolutional neural networks (CNNs) for the direct estimate of 3D bounding box parameters. Detecting 3D objects using a single camera to infer their 3D position, size, and orientation from one image has garnered considerable attention in recent years due to its cost-effectiveness, low power consumption, and ease of deployment in practical applications [36–40]. This method facilitates end-to-end training and exhibits practicality; however, its exclusive dependence on individual photos presents difficulties in precisely determining the 3D position, dimensions [41, 42], and orientation of objects, absent supplementary depth maps or point cloud data. Factors like as occlusion, fluctuations in viewpoint, and alterations in lighting conditions may adversely affect detection accuracy, while dependence on a single camera for depth perception may compromise overall performance [43–46]. Nevertheless, owing to the lack of enough 3D information in monocular images, researchers have concentrated on using depth information acquired from depth estimation tasks, notwithstanding the intrinsic challenges associated with monocular depth estimation. A significant method is prior-guided monocular 3D object detection [31–35] and [47–59], which persistently investigates the incorporation of concealed prior knowledge regarding object shapes and scene geometry in images to address the difficulties associated with monocular 3D object detection. Effectively integrating past knowledge is essential for reducing uncertainty and addressing the ill-posed characteristics of monocular 3D object detection challenges. Integrating pre-trained networks or auxiliary tasks, previous knowledge offers significant insights into object shape detection, geometric consistency, temporal limitations, and segmentation information [55–59]. The shape of an item provides significant insights into its appearance and structure,

Figure 2. The organization of this review paper.

enhancing the precision of spatial inference and pose estimation [52]. Geometric consistency facilitates comprehension of the positioning relationships among objects, hence enhancing detection consistency and robustness. Previous algorithms [50, 51] have notably exhibited diverse methodologies for leveraging prior information to enhance object detection. As our comprehension and utilisation of existing information advances, substantial advancements are expected in monocular 3D object detection, heralding breakthroughs and prospects in the fields of computer vision and intelligent systems.

Conversely, an alternative method of depth-assisted monocular 3D object detection utilizes depth estimation networks to improve detection precision with monocular pictures [60–68]. This methodology incorporates depth information to overcome the constraints of classic monocular methods, aiming for enhanced detection accuracy. Nonetheless, a possible disadvantage of depth-assisted monocular detection is the performance disparity between pseudo-LiDAR representations and LiDAR-based detectors resulting from inaccuracies in image-to-LiDAR conversion. Initiatives to close this gap involve transferring intricate structure information from point clouds to enhance monocular image detection.

2.2 Stereo-Based 3D Object Detection

Stereo-based 3D object detection utilizes the distinct functionalities of stereo cameras to ascertain depth through the examination of a dual image set. This technique presents a viable approach for precise depth estimation, a crucial element absents in monocular configurations. Notwithstanding these benefits, stereo-based methods continue to underperform relative to LiDAR-based techniques. Furthermore, investigations into 3D object identification utilizing stereo pictures are still rather scarce. One method in stereo-based detection entails modifying conventional 2D object detection frameworks. Stereo R-CNN [59] use a 2D detector to generate regions of interest (RoIs) in both the left and right images, thereafter estimating 3D object properties derived from these RoIs. Numerous following studies have embraced this paradigm [70], [71–77]. Another approach utilizes pseudo-LiDAR representations derived from the anticipated disparity map of stereo images. This approach converts depth maps into point clouds similar to LiDAR data. Wang et al. [78] presented the Pseudo-LiDAR concept, necessitating models to estimate depth to facilitate detection. Subsequent investigations have refined this methodology by integrating supplementary color data [79], auxiliary tasks [80–82], and coordinate

transformation frameworks [83], [84]. Ma et al. [84] introduced PatchNet, which questions the necessity of pseudo-LiDAR by recording 3D coordinates for every pixel. This observation indicates that the effectiveness of pseudo-LiDAR is attributed to its coordinate transformation rather than the point cloud representation itself. These methodologies utilize stereo imaging, employing both sparse and rich information, encompassing semantic and geometric clues, to precisely detect and localize 3D objects. Stereo R-CNN [84] enhances the Faster R-CNN architecture for stereo inputs, attaining state-of-the-art outcomes by integrating additional branches for sparse key point and perspective prediction. These advancements have resulted in significant enhancements in performance measures. In [86], it presents a Deep Stereo Geometry Network (DSGN) that uses a differentiable volumetric representation to efficiently encode three-dimensional geometric structures. This strategy substantially narrows the performance disparity between image-based and LiDAR-based techniques while also yielding encouraging accuracy outcomes. Moreover, real-time capabilities are highlighted in methodologies such as that proposed in [87], which not only uphold excellent accuracy but also exhibit notable efficiency regarding runtime. Moreover, innovative methodologies as R-CNN [87] emphasize instance disparity estimation and statistical shape modelling, yielding competitive results even in the absence of LiDAR ground truth during training. These procedures signify substantial advancements in stereo-based 3D object detection, demonstrating their applicability in real-world scenarios and reducing the performance disparity with LiDAR-based techniques. Although stereo-based 3D object detection demonstrates potential, it necessitates additional investigation and enhancement to close the performance disparity with LiDAR-based techniques. Optimising disparity estimation and utilizing supplementary information could significantly advance the discipline [88–90].

2.3 Multi-view 3D object Detection

Multi-view 3D object detection has undergone significant progress, especially in autonomous driving, where the integration of LiDAR point cloud data with RGB pictures has become standard practice [89], [90]. Fundamental to these methodologies is the implementation of panoramic Birds Eye View (BEV) techniques, which obviate the dependence on high-precision maps and enhance detection from 2D to 3D space. Techniques such as the MV3D network encode sparse 3D point clouds using a compact multi-view representation, resulting in enhanced performance in 3D localization and detection accuracy [91]. Furthermore, end-to-end Multiview fusion (MVF) algorithms integrate bird's-eye and perspective views, efficiently using complementing information to improve detection accuracy, particularly for distant or small objects [92].

Query-based multi-view approaches priorities the capture of relationships between regions and views, enhance the content of individual view images, and amalgamate them into distinctive 3D object representations [93]. Methods like dominating set clustering and pooling enhance performance by grouping comparable views and aggregating information, attaining cutting-edge outcomes on benchmark datasets [94–97]. These varied techniques emphasize the significance of multi-view information in 3D object detection, offering substantial improvements in the capabilities of autonomous driving systems for increased road safety and navigation. Depth-based methodologies, such as LSS [98], enable the conversion from two-dimensional to three-dimensional space via depth distribution. They forecast the depth distribution of two-dimensional characteristics and subsequently convert them to voxel space, enabling the following transition to bird's-eye view space. CaDDN, using LSS, employs actual ground truth depth values to improve prediction precision. Subsequent research, such as BEVDet and BEVDepth [25], has enhanced this methodology, improving the precision of object detection in the BEV domain. Drawing from Transformer technology, query-based approaches extract 2D spatial data from 3D environments. DETR3D [99] introduces 3D object queries to consolidate multi-view characteristics by projecting picture features into 2D space with learnt 3D reference points. PETR [89], leveraging ideas from DETR and DETR3D, employs an implicit positional encoding technique to generate the BEV space [95]. In [96], it introduces a Multi-view Labelling Object Detector (MLOD) that combines RGB pictures and LiDAR point clouds for effective feature fusion, attaining superior performance in 3D object detection. Alternative methodologies, such as MEMR [97], advocate for multi-view data fusion strategies grounded in regularization techniques, exhibiting significant enhancements in accuracy rates. These advancements jointly enhance the progression of multi-view 3D object identification, facilitating the development

of more robust and efficient autonomous driving systems [98, 99]. The entire state of the art approaches has been summarized into Table 1 along with main techniques and limitations.

3 Most trending approaches & Technologies

3.1 Point Based 3D Object Detection

Point-based 3D object recognition has become a leading method in utilising point clouds in deep learning, as demonstrated by substantial research contributions [100–103]. This methodology is unique in that it processes raw point clouds directly, avoiding the preprocessing steps typical of voxel-based methods. The utilisation of raw data guarantees the preservation of complex details and the original structure of point clouds, which is essential for extracting detailed features vital for precise object detection. In contrast to methods that convert point clouds into other forms like voxels or pictures, point-based algorithms preserve the original data's integrity, utilizing it to capture the subtle properties critical for detection tasks. This methodology is underpinned by seminal research in point cloud processing, as referenced in [100] and [101], establishing a framework for the effective management of raw point collections. The efficacy of point-based detection systems depends on balancing the density of contextual points with the size of the context radius for feature extraction. The primary elements propelling these developments are advanced point cloud sampling methods and complex feature learning systems, each playing a distinct role in the detection process [102, 103].

A fundamental element of point-based detection is the effective sampling of point clouds, which seeks to minimize processing requirements while maintaining crucial geometric and semantic data. Farthest Point Sampling (FPS), emphasized in foundational studies such as the PointNet++ framework [101], is distinguished for selecting points that provide optimal spatial coverage, hence enhancing the efficiency and efficacy of the detection process. Methods like PointRCNN [104] employ FPS to produce 3D suggestions directly from unprocessed point clouds, subsequently refining these proposals by incorporating semantic and spatial characteristics to improve detection precision. Notwithstanding its benefits, FPS encounters difficulties, such as the incorporation of extraneous points and the management of uneven point distributions among objects. Innovations designed to address these challenges encompass segmentation-guided sampling [105, 106], random and feature space sampling [107, 108], and voxel-based sampling methods [109–118], each presenting unique approaches to optimize computational resources on points that are more representative of objects of interest.

3.1.1 Advanced Feature Learning in Point-based Detection

At the core of point-based detection methods is the extraction of rich, discriminative features from raw point clouds. This extraction process benefits from neural network architectures that are specifically designed to be invariant to the order of points, capable of recognizing local geometries, and adept at integrating contextual information from varying perspectives.

- **PointNet-based Approaches**: PointNet and its successors [107, 111] have introduced the concept of set abstraction, a technique that down samples point clouds to aggregate local information effectively. For instance, PointRCNN [104] not only employs set abstraction for semantic segmentation but also refines spatial features to concentrate on significant areas for object detection, showcasing the synergy between semantic understanding and spatial precision.
- **Graph Neural Networks (GNN)**: GNN-based methods [110], [112–114], represent a dynamic evolution in feature learning, constructing adaptive graphs that encapsulate the complex spatial relationships within point clouds. This approach allows for detailed modeling of both local and global contexts, enhancing the detection systems ability to interpret the structural nuances of point clouds. GNNs, like Point-GNN [110], have demonstrated their efficacy in 3D object detection by effectively leveraging graph-based learning mechanisms.
- **Transformer Architectures**: The advent of transformer-based methods [115], [116] has introduced a new dimension to feature learning from point clouds, capitalizing on the transformer architectures capacity to model intricate interactions and long-range dependencies [117]. These methods employ attention mechanisms to dissect point clouds across different scales, facilitating a detailed analysis at both the micro (object) and macro (scene) levels. Despite their computational intensity, transformers represent a promising avenue for enhancing the depth and breadth of feature extraction from point

Table 1. Summary related to state-of-the-art approaches.

Method	Technique	Papers	Limitations
Monocular 3D Object Detection	Key Point Estimation and Object Key points	[28], [36], [29]	Limited accuracy in cluttered scenes Sensitivity to occlusion and object scale variations Computational complexity
	Geometric Reasoning and Depth Estimation	[60], [61], [39], [45], [40], [62], [63]	Reliance on accurate depth estimation, which can be challenging. Difficulty in handling varying lighting conditions and surface properties Limited performance in dynamic scenes
	Leveraging Geometry and Kinematics	[42], [33]	Dependency on accurate motion estimation Limited applicability to stationary scenes Challenges in handling fast-moving objects
	Network Architectures and Learning Strategies	[31], [38], [37]	Vulnerability to overfitting due to complex architectures Limited generalization to diverse environments and object categories Computational resource requirements
	Uncertainty and Ambiguity Handling	[30]	Difficulty in disentangling object occlusions Limited performance in scenes with complex object interactions Challenges in handling ambiguous depth estimations
	Cross-Task and Cross-Domain Generalization	[51], [53]	Need for extensive labeled data for cross-task generalization Limited performance in unseen domains without domain adaptation techniques Complexity in integrating multi-task learning objectives
	Semantic Segmentation and Contextual Information	[35], [64], [65]	Reliance on accurate semantic segmentation, which can be affected by scene complexity Difficulty in handling semantic ambiguities and misclassifications Challenges in incorporating long-range contextual information.
Stereo-Based 3D Object Detection	Stereo R-CNN	[85]	Requires refinement to bridge the performance gap with LiDAR-based methods.
	Disp R-CNN	[70], [73], [88]	Could improve performance with further optimization for real-world applications.
	DSGN	[86]	Potential limitations regarding real-time capabilities and scalability in complex environments.
	LIGA-Stereo	[75], [76]	May face challenges in accurately representing complex geometric structures.
	Stereo VoVNet-CNN	[77], [78]	Further validation is needed for its effectiveness in diverse scenarios and datasets.
Multi-View Based 3D Object Detection	Depth-based	[81]	Limited applicability to specific camera setups.
	Multi-View Fusion	[89], [90]	Limited performance in complex urban environments with dense traffic. Sensitivity to variations in LiDAR point cloud density.
	Query-based	[91], [92], [99]	Difficulty in accurately localizing objects with occlusions in multiple views. Challenges in handling large volumes of multi-view data efficiently.
	MLOD	[94]	Limited effectiveness in scenarios with significant occlusions or clutter.
	3D-R2N2	[95]	Challenges in accurately reconstructing objects with complex shapes.
	Multi-view ensemble manifold regularization	[96], [97] ,[101], [103],[104]	Difficulty in effectively capturing inter-view relationships in high-dimensional feature spaces.

clouds [118].

3.1.2 Grid Point Based 3d object detection

Grid-based 3D object identification techniques represent a leading edge in the evolution of autonomous driving systems, providing a reliable foundation for the rapid and precise processing of point cloud data. These methods, essential for the secure and dependable functioning of autonomous cars, leverage the discretization of

point clouds into grid representations, enabling feature extraction and object recognition with exceptional accuracy. The overarching foundation of grid-based detection entails the rasterization of point cloud data into discrete grid structures, such as voxels, pillars, or birds-eye view (BEV) feature maps [119]. Conventional 2D convolutional neural networks (CNNs) or specialized 3D sparse neural networks are employed to extract features from grid representations, facilitating precise detection of 3D objects [120–126]. Diverse grid-based representations provide distinct methodologies for encoding point cloud data, each with specific advantages and uses [125]. Voxels represent three-dimensional cubes that encompass points, facilitating effective voxelization and subsequent feature extraction from non-empty voxel cells [125]. Pillars give a specialized representation by aggregating features from points through methods such as PointNet encoding, achieving a compromise between efficacy and efficiency [127]. BEV feature maps, which are dense 2D representations encapsulating point cloud data within pixel regions, provide a flexible method for extracting object features [122].

Grid-based neural networks include both 2D CNNs and 3D sparse neural networks, designed to efficiently analyze various grid representations [126, 127]. 2D CNN architectures, derived from effective designs in 2D object detection, excel in processing BEV feature maps, whereas 3D sparse neural networks utilize specialized convolutional operators for the efficient processing of non-empty voxels [122], [126]. In the realm of autonomous driving systems, grid-based methodologies, especially those employing grid point representations, provide significant benefits for efficiency and precision [128–130]. CenterPoint attains exceptional outcomes with a mean Average Precision (mAP) of 90.5% on the nuScenes dataset and 88.7% on the Waymo dataset, underscoring the effectiveness of grid-based methodologies in practical applications [131]. Nonetheless, issues concerning the selection of grid cell size and the corresponding trade-offs between grid resolution and memory usage continue to be significant research priorities [119]. Enhancing grid-based techniques to achieve a balance between effectiveness and efficiency is crucial for the progression of 3D object detection in autonomous driving applications [126, 127] and [119].

3.1.3 Point-Voxel-Based 3D Object Detection

Point-voxel-based three-dimensional object identification methodologies signify a progression in autonomous driving perception systems. These strategies reconcile point-based and voxel-based approaches, leveraging the advantages of both paradigms while alleviating their respective shortcomings. This integration seeks to capitalize on the advantages of both approaches while mitigating their intrinsic limits, hence improving the accuracy and reliability of object detection in dynamic situations. Point-based approaches effectively capture high-resolution spatial information [132], although frequently encounter challenges related to computational complexity, particularly when processing sparse data [4], [133–136]. Conversely, voxel-based methods offer a structured data representation, hence improving computational efficiency. Nonetheless, they may forfeit intricate spatial information as a result of the discretization process. Point-voxel (PV) methods seek to achieve a balance by combining the detailed information acquisition capabilities of point-based approaches with the computing efficiency of voxel-based techniques [141–143].

The main aim of PV-based approaches is to enable feature interaction between voxels and points via point-to-voxel or voxel-to-point transformations. These methods provide comprehensive analysis of point cloud data, encompassing both global structures and micro-geometric features essential for the safety of autonomous driving systems. Photovoltaic-based techniques can be classified into two primary approaches; a notable direction involves the creation of single-stage and two-stage detection frameworks, each providing distinct benefits in acquiring intricate spatial information while addressing computing complexity [144].

Single-stage architectures aim to cohesively incorporate point and voxel properties within backbone networks [144, 145]. By utilising the detailed geometric information of points and the computational efficiency of voxels, these frameworks offer improved feature extraction and detection accuracy [146, 147]. Significant innovations encompass point-voxel convolutions and auxiliary point-based networks, which enhance the integration of point-voxel data. Conversely, two-stage frameworks utilize a multi-step methodology, implementing distinct data representations at each phase to enhance object detection outcomes [148]. Voxel-based detection frameworks first produce a collection of 3D item ideas, establishing a basis for further refining. During the refining phase,

Table 2. Summary related to most trending approaches and technologies.

Method	Technique	Paper Name(s)	Limitations
Point-Based 3d Object Detection	Point-based	[108], [104]	Limited performance in dense and cluttered environments. Lack of contextual information. Limited scalability to large scenes.
	Graph neural	[109], [110]	Complexity in defining and designing graph structures. Computationally expensive for large graphs. Difficulty in capturing long-range dependencies.
	Point cloud -based	[112], [118]	Computationally intensive due to processing large point clouds. Limited performance in handling irregular point densities. Limited capability to handle objects with complex shapes. Dependency on accurate object center estimation.
	Transformer-based	[116], [114]	Computationally expensive due to self-attention mechanisms. Difficulty in capturing fine-grained spatial relationships. Difficulty in capturing long-range dependencies. Limited scalability to large scenes.
	Feature-based	[105], [103]	Limited performance in preserving fine details. Difficulty in handling non-uniform point densities. Limited scalability to large point clouds. Dependency on accurate feature interpolation methods.
Grid Based 3d Object Detection	3D Fully Convolutional Network	[133], [121]	Limited to vehicle detection, may not generalize well to other object types
	Birdnet	[122]	Limited to bird detection, may not generalize well to other object types
	PointPillars	[127], [128]	Requires substantial computational resources
	Vote3Deep	[120]	Limited to fast object detection, may not perform well in scenarios with complex environments
	Voting for Voting	[119]	May not scale well to large point cloud datasets
	From Points to Parts	[129]	Limited to detecting objects with distinguishable parts
	SECOND	[126]	May struggle with dense scenes or cluttered environments
	HDNet	[124]	Relies on availability of high-definition maps
	PIXOR	[123]	May struggle with detecting small or occluded objects
	Center-based 3D Object Detection and Tracking	[131]	May have challenges with tracking objects in crowded scenes
Point- Voxel Based 3d detection	Voxel-to-Point Decoder	[137]	Lack of fine spatial details in voxels
	VoxelNet, PV-RCNN++	[125], [132], [134], [133], [135], [148], [152]	Requires substantial computational resources May struggle with detecting small or occluded objects Computational complexity for sparse data
	Pyramid R-CNNLiDAR R-CNN,	[146], [141]	Limited to monocular input; lacks stereo vision for depth perception, relies on LiDAR data; may be affected by occlusions and adverse weather conditions
	RTS3D	[143]	Requires stereo camera setup; may struggle in complex lighting conditions
	M3DSSD	[144]	Limited to monocular input; may lack depth perception in certain scenarios
	Hollow-3D R-CNN	[146]	Limited to monocular input; may lack depth perception in certain scenarios
	PP-RCNN, PillarNext	[150], [151]	Limited to LiDAR point cloud input; may struggle with sparse data
	SVGA -NeHVPR	[145], [139]	Limited by voxel-based representation; may lose fine details
	PVNAS	[142]	May require extensive computational resources for architecture search
	M3DETR	[147]	May be computationally intensive due to transformer-based architecture
	PVGNet	[138]	Limited to voxel-based representation; may struggle with fine-grained details

essential points are extracted from the input point cloud, and innovative point operators are utilized to further improve detection accuracy [149–151]. Techniques like RoI-grid pooling and point-wise attention are essential for enhancing object suggestions and augmenting the overall robustness of detection outcomes. The incorporation of these advancements into point-voxel-based algorithms promises enhanced accuracy and efficiency in 3D object detection for applications like autonomous driving [152]. However, obstacles persist, especially in reconciling detection precision with processing economy. Nevertheless, continuous research endeavors persist in advancing point-voxel-based detection, enhancing safety perception and decision-making for autonomous systems. Table 2 summarizes all such techniques along with their limitations on next page.

4 Comparative Analysis of Camera, LiDAR, and Radar Sensors

4.1 Utilization of Camera

Cameras are an economical alternative for applications that incorporate radar and imaging technology. They are frequently utilised in forward collision warning systems, lane departure warning systems, traffic sign recognition systems, parking assistance systems, and blind spot monitoring systems.

Cameras function by exposing photosensitive cells to light, resulting in a photoelectric effect that produces electrical charges. The picture signals are conveyed to the respective analogue signal processing unit and digital-to-analog conversion circuit. Cameras are categorized into monocular and binocular types, with monocular cameras offering a perspective of 50° to 60° with visual ranges of 100m to 200m. Binocular cameras can replicate human visual perception for three-dimensional imaging, enhance object recognition accuracy, and acquire distance and velocity data via algorithms. In contrast to LiDAR, cameras may utilize natural daylight and identify distant objects with superior resolution and reduced costs in well-lit environments [154, 155]. Cameras are vulnerable to variations in weather and lighting, and existing technology complicates the identification of distant objects in static images.

4.2 Utilization of LiDAR

LiDAR technology enables real-time environmental data sensing and the generation of high-definition 3D visuals. LiDAR employs pulsed modulated light to produce signals, track the time interval between emitted and reflected light, and simultaneously scan or measure several light beams to ascertain range and collect angular data [156]. This sensor system is the most accurate among regularly used environmental measurement choices, including rapid response, expanded detection range, and high precision. Noise is inevitably composed of the point cloud data gathered by the onboard LiDAR system due to several factors such as acceleration, deceleration, and alterations in driving direction [157]. Most LiDAR systems may provide strength data on the reflected pulse, reflecting the energy returned from the target surface and other object characteristics. The quantity of information conveyed by the light from the laser point cloud will, however, vary considerably according on the LiDAR systems, atmospheric conditions, and other specific circumstances.

Furthermore, LiDAR point cloud data is devoid of item category information, complicating automatic recognition and feature extraction. LiDAR's high cost for mass manufacturing may result in market trends such as the development of LiDAR with restricted detection ranges for particular applications and the use of solid-state LiDAR to lower expenses. These trends seek to enhance performance and save expenses in the LiDAR market [158–160].

4.3 Utilization of Radar

Radar sensors are crucial in autonomous cars for navigation, obstacle detection, and collision prevention. They produce radio waves throughout the microwave or radio frequency spectrum, specifically from 24 GHz to 77 GHz, and analyze reflections to discern the vehicle's surroundings [160, 161]. These sensors are engineered to function effortlessly in all locations and weather situations, guaranteeing uninterrupted and dependable detection capabilities. It generally functions within designated frequency bands, including the X-band (8-12 GHz), Ku-band (12-18 GHz), or mmWave band (24-77 GHz). These bands provide differing degrees of performance regarding range, resolution, and vulnerability to interference. Radar sensors can identify objects both in proximity and at a distance; they are essential for recognizing stationary barriers as well as tracking dynamic entities and forecasting their trajectories, hence improving safety and comfort for passengers and road users [162–165].

Cameras, LiDAR, and radar sensors are essential elements of perception systems in autonomous vehicles, delivering vital information regarding the

Table 3. Comparison for the utilization of Camera, Radar and LiDAR Sensors for different purpose.

Aspect	Camera	Radar	LiDAR
Object Detection	Good	Good	Excellent
Object Classification	Good	Fair	Fair to Good
Distance Estimation	Fair to Good	Excellent	Excellent
Edge Detection	Good	Fair to Good	Excellent
Lane Tracking	Fair to Good	Excellent	Fair to Good
Visibility Range	Limited	Long range	Moderate to Long range
Poor Weather Performance	Poor	Excellent	Moderate to Good
Low Light Performance	Poor	Fair to Good	Moderate to Good

vehicle's environment. Table 3 delineates the functions of cameras, LiDAR, and radar sensors, which exhibit varying performance metrics contingent upon range, meteorological conditions, data density, and cost.

4.4 Comparative analysis of Radar and LiDAR

Research comparing radar and LiDAR [162] for SLAM (Simultaneous Localization and Mapping) applications assesses the efficacy of both sensors utilizing two advanced SLAM algorithms: NDT-OM fuser and Gmapping. The Mechanical Pivoting Radar (MPR) utilised in their experiment was designed at Fraunhofer FHR, incorporating a 2D millimeter-wave radar affixed to a pivoting motor with a rotational velocity of 2.5Hz. The MPR offered a measurement precision of $\pm$3.75 cm within a range of 19 m, yielding 200 range readings for each complete rotation of the antenna. For comparison, they employed the Velodyne VLP-16, a 3D range scanner with a 100-meter range and a channel distribution of 2.00° between channels, providing a range accuracy of $\pm$3 cm [165]. The assessment technique utilized was the Relative Pose Error (RPE) method introduced in [167]. for the KITTI dataset. This method assessed the displacement at each posture relative to a surrounding neighborhood of poses, treating translation and rotation errors independently. The experimental results presented in the table elucidated the efficacy of radar and LiDAR sensors in SLAM applications [168]. The Velodyne LiDAR exhibited enhanced precision in calculating robot trajectories and constructing environmental maps relative to the MPR radar. Gmapping typically produced trajectory estimates that were more aligned with the ground truth endpoint than the NDT-OM fuser when utilizing the MPR, however the NDT-OM fuser exhibited marginally superior performance with the Velodyne LiDAR. One may see Table 4 that summarizes the performance of Radar and LiDAR [169].

A comparative research study [162] on heuristic fusion with adaptive gating versus track-to-track fusion algorithms for sensor fusion in forward vehicle tracking, utilizing simulated data from camera and radar sensors, indicates that the track-to-track fusion algorithm surpassed the heuristic fusion with adaptive gating algorithm, achieving an average accuracy enhancement of 29.39% across multiple scenarios [170]. The track-to-track fusion technique demonstrated a decrease in data association failures, as evidenced by the OSPA distance graphs. In the evaluation of camera and radar sensor fusion performance, both methodologies demonstrated efficacy across various scenarios. Radar sensor fusion exhibited comparable or marginally superior performance in scenarios such as linear motion, creation, and elimination. Camera sensor fusion demonstrated superior accuracy enhancement rates in situations including stopping and curved motion. The results indicate the superiority of the track-to-track fusion method for forward vehicle tracking applications, with ramifications for practical vehicle systems [161–165].

5 Integration of Camera and LiDAR Sensor

The integration of camera and LiDAR sensors in autonomous cars promises to substantially improve perception capabilities, especially in difficult environmental situations. Researchers [164–170] seek to enhance perception systems by integrating the advantages of both cameras, which offer extensive visual information, and LiDAR, known for its precise 3D mapping capabilities, therefore addressing the limits of individual sensors.

In these studies, researchers performed trials to assess object detection performance in autonomous vehicle sensors, both separately and in combination. They examine the constraints of current object detecting methods, especially under adverse lighting

Table 4. Performance comparison between Radar and LiDAR

Sensor	SLAM	Accuracy	Translation Error	Orientation Error	Map Accuracy
Radar (MPR)	NDT-OM fuser	High	Similar to LiDAR	Similar to LiDAR	Inferior
	Gmapping	High	Similar to LiDAR	Similar to LiDAR	Inferior
LiDAR (Velodyne)	NDT-OM fuser	High	Similar to Radar	Similar to Radar	Superior
	Gmapping	High	Similar to Radar	Similar to Radar	Superior

and meteorological circumstances. Conventional object detection systems predominantly depend on vision sensors, which may falter in low visibility conditions [171]. To tackle these problems, some studies advocate for sensor fusion systems that integrate thermal infrared cameras with LiDAR sensors. Thermal infrared cameras perform exceptionally well in low visibility environments, but LiDAR sensors deliver accurate three-dimensional spatial data. The integration of these sensors seeks to enhance the reliability of object detection, particularly in areas with limited visibility [164], [167]. Prior to integration, the documents emphasize the necessity for precise extrinsic parameter calibration between the thermal infrared camera and the LiDAR sensor. Conventional calibration techniques may not be immediately applicable to thermal infrared cameras due to difficulties in precisely extracting three-dimensional shapes. A article presents an innovative calibration technique employing a 3D calibration target and feature point extraction to accurately align the sensors. During daylight conditions, the object detection Average Precision (AP) for the visual camera and LiDAR sensor is 56.167%, but for the thermal infrared camera and LiDAR sensor, it is 55.914%. During dark conditions, the visual camera and LiDAR sensor exhibit a performance level with an Average Precision (AP) of 49.878%, but the thermal infrared camera and LiDAR sensor get an AP of 57.516% [164].

5.1 Probabilistic Sensor Fusion Approach

A different method employs an innovative probabilistic sensor fusion technique to amalgamate semantic information from cameras with accurate 3D data from LiDAR sensors, thereby generating three-dimensional semantic voxelized maps for autonomous cars in urban settings [165], [168]. This methodology tackles essential difficulties such as sensor synchronization, motion distortion compensation, occlusion management, and uncertainty propagation. The effectiveness of this pipeline is confirmed through stringent experiments utilizing datasets that include pictures, point clouds, and odometry data. The experimental evaluation assesses three LIDAR-image projection methodologies: direct projection, motion-corrected projection, and projection with occlusion management, using ground truth labels from 20 LIDAR scans as a benchmark. The results provide strong evidence for the efficacy of the proposed approach, especially with the integration of motion correction and occlusion handling techniques. Quantitative metrics, like as recall and precision, act as essential indicators of performance improvement. The occlusion management technique exhibits a significant reduction in the mislabeling of occluded points, leading to a 7% drop in the quantity of labelled points per scan relative to direct projection [165].

5.2 Object Detection and Ranging System

A vision system that is specifically designed for autonomous driving is another key contribution. This system places an emphasis on the vital need for exact object recognition, categorization, and ranging under driving settings that are both complicated and unpredictable. This sensor fusion technique takes advantage of the complimentary characteristics of camera and LiDAR sensors in order to improve the resilience and accuracy of the system [166], [169]. While cameras are excellent at recognizing color, form, and different types of objects, they are not capable of detecting depth. In contrast, LiDAR is capable of producing accurate three-dimensional mapping, but it has difficulty detecting objects due to the presence of sparse point clouds and a wide range of geometric shapes. The method of fusion involves projecting LiDAR point clouds onto camera images in order to integrate information in a seamless manner. This process is confirmed by employing algorithms such as YOLOv3 and Point Pillars.

5.3 Enhancing Accuracy and Real-time performance

By combining the raw data from LiDAR and video sensors, additional research is being conducted to investigate the possibility of improving obstacle detection in autonomous cars [169], [170].

Inaccuracies in perception, sensor calibration, and data synchronization are all subjects that are addressed by the methodology. These constraints are intended to be overcome by the fusion, which also intends to improve perceptual accuracy. The procedure involves the calibration of both intrinsic and extrinsic sensors, which is validated by the analysis of real-time data. It also enables the identification of obstacles, classification of obstacles, and assessment of depth. Sensor data fusion has been shown to increase accuracy in various tasks, with considerable reductions in translational, rotational, and projection errors, as demonstrated by the results of the experiments. Real-time perception is made possible for autonomous vehicles through the combination of raw data from LiDAR and image sensors, which dramatically minimizes the amount of time required for processing.

5.4 Indoor Mapping & Combined Sensor Efficacy

Another innovative method involves data fusion mapping using single-line LiDAR, depth camera, and IMU sensors to overcome limitations in single-sensor mapping in indoor environments. Individual sensors faced challenges such as sparse point cloud data from LiDAR, limitations in weak light or texture less environments for the depth camera, and IMU drift over time as shown in Figure 3. This approach compensates for errors caused by LiDAR measurements through fusion of environmental feature data from LiDAR and pseudo laser data from the depth camera, and pose information from the IMU using Kalman Filtering [170]. The global map generated after fusion ensures mapping accuracy and provides a more comprehensive environmental map for mobile robots, reducing the possibility of collisions with obstacles during operation.

The analysis of combined sensor effectiveness, as evidenced in multiple studies, highlights the integration of thermal infrared cameras with LiDAR sensors for object detection in autonomous cars. These investigations underscore the difficulties encountered by conventional object detecting systems, especially under unfavorable lighting and meteorological circumstances. Researchers advocate for sensor fusion systems to improve the reliability of object detection, particularly in low-visibility conditions. Accurate calibration between thermal infrared cameras and LiDAR sensors is essential prior to integration, necessitating calibration techniques to overcome obstacles in precisely extracting 3D shapes [170]. The Table 5 presents diverse efficacy results from the tests,

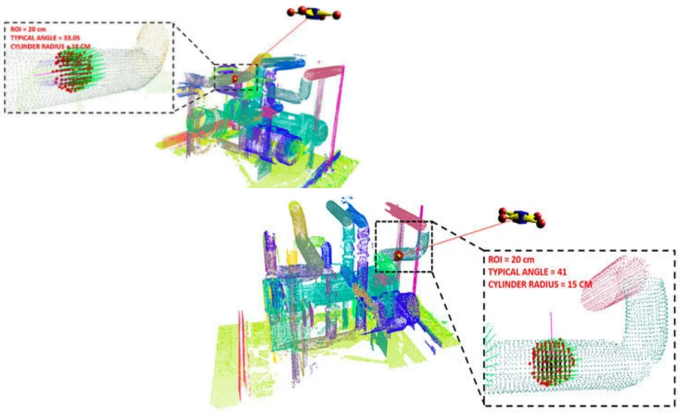

Figure 3. Indoor mapping and combined sensor.

indicating promising outcomes. In daylight conditions, the amalgamation of thermal infrared cameras with LiDAR sensors achieves object detection performance on par with the combination of visual cameras and LiDAR sensors. In low-light circumstances, thermal infrared cameras and LiDAR sensors outperform visual cameras and LiDAR sensors, underscoring the efficacy of thermal infrared sensors in enhancing object identification reliability in difficult lighting scenarios. The findings highlight the effectiveness of sensor fusion methods in improving object identification abilities in autonomous cars, especially in challenging weather situations. Table 5 summarizes further insights into the usefulness of coupled sensor systems, highlighting enhanced results for their effectiveness.

6 Benefits of Integrating LiDAR and Camera

The integration of camera and LiDAR systems in autonomous cars offers significant advantages by utilizing their complementary strengths and mitigating their unique limitations. This amalgamation fosters a more dependable perceptual system. It improves perception by integrating high-resolution images from cameras with the intricate 3D point clouds produced by LiDAR sensors [168]. Cameras provide visual data on item characteristics, including look, color, and texture, whilst LiDAR delivers precise distance measurements, culminating in a comprehensive system for object detection and classification. Furthermore, it enhances object detection. Cameras excel at identifying visual characteristics but may encounter difficulties in low-light or adverse visibility circumstances [169, 170]. LiDAR, impervious to lighting conditions, provides accurate depth information, enhancing the reliability of object detection across diverse environmental scenarios. Furthermore, it guarantees resilience under unfavorable circumstances. LiDAR operates

Table 5. Summary related to Combined Sensor Technologies and Approaches.

Paper	Limitations of Individual Sensors	Technique	Improvements
Object Detection in AV Sensors [161]	Reliance on vision sensors, limited visibility conditions	Sensor fusion, Extrinsic calibration	-Improving object detection reliability in poor visibility scenarios for day and night.
Probabilistic Sensor Fusion Approach for Avs [162]-[163]	Sensor synchronization, motion distortion, occlusion handling, uncertainty propagation	Sensor fusion, Calibration, Motion correction	-Improved recall and precision metrics. -Reduced mislabeling of occluded points. -Enhanced accuracy in object recognition and semantic mapping, addressing challenges of motion distortion and occlusions.
Multimodal Object Detection and Ranging [164]	Lack of depth perception in cameras, sparse point clouds from LiDAR	Sensor fusion, Camera-LiDAR calibration	-Successful detection and ranging in dynamic driving environments. -Achieving heightened perception accuracy. -Precise fusion through meticulous calibration. -Real-time processing for successful detection.
Sensor Fusion in Autonomous Vehicles [165]	Adverse weather, low light, errors in position estimation	Sensor fusion, Odometry, Kalman filters	Enhancing estimation accuracy, reducing errors in AV performance, seamless integration of sensor data.
LiDAR and Camera Raw Data Sensor Fusion [166]	Sparse point clouds from LiDAR, limitations in weak light for cameras	Sensor fusion, Intrinsic and extrinsic calibration	-Reduction in processing time for real-time perception. -Enhancing obstacle detection, classification, and depth estimation accuracy. -Real-time processing for improved autonomous vehicle (AV) perception.
Data Fusion Mapping Using LiDAR, Depth Camera, and IMU Sensors [167]-[168]	Sparse point clouds from LiDAR, limitations in weak light or texture less environments for depth camera, IMU drift over time	Data fusion, Kalman Filtering	-Significant improvements in mapping accuracy. -Global map generation ensuring accuracy. -Comprehensive environmental mapping. -Overcoming limitations of single-sensor mapping. -Enhancing map construction accuracy. -Reducing collisions with obstacles during operation.

independently of light conditions or weather by assessing the time-of-flight of laser pulses. Cameras provide contextual elements such as color and texture, facilitating object identification in conditions of fog, rain, or darkness. The incorporation of these sensors improves overall performance and dependability [171].

Furthermore, it offers precise geographical mapping. LiDAR produces rich three-dimensional point clouds with accurate spatial data, whilst cameras provide intricate visual information. The integration of these elements facilitates precise environmental mapping, crucial for navigation and obstacle evasion. Fifthly, it enhances redundancy and reliability. Should one sensor encounter problems, such as obstructions or malfunctions, the other might provide compensation. For instance, if a camera encounters glare or shadows, LiDAR maintains depth information, guaranteeing consistent functionality and safety. Sixthly, it is economically advantageous. Although LiDAR technology is costly, cameras are comparatively economical. Combining both diminishes total system expenses while preserving high performance, hence enhancing the accessibility and scalability of autonomous systems. Ultimately, it provides diversity and adaptability. Cameras record visual data such as colors and textures, while LiDAR delivers accurate spatial information. This combination enables the system to address a range of applications, from autonomous driving to environmental monitoring, by meeting varied constraints and requirements [171].

7 Affordable Alternatives

The advancement of sensor technology for autonomous vehicles and advanced driver assistance systems (ADAS) has led to the exploration of many cost-effective alternatives to LiDAR, mostly focusing on camera and radar sensor technologies. These LiDAR alternatives provide substantial advantages in terms of cost-effectiveness and usefulness, although they come with inherent constraints. This section examines the technical analysis and performance implications of diverse sensors, substantiated by recent research and conclusions in the domain [172]. Previous studies have investigated the effectiveness of camera and radar systems under various driving situations, providing a foundation for current beliefs about their viability as economical alternatives to

LiDAR. A study [173] suggested that using these sensors could deliver sufficient environmental awareness for autonomous vehicles at a significantly reduced cost compared to LiDAR systems. The working hypothesis posits that despite the inherent limitations of each sensor, their combined application through sophisticated sensor fusion algorithms can achieve performance comparable to that of LiDAR. Camera-based systems are extensively employed for their enhanced resolution and ability to record complex visual details. These systems utilize sophisticated computer vision algorithms, including convolutional neural networks (CNNs) and deep learning techniques, to examine and classify objects, lane markings, traffic signals, and pedestrians. Recent developments in camera technology, such as high dynamic range (HDR) pictures and greater low-light performance, have boosted their functionalities. A study by [23] demonstrates that multi-camera systems significantly improve the field of view and depth perception, rendering them more useful for intricate driving situations.

Camera systems have difficulties in inclement weather and low-light environments. Precipitation, fog, and glare can impair image quality, thus diminishing the effectiveness of computer vision algorithms [174]. To resolve these challenges, proposed solutions include polarizing filters and sophisticated image processing techniques, such as noise reduction and contrast enhancement algorithms. Polarizing filters can diminish glare from wet road surfaces and enhance image quality in inclement weather, while sophisticated image processing techniques can alleviate problems associated with low-light and fog, so enhancing the overall dependability of camera-based systems in adverse situations. Radar sensors operate by producing radio waves and examining their reflections off objects, guaranteeing dependable operation in many environmental situations, including fog, rain, and darkness. The capacity of radar to precisely evaluate object velocity is very advantageous for collision avoidance systems. The cost-efficiency of radar, along with recent developments in high-resolution imaging radar technologies, such as Frequency-Modulated Continuous Wave (FMCW) radar, renders it a compelling option. A study [174] demonstrates that contemporary radar systems can get significant resolution enhancements through synthetic aperture radar (SAR) methodologies, improving their capacity to differentiate between closely positioned objects. Notwithstanding these developments, radar sensors frequently yield inferior spatial resolution compared to LiDAR, so constraining their capacity to generate intricate 3D representations of the environment.

Sensor fusion methodologies are utilized to alleviate this constraint. Research [153] indicates that the amalgamation of radar and camera data merges the high resolution of cameras with the dependability of radar, yielding a more resilient and precise perception system. This method employs deep learning models that can analyze multi-sensor data, enhancing object recognition and classification precision. Previous investigations and the foundational premise suggest that a multi-sensor strategy, combining cameras and radar, can mitigate the limitations inherent in the exclusive use of individual sensors. Algorithms for sensor fusion, particularly those utilizing deep learning frameworks like YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector), are being designed to effectively integrate data from these sensors. These algorithms can evaluate extensive datasets in real-time, providing critical environmental awareness necessary for autonomous driving [173].

Radar and camera technologies offer considerable advantages over LiDAR in terms of cost-effectiveness. Cameras are generally economical, and even advanced radar sensors are far more affordable than LiDAR systems [174]. The diminished cost of these sensors can decrease the total expenditure of autonomous vehicle systems, enhancing their accessibility for general use. Future research directions include the enhancement of sensor fusion methodologies, the fortification of camera systems under adverse conditions, and the advancement of radar technology for superior resolution and enhanced environmental awareness. Investigating advanced machine learning techniques for sensor fusion, such as Transformer models and Graph Neural Networks (GNNs), has the potential to enhance the precision and dependability of these systems. Furthermore, investigating meta-learning and transfer learning may enhance the adaptability of these systems to novel contexts and situations [175].

8 Technical Discussion on Limitations & Shortcomings

The incorporation of many sensors in autonomous vehicles is essential for attaining dependable and precise environmental perception. This integration poses numerous problems and limitations stemming from the fundamental disparities in

sensor technologies, data processing necessities, and operational situations. A significant problem is the intricacy of sensor fusion. Integrating data from LiDAR, radar, and cameras necessitates meticulous calibration and synchronization, given that each sensor functions on varying physical principles and generates data in unique formats [175]. This integration necessitates advanced algorithms adept at handling temporal and spatial inconsistencies to provide precise item detection and localization. Misalignment or synchronization issues may result in unreliable outcomes, jeopardizing the vehicle's safety and performance. Moreover, real-time data processing presents a considerable hurdle. Autonomous vehicles rely on swift decision-making for secure navigation, and the substantial data generated by high-resolution sensors such as LiDAR and cameras necessitates robust processing units and refined algorithms. The requirement for substantial processing capacity not only elevates energy consumption but also contributes to the overall expense and intricacy of the autonomous system [152, 153].

Environmental variables exacerbate the complexities of sensor integration. Although LiDAR and radar are less influenced by lighting conditions, they may encounter difficulties in severe weather, like heavy rain, snow, or fog, which might compromise data quality. Cameras, as passive sensors, exhibit heightened sensitivity to fluctuations in lighting and atmospheric conditions, which may impair visibility and diminish image quality. Integrating data from various sensors necessitates resilient algorithms to counteract environmental influences. The substantial expense associated with the implementation and upkeep of various sensor systems, especially LiDAR, constrains access to sophisticated autonomous systems, hindering their acceptance and elevating operational costs due to continuous maintenance and calibration demands. Privacy issues emerge with camera usage, as they record intricate visual information that may encompass identifying personal data [149]. Reconciling environmental awareness with privacy concerns necessitates meticulous planning and data management techniques. Scalability is a significant concern, as handling extensive data from numerous sensors while maintaining real-time processing and decision-making requires sophisticated communication infrastructure and resilient data management systems. Notwithstanding considerable progress in 3D object detection technologies, contemporary state-of-the-art methodologies face numerous constraints that impact their efficacy and wider use. These constraints encompass sensor capability, data processing efficiency, environmental adaptability, and computational requirements [123].

LiDAR sensors, despite generating high-density point clouds, pose issues related to computational complexity and resource requirements. Real-time processing and interpretation of extensive data necessitates sophisticated gear, which may be impractical in resource-limited settings. Moreover, LiDAR data may be sparse and partial, particularly at extended distances or in the presence of obstructions, hence complicating the detection and classification processes. While less influenced by lighting conditions, LiDAR performance may deteriorate in inclement weather such as heavy rain, snow, or fog. The elevated expense of LiDAR sensors continues to pose a substantial obstacle, hindering widespread adoption despite declining prices. Radar sensors, although resilient in diverse environmental circumstances, generally provide worse resolution relative to LiDAR and cameras. This diminished resolution can impede the identification and distinction of small or closely positioned items, hence impacting the precision of 3D object detection [134]. Moreover, radar's capacity to deliver specific information regarding item forms and dimensions is constrained, affecting applications that necessitate accurate object detection and classification. Signal interference and crosstalk, especially in settings with numerous radar systems, can adversely impact detection accuracy [135], [163].

Cameras depend on ambient illumination, rendering them vulnerable to fluctuating lighting conditions. Factors such as insufficient illumination, shadows, and glare can considerably affect image quality and detection precision. Cameras produce 2D images devoid of intrinsic depth sense, hence confounding the assessment of distance and size. Stereo vision techniques provide limited depth information but are generally less precise and more computationally demanding than LiDAR. Cameras are particularly susceptible to inclement weather, which can obstruct visibility and diminish image quality [170].

9 Future Directions and Recommendations

When it comes to achieving dependable and precise environmental perception, the incorporation of several sensors into autonomous cars is absolutely necessary. Nevertheless, in order to ensure the progression of technology, it is necessary to conquer the difficulties that are related with this integration.

Enhancing sensor fusion techniques should be the primary emphasis of future research in order to better manage the difficulties involved in merging data from LiDAR technologies, radar, and cameras. Specifically, this entails the development of more sophisticated algorithms for calibration and synchronization, the resolution of temporal and spatial inconsistencies, and the enhancement of the robustness of data fusion in order to guarantee accurate object detection and localization. Furthermore, it is of the utmost importance to effectively solve the difficulty of real-time data processing. For the purpose of managing the huge amounts of data that are produced by high-resolution sensors, this can be accomplished through the development of algorithms that are more effective and by the utilisation of developments in hardware acceleration, such as graphics processing units (GPUs) and specialized processors [11–15]. Additionally, the investigation of edge computing and distributed processing strategies can be of assistance in lowering latency and increasing the speed of decision-making. In particular, environmental adaptation continues to be a key concern, especially when poor weather conditions are present. In the future, research should concentrate on either building adaptive algorithms that can react to changing environmental elements or designing sensors that are more resistant to the conditions that are now being encountered [25–35]. Furthermore, in order to facilitate wider use, it is necessary to achieve a reduction in the high prices associated with advanced sensors, particularly LiDAR. The goal of research should be to find alternatives that are more cost-effective and to investigate improvements in the fabrication of sensors. The protection of privacy, in particular with regard to camera data, necessitates the implementation of sophisticated data management systems that strike a compromise between environmental perception and privacy concerns [166]. This comprises the implementation of procedures that protect individuals' privacy and the guaranteeing of conformity with rules. Last but not least, it is essential to address scalability difficulties by creating effective communication infrastructure and data management systems in order to handle large amounts of data while preserving the ability to analyze it in real time [170–175].

10 Conclusion

The combination of LiDAR, radar, and cameras is crucial for ensuring dependable and precise environmental perception in autonomous vehicles. This integration, while utilizing the advantages of several sensors, poses several obstacles. Sensor fusion's complexity necessitates advanced algorithms to process data from several sensors, each functioning on distinct physical principles and data formats. Misalignment or synchronization issues might compromise object detection accuracy and localization, hence affecting vehicle performance and safety. Furthermore, handling the extensive data streams produced by high-resolution sensors requires substantial computer resources, rendering real-time processing a formidable issue. Environmental considerations exacerbate integration challenges; LiDAR and radar efficacy diminishes in inclement weather, whilst cameras are influenced by variations in lighting and atmospheric conditions. A notable obstacle to widespread adoption is the elevated expense of sensors, especially LiDAR. Financial considerations significantly influence the viability of extensive deployment of autonomous systems. In addition to technical and budgetary obstacles, data privacy poses a significant barrier, especially in the management of camera data, which may unintentionally record sensitive personal information. Adherence to data protection standards, including the General Data Protection Regulation (GDPR) in Europe, is essential for mitigating these issues. Autonomous vehicles may need to employ real-time data anonymisation methods or restrict data retention to prevent incorrect storage or misuse of personal information. Measures such as encrypted data storage and restricting image collection to non-identifiable information are now being investigated to alleviate privacy problems. Despite progress in 3D object detection technologies, existing methodologies continue to encounter constraints concerning sensor performance, data processing efficacy, and environmental adaptability. Mitigating these limits via focused research and development is essential for enhancing sensor integration, augmenting reliability, decreasing costs, and progressing the future of autonomous systems. Furthermore, continuous initiatives to mitigate privacy threats via technological innovations and legal structures will guarantee that autonomous vehicle systems function in a way that honours individual privacy while delivering strong and efficient performance.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgement

This work was supported by Interdisciplinary Research Centre for Aviation and Space Exploration (IRCASE), King Fahd University of Petroleum and Minerals (KFUPM, Kingdom of Saudi Arabia).

References

[1] Banham, M. R., & Katsaggelos, A. K. (1997). Digital image restoration. *IEEE signal processing magazine*, 14(2), 24-41. [CrossRef]

[2] Bao, W., Xu, B., & Chen, Z. (2019). Monofenet: Monocular 3d object detection with feature enhancement networks. *IEEE Transactions on Image Processing*, 29, 2753-2765. [CrossRef]

[3] Barabas, I., Todoruţ, A., Cordoş, N., & Molea, A. (2017, October). Current challenges in autonomous driving. In *IOP conference series: materials science and engineering* (Vol. 252, No. 1, p. 012096). IOP Publishing. [CrossRef]

[4] Li, J., Yang, B., Chen, C., Huang, R., Dong, Z., & Xiao, W. (2018). Automatic registration of panoramic image sequence and mobile laser scanning data using semantic features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136, 41-57. [CrossRef]

[5] Liao, Y., Li, J., Kang, S., Li, Q., Zhu, G., Yuan, S., ... & Yang, B. (2023). SE-Calib: Semantic Edge-Based LiDAR–Camera Boresight Online Calibration in Urban Scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-13. [CrossRef]

[6] Wang, J. G., & Zhou, L. B. (2018). Traffic light recognition with high dynamic range imaging and deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 20(4), 1341-1352. [CrossRef]

[7] Melotti, G., Premebida, C., Gonçalves, N. M. D. S., Nunes, U. J., & Faria, D. R. (2018, November). Multimodal CNN pedestrian classification: a study on combining LIDAR and camera data. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 3138-3143). IEEE. [CrossRef]

[8] Wang, K., Ma, S., Ren, F., & Lu, J. (2021). SBAS: Salient bundle adjustment for visual SLAM. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-9. [CrossRef]

[9] Kurihata, H., Takahashi, T., Ide, I., Mekada, Y., Murase, H., Tamatsu, Y., & Miyahara, T. (2005, June). Rainy weather recognition from in-vehicle camera images for driver assistance. In *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.* (pp. 205-210). IEEE. [CrossRef]

[10] Webster, D. D., & Breckon, T. P. (2015, September). Improved raindrop detection using combined shape and saliency descriptors with scene context isolation. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 4376-4380). IEEE. [CrossRef]

[11] Zhang, W., Wang, Z., & Change Loy, C. Multi-modality cut and paste for 3d object detection. arXiv 2020. arXiv preprint arXiv:2012.12741.

[12] Filgueira, A., González-Jorge, H., Lagüela, S., Díaz-Vilariño, L., & Arias, P. (2017). Quantifying the influence of rain in LiDAR performance. *Measurement*, 95, 143-148. [CrossRef]

[13] Rasshofer, R. H., Spies, M., & Spies, H. (2011). Influences of weather phenomena on automotive laser radar systems. *Advances in radio science*, 9, 49-60. [CrossRef]

[14] Abro, G. E. M., Zulkifli, S. A. B., Kumar, K., El Ouanjli, N., Asirvadam, V. S., & Mossa, M. A. (2023). Comprehensive review of recent advancements in battery technology, propulsion, power interfaces, and vehicle network systems for intelligent autonomous and connected electric vehicles. *Energies*, 16(6), 2925. [CrossRef]

[15] Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., ... & Dietmayer, K. (2020). Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3), 1341-1360. [CrossRef]

[16] Wei, Z., Zhang, F., Chang, S., Liu, Y., Wu, H., & Feng, Z. (2022). Mmwave radar and vision fusion for object detection in autonomous driving: A review. *Sensors*, 22(7), 2542. [CrossRef]

[17] Svenningsson, P., Fioranelli, F., & Yarovoy, A. (2021, May). Radar-pointgnn: Graph based object recognition for unstructured radar point-cloud data. In *2021 IEEE Radar Conference (RadarConf21)* (pp. 1-6). IEEE. [CrossRef]

[18] Ulrich, M., Braun, S., Köhler, D., Niederlöhner, D., Faion, F., Gläser, C., & Blume, H. (2022, October). Improved orientation estimation and detection with hybrid object detection networks for automotive radar. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 111-117). IEEE. [CrossRef]

[19] Kim, Y., Choi, J. W., & Kum, D. (2020, October). Grif net: Gated region of interest fusion network for robust 3d object detection from radar point cloud and monocular image. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 10857-10864). IEEE. [CrossRef]

[20] Chadwick, S., Maddern, W., & Newman, P. (2019, May). Distant vehicle detection using radar and vision. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 8311-8317). IEEE. [CrossRef]

[21] Nobis, F., Geisslinger, M., Weber, M., Betz, J., & Lienkamp, M. (2019, October). A deep learning-based radar and camera sensor fusion architecture for object detection. In *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)* (pp. 1-7). IEEE. [CrossRef]

[22] John, V., & Mita, S. (2019). RVNet: Deep sensor fusion of monocular camera and radar for image-based obstacle detection in challenging environments. In

Image and Video Technology: 9th Pacific-Rim Symposium, PSIVT 2019, Sydney, NSW, Australia, November 18–22, 2019, Proceedings 9 (pp. 351-364). Springer International Publishing. [CrossRef]

[23] Li, L. Q., & Xie, Y. L. (2020, December). A feature pyramid fusion detection algorithm based on radar and camera sensor. In *2020 15th IEEE International Conference on Signal Processing (ICSP)* (Vol. 1, pp. 366-370). IEEE. [CrossRef]

[24] Chang, S., Zhang, Y., Zhang, F., Zhao, X., Huang, S., Feng, Z., & Wei, Z. (2020). Spatial attention fusion for obstacle detection using mmwave radar and vision sensor. *Sensors*, 20(4), 956. [CrossRef]

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

[26] Li, Y., Zeng, K., & Shen, T. (2023). CenterTransFuser: radar point cloud and visual information fusion for 3D object detection. *EURASIP Journal on Advances in Signal Processing*, 2023(1), 7. [CrossRef]

[27] Long, Y., Kumar, A., Morris, D., Liu, X., Castro, M., & Chakravarty, P. (2023, June). RADIANT: Radar-image association network for 3D object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 2, pp. 1808-1816). [CrossRef]

[28] Li, P., Zhao, H., Liu, P., & Cao, F. (2020, August). Rtm3d: Real-time monocular 3d detection from object keypoints for autonomous driving. In *European Conference on Computer Vision* (pp. 644-660). *Cham: Springer International Publishing*. [CrossRef]

[29] Zhang, Y., Lu, J., & Zhou, J. (2021). Objects are different: Flexible monocular 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3289-3298). [CrossRef]

[30] Simonelli, A., Bulo, S. R., Porzi, L., López-Antequera, M., & Kontschieder, P. (2019). Disentangling monocular 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1991-1999). [CrossRef]

[31] Brazil, G., & Liu, X. (2019). M3d-rpn: Monocular 3d region proposal network for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9287-9296). [CrossRef]

[32] Cai, Y., Li, B., Jiao, Z., Li, H., Zeng, X., & Wang, X. (2020, April). Monocular 3d object detection with decoupled structured polygon estimation and height-guided depth estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 10478-10485). [CrossRef]

[33] Chen, H., Huang, Y., Tian, W., Gao, Z., & Xiong, L. (2021). Monorun: Monocular 3d object detection by reconstruction and uncertainty propagation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10379-10388). [CrossRef]

[34] Chen, Y., Tai, L., Sun, K., & Li, M. (2020). Monopair: Monocular 3d object detection using pairwise spatial relationships. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12093-12102). [CrossRef]

[35] Heylen, J., De Wolf, M., Dawagne, B., Proesmans, M., Van Gool, L., Abbeloos, W., ... & Reino, D. O. (2021). Monocinis: Camera independent monocular 3d object detection using instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 923-934). [CrossRef]

[36] Liu, Z., Wu, Z., & Tóth, R. (2020). Smoke: Single-stage monocular 3d object detection via keypoint estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 996-997). [CrossRef]

[37] Liu, L., Lu, J., Xu, C., Tian, Q., & Zhou, J. (2019). Deep fitting degree scoring network for monocular 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1057-1066). [CrossRef]

[38] Luo, S., Dai, H., Shao, L., & Ding, Y. (2021). M3dssd: Monocular 3d single stage object detector. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6145-6154). [CrossRef]

[39] Wang, T., Zhu, X., Pang, J., & Lin, D. (2021). Fcos3d: Fully convolutional one-stage monocular 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 913-922). [CrossRef]

[40] Lu, Y., Ma, X., Yang, L., Zhang, T., Liu, Y., Chu, Q., ... & Ouyang, W. (2021). Geometry uncertainty projection network for monocular 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3111-3121). [CrossRef]

[41] Mousavian, A., Anguelov, D., Flynn, J., & Kosecka, J. (2017). 3d bounding box estimation using deep learning and geometry. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 7074-7082). [CrossRef]

[42] Brazil, G., Pons-Moll, G., Liu, X., & Schiele, B. (2020). Kinematic 3d object detection in monocular video. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16* (pp. 135-152). Springer International Publishing. [CrossRef]

[43] Simonelli, A., Bulo, S. R., Porzi, L., Ricci, E., & Kontschieder, P. (2020). Towards generalization across depth for monocular 3d object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16* (pp. 767-782). Springer International Publishing.

[CrossRef]

[44] Li, B., Ouyang, W., Sheng, L., Zeng, X., & Wang, X. (2019). Gs3d: An efficient 3d object detection framework for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1019-1028). [CrossRef]

[45] Qin, Z., Wang, J., & Lu, Y. (2019, July). Monogrnet: A geometric reasoning network for monocular 3d object localization. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 8851-8858). [CrossRef]

[46] Shi, X., Chen, Z., & Kim, T. K. (2020). Distance-normalized unified representation for monocular 3d object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16* (pp. 91-107). Springer International Publishing. [CrossRef]

[47] Hu, H. N., Cai, Q. Z., Wang, D., Lin, J., Sun, M., Krahenbuhl, P., ... & Yu, F. (2019). Joint monocular 3D vehicle detection and tracking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 5390-5399). [CrossRef]

[48] Ku, J., Pon, A. D., & Waslander, S. L. (2019). Monocular 3d object detection leveraging accurate proposals and shape reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11867-11876). [CrossRef]

[49] Lian, Q., Ye, B., Xu, R., Yao, W., & Zhang, T. (2022). Exploring geometric consistency for monocular 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1685-1694). [CrossRef]

[50] Zeeshan Zia, M., Stark, M., & Schindler, K. (2014). Are cars just 3d boxes?-jointly estimating the 3d shape of multiple objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3678-3685). [CrossRef]

[51] Chabot, F., Chaouch, M., Rabarisoa, J., Teuliere, C., & Chateau, T. (2017). Deep manta: A coarse-to-fine many-task network for joint 2d and 3d vehicle analysis from monocular image. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2040-2049). [CrossRef]

[52] He, T., & Soatto, S. (2019, July). Mono3d++: Monocular 3d vehicle detection with two-scale 3d hypotheses and task priors. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 8409-8416). [CrossRef]

[53] Rogage, K., & Doukari, O. (2024). 3D object recognition using deep learning for automatically generating semantic BIM data. *Automation in Construction*, 162, 105366. [CrossRef]

[54] Manhardt, F., Kehl, W., & Gaidon, A. (2019). Roi-10d: Monocular lifting of 2d detection to 6d pose and metric shape. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2069-2078). [CrossRef]

[55] Beker, D., Kato, H., Morariu, M. A., Ando, T., Matsuoka, T., Kehl, W., & Gaidon, A. (2020). Monocular differentiable rendering for self-supervised 3d object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16* (pp. 514-529). Springer International Publishing. [CrossRef]

[56] Zakharov, S., Kehl, W., Bhargava, A., & Gaidon, A. (2020). Autolabeling 3d objects with differentiable rendering of sdf shape priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12224-12233). [CrossRef]

[57] Jörgensen, E., Zach, C., & Kahl, F. (2019). Monocular 3d object detection and box fitting trained end-to-end using intersection-over-union loss. *arXiv preprint arXiv:1906.08070*. [CrossRef]

[58] Naiden, A., Paunescu, V., Kim, G., Jeon, B., & Leordeanu, M. (2019, September). Shift r-cnn: Deep monocular 3d object detection with closed-form geometric constraints. In *2019 IEEE international conference on image processing (ICIP)* (pp. 61-65). IEEE. [CrossRef]

[59] Shi, X., Ye, Q., Chen, X., Chen, C., Chen, Z., & Kim, T. K. (2021). Geometry-based distance decomposition for monocular 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 15172-15181). [CrossRef]

[60] Wang, T., Pang, J., & Lin, D. (2022, October). Monocular 3d object detection with depth from motion. In *European Conference on Computer Vision* (pp. 386-403). Cham: Springer Nature Switzerland. [CrossRef]

[61] Wang, Y., Chao, W. L., Garg, D., Hariharan, B., Campbell, M., & Weinberger, K. Q. (2019). Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8445-8453). [CrossRef]

[62] You, Y., Wang, Y., Chao, W. L., Garg, D., Pleiss, G., Hariharan, B., ... & Weinberger, K. Q. (2019). Pseudo-lidar++: Accurate depth for 3d object detection in autonomous driving. *arXiv preprint arXiv:1906.06310*. [CrossRef]

[63] Ding, M., Huo, Y., Yi, H., Wang, Z., Shi, J., Lu, Z., & Luo, P. (2020). Learning depth-guided convolutions for monocular 3d object detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition workshops* (pp. 1000-1001). [CrossRef]

[64] Weng, X., & Kitani, K. (2019). Monocular 3d object detection with pseudo-lidar point cloud. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 0-0). [CrossRef]

[65] Wang, L., Du, L., Ye, X., Fu, Y., Guo, G., Xue,

X., ... & Zhang, L. (2021). Depth-conditioned dynamic message propagation for monocular 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 454-463). [CrossRef]

[66] Ma, X., Wang, Z., Li, H., Zhang, P., Ouyang, W., & Fan, X. (2019). Accurate monocular 3d object detection via color-embedded 3d reconstruction for autonomous driving. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6851-6860). [CrossRef]

[67] Park, D., Ambrus, R., Guizilini, V., Li, J., & Gaidon, A. (2021). Is pseudo-lidar needed for monocular 3d object detection?. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3142-3152). [CrossRef]

[68] Ma, X., Liu, S., Xia, Z., Zhang, H., Zeng, X., & Ouyang, W. (2020). Rethinking pseudo-lidar representation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16* (pp. 311-327). Springer International Publishing. [CrossRef]

[69] Chang, J., & Wetzstein, G. (2019). Deep optics for monocular depth estimation and 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10193-10202). [CrossRef]

[70] Li, P., Chen, X., & Shen, S. (2019). Stereo r-cnn based 3d object detection for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7644-7652). [CrossRef]

[71] Sun, J., Chen, L., Xie, Y., Zhang, S., Jiang, Q., Zhou, X., & Bao, H. (2020). Disp r-cnn: Stereo 3d object detection via shape prior guided instance disparity estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10548-10557). [CrossRef]

[72] Liu, Y., Wang, L., & Liu, M. (2021, May). Yolostereo3d: A step back to 2d for efficient stereo 3d detection. In *2021 IEEE international conference on Robotics and automation (ICRA)* (pp. 13018-13024). IEEE. [CrossRef]

[73] Qin, Z., Wang, J., & Lu, Y. (2019). Triangulation learning network: from monocular to stereo 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7615-7623). [CrossRef]

[74] Chen, Y., Liu, S., Shen, X., & Jia, J. (2020). Dsgn: Deep stereo geometry network for 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12536-12545). [CrossRef]

[75] Guo, X., Shi, S., Wang, X., & Li, H. (2021). Liga-stereo: Learning lidar geometry aware representations for stereo-based 3d detector. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3153-3163). [CrossRef]

[76] Guo, X., Wang, S. S. X., & Li, H. Supplementary Materials of LIGA-Stereo: Learning LiDAR Geometry Aware Representations for Stereo-based 3D Detector. [CrossRef]

[77] Su, K., Yan, W., Wei, X., & Gu, M. (2022). Stereo VoVNet-CNN for 3D object detection. *Multimedia Tools and Applications*, 81(25), 35803-35813. [CrossRef]

[78] Xu, Z., Zhang, W., Ye, X., Tan, X., Yang, W., Wen, S., ... & Huang, L. (2020, April). Zoomnet: Part-aware adaptive zooming neural network for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 12557-12564). [CrossRef]

[79] Shi, Y., Guo, Y., Mi, Z., & Li, X. (2022). Stereo CenterNet-based 3D object detection for autonomous driving. *Neurocomputing*, 471, 219-229. [CrossRef]

[80] Chen, L., Sun, J., Xie, Y., Zhang, S., Shuai, Q., Jiang, Q., ... & Zhou, X. (2021). Shape prior guided instance disparity estimation for 3d object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 5529-5540. [CrossRef]

[81] Peng, W., Pan, H., Liu, H., & Sun, Y. (2020). Ida-3d: Instance-depth-aware 3d object detection from stereo vision for autonomous driving. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13015-13024). [CrossRef]

[82] Peng, X., Zhu, X., Wang, T., & Ma, Y. (2022). Side: Center-based stereo 3d detector with structure-aware instance depth estimation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 119-128). [CrossRef]

[83] Qian, R., Garg, D., Wang, Y., You, Y., Belongie, S., Hariharan, B., ... & Chao, W. L. (2020). End-to-end pseudo-lidar for image-based 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5881-5890). [CrossRef]

[84] Liu, Y., Yixuan, Y., & Liu, M. (2021). Ground-aware monocular 3d object detection for autonomous driving. *IEEE Robotics and Automation Letters*, 6(2), 919-926. [CrossRef]

[85] Peng, L., Liu, F., Yu, Z., Yan, S., Deng, D., Yang, Z., ... & Cai, D. (2022, October). Lidar point cloud guided monocular 3d object detection. In *European conference on computer vision* (pp. 123-139). Cham: Springer Nature Switzerland. [CrossRef]

[86] Wang, X., Yin, W., Kong, T., Jiang, Y., Li, L., & Shen, C. (2020, April). Task-aware monocular depth estimation for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 12257-12264). [CrossRef]

[87] Ye, X., Du, L., Shi, Y., Li, Y., Tan, X., Feng, J., ... & Wen, S. (2020). Monocular 3d object detection via feature domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28,*

2020, Proceedings, Part IX 16 (pp. 17-34). Springer International Publishing. [CrossRef]

[88] Wang, L., Zhang, L., Zhu, Y., Zhang, Z., He, T., Li, M., & Xue, X. (2021). Progressive coordinate transforms for monocular 3d object detection. *Advances in Neural Information Processing Systems*, 34, 13364-13377. [CrossRef]

[89] Meng, H., Li, C., Chen, G., & Chen, L. (2023). Efficient 3D Object Detection Based on Pseudo-LiDAR Representation. *IEEE Transactions on Intelligent Vehicles*. [CrossRef]

[90] Tao, C., Cao, C., Cheng, H., Gao, Z., Luo, X., Zhang, Z., & Zheng, S. (2023). An efficient 3D object detection method based on fast guided anchor stereo RCNN. *Advanced Engineering Informatics*, 57, 102069. [CrossRef]

[91] Xia, Y., Shi, L., Ding, Z., Henriques, J. F., & Cremers, D. (2024). Text2loc: 3d point cloud localization from natural language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14958-14967). [CrossRef]

[92] Königshof, H., Salscheider, N. O., & Stiller, C. (2019, October). Realtime 3d object detection for automated driving using stereo vision and semantic information. In *2019 IEEE Intelligent Transportation Systems Conference* (ITSC) (pp. 1405-1410). IEEE. [CrossRef]

[93] Tao, C., He, H., Xu, F., & Cao, J. (2021). Stereo priori RCNN based car detection on point level for autonomous driving. *Knowledge-Based Systems*, 229, 107346. [CrossRef]

[94] Chen, X., Ma, H., Wan, J., Li, B., & Xia, T. (2017). Multi-view 3d object detection network for autonomous driving. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1907-1915). [CrossRef]

[95] Zhou, Y., Sun, P., Zhang, Y., Anguelov, D., Gao, J., Ouyang, T., ... & Vasudevan, V. (2020, May). End-to-end multi-view fusion for 3d object detection in lidar point clouds. In *Conference on Robot Learning* (pp. 923-932). PMLR. [CrossRef]

[96] Rubino, C., Crocco, M., & Del Bue, A. (2017). 3d object localisation from multi-view image detections. *IEEE transactions on pattern analysis and machine intelligence*, 40(6), 1281-1294. [CrossRef]

[97] Yang, Z., & Wang, L. (2019). Learning relationships for multi-view 3D object recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 7505-7514). [CrossRef]

[98] Wang, C., Pelillo, M., & Siddiqi, K. (2019). Dominant set clustering and pooling for multi-view 3d object recognition. *arXiv preprint arXiv:1906.01592*. [CrossRef]

[99] Deng, J., & Czarnecki, K. (2019, October). MLOD: A multi-view 3D object detection based on robust feature fusion method. In *2019 IEEE intelligent transportation systems conference* (ITSC) (pp. 279-284). IEEE. [CrossRef]

[100] Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S. (2016). 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14* (pp. 628-644). Springer International Publishing. [CrossRef]

[101] Ku, J., Pon, A. D., Walsh, S., & Waslander, S. L. (2019, November). Improving 3d object detection for pedestrians with virtual multi-view synthesis orientation estimation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS) (pp. 3459-3466). IEEE. [CrossRef]

[102] Hong, C., Yu, J., You, J., Chen, X., & Tao, D. (2015). Multi-view ensemble manifold regularization for 3D object recognition. *Information sciences*, 320, 395-405. [CrossRef]

[103] Philion, J., & Fidler, S. (2020). Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16* (pp. 194-210). Springer International Publishing. [CrossRef]

[104] Wang, Y., Guizilini, V. C., Zhang, T., Wang, Y., Zhao, H., & Solomon, J. (2022, January). Detr3d: 3d object detection from multi-view images via 3d-to-2d queries. In *Conference on Robot Learning* (pp. 180-191). PMLR. [CrossRef]

[105] Lin, J., Rickert, M., & Knoll, A. (2021, May). Deep hierarchical rotation invariance learning with exact geometry feature representation for point cloud classification. In 2021 IEEE international conference on robotics and automation (ICRA) (pp. 9529-9535). IEEE. [CrossRef]

[106] Zhang, K., Hao, M., Wang, J., Chen, X., Leng, Y., de Silva, C. W., & Fu, C. (2021, November). Linked dynamic graph cnn: Learning through point cloud by linking hierarchical features. In *2021 27th international conference on mechatronics and machine vision in practice* (*M2VIP*) (pp. 7-12). IEEE. [CrossRef]

[107] Zhang, J., Liu, J., Liu, X., Wei, J., Cao, J., & Tang, K. (2021). Feature interpolation convolution for point cloud analysis. *Computers & Graphics*, 99, 182-191. [CrossRef]

[108] Shi, S., Wang, X., & Li, H. (2019). Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 770-779). [CrossRef]

[109] Liu, Z., Tang, H., Lin, Y., & Han, S. (2019). Point-voxel cnn for efficient 3d deep learning. *Advances in neural information processing systems*, 32. [CrossRef]

[110] Chen, C., Chen, Z., Zhang, J., & Tao, D. (2022, June). Sasa: Semantics-augmented set abstraction for

point-based 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 1, pp. 221-229). [CrossRef]

[111] Ngiam, J., Caine, B., Han, W., Yang, B., Chai, Y., Sun, P., ... & Vasudevan, V. (2019). Starnet: Targeted computation for object detection in point clouds. *arXiv preprint arXiv:1908.11069*. [CrossRef]

[112] Yang, Z., Sun, Y., Liu, S., & Jia, J. (2020). 3dssd: Point-based 3d single stage object detector. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11040-11048). [CrossRef]

[113] Yang, H., Liu, Z., Wu, X., Wang, W., Qian, W., He, X., & Cai, D. (2022, October). Graph r-cnn: Towards accurate 3d object detection with semantic-decorated local graph. In *European Conference on Computer Vision* (pp. 662-679). Cham: Springer Nature Switzerland. [CrossRef]

[114] NShi, W., & Rajkumar, R. (2020). Point-gnn: Graph neural network for 3d object detection in a point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1711-1719). [CrossRef]

[115] Zhou, D., Fang, J., Song, X., Liu, L., Yin, J., Dai, Y., ... & Yang, R. (2020). Joint 3d instance segmentation and object detection for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1839-1849). [CrossRef]

[116] He, Q., Wang, Z., Zeng, H., Zeng, Y., & Liu, Y. (2022, June). Svga-net: Sparse voxel-graph attention network for 3d object detection from point clouds. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 1, pp. 870-878). [CrossRef]

[117] Zarzar, J., Giancola, S., & Ghanem, B. (2019). PointRGCN: Graph convolution networks for 3D vehicles detection refinement. *arXiv preprint arXiv:1911.12236*. [CrossRef]

[118] Feng, M., Gilani, S. Z., Wang, Y., Zhang, L., & Mian, A. (2020). Relation graph network for 3D object detection in point clouds. *IEEE Transactions on Image Processing*, 30, 92-107. [CrossRef]

[119] Pan, X., Xia, Z., Song, S., Li, L. E., & Huang, G. (2021). 3d object detection with pointformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7463-7472). [CrossRef]

[120] Liu, Z., Zhang, Z., Cao, Y., Hu, H., & Tong, X. (2021). Group-free 3d object detection via transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2949-2958). [CrossRef]

[121] Fayyad, J., Jaradat, M. A., Gruyer, D., & Najjaran, H. (2020). Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors*, 20(15), 4220. [CrossRef]

[122] Wang, Q., Chen, J., Deng, J., & Zhang, X. (2021). 3D-CenterNet: 3D object detection network for point clouds with center estimation priority. *Pattern Recognition*, 115, 107884. [CrossRef]

[123] Wang, D. Z., & Posner, I. (2015, July). Voting for voting in online point cloud object detection. In *Robotics: science and systems* (Vol. 1, No. 3, pp. 10-15). [CrossRef]

[124] Engelcke, M., Rao, D., Wang, D. Z., Tong, C. H., & Posner, I. (2017, May). Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1355-1361). IEEE. [CrossRef]

[125] Cui, Y., Zhang, Y., Dong, J., Sun, H., Chen, X., & Zhu, F. (2024). Link3d: Linear keypoints representation for 3d lidar point cloud. *IEEE Robotics and Automation Letters*. [CrossRef]

[126] Bai, L., Li, Y., Cen, M., & Hu, F. (2021). 3D instance segmentation and object detection framework based on the fusion of LIDAR remote sensing and optical image sensing. *Remote Sensing*, 13(16), 3288. [CrossRef]

[127] Wang, B., Zhu, M., Lu, Y., Wang, J., Gao, W., & Wei, H. (2021). Real-time 3D object detection from point cloud through foreground segmentation. *IEEE Access*, 9, 84886-84898. [CrossRef]

[128] Yang, B., Liang, M., & Urtasun, R. (2018, October). Hdnet: Exploiting hd maps for 3d object detection. In *Conference on Robot Learning* (pp. 146-155). PMLR. [CrossRef]

[129] Zhou, Y., & Tuzel, O. (2018). Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4490-4499). [CrossRef]

[130] Yan, Y., Mao, Y., & Li, B. (2018). Second: Sparsely embedded convolutional detection. *Sensors*, 18(10), 3337. [CrossRef]

[131] Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., & Beijbom, O. (2019). Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12697-12705). [CrossRef]

[132] Wang, Y., Fathi, A., Kundu, A., Ross, D. A., Pantofaru, C., Funkhouser, T., & Solomon, J. (2020). Pillar-based object detection for autonomous driving. In *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16 (pp. 18-34). Springer International Publishing. [CrossRef]

[133] Shi, S., Wang, Z., Shi, J., Wang, X., & Li, H. (2020). From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. *IEEE transactions on pattern analysis and machine intelligence*, 43(8), 2647-2664. [CrossRef]

[134] Li, B. (2017, September). 3d fully convolutional network for vehicle detection in point cloud. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1513-1518). IEEE. [CrossRef]

[135] Yin, T., Zhou, X., & Krahenbuhl, P. (2021). Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11784-11793). [CrossRef]

[136] Mao, J., Xue, Y., Niu, M., Bai, H., Feng, J., Liang, X., ... & Xu, C. (2021). Voxel transformer for 3d object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3164-3173). [CrossRef]

[137] Deng, J., Shi, S., Li, P., Zhou, W., Zhang, Y., & Li, H. (2021, May). Voxel r-cnn: Towards high performance voxel-based 3d object detection. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 2, pp. 1201-1209). [CrossRef]

[138] Song, Z., Wei, H., Jia, C., Xia, Y., Li, X., & Zhang, C. (2023). VP-Net: Voxels as points for 3-D object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-12. [CrossRef]

[139] Wang, H., Chen, Z., Cai, Y., Chen, L., Li, Y., Sotelo, M. A., & Li, Z. (2022). Voxel-RCNN-complex: An effective 3-D point cloud object detector for complex traffic conditions. *IEEE Transactions on Instrumentation and Measurement*, 71, 1-12. [CrossRef]

[140] Sheng, H., Cai, S., Liu, Y., Deng, B., Huang, J., Hua, X. S., & Zhao, M. J. (2021). Improving 3d object detection with channel-wise transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2743-2752). [CrossRef]

[141] Li, J., Dai, H., Shao, L., & Ding, Y. (2021, October). From voxel to point: IoU-guided 3D object detection for point cloud with voxel-to-point decoder. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 4622-4631). [CrossRef]

[142] Miao, Z., Chen, J., Pan, H., Zhang, R., Liu, K., Hao, P., ... & Zhan, X. (2021). Pvgnet: A bottom-up one-stage 3d object detector with integrated multi-level features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3279-3288). [CrossRef]

[143] Noh, J., Lee, S., & Ham, B. (2021). Hvpr: Hybrid voxel-point representation for single-stage 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14605-14614). [CrossRef]

[144] Guan, T., Wang, J., Lan, S., Chandra, R., Wu, Z., Davis, L., & Manocha, D. (2022). M3detr: Multi-representation, multi-scale, mutual-relation 3d object detection with transformers. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 772-782). [CrossRef]

[145] Mao, J., Niu, M., Bai, H., Liang, X., Xu, H., & Xu, C. (2021). Pyramid r-cnn: Towards better performance and adaptability for 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2723-2732). [CrossRef]

[146] Liu, Z., Tang, H., Zhao, S., Shao, K., & Han, S. (2021). Pvnas: 3d neural architecture search with point-voxel convolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 8552-8568. [CrossRef]

[147] Li, P., Su, S., & Zhao, H. (2021, May). Rts3d: Real-time stereo 3d detection from 4d feature-consistency embedding space for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 3, pp. 1930-1939). [CrossRef]

[148] Zhang, R., Qiu, H., Wang, T., Guo, Z., Cui, Z., Qiao, Y., ... & Gao, P. (2023). MonoDETR: Depth-guided transformer for monocular 3D object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9155-9166). [CrossRef]

[149] Lu, B., Sun, Y., & Yang, Z. (2023). Voxel Graph Attention for 3D Object Detection from Point Clouds. *IEEE Transactions on Instrumentation and Measurement*. [CrossRef]

[150] Deng, J., Zhou, W., Zhang, Y., & Li, H. (2021). From multi-view to hollow-3D: Hallucinated hollow-3D R-CNN for 3D object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(12), 4722-4734. [CrossRef]

[151] Zhang, Y., Chen, J., & Huang, D. (2022). Cat-det: Contrastively augmented transformer for multi-modal 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 908-917). [CrossRef]

[152] Shi, S., Jiang, L., Deng, J., Wang, Z., Guo, C., Shi, J., ... & Li, H. (2023). PV-RCNN++: Point-voxel feature set abstraction with local vector representation for 3D object detection. *International Journal of Computer Vision*, 131(2), 531-551. [CrossRef]

[153] Wu, P., Gu, L., Yan, X., Xie, H., Wang, F. L., Cheng, G., & Wei, M. (2023). PV-RCNN++: semantical point-voxel feature interaction for 3D object detection. *The Visual Computer*, 39(6), 2425-2440. [CrossRef]

[154] Tu, J., Wang, P., & Liu, F. (2021, July). Pp-rcnn: Point-pillars feature set abstraction for 3d real-time object detection. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE. [CrossRef]

[155] Li, J., Luo, C., & Yang, X. (2023). PillarNeXt: Rethinking network designs for 3D object detection in LiDAR point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 17567-17576). [CrossRef]

[156] Hu, J. S., Kuai, T., & Waslander, S. L. (2022). Point density-aware voxels for lidar 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8469-8478). [CrossRef]

[157] Geng, K., Dong, G., Yin, G., & Hu, J. (2020). Deep dual-modal traffic objects instance segmentation method using camera and lidar data for autonomous driving. *Remote Sensing*, 12(20), 3274. [CrossRef]

[158] Ignatious, H. A., & Khan, M. (2022). An overview of sensors in Autonomous Vehicles. *Procedia Computer Science*, 198, 736-741. [CrossRef]

[159] Vargas, J., Alsweiss, S., Toker, O., Razdan, R., & Santos, J. (2021). An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors*, 21(16), 5397. [CrossRef]

[160] Cartenì, A. (2020). The acceptability value of autonomous vehicles: A quantitative analysis of the willingness to pay for shared autonomous vehicles (SAVs) mobility services. *Transportation Research Interdisciplinary Perspectives*, 8, 100224. [CrossRef]

[161] Sakib, S. M. (2022). LiDAR Technology-An Overview. *IUP Journal of Electrical & Electronics Engineering*, 15(1).

[162] Bastos, D., Monteiro, P. P., Oliveira, A. S., & Drummond, M. V. (2021, February). An overview of LiDAR requirements and techniques for autonomous driving. In *2021 Telecoms Conference* (*ConfTELE*) (pp. 1-6). IEEE. [CrossRef]

[163] Royo, S., & Ballesta, M. (2019). An overview of imaging lidar sensors for autonomous vehicles. [CrossRef]

[164] Thomä, R., Dallmann, T., Jovanoska, S., Knott, P., & Schmeink, A. (2021, March). Joint communication and radar sensing: An overview. In *2021 15th European Conference on Antennas and Propagation* (*EuCAP*) (pp. 1-5). IEEE. [CrossRef]

[165] Paterniani, G., Sgreccia, D., Davoli, A., Guerzoni, G., Di Viesti, P., Valenti, A. C., ... & Boriani, G. (2023). Radar-based monitoring of vital signs: A tutorial overview. Proceedings of the IEEE, 111(3), 277-317. [CrossRef]

[166] Mielle, M., Magnusson, M., & Lilienthal, A. J. (2019, September). A comparative analysis of radar and lidar sensing for localization and mapping. In *2019 European Conference on Mobile Robots* (*ECMR*) (pp. 1-6). IEEE. [CrossRef]

[167] Kim, K. E., Lee, C. J., Pae, D. S., & Lim, M. T. (2017, October). Sensor fusion for vehicle tracking with camera and radar sensor. In *2017 17th International Conference on Control, Automation and Systems* (*ICCAS*) (pp. 1075-1077). IEEE. [CrossRef]

[168] Abro, G. E. M., Abdullahi, M. S., Ganasan, J., & Ricky, S. K. (2021). Prototyping an IoT-enabled Autonomous Unmanned Ground Vehicle Using SLAM. *International Journal of Control Systems and Robotics*, 6. [CrossRef]

[169] Pravallika, A., Hashmi, M. F., & Gupta, A. (2024). Deep Learning Frontiers in 3D Object Detection: A Comprehensive Review for Autonomous Driving. *IEEE Access*. [CrossRef]

[170] Berrio, J. S., Shan, M., Worrall, S., & Nebot, E. (2021). Camera-LIDAR integration: Probabilistic sensor fusion for semantic mapping. *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 7637-7652. [CrossRef]

[171] Khan, D., Baek, M., Kim, M. Y., & Han, D. S. (2022, October). Multimodal Object Detection and Ranging Based on Camera and Lidar Sensor Fusion for Autonomous Driving. In *2022 27th Asia Pacific Conference on Communications* (*APCC*) (pp. 342-343). IEEE. [CrossRef]

[172] Das, D., Adhikary, N., & Chaudhury, S. (2022, September). Sensor fusion in autonomous vehicle using LiDAR and camera Sensor. In *2022 IEEE 10th Region 10 Humanitarian Technology Conference* (*R10-HTC*) (pp. 336-341). IEEE. [CrossRef]

[173] Mendez, J., Molina, M., Rodriguez, N., Cuellar, M. P., & Morales, D. P. (2021). Camera-LiDAR multi-level sensor fusion for target detection at the network edge. *Sensors*, 21(12), 3992. [CrossRef]

[174] Thakur, A., & Rajalakshmi, P. (2023, July). LiDAR and Camera Raw Data Sensor Fusion in Real-Time for Obstacle Detection. In *2023 IEEE Sensors Applications Symposium* (*SAS*) (pp. 1-6). IEEE. [CrossRef]

[175] Ai, C., Qi, Z., Zheng, L., Geng, D., Feng, Z., & Sun, X. (2021, March). Research on mapping method based on data fusion of lidar and depth camera. In *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering* (*AEMCSE*) (pp. 360-365). IEEE. [CrossRef]

Ghulam E Mustafa Abro earned his B.S. in Electronic Engineering with honors from Hamdard University, Pakistan, in 2016, followed by M.S. in Control and Automation from Sir Syed University in 2019, and a Ph.D. in Electrical and Electronic Engineering from Universiti Teknologi PETRONAS, Malaysia, in 2023. He is currently a Postdoctoral Fellow at King Fahd University of Petroleum and Minerals (KFUPM) in Saudi Arabia, working in the Interdisciplinary Research Centre for Aviation and Space Exploration. Dr. Abro has nearly a decade of involvement with IEEE, serving in various roles, including conference chair and reviewer for SCI-indexed journals. His diverse research interests span control of underactuated systems, autonomous navigation, robotics, swarm technology, and multi-agent systems. Prior to KFUPM, he held academic and research roles at Hamdard University, Universiti Teknologi PETRONAS, and defense research institutes in Malaysia. (Email: Ghulam.abro@kfupm.edu.sa; mustafa.abro@ieee.org)

Zain Anwar Ali earned his B.S. in Electronic Engineering from Sir Syed University of Engineering and Technology (SSUET), Karachi, in 2009, followed by an M.S. in Industrial Control and Automation from Hamdard University in 2012, and a Ph.D. in Control Theory and Engineering from Nanjing University of Aeronautics and Astronautics (NUAA) in 2017. He has held academic positions at SSUET and Hamdard University, and conducted Ph.D. research with Nanjing Strong Flight Electronics. Currently, he is an Assistant Professor at Department

of Electronic Engineering Department, Maynooth International Engineering College (MIEC), Maynooth University, Maynooth, Co. Kildare, Ireland. Dr. Ali has published over 73 research articles and is a member of various international engineering bodies. He was twice selected as a Highly Talented Foreign Expert by the Chinese Ministry. He has served as Assistant Editor of SSUET Research Journal and Director of the Continuing Education Program at SSUET, and participates in research collaborations funded by Pakistan's Higher Education Commission (HEC). (Email: Zainanwar.ali@mu.ie)

Summaiya Rajput is a telecommunication graduate from Quaid E Awam University of Engineering and Technology (QUEST), Nawabshah, Pakistan. Her academic background has equipped with skills in project design, management, and data analysis. Recognized for resourcefulness and a positive approach. As enthusiastic about applying her knowledge to contribute to the evolving landscape of technology she is looking forward to excelling further in exploring innovative solutions at the forefront of robotics, computer vision and object detection and recognition. She is currently looking forward for an opportunity to pursue MS Studies under any funded project at abroad.(Email: summaiya.rajput@gmail.com)