

A Review on Computer Vision-Based Object and Safe Navigation Zone Identification for Autonomous Vehicles and Advanced Driver Assistance Systems (ADAS)

Vitor Augusto da Rosa Pereira[1(✉)], Rafael Alceste Berri[2], and Fernando Santos Osório[1]

[1] University of São Paulo, São Carlos, São Paulo 13566-590, Brazil
{vitor.arp,fosorio}@usp.br
[2] Federal University of Rio Grande, Rio Grande, Rio Grande do Sul 96203-900, Brazil
rafaelberri@furg.br

Abstract. Road safety is a global challenge, with most traffic accidents being attributed to human errors and recklessness, such as distractions and bad decision-making. ADAS play a crucial role in reducing traffic accidents caused by these mistakes, improving road safety. Computer Vision and Artificial Intelligence are essential for perception in mobile robots and autonomous vehicles, enabling them to analyze and interpret their surroundings. This paper presents a systematic review of Computer Vision and Artificial Intelligence techniques for object detection and safe navigation zone identification in Autonomous Vehicles and ADAS. The analysis compares the selected state-of-the-art works, highlighting their methods, accuracy, strengths, and limitations. In addition, the review compares multiple benchmark datasets employed across the studies, analyzing their characteristics and application scope in perception tasks. The results show advancements and considerable accuracy in object detection and safe navigation zone identification using semantic segmentation and bounding box methods, together with monocular depth estimation and multitask models that increase 3D perception and efficiency. Some approaches are able to run on embedded devices and achieve considerable FPS rate, allowing soft real-time performance. However, challenges like high computational costs and adverse weather conditions remain, showing that sensor fusion along with future research is necessary to improve robustness.

Keywords: Traffic objects · Safe navigation zone · Computer vision · Artificial intelligence · Autonomous vehicles · ADAS

1 Introduction

Driving, although a routine activity, carries several risks that can result in serious and even fatal accidents. A study conducted in the United States, using data

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
O. Gervasi et al. (Eds.): ICCSA 2025, LNCS 15650, pp. 114–131, 2025.
https://doi.org/10.1007/978-3-031-96962-1_8

from the period between 2005 and 2007 by the National Motor Vehicle Crash Causation Survey (NMVCCS), aimed to identify the main factors causing traffic accidents. Approximately 94% of accidents were attributed to human errors, including distractions, poor decision-making, and physiological factors such as drowsiness [29].

In this context, Autonomous Vehicles (AVs) and Advanced Driver Assistance Systems (ADAS) play a crucial role and have great potential in reducing traffic accidents, which are mainly caused by human errors, thus aiming to improve road safety [1]. Over the years, these systems have evolved significantly, driven by the continuous advancement of the technologies that compose them [32]. Moreover, their development has also driven the need for robust techniques to perceive and interpret the environment around the vehicle. Object detection and safe navigation zone identification are key for the safety of autonomous and assisted driving systems.

In mobile robots and AVs, Computer Vision and Artificial Intelligence are fundamental for robot perception, enabling them to analyze images and videos to interpret information about the objects on their surroundings [30]. This capability is directly linked to scene understanding, which involves identifying, locating, and describing elements in the environment, as well as their characteristics and dynamics. Research on new techniques has grown exponentially over the past decade due to its potential impact on various applications. Advances in Deep Learning, the availability of open-source datasets, and the increasing capacity of computational resources have driven the rapid evolution of scene understanding techniques [11].

The objective of this research is to review and analyze state-of-the-art Computer Vision and Artificial Intelligence techniques, mainly focused on RGB cameras for object detection and safe navigation zone identification in Autonomous Vehicles and ADAS.

In the planning phase of the systematic literature review [25], two research questions were defined:

1. Which Artificial Intelligence and Computer Vision techniques have been investigated for developing ADAS and autonomous vehicle systems focused on object and navigable area detection?
2. What types of objects are detected (e.g., vehicles, pedestrians)? Which sensors are utilized? What techniques are employed to complement detection? What actions do the proposed systems take? What limitations do the studies identify?

The literature searches were conducted using IEEE Xplore[1], Scopus[2], and Google Scholar[3], through the keywords "Traffic Object Detection", "Road Obstacle Detection", "Drivable Area Detection", "Computer Vision", "Artificial Intelligence", "ADAS", and "Autonomous Vehicles".

[1] Access the IEEE Xplore website at https://ieeexplore.ieee.org/Xplore/home.jsp.
[2] Access the Scopus website at https://www.scopus.com/home.uri.
[3] Access the Google Scholar website at https://scholar.google.com/.

The papers were first filtered by language, including only those written in English, and by publication year, considering only studies published since 2019. The articles were then selected based on predefined inclusion and exclusion criteria. To be included, a study had to be a primary research article focused on computer vision applications for object and navigation zone detection in vehicles. Exclusion criteria consisted of duplicate studies, studies not utilizing computer vision methods, research not involving land vehicles, extended abstracts, and articles that did not address the use of computer vision for traffic object detection in vehicles, as these were considered out of scope.

An initial set of 516 articles was identified and evaluated according to the defined criteria. Articles were first selected by title, limiting the set to 255. Next, abstracts were reviewed, reducing the number to 153 studies. Subsequently, the introductions and conclusions were read, reducing the selection to 55 articles. The full texts of these articles were then reviewed, leading to 19 studies being selected for inclusion in the review. Parsifal[4] was used to organize, select, and analyze studies in the systematic literature review.

This paper is organized as follows. Section 2 provides an overview of the state-of-the-art techniques in object detection and safe navigation zone identification using computer vision. Section 3 discusses the methods and techniques utilized in the reviewed studies. Finally, Sect. 4 summarizes the conclusions and future works from this research.

2 Object and Safe Navigation Zone Detection

This section presents studies from the past six years, identified through a systematic literature review.

2.1 Unknown Obstacle Detection

The detection of unknown obstacles and anomalies (Fig. 1) refers to the identification of objects on the track that are outside the training set of classification networks, also known as out-of-distribution (OoD) objects.

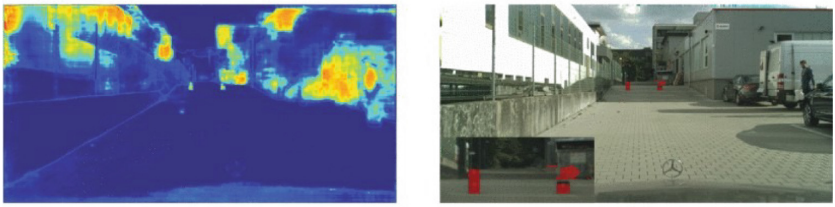


Fig. 1. Unknown/anomalous obstacle detection [7].

[4] Avaliable at https://parsif.al/.

Ohgushi, Horiguchi, and Yamanaka [21] propose a method to detect unknown or complex objects (such as debris, boxes, tires, among others) that are not present in the training set. To achieve this, the authors use two datasets: the Lost and Found Dataset and their own dataset, the Highway Anomaly Dataset. The Lost and Found Dataset is designed for anomaly detection in urban scenarios, containing street images with unexpected objects such as cones and small obstacles. Its labels include pixel-wise segmentation masks for anomalous objects and semantic annotations inherited from the Cityscapes Dataset, covering categories such as "road" and "vehicle". On the other hand, the Highway Anomaly Dataset focuses on videos and images of highways with atypical events, such as stopped vehicles, pedestrians, or animals on the road. The labels are mostly binary ("normal" or "anomalous"), with some being categorized by type of anomaly, including annotations in bounding boxes or timestamps for precise localization of events. The authors use an unsupervised semantic segmentation technique through the ICNet network. Based on the semantic map generated by the network, entropy and perceptual loss are calculated, which, when combined, result in an anomaly map. The approach achieved an accuracy of 0.96 AUC-ROC on the Lost and Found Dataset and an F-measure of 0.452 on the Highway Anomaly Dataset.

In the work of Ci et al. [7], the detection of anomalous objects is performed using RGB images from two datasets: the Lost and Found Dataset and RoadObstacle21. The latter is a dataset focused on images of segmented obstacles, with the road defined as the region of interest. The method combines various techniques, such as semantic segmentation, image resynthesis, monocular depth estimation, and dissimilarity networks, to detect anomalies and unknown obstacles. Anomaly detection achieved an average precision (AP) of 80.04% on the Lost and Found Dataset, with a false positive rate (FPR) of 5.31%. On the RoadObstacle21 dataset, the precision was 70.93%, with an FPR of 4.16%.

The approach proposed by Shoeb et al. [27] presents a modular method for the detection, tracking, and text-based retrieval of out-of-distribution (OoD) objects in videos of road scenarios. The system uses Mask2Former, trained on the Cityscapes dataset, to perform semantic segmentation. Objects are tracked across consecutive frames based on IoU and geometric centers. Additionally, the CLIP model enables textual queries, such as "dog", to retrieve relevant sequences. Tested on three distinct datasets—SOS (Street Obstacle Sequences), WOS (Wuppertal Obstacle Sequences), and CWL (Carla-WildLife)—the method achieved an AUPRC (Area Under the Precision-Recall Curve) above 75% in test cases, reaching 81% on the SOS dataset. Furthermore, the false positive rate (FPR95) was reduced to less than 1%, demonstrating the feasibility of detecting out-of-distribution objects.

In the study conducted by Vojir et al. [36], the JSR-Net was proposed, a network that combines partial image reconstruction and semantic segmentation for anomaly and unknown obstacle detection on roads. The model achieved an AP of 94.4% on the Road Anomaly dataset, as well as 79.4% on Lost and Found, 84.0% on Road Obstacles, and 85.5% on FishyScapes. Additionally, it recorded the lowest false positive rate (FPR95 = 4.4% on average). The approach enables

the detection of out-of-distribution objects without requiring a complete retraining process, enhancing generalization. Nonetheless, the solution faces challenges in detecting fine structures and image artifacts.

2.2 Traffic Object Detection

The detection of elements in traffic scenarios, as illustrated in Figs. 2 and 3, consists of identifying and locating key objects for autonomous driving, such as vehicles, pedestrians, road signs, navigable zones, and obstacles.



Fig. 2. Lane segmentation and vehicle detection [14].

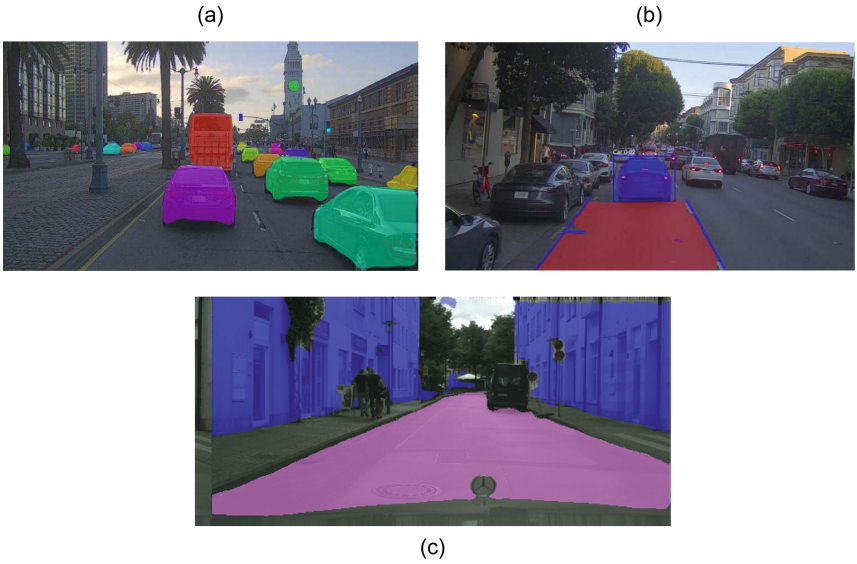


Fig. 3. (a) Instance segmentation for vehicle detection, (b) Lane segmentation and closest object detection [20], (c) Segmentation and width estimation of unmarked roads [22].

In an article by Ljepić et al. [18], a low-cost algorithm was proposed for obstacle detection on the roadway and automatic speed adjustment using only

RGB cameras. The method combines traditional computer vision algorithms, such as Canny and RANSAC, with simulations conducted in Carla. In tests with 152 images, the misclassification rate of vehicles as pedestrians was 6.58%, while the error rate in pedestrian detection was 5.26%. Speed adjustment showed limitations, with successful stops only at 10 km/h due to the low frame processing rate.

Ezzat et al. [10] present a forward collision warning system that combines computer vision and ultrasonic sensors. The method uses YOLOv3 Tiny for real-time vehicle detection, estimating depth through multivariate regression based on the area of bounding boxes and assigning vehicles to specific road lanes using a geometric analysis of the lower corners of the boxes. This technique divides the road into sections (right, left, and emergency). The system achieved 93% accuracy in vehicle detection and 80% in depth estimation, with a performance of up to 23 FPS on the NVIDIA Jetson TX2. Ultrasonic sensors complemented the detection in blind spots, dividing them into safe, warning, and critical zones.

The study by Shyam, Yoon, and Kim [28] presents a weakly supervised approach for object detection and lane marking recognition. The method uses pre-trained networks on the NuScenes and BDD100K datasets to generate pseudo labels for obstacles, integrating this data with lane detection datasets (CULanes, TUSimple, and CurveLanes) for training multi-task models. Two architectures were proposed: one based on CNNs and the other on Transformers. The CNN-based approach uses pixel classification to detect and group lane lines while employing an object detector, both sharing the same encoding backbone. Meanwhile, the Transformer-based approach uses dual decoders, one estimating object bounding boxes and the other calculating polynomial coefficients for lane lines. The Transformer-based approach achieved overall better performance, with an mAP of 81.2% for object detection and an IoU of 87.1% for lane segmentation, whereas the CNN-based approach achieved an mAP of 78.5% and an IoU of 85.4%. Computational efficiency was also analyzed, highlighting the advantage of the CNN architecture, which consumed 112 GFLOPs, compared to the Transformer-based architecture, which consumed 135 GFLOPs.

Usman et al. [35] propose combining instance segmentation and semantic segmentation techniques to improve object detection in autonomous vehicles and create a robust representation of the vehicle's surroundings. The approach integrates the Xception semantic segmentation model, trained on the ADE20K dataset, with the Mask R-CNN model, trained on the COCO dataset, for instance segmentation. The object classification network achieved 92% accuracy and a processing speed of 30 frames per second.

Chen [6] developed a multi-task model based on YOLOv5, integrating object detection and road segmentation. The network utilizes a multi-task branch, with one instance for classification using bounding boxes and another instance for semantic segmentation through a UNet decoder. Training utilized weighted loss strategies (cross-entropy loss) to balance the detection and segmentation tasks. The model achieved 98.5% mean average precision (mAP) for classifica-

tion, 65.96% mean intersection over union (mIoU) for segmentation, operating at a rate of 57 FPS.

In the article by Thombre et al. [33], two methods were proposed to estimate the distance of vehicles and obstacles using a monocular camera, aiming to develop solutions for ADAS. The first method, based on the camera's optical principles, uses parameters such as focal length and the width of object bounding boxes to estimate depth. The second method, Depth Anything, employs deep learning to generate depth maps, and it was integrated into a regression model to enhance the results, using data from a LiDAR sensor as ground truth. The models Focal Length, Depth Anything (raw), and Depth Anything (with regression) were compared. The Depth Anything with regression method achieved the best performance, obtaining an RMSE of 1.542, while the Depth Anything (raw) and Focal Length methods obtained RMSE values of 2.857 and 2.947, respectively.

The work by Kenk et al. [12] aims to investigate the limitations of deep learning models based on the YOLO architecture in detecting road lanes and objects under adverse conditions (such as sharp curves, dense fog, sandstorms, heavy rain, and snowfall), seeking to provide guidance for developing more robust solutions. The YOLO algorithm was adapted for real-time lane segmentation and vehicle detection using an approach that divides the image into grids with fixed anchors to predict lane positions. The solution was evaluated using the DAWN dataset, which contains images and traffic data under various weather conditions. The study found that lane detection performance was impaired in sharp curves due to the difficulty of representing curved lanes with bounding boxes. In heavy rain conditions, reflections and distortions on the asphalt caused false positives. The evaluation metrics for detection included mean Average Precision (mAP), Recall, Precision, and False Positive Rate (FPR). Under challenging conditions, such as a rainy scenario, the method achieved 26.06% Precision, 31.38% Recall, 26.06% detection rate, and an FPR of 68.61%.

The method by Megalingam et al. [20] performs dynamic lane segmentation for autonomous vehicles by combining the Hough transform, FCN-8, and R-CNN. The Hough transform detects lanes in straight road segments, while FCN-8 segments the drivable area. R-CNN performs instance segmentation, identifying and locating obstacles in the scene. The model achieved 97% accuracy in lane segmentation, even on curves and under low lighting conditions, and 94% accuracy in detecting small or partially occluded obstacles. Distance estimates were derived from the area of bounding boxes, while vehicle speed adjustments were determined based on the computed distances. Despite the results, limitations were noted in scenarios with low visibility.

In the proposal by Bruno and Osório [3], a pedestrian detection and tracking system for autonomous vehicles and ADAS was developed, combining YOLOv7-DeepSORT and Kalman Filter on embedded hardware (Jetson Nano). After 20,000 training epochs, the model stabilized, achieving 91.85% mAP and 78.2% IoU in detection using bounding boxes. The frame rate during inference ranged between 5 and 10 FPS, allowing real-time operation but with limitations at higher speeds. The approach demonstrated effectiveness in urban scenarios but

faces challenges under adverse conditions and in the absence of additional depth sensors.

In the system proposed by Li et al. [14], a multitask model integrating CNNs and Transformers for visual perception in autonomous vehicles simultaneously performs object detection, lane segmentation, and drivable area segmentation. The model was evaluated on the BDD100K dataset and achieved an mAP50 of 84.6% for object detection, an mIoU of 91.6% for drivable area segmentation, and an IoU of 32.3% for lane segmentation, outperforming networks like YOLOP and HybridNets. Additionally, it operated at 50 FPS, enabling real-time inference. However, the model has a higher computational cost due to the CNN-Transformer fusion and requires validation in extreme scenarios such as fog and heavy rain.

In the research by Sousa et al. [31], a Deep Learning-based ADAS system was developed for lane segmentation, object detection, and traffic sign classification, optimized for embedded devices NVIDIA Jetson Nano and TX2 NX. The datasets used included TuSimple and CULane, focused on lane segmentation, as well as a custom dataset created by the authors for traffic sign classification. The lane segmentation model, using U-Net and ResNet18, achieved 66.02% accuracy and an IoU of 0.6363, operating at 17.34 FPS on the Jetson Nano and 30.59 FPS on the TX2 NX. Object detection using TrafficCamNet obtained an IoU of 0.7997 at 640×480 resolution. Overall, the TX2 NX outperformed the Nano, achieving 27 FPS compared to the Nano's 15 FPS, making it a more suitable option for real-time applications. However, factors such as adverse lighting conditions and occlusions impacted the accuracy of the system.

As presented by Schieber et al. [26], a Pyramid Fusion Network was proposed, a multimodal and multi-scale fusion approach for 3D semantic segmentation in autonomous vehicles, combining data from RGB cameras and LiDAR. The architecture incorporates a Pyramid Fusion Backbone for hierarchical feature fusion and a Pyramid Fusion Head for final segmentation refinement. Tested on the SemanticKITTI and PandaSet datasets, the solution achieved an mIoU of 61.9%, outperforming previous methods such as Fusion3DSeg (61.8%) and LaserNet++ (56.2%). Additionally, adopting EfficientPS as the camera backbone increased the mIoU to 67.8%. The proposed method faces challenges such as high computational cost and reliance on sensor overlap. Future optimizations are suggested, including radar data integration to enhance perception in adverse conditions.

According to Tran et al. [34], a modular ADAS system based on a monocular camera and Jetson devices was developed to detect vehicles, pedestrians, traffic signs, and perform lane segmentation in real time. The model utilized Scaled-YOLOv4 for object detection and Ultra-Fast Deep Lane Detection for lane segmentation. Tested on the KATECH Dataset (Korea Automotive Technology Institute), the system achieved an mAP of 84.6%, surpassing previous methods such as YOLOP (76.5%) and DLT-Net (68.4%). Execution on the Jetson AGX Xavier reached 50 FPS, while on the Jetson Nano, it operated at 15 FPS.

In the investigation by Zhao et al. [41], a method for nighttime pedestrian detection was proposed, combining infrared images and millimeter-wave radar. The multimodal fusion approach utilizes an enhanced YOLOv5 (integrating GhostNet, SE, and SPPF) for visual detection, while an Extended Kalman Filter tracks pedestrians using radar data. A coordinate conversion process aligns the sensors, enabling decision-level fusion. Tested on the CVC-09 Infrared Dataset, the proposed method achieved 93.9% accuracy and 90% mAP.

In the work conducted by Palafox et al. [22], the SemanticDepth model was proposed, combining semantic segmentation and monocular depth estimation to enable autonomous vehicle navigation on roads without lane markings. The method generates a semantic 3D point cloud to measure road width and fence-to-fence distance, a term that, in this context, refers to walls or barriers. This is achieved by identifying the extreme points of the segmented road and projecting them into 3D based on the Monodepth disparity map. Measurements are obtained by converting disparities into absolute depths, considering known camera parameters such as focal length and virtual baseline. Tested on the Munich Test Set, Cityscapes, and Roborace datasets, the model achieved a mean absolute error of 0.48 m for road width and 0.91 m for fence-to-fence width. The solution has a high computational cost (approx. 0.64 s inference time per image) and depends on segmentation quality. Additionally, the Monodepth model requires manual scale adjustments, affecting the accuracy of 3D reconstruction.

3 Discussion

Based on the analysis of the selected articles, the distribution of corresponding authors by country is as follows: Germany (3 articles), China (3), India (3), Brazil (2), Egypt (2), South Korea (2), Croatia (1), the Czech Republic (1), Japan (1), and Taiwan (1). Figure 4 illustrates this distribution on a world map. Asia appears with the highest number of articles (10), followed by Europe (4), with Africa and South America each contributing with 2 publications.

Figure 5 displays the number of articles published per year over the past six years, with notable growth observed in 2021 and 2023.

In Sect. 2, state-of-the-art research related to autonomous vehicle perception and navigation was introduced, focusing on the detection of objects in traffic scenes and unknown obstacles. Table 1 presents a summary of the related works, detailing the type of detection, the sensors used, the approaches applied, the complementary methods, and the system's actuation, when available. Additionally, the limitations identified in each study are also highlighted.

From the review, it was observed that all works used RGB cameras as the primary sensor, except for one, which combined an infrared camera with radar [41]. This is due to the richness of information that cameras can capture (e.g., colors, textures, shapes), making them essential for visual perception tasks such as object detection, pattern recognition, and scene interpretation. Additionally, RGB cameras have a relatively low cost and enable the execution of various computer vision algorithms, making them a favorable option for autonomous

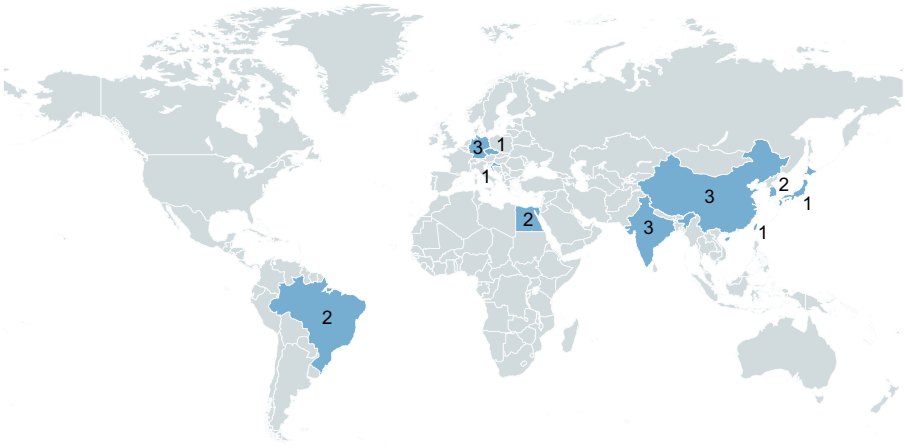


Fig. 4. Countries of the Corresponding Authors.

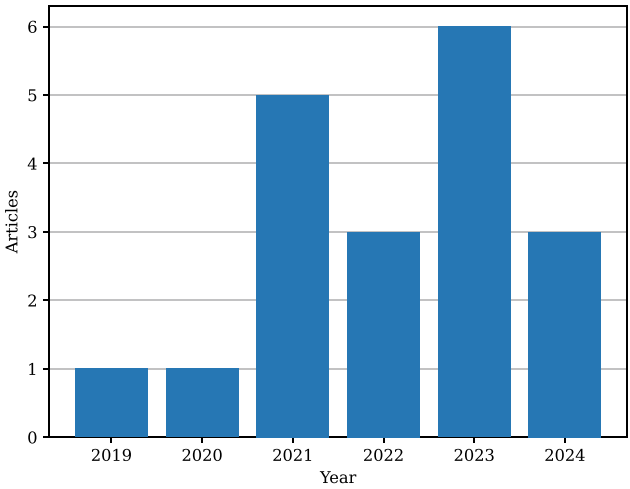


Fig. 5. Number of articles published by year.

vehicle and ADAS systems. Some approaches employ additional sensors through sensor fusion, such as ultrasonic, radar, and LiDAR [10,26,33], to enhance depth perception. 3D perception, mainly provided by sensors like LiDAR and radar, allows for spatial understanding of the environment, enabling the generation of 3D maps and point clouds. The integration of these 3D data with the visual information from RGB cameras enriches the environmental representation.

Most studies focus on identifying the navigable zone, pedestrians, and vehicles, as these are considered the most critical tasks for the safe navigation of autonomous vehicles. However, the detection of unknown obstacles and traffic

Table 1. Comparative table of selected articles.

		Ohgushi, Horiguchi and Yamanaka [21]	Ci et al. [7]	Ljepic et al. [18]	Ezzat et al. [10]	Shyam, Yoon and Kim [28]	Shoeb et al. [27]	Usman et al. [35]	Chen [6]	Thombre et al. [33]	Kenk et al. [12]	Megalingam et al. [20]	Bruno and Osório [3]	Li et al. [14]	Vojir et al. [36]	Sousa et al. [31]	Schieber et al. [26]	Tran et al. [34]	Zhao et al. [41]	Palafox et al. [22]
Detection	Unknown or anomalous obstacles	X	X				X								X					
	Traffic signs and road safety components					X										X	X	X		
	Safe navigation zone	X	X	X	X	X	X		X		X	X		X	X	X	X	X		X
	Pedestrians			X		X		X	X	X	X	X	X	X		X	X	X	X	
	Vehicles			X	X	X		X	X	X	X	X		X		X	X	X		
Approach	Semantic segmentation	X	X				X	X	X			X		X	X	X	X			X
	Monocular depth estimation		X							X										X
	Digital image processing			X								X								
	Bounding boxes				X	X			X	X	X	X	X	X		X		X	X	
	Instance segmentation							X				X								
	Row-wise classification for lane marker detection					X												X		
	Polynomial coefficients for lane detection					X														
Complementary methods	Anomaly maps	X	X												X					
	Image reconstruction	X													X					
	Point cloud 3D reconstruction																			X
	Geometric analysis				X		X			X										
	Multitask models					X			X		X			X						
	Pyramid fusion networks																X			
	Hough transform											X								
	Object tracking												X						X	
	Kalman filter												X						X	
	Decision-level sensor fusion																		X	
	Regression with depth data									X										
	Depth estimation using bounding boxes				X					X		X				X				
	Statistical road width estimation																			X
	Not applicable			X				X	X							X		X		
Actuation	Speed/braking control			X	X							X								
	Not applicable	X	X			X	X	X	X	X	X		X	X	X	X	X	X	X	X
Limitations	Occlusions	X														X				
	Computationally expensive	X	X	X				X	X	X			X	X			X	X		X
	Image quality				X										X					
	Shadows					X														
	Lack of lane markers					X														
	Objects out of the road (false positives)						X													
	Low-visibility conditions											X	X			X				
	Adverse weather conditions							X			X		X	X					X	
	Uncommon or out-of-distribution objects								X											
	Curves					X					X									
Sensors	RGB camera	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
	Infrared camera																		X	
	Ultrassonic sensor				X															
	Radar																		X	
	LiDAR									X							X			

signs is also addressed, as they play a crucial role in ensuring safety and supporting the vehicle's decision-making in dynamic scenarios.

Regarding the approaches, semantic segmentation [7,21,22,26,27,35,36] and detection through bounding boxes [3,10,12,28,33,34,41] techniques stood out.

Semantic segmentation plays a key role in scene interpretation and the identification of complex elements, such as lanes, navigable zones, buildings, and walls, as well as the detection of anomalous obstacles. On the other hand, bounding boxes are essential for recognizing and tracking well-defined traffic objects, such as vehicles, pedestrians, cyclists, motorcyclists, and road signs. Some authors explore the integration of semantic segmentation and bounding box detection within the same system [6,14,20,31], either by combining different models or through multitask approaches. This fusion of techniques enables a more robust perception of the environment, leveraging semantic segmentation to identify surfaces and navigable regions, while bounding boxes detect specific objects.

The use of techniques such as instance segmentation [20,35] is also notable, as it not only identifies and segments different objects in a scene but also distinguishes them individually, providing more detailed segmentation. Additionally, the application of monocular depth estimation techniques using RGB cameras [6,7,22] has been observed. This approach allows for the inference of three-dimensional information from a single image. Given the high cost of depth sensors like LiDAR, AI algorithms for monocular depth estimation emerge as a viable alternative for inferring relative distances in a scene and enabling 3D reconstructions. Other techniques specifically used for lane and road marking detection include digital image processing [18,20] and line-based classification models [28,33].

Some of the developed systems employ complementary methods to refine analysis and interpret the data generated by the primary approaches. Among them is the use of distance estimation via bounding boxes [10,20,31,33], a technique that correlates the size of bounding boxes with the distance of vehicles relative to the camera. Complementary geometric analysis techniques were applied to enhance object tracking, lane detection, and distance estimation using the camera's intrinsic parameters [10,27,33]. Many approaches utilized multitask models [6,12,14,28], capable of simultaneously performing lane segmentation or detection and identifying traffic objects. This optimizes computational resources by using a single model instead of multiple distinct models, each dedicated to a specific function. Finally, anomaly maps were developed to detect unknown or out-of-distribution objects. An anomaly map highlights unexpected regions in a scene by comparing the original image with a version reconstructed by neural networks. Differences measured using metrics such as perceptual loss and entropy indicate potential anomalies. After refinements, this map is used to detect obstacles [7,21,36].

Only a small number of the studied systems proposed any type of action, such as longitudinal speed control or braking [7,10,20]. On the other hand, most approaches focused exclusively on developing perception techniques without implementing control mechanisms.

Among the identified limitations, high computational cost stood out, reflecting the complexity of perception tasks in autonomous vehicles. This limitation was mainly observed in approaches that employed semantic and instance segmentation, monocular depth estimation, and multitask models due to their process-

ing demands. Similarly, extreme weather conditions and low-visibility scenarios were also recurring challenges, affecting the performance of solutions primarily based on RGB cameras, which rely heavily on lighting quality and environmental visibility. However, in such scenarios, the fusion of infrared cameras with radar helps mitigate these limitations by combining thermal data with depth information and object detection, providing a complementary perception to traditional RGB cameras [41].

3.1 Datasets

Diverse datasets were studied to implement and evaluate the system. This section provides a brief analysis of these databases. Table 2 summarizes the datasets, including their names, publication years, sensors used for data collection, dataset tasks based on annotations, and access links. The datasets were either publicly available or made public by their authors.

Most datasets employ monocular or stereo cameras, with only a few incorporating LiDAR. SemanticKITTI [2] provides LiDAR data exclusively, while PandaSet [37] and nuScenes [4] combine LiDAR with Inertial Measurement Unit (IMU) and GPS data. CARLA-WildLife [19] is a synthetic video dataset created using CARLA driving simulator [9], so it does not rely on any physical sensors.

Regarding the tasks provided, some focus on segmenting unknown obstacles (also known as anomalous or out-of-distribution obstacles) and the safe Navigation Zone (NZ) [5,19,24]. In these datasets, pixel-level classes typically include "obstacle" and "region of interest" The "region of interest" class represents the road or drivable area.

Object detection tasks in both 2D and 3D use bounding box (bbox) annotations: 2D bboxes for images and 3D bboxes for spatial data. For segmentation tasks, 2D data employs pixel-wise masks, while 3D data uses point cloud segmentation. The nuScenes dataset [4] stands out as one of the most comprehensive benchmarks, supporting tasks such as 2D/3D object detection, segmentation, tracking, and lane detection, though it does not include unknown obstacle detection. The dataset incorporates LiDAR data (enabling 3D detection and segmentation) along with IMU and GPS measurements for odometry tracking. As well as BDD100K [40], that similarly supports multiple perception tasks including 2D object detection, segmentation, tracking, and lane detection, while using only camera data.

The authors evaluated three lane detection specific datasets: CULanes [23], TUSimple [39], and CurveLanes [38]. Additionally, they incorporated general purpose 2D segmentation datasets including COCO [15] and ADE20K [42], aiming to create a unified framework for multi-task integration. Also, the DAWN dataset [13], focused in 2D bbox vehicle detection under adverse weather conditions.

For a comprehensive survey of autonomous driving datasets covering perception, planning, and control, see [17].

Table 2. Vision-based datasets.

Dataset	Year	Sensors	Tasks	Website
Lost and Found [24]	2016	Stereo camera	UO Seg, NZ Seg	Link
RoadObstacle21 [5]	2021	Camera	UO Seg, NZ Seg	Link
Road Anomaly [16]	2019	Camera	UO Seg	Link
Cityscapes [8]	2016	Stereo camera	2D Det, 3D Det, 2D Seg	Link
Obstacle Sequences [19]	2022	Camera	UO Seg, NZ Seg	Link
CARLA-WildLife [19]	2022	-	UO Seg, NZ Seg, T	Link
nuScenes [4]	2019	LiDAR, Radar, Camera, IMU, GPS	2D Det, 3D Det, 2D Seg, 3D Seg, T, LD	Link
BDD100K [40]	2020	Camera	2D Det, 2D Seg, T, LD	Link
CULanes [23]	2018	Camera	LD	Link
TUSimple [39]	2020	Camera	LD	Link
CurveLanes [38]	2020	Camera	LD	Link
COCO [15]	2015	Camera	2D Seg	Link
ADE20K [42]	2017	Camera	2D Seg	Link
DAWN [13]	2020	Camera	2D Det	Link
SemanticKITTI [2]	2019	LiDAR	3D Seg	Link
PandaSet [37]	2021	LiDAR, Camera, IMU, GPS	3D Det, 3D Seg, T	Link

[1] Det: Detection
[2] Seg: Segmentation
[3] T: Tracking
[4] LD: Lane Detection
[5] UO: Unknown Obstacle
[6] NZ: Navigation Zone

4 Conclusions and Future Work

This paper presents a literature review on object and navigation area detection using Computer Vision and Artificial Intelligence, focusing on applications for Autonomous Vehicles and ADAS. The study of state-of-the-art approaches showed important advancements in perception techniques, including object detection, segmentation, and distance estimation. These methodologies have demonstrated great accuracy on detecting and tracking navigable zones, pedestrians, vehicles, and road obstacles, which contributes to the improvement of safety in autonomous navigation.

Deep Learning and Vision Transformer-based methods, like semantic segmentation, play a crucial role in scene understanding. These models can detect and interpret complex elements and shapes, such as lanes, walls, and navigable or non-navigable areas, as well as objects that cannot be effectively represented

using bounding boxes. Another important AI technique is monocular depth estimation, which provides a relative sense of depth in a scene using just a single image. Moreover, multitask models have shown to achieve efficiency by combining multiple perception tasks into a single framework. It is also noticed the use of more traditional techniques like image processing to detect lanes and geometric and statistical analysis to estimate objects distance.

Despite these advancements, some challenges remain. Many of the reviewed methods are computationally expensive, making them difficult to implement in real-time on embedded systems like intelligent vehicles or mobile robots. However, some approaches have demonstrated potential by running on embedded devices and achieving a considerable FPS rate, enabling soft real-time performance. Additionally, camera based systems are affordable and can capture detailed scene information, but they face difficulties in extreme weather and low light conditions. To overcome these limitations, combining data from LiDAR, radar, or infrared sensors has shown positive results.

In future research we propose the development of a computer vision system based on Deep Learning and Vision Transformers (ViTs), seeking to fuse semantic segmentation (2D data) and monocular depth estimation (3D data) to improve the detection of navigation zones, objects, and obstacles on roadways. The goal is to advance environmental perception and scene understanding, thus increasing the safety and reliability of autonomous vehicles and ADAS.

Acknowledgement. We thank PPG-CCMC[5](PPG-CCMC is the acronym for *Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional* in Portuguese.), University of São Paulo and Federal University of Rio Grande, which made this work possible. This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) - Finance Code 001.

References

1. Aleksa, M., Schaub, A., Erdelean, I., Wittmann, S., Soteropoulos, A., Fürdös, A.: Impact analysis of advanced driver assistance systems (ADAS) regarding road safety-computing reduction potentials. Eur. Transp. Res. Rev. **16**(1), 39 (2024)
2. Behley, J., et al.: Semantickitti: a dataset for semantic scene understanding of lidar sequences. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9297–9307 (2019)
3. Bruno, D.R., Osório, F.S.: Real-time pedestrian detection and tracking system using deep learning and kalman filter: Applications on embedded systems in advanced driver assistance systems. In: 2023 Latin American Robotics Symposium (LARS), 2023 Brazilian Symposium on Robotics (SBR), and 2023 Workshop on Robotics in Education (WRE), pp. 549–554 (2023). https://doi.org/10.1109/LARS/SBR/WRE59448.2023.10333032
4. Caesar, H., et al.: nuscenes: A multimodal dataset for autonomous driving. arXiv preprint arXiv:1903.11027 (2019)
5. Chan, R., et al.: Segmentmeifyoucan: a benchmark for anomaly segmentation. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2021). https://doi.org/10.48550/arXiv.2104.14812

6. Chen, H.M.: Multitask yolo: Versatile perception network for autonomous driving. In: 2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 46–51 (2023). https://doi.org/10.1109/ICMEW59549.2023.00014
7. Ci, W., Xuan, J., Lin, R., Lu, S.: A novel method for road anomaly objects detection in the traffic environment with multi-mechanism fusion. IEEE Access **12**, 28369–28381 (2024). https://doi.org/10.1109/ACCESS.2024.3359695
8. Cordts, M., et al.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
9. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning, pp. 1–16 (2017)
10. Ezzat, A., Ibrahim, A.M., Younis, M., Hassan, R.M., Darweesh, M.S.: Demonstration of forward collision warning system based on real-time computer vision. In: 2020 16th International Computer Engineering Conference (ICENCO), pp. 47–50 (2020). https://doi.org/10.1109/ICENCO49778.2020.9357374
11. Iosifidis, A., Tefas, A.: Deep learning for robot perception and cognition. Academic Press (2022). https://doi.org/10.1016/C2020-0-02902-6
12. Kenk, M.A., Elsaidy, M., Hassaballah, M., Mansour, M.B.A.: Driving perception in challenging road scenarios: an empirical study. In: 2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), pp. 1–6 (2023). https://doi.org/10.1109/AICCSA59173.2023.10479343
13. Kenk, M.A., Hassaballah, M.: Dawn: vehicle detection in adverse weather nature dataset. arXiv preprint arXiv:2008.05402 (2020)
14. Li, J., Ke, X., Wang, Z., Wan, J., Tan, G.: Cutransnet: transformers to make strong encoders for multi-task vision perception of autonomous driving. In: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7385–7389 (2024). https://doi.org/10.1109/ICASSP48485.2024.10445836
15. Lin, T.Y., et al.: Microsoft coco: Common objects in context (2015). https://arxiv.org/abs/1405.0312
16. Lis, K., Nakka, K., Fua, P., Salzmann, M.: Detecting the unexpected via image resynthesis (2019). https://arxiv.org/abs/1904.07595
17. Liu, M., et al.: A survey on autonomous driving datasets: Statistics, annotation quality, and a future outlook. IEEE Transactions on Intelligent Vehicles (2024)
18. Ljepić, S., Grbić, R., Kovačević, J., Krunić, M.: Detecting obstacles within the driving lane and vehicle speed adjustment. In: 2021 Zooming Innovation in Consumer Technologies Conference (ZINC), pp. 225–230 (2021). https://doi.org/10.1109/ZINC52049.2021.9499311
19. Maag, K., Chan, R., Uhlemeyer, S., Kowol, K., Gottschalk, H.: Two video data sets for tracking and retrieval of out of distribution objects. In: Proceedings of the Asian Conference on Computer Vision (ACCV), pp. 3776–3794 (December 2022)
20. Megalingam, R.K., Rudravaram, G., Kumar, D.V., Deepika, A.S., Smaran, K.S.: Dynamic lane segmentation for autonomous vehicles using neural networks. In: 2022 5th International Conference on Advances in Science and Technology (ICAST), pp. 444–449 (2022). https://doi.org/10.1109/ICAST55766.2022.10039575
21. Ohgushi, T., Horiguchi, K., Yamanaka, M.: Road obstacle detection method based on an autoencoder with semantic segmentation. In: Ishikawa, H., Liu, C.-L., Pajdla, T., Shi, J. (eds.) ACCV 2020. LNCS, vol. 12627, pp. 223–238. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-69544-6_14

22. Palafox, P.R., Betz, J., Nobis, F., Riedl, K., Lienkamp, M.: Semanticdepth: fusing semantic segmentation and monocular depth estimation for enabling autonomous driving in roads without lane lines. Sensors **19**(14), 3224 (2019). https://doi.org/10.3390/s19143224
23. Pan, X., Shi, J., Luo, P., Wang, X., , Tang, X.: Spatial as deep: Spatial CNN for traffic scene understanding. In: AAAI Conference on Artificial Intelligence (AAAI) (February 2018)
24. Pinggera, P., Ramos, S., Gehrig, S., Franke, U., Rother, C., Mester, R.: Lost and found: detecting small road hazards for self-driving vehicles (2016). https://arxiv.org/abs/1609.04653
25. Scannavino, K.R.F., Nakagawa, E.Y., Fabbri, S.C.P.F., Ferrari, F.C.: Revisão sistemática da literatura em engenharia de software: teoria e prática (2017)
26. Schieber, H., Duerr, F., Schoen, T., Beyerer, J.: Deep sensor fusion with pyramid fusion networks for 3d semantic segmentation. In: 2022 IEEE Intelligent Vehicles Symposium (IV), pp. 375–381 (2022). https://doi.org/10.1109/IV51971.2022.9827113,
27. Shoeb, Y., Chan, R., Schwalbe, G., Nowzard, A., Güney, F., Gottschalk, H.: Have we ever encountered this before? retrieving out-of-distribution road obstacles from driving scenes (2023). https://arxiv.org/abs/2309.04302
28. Shyam, P., Yoon, K.J., Kim, K.S.: Weakly supervised approach for joint object and lane marking detection. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 2885–2895 (2021). https://doi.org/10.1109/ICCVW54120.2021.00323
29. Singh, S.: Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Traffic Safety Facts (2018). https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812506
30. Soori, M., Arezoo, B., Dastres, R.: Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cogn. Robot. **3**, 54–70 (2023)
31. Sousa, F.L.M.d., Silva, M.J.d., Santos, R.C.C.d.M., Silva, M.C., Oliveira, R.A.R.: Deep-learning-based embedded adas system. In: 2021 XI Brazilian Symposium on Computing Systems Engineering (SBESC), pp. 1–8 (2021). https://doi.org/10.1109/SBESC53686.2021.9628316
32. StraitsResearch: ADAS and autonomous driving component market segmentation, demand, growth, forecast to 2030 (2022). https://straitsresearch.com/report/adas-and-autonomous-driving-component-market
33. Thombre, A., Rai, A.K., Dumka, L., Agarwal, A.: Object distance estimation from a single moving camera for advanced driver assistance system. In: 2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp. 1–6 (2024). https://doi.org/10.1109/CONECCT62155.2024.10677218
34. Tran, D., Pham, L.H., Nguyen, H.H., Tran, T., Jeon, H.J., Jeon, J.W.: Universal detection-based driving assistance using a mono camera with jetson devices. IEEE Access **10**, 59400–59412 (2022). https://doi.org/10.1109/ACCESS.2022.3179999
35. Usman, M., K, T.A., Ahmed, M.R., Gudodagi, R., T, N.K.: Exploiting the joint potential of instance segmentation and semantic segmentation in autonomous driving. In: 2023 International Conference for Advancement in Technology (ICONAT), pp. 1–7 (2023). https://doi.org/10.1109/ICONAT57137.2023.10080167
36. Vojir, T., Šipka, T., Aljundi, R., Chumerin, N., Reino, D.O., Matas, J.: Road anomaly detection by partial image reconstruction with segmentation coupling. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 15631–15640 (2021). https://doi.org/10.1109/ICCV48922.2021.01536

37. Xiao, P., et al.: Pandaset: Advanced sensor suite dataset for autonomous driving (2021). https://arxiv.org/abs/2112.12610
38. Xu, H., Wang, S., Cai, X., Zhang, W., Liang, X., Li, Z.: Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending. In: ECCV (2020)
39. Yoo, S., et al.: End-to-end lane marker detection via row-wise classification (2020). https://arxiv.org/abs/2005.08630
40. Yu, F., et al.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning (2020). https://arxiv.org/abs/1805.04687
41. Zhao, W., Wang, T., Tan, A., Ren, C.: Nighttime pedestrian detection based on a fusion of visual information and millimeter-wave radar. IEEE Access **11**, 68439–68451 (2023). https://doi.org/10.1109/ACCESS.2023.3291398
42. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ade20k dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 633–641 (2017)