Information Fusion 113 (2025) 102644

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

Terrain detection and segmentation for autonomous vehicle navigation: A state-of-the-art systematic review

Md Mohsin Kabir [a,b,*], Jamin Rahman Jim [c], Zoltán Istenes [a]

[a] Faculty of Informatics, Eötvös Loránd University, Budapest, 1117, Hungary
[b] Superior Polytechnic School, University of Girona, Girona, 17001, Spain
[c] Advanced Machine Intelligence Research Lab, Dhaka, 1216, Bangladesh

ARTICLE INFO

Keywords:
Autonomous vehicles
Terrain detection
Segmentation techniques
Machine learning
Deep learning
Computer vision

ABSTRACT

This review comprehensively investigates the current state and emerging trends of autonomous vehicle terrain detection and segmentation. By systematically reviewing literature from various databases, this study outlines the evolution of detection and segmentation techniques from traditional computer vision methods to advanced machine learning and deep learning approaches. It identifies critical technological advancements, evaluates their performance, and discusses the challenges faced under various environmental conditions, data acquisition, and integration with vehicle systems. This study also highlights the need for standardized benchmarks and datasets to facilitate the development and testing of robust terrain detection systems. This review encompasses terrain detection and segmentation in structured environments, such as urban roads and highways, and unstructured environments, including rural paths and off-road terrains, to comprehensively analyze autonomous vehicle navigation challenges. By analyzing recent research findings, this review provides insights into future directions for overcoming these limitations and fostering innovation in the autonomous driving domain.

1. Introduction

The journey towards autonomous driving has transformed from a futuristic dream to a tangible reality, rapidly evolving over the past decade [1]. Essentially, the key to this transformation is how well vehicles can figure out and move around in their surroundings, which heavily depends on the accurate detection and segmentation of terrains. This capability is a technical requirement and foundation for ensuring the safety, efficiency, and reliability of autonomous vehicles [2].

This review systematically explored terrain detection and segmentation technologies in structured and unstructured environments. In structured environments such as urban roads and highways, the focus is on evaluating how existing technologies navigate well-defined paths and traffic regulations [3]. Conversely, in unstructured environments such as rural paths and off-road terrains, this review assessed the challenges and solutions for navigating areas without clear demarcations or predictable conditions [4]. By addressing these diverse settings, this review highlights the versatility and adaptability of the current approaches for handling varied terrain complexities, which are crucial for advancing autonomous vehicle navigation technologies.

Autonomous systems technology has evolved rapidly, driven by advancements in sensors, algorithms, and computational power [5]. The ability of a vehicle to accurately detect the terrain and segment various elements within a driving environment is central to this evolution [6]. Terrain detection in autonomous driving refers to the process by which a vehicle identifies the type of ground surface being navigated. This capability is critical for adjusting the driving strategies to accommodate asphalt, gravel, mud, or snow conditions. Sophisticated sensor systems and machine learning algorithms have led to recent advancements in terrain detection technology [7]. LIDAR (Light Detection and Ranging) sensors, combined with radar and high-resolution cameras, provide detailed information about the vehicle's surroundings, enabling precise detection of various terrain types. Machine learning models and deep neural networks have significantly improved the accuracy of terrain detection by learning from vast terrain images and sensor data datasets [8]. These models can distinguish subtle differences in terrain textures and patterns that traditional algorithms cannot easily discern. Furthermore, integrating GPS and geographic information system (GIS) data helps predict terrain types ahead of the vehicle, thereby enhancing the predictive capabilities of autonomous driving systems [9].

Terrain segmentation extends the concept of terrain detection by identifying different terrain types and segmenting the visual field into distinct categories based on surface type. This segmentation is crucial

* Corresponding author at: Faculty of Informatics, Eötvös Loránd University, Budapest, 1117, Hungary.
E-mail address: mdmkabi@gmail.com (M.M. Kabir).

https://doi.org/10.1016/j.inffus.2024.102644
Received 8 April 2024; Received in revised form 13 August 2024; Accepted 15 August 2024
Available online 19 August 2024

Table 1
Complexities in terrain detection and segmentation for autonomous driving.

Category	Description	Example challenges
Environmental	Challenges arising from varying and unpredictable outdoor conditions.	– Lighting variations (e.g., bright sunlight, night-time) – Weather conditions (e.g., rain, fog, snow) – Seasonal changes affecting road visibility
Technical	Challenges related to the technologies and methodologies used in detection and segmentation.	– Real-time processing requirements – Accuracy and reliability under diverse scenarios – Algorithmic limitations in complex scenes
Data-related	Challenges associated with the data used for training and testing the systems.	– Availability of high-quality, annotated datasets – Diversity in data to cover various driving environments – Data privacy and security concerns
Integration	Challenges involved in integrating terrain detection and segmentation systems with other vehicle systems.	– Sensor fusion for enhanced reliability, combining data from multiple sensors (e.g., LIDAR, cameras, radar) – Coordination with other perception systems (e.g., obstacle detection) – Scalability and adaptability to different vehicle models

for autonomous vehicles to make informed speed, steering, and path-planning decisions. The newest technologies in terrain segmentation leverage deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Semantic Segmentation models [10]. These models were trained on extensively labeled datasets to classify each pixel of an image into different terrain categories accurately. Recent innovations include Generative Adversarial Networks (GANs), which enhance the resolution and accuracy of real-time segmentation maps [11]. In addition, transfer learning techniques allow these models to adapt to new, unseen environments with minimal additional training [12]. This adaptability is particularly beneficial for autonomous vehicles operating under varied and unpredictable terrain conditions, thereby ensuring safe and efficient navigation across different landscapes.

However, terrain detection and segmentation are also complex [13]. From the unpredictability of weather conditions to the diverse urban and rural landscapes, these technologies must accurately interpret various scenarios. These challenges extend beyond environmental conditions and include issues related to data quality and availability. In addition, there are concerns regarding the robustness of algorithms against real-world variability and the integration of these technologies into a unified driving system [14]. Table 1 provides a primary understanding of the terrain detection and segmentation complexities under autonomous driving.

Recognizing these challenges, this literature review aims to dissect current terrain detection and segmentation technologies for autonomous vehicles. We aim to catalog the emerging trends and technologies shaping the field and critically analyze the remaining challenges. In doing so, we aim to integrate where we stand and where the gaps lie in our collective knowledge and technological capabilities. This systematic literature review aimed to:

- Explore the spectrum of technologies and methodologies employed in terrain detection and segmentation for autonomous vehicles.
- Evaluate the effectiveness and limitations of current approaches under varied environmental conditions.
- Identifying critical challenges hindering the advancement of terrain detection systems. Emerging trends and potential solutions for addressing these challenges have been highlighted.
- Suggest areas for future research and development to enhance the reliability of autonomous driving technologies.

Additionally, this review examines studies published within the last six years (2018–present), focusing on peer-reviewed journals, conference proceedings, and reputable book chapters. The selection aims to capture the evolution of terrain detection and segmentation technologies and their application in autonomous driving. By setting these boundaries, the review ensures relevance and comprehensiveness while providing insights into the field's research and development trajectory.

The need for this review stems from the rapid pace at which autonomous driving technologies have been advancing. With the exponential growth in the field, there is a critical need to consolidate knowledge, synthesize recent findings, and identify gaps in the literature. This review provides a foundation for researchers, engineers, and developers by providing an updated analysis of the technologies, challenges, and opportunities, thereby guiding future research and development. Moreover, no comprehensive study has been conducted in this domain. An examination of recent literature underscores a notable gap in comprehensive surveys that specifically target the domain of terrain detection and segmentation for autonomous driving. Table 2 highlights the three most pertinent surveys in this domain and their contributions. A systematic review revealed sporadic coverage: a 2021 survey [15] focused on terrain identification for autonomous robots without addressing dataset preprocessing or autonomous driving challenges, whereas a 2022 study [16] delved into vision-based semantic segmentation, albeit with a narrow focus on vision techniques, and a 2023 survey [17] concentrated on 3D object detection, sidestepping the nuanced challenges of terrain segmentation. This landscape highlights the urgent need for a dedicated survey that combines the disparate threads of terrain detection and segmentation research with a keen focus on datasets, preprocessing, and integrating these technologies into autonomous driving systems. Our survey aims to fill this critical void by offering a comprehensive overview that is conspicuously absent in the current body of literature.

The remainder of this paper is organized as follows. Section 2 outlines the methodology for selecting and analyzing the literature. Section 3 presents an overview of the technologies used in terrain detection and segmentation and offers a comparative analysis of their strengths and weaknesses. Section 4 examines the challenges encountered in the field and discusses technical, data-related, and evaluative aspects. Section 5 highlights applications that illustrate the practical implications of these technologies. Finally, Sections 6 and 7 summarize the discussion, conclude the review, summarize the key findings, and suggest directions for future research.

2. Research methodology

This section details the Systematic Literature Review (SLR) approach, structured according to guidelines established by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [18]. Our methodology ensures a thorough and impartial collection, analysis, and synthesis of relevant literature and includes the following stages: Search Strategy, Selection Criteria, Data Extraction, and Quality Assessment. PRISMA flow diagram. Fig. 1 summarizes the literature screening and selection stages. Initially, 507 articles were identified using the bibliographic databases. After screening the titles and abstracts, 127 articles were excluded because of irrelevance and duplication. Further full-text review of the remaining 380 articles excluded 161 more articles because of incomplete information or failure to meet our specific criteria related to scope, novelty, and methodological rigor. This process culminated in 219 articles that were deemed eligible for in-depth analysis, with a final refined set of 122 selected articles

Table 2
Summary of key surveys in terrain detection and segmentation for autonomous driving.

Ref.	Domain	Systematic review	Dataset & preprocessing	Technological advancements	Results analysis	Challenges & future directions	Contribution
[15] (2021)	Terrain identification and classification.	✓	✗	✗	✓	✗	Provides a comprehensive review of recent developments in terrain identification, classification, parameter estimation, and control strategies for autonomous robots.
[16] (2022)	Vision-based semantic segmentation in scene understanding	✓	✓	✗	✓	✓	Analysis of the current achievements, challenges, and outlooks in vision-based semantic segmentation for scene understanding in autonomous driving.
[17] (2023)	3D object detection for autonomous driving	✓	✓	✓	✗	✓	A comprehensive survey of image-based 3D object detection for autonomous driving, including summarizing pipelines, analyzing components, introducing new taxonomies, and identifying current challenges and future research directions.

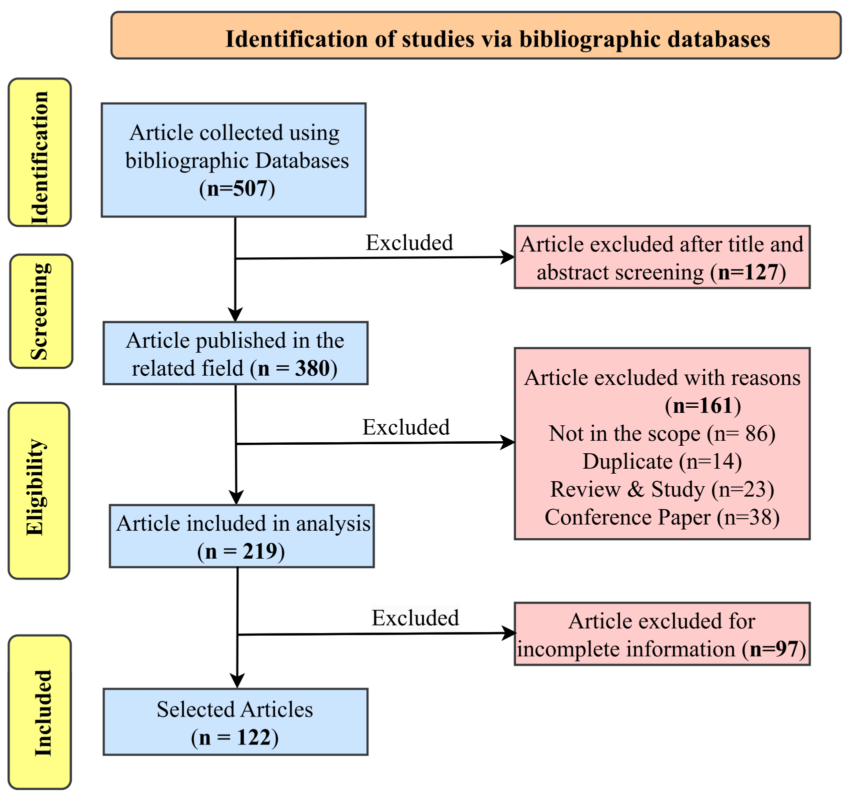

Fig. 1. PRISMA flow diagram: Synthesizing research methodology.

forming the core of our systematic review. This rigorous selection process ensured a comprehensive and unbiased assessment of the current landscape and the emerging trends in the field.

2.1. Search strategy

A systematic search was conducted in databases such as IEEE Xplore, ScienceDirect, Scopus, and Google Scholar to capture the broad scope of terrain detection and segmentation in autonomous vehicles. The search covered peer-reviewed journal articles and conference proceedings published in English over the past six years. Keywords were strategically combined using Boolean operators to maximize the relevance and comprehensiveness of the search. Fig. 2 presents a word cloud derived from bibliographic data on autonomous vehicle terrain detection and segmentation, from 2018 to 2024, using Scopus as the source. This visualization highlights the frequency and significance of key terms, with larger text indicating more prevalent topics in the field. The prominence of terms like “Deep Learning”, “Machine Learning”, and “Semantic Segmentation” underscores the shift to advanced computational techniques. Similarly, “Computer Vision”, “Obstacle Avoidance”, and “Object Detection” reflect core research focus that is vital for improving autonomous navigation in complex environments. The inclusion of “Automotive Radar”, “Data Fusion”, and “Sensor Fusion” points to crucial integration efforts in-vehicle systems to enhance detection and decision-making capabilities. This word cloud effectively summarizes current trends and pinpoints areas for future research, providing a clear snapshot of the field’s evolution and technological landscape.

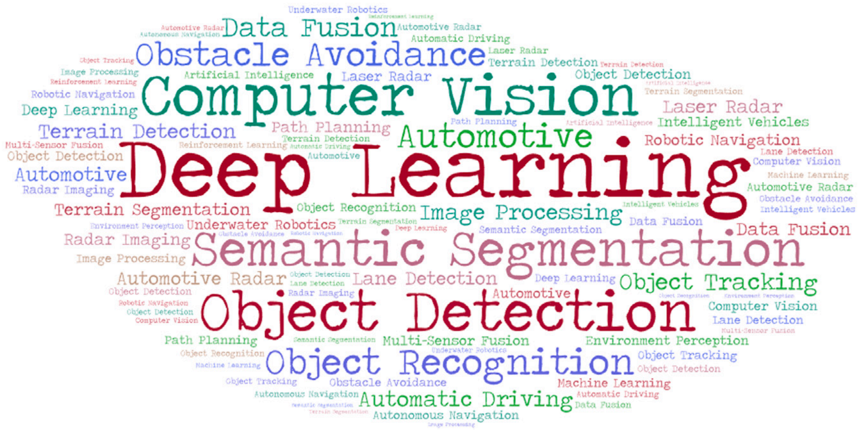

Fig. 2. Word cloud of co-occurring keywords in bibliographic data (2018–2024) based on scopus analysis.

Table 3
Inclusion and exclusion criteria for systematic literature review.

Inclusion criteria	Exclusion criteria
– Studies published in the last six years have focused on terrain detection and segmentation in autonomous vehicles. – Research using AI, machine learning, or deep learning with empirical performance evaluation. – English-language peer-reviewed articles that offer a new comparative analysis or integration of terrain detection with vehicle systems. – Studies include ethical discussions, real-world testing, simulation results, and open-source code or datasets.	– Studies older than six years (with exceptions) or unrelated to terrain detection in autonomous vehicles. – Non-peer-reviewed articles, or those lacking empirical results or proper evaluation. – Articles in languages other than English, preliminary reports, or abstracts without full-text availability. – Papers focusing on indoor systems and high-level decision-making without direct relevance or lacking detailed outcomes (e.g., conference abstracts, editorials).

2.2. Selection criteria

The selection criteria are streamlined for clarity and focus on recent, high-quality research pivotal to advancing terrain detection and segmentation technologies. A summary of the inclusion and exclusion criteria is shown in Table 3.

2.3. Data extraction

For each qualifying study, we extracted critical data to synthesize the current state of terrain detection and segmentation technologies, including publication details, study objectives, methodologies, sensor types, main findings, and identified challenges.

2.4. Quality assessment

Quality assessment was performed to ensure the credibility and relevance of the included studies. This assessment emphasized the relevance of the findings, methodological rigor, innovation, and implications.

3. Technological advancements

The advancement of autonomous vehicle (AV) technology significantly depends on the accurate detection and segmentation of terrain. This capability ensures that the AVs can navigate safely and efficiently in diverse environments. This section examines the core technologies that underpin terrain detection and segmentation. This section covers the sensors and data acquisition methods, databases, data processing techniques, detection and segmentation algorithms, and how these technologies are integrated into vehicle systems.

3.1. Sensors and data acquisition

Sensors are crucial for autonomous vehicle navigation and provide raw terrain detection and segmentation data. Table 4 provides a brief overview of the different sensors used in autonomous driving for terrain detection and segmentation, including the LiDAR, Radar, and Cameras. These sensors capture detailed environmental features, enabling vehicles to intelligently interpret and interact with their surroundings.

3.2. Databases

In this section, we explore diverse datasets that is crucial for advancing the research on terrain detection and segmentation in autonomous vehicles. This exploration is essential for understanding the various environmental scenarios, sensor modalities, and annotation diversities of these datasets, each contributing uniquely to the development, testing, and enhancement of terrain detection algorithms. From real-world driving scenes in CODA [19] to the complex urban landscapes of the Cityscapes [20] dataset, these collections provide invaluable resources for developing sophisticated and resilient autonomous driving technologies. The datasets vary widely, covering urban and highway scenarios and challenging off-road conditions, as observed in the RUGD [21] and RELLIS-3D [22] datasets. They include data from various sensors, including LiDAR, radar, and cameras, and offer diverse annotations from object detection to semantic segmentation. This comprehensive dataset repository is instrumental for researchers and practitioners in the field, facilitating the development of algorithms capable of navigating the complexities of real-world environments with unprecedented accuracy and reliability. Table 5 presents the frequently used databases in this domain, highlighting their significance and broad applications in the field.

Table 4
Popular sensors used for data collection in the terrain detection and segmentation domain.

Sensor type	Primary use	Advantages	Limitations
LiDAR	High-resolution mapping of the environment	High accuracy works in various lighting conditions	High cost, complex data processing
Radar	Detection of objects and terrain under various weather conditions	Works in poor weather, long-range detection	Low resolution
Cameras	Visual recognition of terrain features	Rich data, low cost	Affected by lighting and weather conditions
Others (Ultrasonic, Infrared)	Close range detection and simple terrain features recognition	Low cost, useful for specific applications like parking	Limited range, affected by environmental conditions

3.3. Processing techniques

Data preprocessing is a crucial step in terrain detection and segmentation tasks, particularly in the context of autonomous driving, where the accuracy and reliability of terrain analysis can significantly affect vehicle performance and safety. Some popular data preprocessing techniques used for terrain detection and segmentation are as follows:

3.3.1. Noise reduction

Filtering: Filters such as Gaussian blur [69], median filter, or bilateral filters [70] are used to reduce noise and smooth the data. This is particularly important for LiDAR and radar data, which may contain noise owing to environmental factors or sensor inaccuracies.

Ground Removal: In LiDAR data, the preprocessing step of removing ground points is primarily applied in structured environments such as urban or highway settings, where the focus is on identifying obstacles and terrain irregularities [71]. Techniques such as the Random Sample Consensus (RANSAC) [72] algorithm are effective for ground plane estimation and removal in scenarios where the ground is relatively flat and smooth. However, it is important to note that in unstructured outdoor or field scenarios, where terrain segmentation is the objective, the approach to ground detection must be carefully considered. Here, the ground provides crucial information regarding the nature of the terrain, and indiscriminate removal may not be beneficial. This underscores the need for adaptive preprocessing strategies that consider specific characteristics of the analyzed environment.

3.3.2. Normalization

Intensity normalization: Intensity normalization techniques are essential preprocessing steps for standardizing the range of pixel values in images, thereby improving subsequent analysis and processing tasks [73]. Global normalization standardizes the entire dataset to a uniform mean and standard deviation, usually aiming for a Gaussian distribution of intensity values. Local histogram equalization adjusts the intensities in specific regions to enhance contrast, which is beneficial in environments with variable lighting. Gamma correction [74] applies a power-law transformation to amend the non-linear effects of display systems or augment contrast the under dim conditions. Logarithmic transformation balances brightness across images by spreading dark and compressing bright pixels, aiding in managing images with a wide dynamic range [75]. Finally, retinex theory-based methods boost contrast and color fidelity by emulating human vision characteristics [76], which is particularly advantageous for enhancing feature visibility in poorly lit scenes.

Range Data Normalization: Popular techniques for range data normalization address the challenge of varying sensor value ranges by adjusting data to a consistent scale and enhancing the algorithm performance across different scenarios. Min–max normalization [77] typically scales data to a fixed range [0, 1], based on the minimum and maximum values observed, which can be applied on a per-frame basis or across the entire dataset for uniformity. Z-Score Normalization [78], or standardization, transforms data to have zero mean and unit variance, is suitable for algorithms sensitive to data scale, and is often used across the entire dataset to ensure consistency. Robust Scaling, utilizing median and interquartile ranges, offers resilience against outliers and can be adapted to specific frames or scenarios in which the influence of outliers is significant. Quantile Normalization adjusts the data distribution to match a specified distribution such as a uniform or Gaussian distribution, which is beneficial for mitigating sensor noise or outliers in scenario-specific applications. Finally, Unit Vector Transformation, which normalizes data points to unit length, is particularly beneficial for 3D point cloud data from LiDAR, ensuring a consistent representation of directional data and is typically applied per data point or frame.

3.3.3. Data augmentation

Geometric Transformations: Applying rotations, translations, scaling, and flipping can increase the robustness of terrain detection algorithms by providing a more diverse training dataset [79].

Photometric Transformations: Adjusting brightness and contrast and adding synthetic shadows or highlights can help algorithms become more invariant to lighting conditions [80].

3.3.4. Feature extraction

Edge Detection: Techniques such as Sobel, Canny, or Laplacian filters highlight the edges in images, which are useful for distinguishing between different terrain types [81].

Texture Features: Extracting texture features, such as those from Local Binary Patterns (LBP) or Gabor filters, can help identify different terrains based on their surface properties [82].

3.3.5. Dimensionality reduction

Dimensionality reduction plays a pivotal role in preprocessing data for terrain detection and segmentation, particularly when dealing with high-dimensional sensor data from LiDAR, radar, and cameras [83]. This process is essential for mitigating the curse of dimensionality, enhancing algorithmic efficiency, and facilitating data visualization. Among popular techniques, Principal Component Analysis (PCA) is celebrated for its ability to transform data into linearly uncorrelated variables, effectively preserving data variance [84]. Linear Discriminant Analysis (LDA) performed in supervised settings by identifying distinguishing features between classes [85]. To visualize complex, high-dimensional datasets, t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) are preferred because of their ability to maintain the global structure of the data while facilitating dimensionality reduction. Autoencoders, a neural network-based approach, are adept at unsupervised data encoding for dimensionality reduction. Feature Selection Techniques focus on retaining a subset of original features for interpretability. Techniques such as Isomap and Multidimensional Scaling (MDS) are instrumental in preserving geodesic distances and visualizing data similarity. Random Projection offers a computationally efficient alternative that leverages random matrices to project data into lower-dimensional spaces. The nature of the data dictates the selection of these techniques and

Table 5
Summary of key datasets for terrain detection and segmentation in autonomous vehicle research. This table overviews diverse datasets, highlighting their contributions to developing and evaluating terrain detection and segmentation technologies.

Reference	Dataset name	Details	Papers
[19] (2022)	CODA	The dataset consists of 1500 carefully selected real-world driving scenes, each containing four object-level corner cases (on average), spanning more than 30 object categories.	
[23] (2021)	3D-Point Cloud dataset	Depth vision technology in locomotion devices benefit individuals with disabilities, supporting an ecologically friendly lifestyle. The dataset, including point cloud recordings during diverse urban locomotion modes, is valuable for researchers across disciplines. It aids in improving control for human locomotion assistive devices and designing vision-based detection systems for humanoid robots navigating urban terrains.	[24,25]
[21] (2019)	Pinocchio	Pinocchio is an open-source software framework for rigid body dynamics algorithms and their derivatives. It goes beyond standard robotics algorithms, providing features for efficient robot control, planning, and simulation. The paper outlines Pinocchio's design and programming patterns, highlighting its efficiency. Performance evaluation against RBDL, another popular framework, is conducted.	[26,27]
[22] (2019)	RUGD	The Robot Unstructured Ground Driving (RUGD) dataset addresses a gap in autonomous driving research by offering diverse video sequences from a small unmanned robot navigating challenging, unstructured environments. With over 7000 annotated frames, it provides valuable data for improving navigation in off-road scenarios, showcasing unique challenges not covered by existing benchmarks.	[22],
[28] (2021)	RELLIS-3D datasets	RELLIS-3D, an off-road dataset from Texas A&M's Rellis Campus, offers 13,556 LiDAR scans and 6235 images. It challenges existing algorithms with class imbalance and environmental complexities, highlighting the need for advanced models in off-road navigation research.	[29,30]
[31] (2017)	ADE20K	The paper introduces ADE20K, a richly annotated dataset for scene parsing in computer vision, covering scenes, objects, and object parts. It includes a benchmark with 150 classes, evaluating segmentation models, and proposing a novel Cascade Segmentation Module for improved parsing accuracy.	[32,33]
[34] (2014)	Microsoft COCO	It features images of complex scenes with 91 easily recognizable objects labeled with per-instance segmentations for precise localization. The dataset, created with extensive crowd worker involvement, comprises 2.5 million labeled instances in 328,000 images.	[35,36]
[20] (2016)	The cityscapes dataset	The Cityscapes dataset are introduced as a semantic urban scene understanding benchmark. Unlike existing datasets, it addresses the complexity of real-world urban scenes. It comprises stereo video sequences from 50 cities, with 5000 images with high-quality pixel-level annotations and 20,000 images with coarse annotations. Cityscapes aims to advance pixel-level and instance-level semantic labeling approaches in urban environments.	[37,38]
[39] (2019)	Semantickitti	A large dataset for laser-based semantic segmentation, derived from the KITTI Vision Odometry Benchmark. Focused on self-driving cars, the dataset provides dense point-wise annotations for a 360-degree field-of-view of automotive LiDAR. Three benchmark tasks are proposed: semantic segmentation using a single scan, semantic segmentation with multiple past scans, and semantic scene completion for anticipating future semantic scenes.	[40,41]
[42] (2018)	HighD	A large-scale naturalistic vehicle trajectory dataset from German highways, encompassing 16.5 h of measurements, 110,000 vehicles, a total driven distance of 45,000 km, and 5600 recorded complete lane changes. The dataset is evaluated for quantity, variety, and scenario coverage.	[43,44]
[45] (2020)	Nuscenes	A dataset incorporating a fully autonomous vehicle sensor suite, including six cameras, five radars, and one lidar, all with a 360-degree field of view. The dataset consists of 1000 scenes, each 20 s long, and is fully annotated with 3D bounding boxes for 23 classes and eight attributes. nuScenes aims to facilitate the training and evaluation of machine learning-based methods for comprehensive object detection and tracking in autonomous vehicles.	[46,47]
[48] (2017)	Mapillary Vistas	The Mapillary Vistas Dataset is a large-scale street-level image dataset with 25,000 high-resolution images annotated into 66 object categories. It includes instance-specific labels for 37 classes, annotated in a dense and fine-grained style using polygons to delineate individual objects. The dataset is five times larger than the fine annotations in Cityscapes. It features images from around the world captured under various conditions, including different weather, seasons, and times of the day. Images come from diverse imaging devices and photographers, ensuring a rich diversity of details and geographic coverage.	[49,50]
[51] (2019)	ApolloScape	ApolloScape is an extensive street-level road scene dataset for self-driving applications, offering 143,906 frames with high-quality ground-truth data. It includes pixel-level semantic segmentation, poses information, and 3D point clouds, providing nearly 15 times more data than existing datasets like KITTI or Mapillary Vistas, with rich labeling for various applications.	[52,53]
[54] (2013)	KITTI Road	The KITTI Road benchmark features 289 training and 290 test images for road and lane estimation in urban environments, divided into three categories and a combined category. It provides manually annotated ground truth for the overall road area and the ego-lane (only in "um" category) but only for training images.	[55,56]

(*continued on next page*)

Table 5 (*continued*).

[57] (2019)	Road Anomaly	This dataset contains images of unusual dangers that can be encountered by a vehicle on the road — animals, rocks, traffic cones, and other obstacles. It aims to test autonomous driving perception algorithms in rare but safety-critical circumstances.	[58,59]
[60] (2018)	DeepGlobe	The DeepGlobe Satellite Image Understanding Challenge focuses on three satellite image analysis tasks. The released datasets aim to serve as benchmarks for future satellite image research. Additionally, since the challenge addresses real-world computer vision problems, these datasets will likely be valuable for creating robust vision algorithms with applications extending beyond remote sensing.	[61,62]
[63] (2016)	Cityscapes	Cityscapes is a database for urban scene semantic analysis, annotating 30 classes in 8 categories, such as humans and vehicles. It includes around 5000 finely and 20,000 coarsely annotated images from 50 cities, selected for their dynamic elements and diverse layouts. The data, captured in good weather, originally comes from video frames.	[64,65]
[66] (2009)	CamVid	CamVid is a database for road and driving scene analysis, featuring 701 frames sampled from five video sequences shot with a 960 × 720 resolution camera on a car's dashboard. Frames, sampled at 1 fps for four sequences and 15 fps for one, are manually annotated with 32 classes, including buildings, vehicles, and road elements.	[67,68]

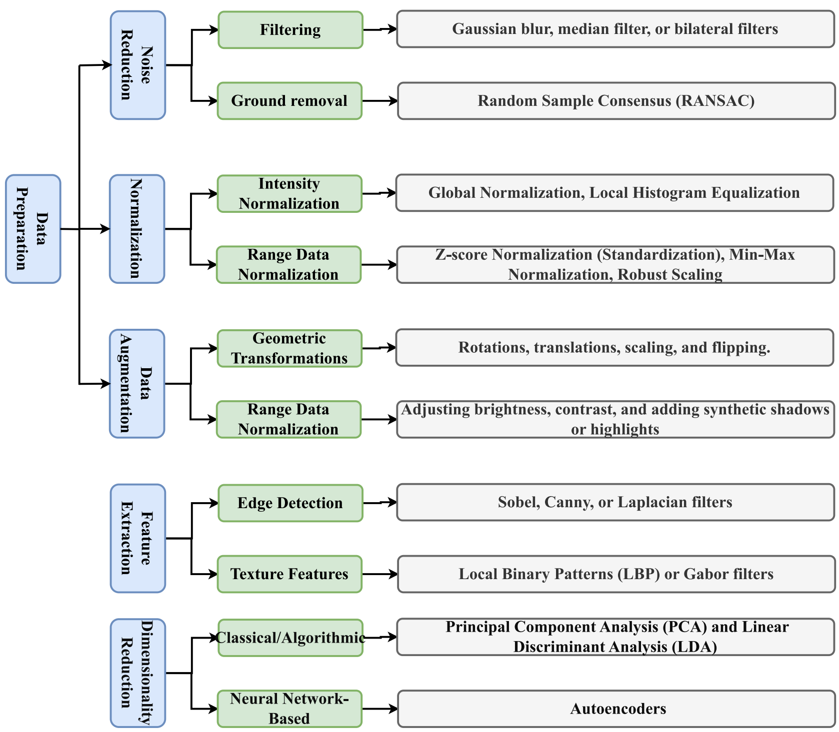

Fig. 3. Popular preprocessing techniques in terrain detection and segmentation.

the specific goals of the analysis, with PCA and LDA favoring linear reductions and t-SNE and UMAP for handling non-linear complexities in datasets.

Fig. 3 shows several commonly employed preprocessing techniques for terrain detection and segmentation.

3.4. Detection algorithms

This section summarizes the evolution and diversity of detection algorithms from 2018 to 2024, showing the significant advancements in object detection technologies. Early models, such as Faster R-CNN [86], laid the groundwork for integrating region proposals directly into convolutional networks, achieving high accuracy but at a computational cost. The introduction of SSD [87] and YOLO [88] marked a shift towards single-pass detection, prioritizing speed and enabling real-time processing, albeit with compromises in detecting small objects. Advances in 3D point cloud analysis were led by PointNet++ [89] and PointCNN [90], which introduced methods for capturing local structures and enabling effective learning from 3D data despite the challenges in computational demand and generalization.

DANet [91] explored weakly supervised object localization, thereby reducing its dependency on extensively labeled datasets. DeepLabv3+ [92] pushes the boundaries of semantic segmentation using a combination of atrous convolution and spatial pyramid pooling, thereby setting high standards for accuracy at the expense of computational resources.

More recent innovations such as BevDet [93] and Detr3D [94] have ventured into 3D object detection from multi-view images, enhancing spatial analysis in autonomous driving but requiring sophisticated calibration. Meta-learning [95] and MFF-Net [96] reflected a trend towards models that adapt quickly to new tasks or integrate multiple data features, highlighting the field's move towards flexibility, adaptation, and enhanced accuracy, albeit with increased complexity and computational needs. These developments underscore the rapid progress in detection technologies, addressing specific challenges and pushing the envelope in object detection and segmentation capabilities. A brief summary and acomparative analysis of these algorithms are presented in Table 6. In addition, a comprehensive exploration of these algorithms is presented in this section

Table 6
Brief summary and comparative analysis of recent detection algorithms commonly used for AV.

Algorithm	Key process	Advantages	Limitations
Faster R-CNN [86] (2015)	Enhances R-CNN by integrating region proposal with the convolutional network.	High accuracy, considers region proposals and feature extraction.	Computationally intensive compared to newer models.
SSD [87] (2016)	Object detection in a single pass, eliminating separate region proposal step.	Faster than R-CNN variants, real-time processing.	Struggles with very small objects, speed-accuracy trade-off.
YOLO [88] (2016)	Divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell.	Extremely fast, suitable for real-time processing.	Less accurate on small objects in earlier versions.
PointNet++ [89] (2017)	Applies PointNet recursively on nested partitions of the input point set.	Captures local structures at multiple scales, effective for 3D point clouds.	Computationally intensive, may miss fine details.
PointCNN [90] (2018)	Transforms input points for convolution, learning from unordered point clouds.	Allows convolutional networks to learn from 3D point clouds effectively.	Computationally expensive transformation, may not generalize well.
DANet [91] (2019)	Uses divergent activation to enhance feature discriminability with weak supervision.	Effective with limited annotations, reduces the need for extensive labeled datasets.	May not achieve the accuracy of fully supervised methods.
DeepLabv3+ [92] (2020)	Combines atrous convolution with spatial pyramid pooling for semantic segmentation.	High segmentation accuracy, especially at object boundaries.	Demands significant computational resources.
BevDet [93] (2021)	Projects detections to bird-eye-view from multi-camera inputs for 3D object detection.	Enhances 3D detection in autonomous driving, comprehensive spatial analysis.	Requires multi-camera calibration and synchronization.
Meta-Learning [95] (2021)	Adapts to new tasks quickly with minimal data by learning the learning process.	Flexible and rapid adaptation to new tasks.	Complex training, careful tuning required.
MFF-Net [96] (2022)	Integrates multiple features to improve detection in community detection.	Combines various data representations for enhanced accuracy.	Increases model complexity and computational demand.
Detr3D [94] (2022)	Simplifies 3D object detection from images by transforming it into a 2D query problem.	Streamlines 3D detection from multi-view images.	Challenges in interpreting complex 3D scenes from 2D projections.

3.4.1. Faster R-CNN

Faster R-CNN [86] is an object detection model that integrates region proposal and object detection into a single, end-to-end framework. The model consists of three main components: a deep convolutional neural network (CNN) for feature extraction, Region Proposal Network (RPN) for generating object proposals, and Fast R-CNN detector for refining these proposals and predicting object classes. The RPN generates proposals by sliding a small network over the convolutional feature map, predicting objectness scores and bounding box regression offsets for the anchor boxes. The loss function for Faster R-CNN combines classification and regression losses, formulated as:

$$L(p_i, t_i) = \frac{1}{N_{\mathrm{cls}}} \sum_i L_{\mathrm{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\mathrm{reg}}} \sum_i p_i^* L_{\mathrm{reg}}(t_i, t_i^*) \tag{1}$$

where p_i is the predicted probability of anchor i being an object, p_i^* is the ground-truth label (1 if the anchor is positive and, 0 otherwise), t_i is the predicted bounding box, t_i^* is the ground-truth bounding box, L_{cls} is the classification loss (e.g., softmax loss), L_{reg} is the regression loss (e.g., smooth L1 loss), N_{cls} is the minibatch size, N_{reg} is the number of anchor locations, and λ is a balancing parameter. This unified framework allows efficient and accurate object detection, significantly improving the speed and performance of previous methods.

Faster R-CNN excels in terrain detection owing to its high accuracy and efficient region proposal generation. Its deep learning architecture ensures precise object localization and classification. However, its drawbacks include high computational demands and memory requirements,which limit the real-time deployment on low-resource devices. Additionally, achieving robust performance may necessitate extensive training data and careful parameter tuning to handle terrain complexity effectively.

3.4.2. SSD

The single shot multiBox detector (SSD) [87] algorithm is a popular method for object detection that discretizes the output space of bounding boxes into a set of default boxes with different aspect ratios and scales per feature map location. During the prediction, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. The SSD model combines the predictions from multiple feature maps with different resolutions to handle objects of various sizes. The loss function used in SSD is a weighted sum of the localization loss (e.g., Smooth L1) and confidence loss (e.g., softmax loss). The overall objective function can be formulated as

$$L(x, c, l, g) = \frac{1}{N} \left(L_{\mathrm{conf}}(x, c) + \alpha L_{\mathrm{loc}}(x, l, g) \right) \tag{2}$$

where L_{conf} is the confidence loss, L_{loc} is the localization loss, x matches default boxes to ground truth, c is the class confidences, l and g are the predicted and ground truth box parameters respectively, N is the number of matched default boxes, and α is a weighting term to balance the two loss components.

SSD is advantageous for terrain detection owing to its real-time processing and multi-scale object detection capability, accommodating diverse terrain features effectively. Its straightforward architecture facilitates easier implementation and optimization. However, SSD may struggle with small object detection accuracy and challenges in handling highly overlapping objects or complex backgrounds typical in terrain scenes. Additionally, its performance relies heavily on the quality and diversity of training data, which can be a limitation for specific terrain types.

3.4.3. YOLO

The You Only Look Once [88] model is a real-time object detection system that frames object detection as a single regression problem, from image pixels to bounding box coordinates and class probabilities. The model divides the image into an $S \times S$ grid, and each grid cell predicts B bounding boxes, confidence scores for these boxes, and C class probabilities. The confidence score is defined as $\mathrm{Pr(Object)} \times \mathrm{IOU}_{\mathrm{pred}}^{\mathrm{truth}}$, where Pr(Object) is the probability of an object being present and $\mathrm{IOU}_{\mathrm{pred}}^{\mathrm{truth}}$ is the Intersection Over Union between the predicted box and the ground truth. The final output prediction for each bounding box is encoded as a vector (x, y, w, h, C) where (x, y) represents the center coordinates relative to the grid cell; w and h are the width and height relative to the image dimensions, respectively; and C is the confidence score. The class probabilities for each cell were also predicted, leading to a final output tensor of shape $(S, S, B \times 5 + C)$.

YOLO's real-time capability and grid-based object detection make it advantageous for terrain detection, ensuring fast processing and spatial context retention. However, it may struggle with small or heavily occluded terrain features,thereby limiting the accuracy of these cases. Specialized terrain types may also require tailored training or adjustments to achieve optimal performance.

3.4.4. PointNet++

PointNet++ [89] is an extension of PointNet that addresses the issue of capturing the local geometric structures in point clouds. It introduces a hierarchical neural network that aggregates features from local regions at increasing scales. The key components include the set abstraction (SA) and feature propagation (FP) modules. The set abstraction module uses farthest point sampling (FPS) to select representative points and extract local features using PointNet. It also applies multi-scale grouping to gather contextual information. The feature propagation module further refines the features by aggregating information from different scales using nearest-neighbor interpolation. Mathematically, the set abstraction module can be represented as:

$$\mathrm{SA}(\{\mathbf{x}_i\}_{i=1}^{N}, \{\mathbf{f}_i\}_{i=1}^{N}) = \{\mathbf{f}'_j\}_{j=1}^{M} \tag{3}$$

where $\{\mathbf{x}_i\}_{i=1}^{N}$ are the input points, $\{\mathbf{f}_i\}_{i=1}^{N}$ are the input features, and $\{\mathbf{f}'_j\}_{j=1}^{M}$ are the output features of the sampled points. This hierarchical approach allows PointNet++ to effectively capture both local and global features from point clouds, thereby enhancing its performance in tasks such as point cloud classification and segmentation.

PointNet++ is effective for terrain detection, leveraging its hierarchical architecture to capture detailed local and global features from the point clouds. It is robust to point permutation and varying configurations but may be computationally intensive with larger datasets and require substantial training data for robust performance across diverse terrains.

3.4.5. PointCNN

PointCNN [90] is a neural network architecture designed for point cloud processing that incorporates local feature learning and permutation invariance through a spatial transformer network (STN) and feature-wise max pooling. It utilizes a density-adaptive convolution operation to dynamically adjust the convolutional receptive field based on local point densities. This was achieved using the following equations:

$$\mathbf{y}_i = \mathrm{MLP}(\mathrm{STN}(\{\mathbf{x}_j - \mathbf{x}_i\}_{j\in\mathcal{N}_i}) \cdot \{\mathbf{x}_j - \mathbf{x}_i\}_{j\in\mathcal{N}_i}) \tag{4}$$

$$\mathbf{y}_i = \mathrm{maxpool}(\{\mathbf{y}_j\}_{j\in\mathcal{N}_i}) \tag{5}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ represent the input and output features at point i, $\mathcal{N}_i$ denotes the neighbors of point i, MLP denotes a multi-layer perceptron, and STN represents the spatial transformer network. These equations enable PointCNN to effectively capture local structures and features from point clouds while maintaining permutation invariance.

PointCNN excels in terrain detection by dynamically adjusting local point densities and maintaining robustness to point order,which is crucial for processing unstructured terrain scans. However, its computational demands and dependence on parameter tuning may limit scalability and require careful optimization for optimal performance on diverse datasets.

3.4.6. DANet

DANet [91] utilized dual attention mechanisms to enhance the feature representation in computer vision tasks. The network incorporates both spatial and channel attention modules to adaptively recalibrate feature maps, leveraging the interdependencies across spatial locations and channel-wise relationships. The spatial attention mechanism recalibrates feature responses by explicitly modeling pairwise dependencies, whereas the channel attention mechanism captures dependencies among different feature channels. This dual attention scheme, represented by the following equations, significantly improves the discriminative power and robustness of the feature representation:

$$\mathbf{F}_{\mathrm{spatial}} = \sigma(\mathrm{softmax}(\mathbf{W}_s\mathbf{X}(\mathbf{W}'_s)^{\top}))\mathbf{X}\mathbf{W}_v \tag{6}$$

$$\mathbf{F}_{\mathrm{channel}} = \sigma(\mathrm{softmax}(\mathbf{W}_c\mathbf{X}))\mathbf{X}\mathbf{W}'_v \tag{7}$$

where $\mathbf{X}$ denotes the input feature maps; $\mathbf{W}_s$, $\mathbf{W}'_s$, $\mathbf{W}_c$, $\mathbf{W}_v$, and $\mathbf{W}'_v$ are learnable parameters, and σ represents the activation function (e.g., sigmoid).

3.4.7. Meta-learning

Meta-learning [95] refers to the process by which a learning algorithm improves its learning process through experience. It involves learning, typically by optimizing the parameters or hyperparameters of another learning algorithm. Formally, meta-learning can be framed as finding the optimal parameters θ^* that minimize the expected loss $\mathbb{E}_{\mathcal{D}}[\mathcal{L}(\theta^*; \mathcal{D})]$, where $\mathcal{D}$ represents the distribution of tasks, $\mathcal{L}$ denotes the loss function, and θ are the parameters of the learning algorithm being optimized. Meta-learning algorithms often operate by leveraging data from multiple tasks to generalize across new, unseen tasks, thereby facilitating more efficient learning and adaptation.

Meta-learning for terrain detection offers rapid adaptation to new terrains using prior knowledge from similar tasks, therebyreducing the data requirements and training time. However, its effectiveness depends on task selection and representation quality and may struggle with highly diverse terrains or novel environments, necessitating careful consideration of task variability and adaptation strategies.

3.4.8. MFF-Net

The Multi-scale Feature Fusion Network (MFF-Net) [96] is a deep learning architecture designed to capture features at different scales for robust image processing tasks. The core idea is to integrate multi-scale feature extraction and fusion into a unified framework. MFF-Net can be mathematically represented as follows:

$$\mathbf{F}_{\mathrm{out}} = \mathcal{F}_{\mathrm{fuse}}(\mathcal{F}_1(\mathbf{X}) \oplus \mathcal{F}_2(\mathbf{X}) \oplus \cdots \oplus \mathcal{F}_n(\mathbf{X})) \tag{8}$$

where $\mathbf{X}$ is the input image, $\mathcal{F}_i(\mathbf{X})$ represents the feature extraction at the ith scale, $\oplus$ denotes the concatenation operation, and $\mathcal{F}_{\mathrm{fuse}}$ is the fusion function that combines the multi-scale features into the final output feature map $\mathbf{F}_{\mathrm{out}}$. This approach enables MFF-Net to leverage the advantages of features at various scales, thereby enhancing its performance in tasks such as object detection and segmentation.

MFF-Net offers significant advantages for terrain detection, such as its ability to capture and integrate features at multiple scales, thus enhancing the ability of the modelto detect and distinguish between various terrain types. This multi-scale feature fusion leads to improved accuracy and robustness, particularly in complex environments with varying textures and patterns. However, the limitations include increased computational complexity and resource requirements owing to multiple feature extraction processes. Additionally, fine-tuning the network for optimal performance across different terrains can be challenging, and the model may struggle with real-time applications owing to its intricate architecture.

Table 7
Brief summary and comparative analysis of segmentation algorithms commonly used for AV.

Algorithm	Key process	Advantages	Limitations
PointFusion [97] (2018)	Combines RGB images and point cloud features for 3D bounding box estimation.	Enhanced accuracy through sensor data fusion.	Dependent on data quality and alignment from different sensors.
ShuffleNet [98] (2018)	Efficient CNN using pointwise group convolution and channel shuffle.	Highly efficient for mobile devices.	May not match the accuracy of complex models on high-resource platforms.
SpiderCNN [99] (2018)	Uses parameterized convolutional filters for point cloud data.	Better captures geometric structures of point clouds.	Increased computational demands owing to filter complexity.
BiSeNet [100] (2018)	Employs spatial and context paths for real-time semantic segmentation.	Fast and accurate segmentation for real-time applications.	Speed-accuracy trade-offs in complex scenes.
PSPNet [101] (2019)	Uses pyramid pooling for context aggregation at different scales.	Excellent at segmenting objects of varying sizes.	Increased computational load from multi-scale feature integration.
SalsaNext [102] (2020)	Uncertainty-aware semantic segmentation for LiDAR point clouds.	Incorporates uncertainty for reliability under varied conditions.	Requires calibration of uncertainty measures.
SETR-YOLOv5N [103] (2022)	Combines SETR and YOLOv5N for low-light lane detection using fractional-order fusion.	Robust in low-light conditions with advanced techniques.	Tailored to a specific application; broader use may need modifications.

3.4.9. Detr3D

Detr3D (DETR for 3D) [94] extends the concept of the original DETR (DEtection TRansformer) to the 3D domain. It formulates 3D object detection as a set prediction problem, directly predicting a fixed number of 3D bounding boxes and their class labels from the input. The core equation involves a bipartite matching loss and a set-based Hungarian loss to enforce a one-to-one correspondence between the predicted and ground truth boxes. The loss function $\mathcal{L}$ can be expressed as:

$$\mathcal{L} = \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{bbox}}\mathcal{L}_{\text{bbox}} + \lambda_{\text{giou}}\mathcal{L}_{\text{giou}} \tag{9}$$

where $\mathcal{L}_{\text{cls}}$ is the classification loss, $\mathcal{L}_{\text{bbox}}$ is the L1 loss for bounding box regression, and $\mathcal{L}_{\text{giou}}$ is the generalized IoU loss for the bounding box quality. The constants λ_{cls}, λ_{bbox}, and λ_{giou} are the weightsthat balance the contribution of each component.

Detr3D's end-to-end training approach simplifies the pipeline for terrain detection by removing the need for separate proposal and post-processing stages and effectively handling complex scenes and varied object sizes through its set-based prediction mechanism. However, it requires high computational resources, has longer training times, and may be inefficient for scenes with large variability in the number of objects. Additionally, it might struggle with small object detection owing to the coarse feature maps produced by transformers.

3.5. Segmentation algorithms

This section briefly overviews the key segmentation algorithms developed between 2018 and 2024 and highlights their principal processes, advantages, and limitations. Algorithms such as PointFusion [97] leverage data fusion from RGB images and point clouds to enhance 3D object detection accuracy, whereas models such as ShuffleNet prioritize the computational efficiency of mobile applications. SpiderCNN [99] and BiSeNet [100] introduced innovative convolutional approaches tailored to point-cloud geometries and real-time semantic segmentation, respectively. PSPNet [101] excels in handling objects of various sizes by aggregating contexts at different scales. SalsaNext [102] introduced uncertainty-aware segmentation for LiDAR point clouds to optimize the reliability under diverse conditions. Finally, SETR-YOLOv5N [103] focuses on low-light lane detection, showcasing the specialized application of segmentation algorithms in challenging visibility conditions and illustrating the breadth of advancements in segmentation techniques aimed at addressing specific challenges within the realm of computer vision. A brief summary and a comparative analysis of these algorithms are presented in Table 7. In addition, a comprehensive exploration of these algorithms is presented in this section.

3.5.1. PointFusion

PointFusion [97] is a method for 3D object detection that integrates both LiDAR point clouds and image data to leverage the strengths of both modalities. It first processes the LiDAR point clouds to extract a set of features F_p, and separately processes the images to extract a set of features F_i. These features were then fused to create a comprehensive feature representation F_f. The fusion process can be mathematically represented as

$$F_f = \text{Fusion}(F_p, F_i) \tag{10}$$

where Fusion denotes a function or operation that combines F_p and F_i to exploit the spatial and semantic information captured by point clouds and images respectively. The fused features F_f are then used by a detection network to predict the 3D bounding boxes of objects. The overall goal of PointFusion is to improve the accuracy and robustness of 3D object detection by effectively combining complementary information from LiDAR and camera sensors.

PointFusion enhances terrain segmentation by combining LiDAR point clouds and image data, thereby improving the spatial resolution and semantic richness for more accurate results. It leverages precise depth information from LiDAR and detailed textures from the images. However, it has limitations such as increased computational complexity, potential misalignment issues between modalities, and reduced performance in environments with sparse point clouds or poor image quality.

3.5.2. ShuffleNet

ShuffleNet [98] is a convolutional neural network architecture designed for efficient computation and reduced parameter count, making it suitable for mobile and embedded devices. It introduces the concept of channel shuffling to enable efficient information exchange between channels while reducing the computational complexity. The architecture utilizes grouped convolutions followed by channel shuffling operations to facilitate communication across feature maps, therebyenhancing the representation power without significantly increasing the computation. This design is expressed succinctly as:

$$\text{ShuffleNet} = \text{Grouped Convolution} + \text{Channel Shuffle} \tag{11}$$

where grouped convolution divides the input channels into groups and processes them separately, and the channel shuffle reorganizes the feature maps to enable cross-group information flow, thereby achieving a balance between the model complexity and computational efficiency.

ShuffleNet is well-suited for terrain segmentation because of its efficient design, which reduces computational demands and memory usage, making it ideal for real-time applications in devices such as drones. By facilitating effective feature interaction through channel shuffling, ShuffleNet maintains competitive performance, albeit with slightly lower accuracy compared to deeper networks such as ResNet or DenseNet, especially when dealing with intricate or varied terrain types. Therefore, while ShuffleNet offers efficiency advantages, its application in terrain segmentation should consider the trade-offs between the computational efficiency and segmentation precision required for specific tasks.

3.5.3. SpiderCNN

SpiderCNN [99] is a convolutional neural network architecture specifically designed for processing point cloud data, leveraging a novel spider convolution operation to capture both local and global geometric features. It introduces a new convolutional kernel that adapts to the non-uniform distribution of points in 3D space, enhancing the network's ability to learn from unordered point sets. The spider convolution operation is defined as

$$\mathrm{SpiderConv}(X) = \sum_{p \in X} \frac{1}{\|X\|} \phi(p) \cdot \psi(p) \tag{12}$$

where X represents the set of input points, p denotes individual points within X, $\|X\|$ is the cardinality of X, and $\phi(p)$ and $\psi(p)$ denote the learnable functions applied to each point p. This formulation allows SpiderCNN to effectively aggregate local and global information from unordered point clouds, making it suitable for tasks such as 3D shape recognition and segmentation.

SpiderCNN excels in terrain segmentation by efficiently processing unordered point cloud data and capturing both local and global geometric features through its unique spider convolution operation. This enables the effective differentiation of terrain features such as hills, valleys, and flat areas. However, it may encounter challenges with very large point clouds owing to computational constraints and may be sensitive to noise levels and data density, necessitating careful preprocessing and parameter tuning for optimal performance.

3.5.4. BiSeNet

BiSeNet [100], short for bilateral segmentation network, was designed to efficiently perform real-time semantic segmentation by leveraging both low-resolution and high-resolution features through a dual-pathway architecture. The core of BiSeNet uses a spatial path (SP) and a context path (CP). The spatial path captures detailed information at a lower resolution, whereas the context path aggregates global context information at a higher resolution. These paths were combined using a weighted fusion module to produce the final segmentation result. The fusion module calculates pixel-wise attention weights based on features from both paths, thereby enhancing the network's ability to handle diverse scales and maintain spatial details crucial for accurate segmentation tasks. The formulation can be succinctly represented as:

$$\mathrm{BiSeNet}(X) = \mathcal{F}_{\mathrm{fuse}}(\mathcal{F}_{\mathrm{sp}}(X), \mathcal{F}_{\mathrm{cp}}(X)) \tag{13}$$

where X represents the input image, $\mathcal{F}_{\mathrm{sp}}$ and $\mathcal{F}_{\mathrm{cp}}$ denote the spatial and context pathways respectively, and $\mathcal{F}_{\mathrm{fuse}}$ represents the fusion module combining the outputs from both paths to produce the final segmentation output. This design efficiently balances the computational cost and segmentation accuracy, making BiSeNet suitable for real-time applications the require robust semantic segmentation.

BiSeNet stands out in terrain segmentation because it effectively combines low-resolution spatial details with high-resolution contextual information, facilitating accurate segmentation at high processing speeds. The dual-pathway design ensures comprehensive feature extraction, which is essential for handling diverse terrain characteristics. However, the performance of BiSeNet may fluctuate based on the quality and diversity of the training data, as well as the challenges posed by complex textures and varying object sizes within terrain environments. Careful consideration of these factors is crucial to optimize their utility in real-world applications.

3.5.5. PSPNet

PSPNet [101] utilizes a pyramid pooling module to effectively capture multi-scale contextual information. The network aggregates the global context by employing pyramid pooling, which partitions the input the feature map into regions of different sizes and then pools features from each region separately. The final feature representation is obtained by concatenating the pooled features with the original feature map. The main equation for the pyramid pooling module in the PSPNet can be summarized as follows:

$$y_i = \mathrm{concat}([G_i(x); x]) \tag{14}$$

where x denotes the input feature map, $G_i(x)$ represents the feature map pooled at different grid sizes i, and y_i is the output feature map after concatenation. This approach enables PSPNet to effectively incorporate both local and global contexts for accurate scene parsing and segmentation.

PSPNet proves advantageous in terrain segmentation by leveraging its pyramid pooling module, which efficiently incorporates multi-scale contextual details for the precise classification of various terrain types. Nevertheless, its intensive computational and memory demands may hinder real-time implementation and deployment of less powerful hardware. Furthermore, the performance of PSPNet can be influenced by dataset quality and the inherent diversity of terrain features, necessitating thoughtful adaptation for robust performance in diverse real-world applications.

3.5.6. SalsaNext

SalsaNext [102] is a state-of-the-art neural network architecture designed for semantic segmentation of 3D point clouds, particularly in autonomous driving scenarios. It incorporates a PointNet-based encoder–decoder structure in which the input point cloud P is first processed by a series of PointNet layers to extract hierarchical features. These features are then decoded to produce per-point semantic labels S. The network utilizes feature propagation and feature expansion modules to enhance context and detail preservation across different scales. Output segmentation S is computed as follows:

$$S = \mathrm{SalsaNext}(P) \tag{15}$$

where SalsaNext represents the entire network's architecture. SalsaNext achieves high segmentation accuracy by effectively capturing both local and global spatial relationships within point cloud data, making it suitable for applications requiring precise scene understanding and object detection in complex environments.

SalsaNext stands out in terrain segmentation because of its ability to accurately capture intricate spatial relationships and contexts at different scales, leveraging its PointNet-based architecture. This makes it particularly effective for applications such as autonomous driving. Challenges include its sensitivity to point cloud quality and density, as well as the computational demands associated with its deep neural network design. Nonetheless, SalsaNext remains a robust and scalable solution to achieve precise terrain segmentation in diverse real-world environments.

3.6. Operational framework of ADAS

Fig. 4 presents an integrative overview of the autonomous vehicle operational framework, categorizing the system functionality into distinct but interconnected components: Sensing the Environment, Processing Steps, Components and Hardware, Integration in the Vehicle, and Edge Computing and Connectivity.

- **Sensing the Environment:** The first tier represents the sensory mechanisms by which the vehicle perceives its surroundings. Cameras can be used to detect signs, vehicles, and pedestrians. LiDAR creates a 3D map of an environment by measuring its distance and shape. Radar is employed to perceive speed and

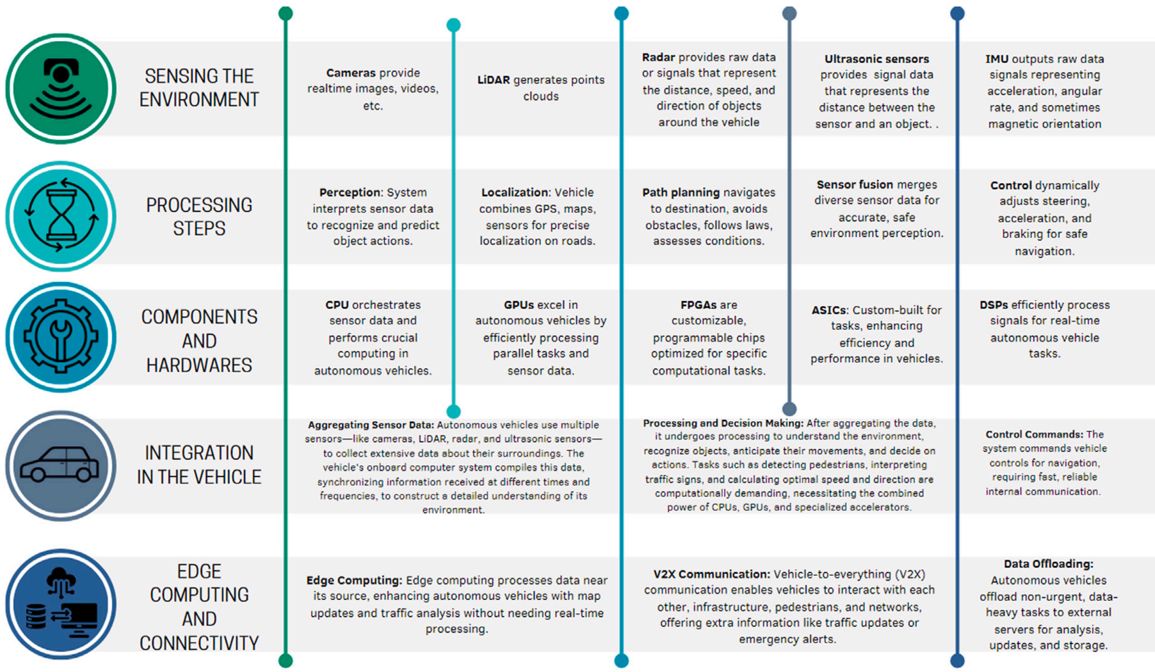

Fig. 4. Operational framework of an autonomous vehicle's driving system.

distance, and operate effectively with poor visibility. Ultrasonic sensors cater to near-range detection, aiding in maneuvers such as parking.

- **Processing Steps:** The second tier describes the processing algorithm. Perception systems analyze sensor data to recognize and predict objective trajectories. Localization combines the GPS and mapping data to achieve precise positioning. Path planning algorithms strategize a vehicle route, consider obstacles, and adhere to traffic rules. Sensor fusion has been highlighted as a method of creating a cohesive perception of the environment.
- **Components and hardwares:** Computational components are described in this section. The CPUs orchestrate sensor data and perform critical operational tasks. GPUs are acknowledged for their parallel processing abilities and their ability to enhance the efficiency of sensor data analysis. FPGAs have been noted for their customization potential in specific computational tasks. ASICs are integrated for their specialized efficiency, and DSPs efficiently process signals in real-time autonomous operations.
- **Integration in the Vehicle:** The framework underscores the crucial integration process within the vehicle, detailing the fusion of sensor data from diverse source cameras, LiDAR, radar, and ultrasonics. This is pivotal for generating a detailed and dynamic representation of a vehicle's surroundings, facilitating informed driving decisions, and directly influencing the vehicle's dynamic control systems. This ensures that the vehicle responds accurately to its environment by adjusting its movement and operation in real-time. Although this integration facilitates comprehensive environmental perception, it also presents significant challenges in synchronizing these inputs with vehicle control systems to ensuring seamless operation and responsive vehicle dynamics under various conditions.
- **Edge Computing and Connectivity:** The final tier addresses additional computing processes conducted near the vehicle, including traffic analysis and updates, without requiring real-time processing. It also mentions V2X communication which allows a vehicle to interact with other vehicles, infrastructure, and networks, enhancing situational awareness and providing updates or emergency alerts. Data offloading to external servers for non-immediate tasks such as deep learning analysis and software updates is depicted.

3.7. Applications

Terrain Detection and Segmentation have revolutionized autonomous vehicle navigation by integrating multiple sensor data sources to accurately perceive and interpret surrounding terrain. By combining inputs from LiDAR, radar, cameras, and other sensors, this technology enables vehicles to detect and classify various terrain types, such as roads, sidewalks, obstacles, and rough terrain, with unprecedented precision. This comprehensive understanding of the environment allows autonomous vehicles to make informed decisions in real-time, ensuring safe and efficient navigation under diverse and challenging conditions. This section discusses diverse terrain detection and segmentation in applications for autonomous vehicle navigation.

- **Safe Navigation:** Terrain Detection and Segmentation play pivotal roles in ensuring safe navigation for autonomous vehicles. By integrating data from various sensors such as LiDAR, radar, and cameras, FETDS provides a comprehensive understanding of the surrounding terrain, including obstacles, road conditions, and terrain topology [104]. This detailed analysis enables autonomous vehicles to make informed decisions in real-time, adjust their navigation strategies to avoid potential hazards and ensure smooth and safe travel. FETDS enhances a vehicle's perception capabilities, contributing to reliable and efficient autonomous navigation in diverse environments while prioritizing safety as a paramount concern.
- **Obstacle Avoidance:** Autonomous vehicles rely on advanced sensor technologies to navigate safely through various terrains and avoid obstacles. Data from multiple sensors such as cameras, LiDAR, and radar can be integrated through sophisticated techniques to comprehensively understand the surroundings of a vehicle comprehensively [105]. This terrain detection and segmentation process enables the vehicle to accurately identify obstacles, differentiate between different types of terrain, and plan appropriate navigation routes in real-time. By combining the

strengths of each sensor modality, autonomous vehicles can navigate challenging environments with greater precision and reliability, thereby ensuring safer journeys for both passengers and pedestrians.

- **Environmental Adaptation:** In autonomous vehicle navigation, advanced technologies such as Terrain Detection and Segmentation play a crucial role in environmental adaptation. By seamlessly integrating data from various sensors and sources, this technology enables vehicles to perceive and understand the surrounding terrain with unprecedented accuracy and detail. This comprehensive understanding empowers autonomous vehicles to make informed decisions in real-time, allowing them to navigate safely and efficiently across diverse landscapes, ranging from urban streets to off-road terrains [106]. With the ability to adapt to changing environmental conditions, vehicles equipped with this technology can optimize their routes, avoid obstacles, and ensure smooth and secure transportation of passengers and cargo.
- **Emergency Response:** In emergency response scenarios, autonomous vehicles rely on advanced technologies to navigate safely through challenging terrain and environments. By integrating sophisticated sensor fusion techniques, these vehicles can accurately detect and segment terrain features, enabling them to make real-time decisions and adapt to rapidly changing conditions [107]. This capability enhances the vehicle's ability to navigate through obstacles, debris, and other hazards, ensuring swift and efficient responses during critical situations such as natural disasters or accidents.
- **Off-Road Exploration:** Autonomous vehicles rely on sophisticated technologies to navigate challenging off-road terrains, which are crucial for exploration missions in remote areas. One such advancement involves a comprehensive system that integrates sensor data from various sources to accurately detect and segment terrain features. By fusing inputs from cameras, LiDAR, radar, and other sensors, this technology creates a detailed map of the environmentand identifies obstacles, rough terrain, and pathways [108]. This enables the vehicle's navigation system to make real-time decisions and adjust its speed and trajectory to traverse diverse landscapes safely. Ultimately, this approach enhances the vehicle's ability to explore off-road environments efficiently and autonomously, thereby opening up new possibilities for scientific research, resource exploration, and disaster response efforts.
- **Construction Site Navigation:** In-construction site navigation for autonomous vehicles. Advanced terrain detection and segmentation play a crucial roles in the navigation of construction sites for autonomous vehicles. By integrating cutting-edge technologies, vehicles can accurately identify and classify various terrain types, such as gravel, sand, and concrete, enabling them to make informed navigation decisions [109]. This capability enhances safety and efficiency by allowing vehicles to adapt their speed and trajectory based on the specific terrain encountered, thereby ensuring smooth and reliable movement in complex construction environments.
- **Agricultural Automation:** Advanced technologies such as Terrain Detection and Segmentation play a crucial role in agricultural automation by enabling autonomous vehicles to navigate effectively. By integrating data from various sensors, agricultural machinery can be empowered accurately interpret terrain features, identify obstacles, and optimize routes for tasks such as planting, spraying, and harvesting [110]. This streamlined approach enhances operational efficiency, boosts productivity, and reduces reliance on human intervention, thereby fostering sustainable and efficient farming practices.
- **Mining and Quarrying:** In mining and quarrying operations, autonomous vehicles rely on advanced technologies to navigate safely and efficiently through challenging terrain. By integrating cutting-edge sensor techniques, these vehicles can analyze diverse data streams from cameras, LiDAR, radar, and other sensors to perceive their surroundings accurately. This comprehensive understanding of the environment enables the accurate identification of obstacle terrain variations and navigational hazards accurately. Autonomous vehicles can plan optimal paths and make informed decisions to ensure smooth and reliable operations through real-time segmentation and analysis of terrain features, such as slopes, gradients, and surface textures. This application enhances safety by minimizing the risk of accidents and improves productivity by optimizing route planning and vehicle performance in rugged mining and quarrying environments.
- **Marine Exploration:** Advanced technologies are pivotal for the navigation of autonomous vehicles in the challenging underwater realms of marine exploration. Terrain detection and segmentation stand out for their ability to amalgamate data from diverse sensors and craft detailed ocean floor maps. By blending information from sonar, LiDAR, and other sources, this system empowers autonomous vehicles to discern and categorize various underwater features such as reefs, trenches, and structures with remarkable accuracy [111]. This capability ensures safe and efficient navigation, facilitating optimal path planning and obstacle avoidance during marine exploration. Moreover, by delivering real-time updates on the surrounding terrain, Terrain Detection and Segmentation elevates the situational awareness of autonomous vehicles, enabling them to swiftly adapt to dynamic underwater conditions and navigate with utmost precision.
- **Disaster Response and Humanitarian Aid:** In disaster response and humanitarian aid efforts, autonomous vehicles play a pivotal role in navigating challenging terrains to deliver aid efficiently. By leveraging advanced terrain detection and segmentation technology, these vehicles can accurately analyze the landscape and identify obstacles, debris, and safe pathways in real-time. This capability enables them to navigate through complex environments such as debris-strewn roads or uneven terrain safely and swiftly, facilitating the delivery of crucial supplies and aid to affected areas [112]. Moreover, by autonomously adapting to changing terrain conditions, these vehicles can optimize their routes, minimize risks, and enhance their overall operational effectiveness, thereby significantly aiding disaster response and humanitarian missions.

4. Current trends

The current trends in terrain detection and segmentation technologies for autonomous vehicles, as listed in Table 8, illustrate various methodologies, datasets, and challenges across the domain. In 2023, Müller et al. [113] utilized the AI4Mars dataset with DeepLabv3, achieving impressive IoUs across various terrains but highlighted the ongoing challenges of limited high-quality training data and the simulation-to-reality gap. Similarly, Panda et al. [114] applied multiple models like ViT-Adapter and HRNet on Agroscapes and Freiburg Forest datasets, showcasing high mIoU percentages, yet acknowledged the inaccuracies stemming from reliance on original record systems.

Firkat et al. [115] used FGSeg on the RELLIS-3D and SemanticKITTI datasets and demonstrated high precision and recall, particularly for horizontal and slope scenes. However, they noted that the sparsity of the point cloud data could overlook small ground obstacles. Steinke et al. [116] combined SemanticKITTI with Airborne Lidar Scanning data using GroundGrid, achieving high accuracy and IoU but facing difficulties with complex terrain variations. Ando et al.'s [117] work with RangeViT on nuScenes and SemanticKITTI datasets marked a significant advancement in 3D semantic segmentation, highlighting the need for enhanced tokenization methods for LiDAR data.

On the detection front, Zhang et al. [118] introduced HCRNS-CT, which reduces mapping time but faces challenges with map construction speed. Acun et al. [119] achieved high detection performance with

Table 8
Summary of the recent research in the AV domain.

Ref.	Year	Domain	Dataset	Model	Performance	Limitations
Müller et al. [113]	2023	Segmentation	AI4Mars	DeepLabv3	Deterministic model IoU: Soil: 93.52%, Bedrock: 80.93%, Sand: 86.96%, Big rock: 60.47%	Limited high-quality data for training, difficulty annotating terrain data, and the sim-to-real gap.
Panda et al. [114]	2023	Segmentation, Detection	Agroscapes, Freiburg Forest	ViT-Adapter, HRNet, ResNeSt, MobileNetV3	Semantic segmentation results (mIoU): ViT-Adapter: 96.43%, HRNet: 95.28%, ResNeSt: 95.34%, MobileNetV3: 94.57%	Faces inaccuracies in accidental data owing to the reliance on the accuracy of the original records system.
Firkat et al. [115]	2023	Segmentation	RELLIS-3D and SemanticKITTI	FGSeg	Horizontal scene: Precision: 96.76%, Recall: 93.35%, F1 score: 93.35%, Time: 51 ms Slope scene: Precision: 95.46%, Recall: 97.67%, F1 score: 96.54%, Time: 49 ms	The sparsity of point cloud data might result in small ground obstacles being overlooked, which could challenge the precise navigation of agricultural robots.
Steinke et al. [116]	2023	Segmentation	SemanticKITTI and Airborne Lidar Scanning (ALS)	GroundGrid	Precision: 96.99%, Recall: 97.65%, F1: 97.32%, Accuracy: 96.60%, IoU: 94.78%	Faces challenges in handling complex terrain variations and the need for further validation across diverse datasets and environmental conditions beyond SemanticKITTI and self-acquired data.
Ando et al. [117]	2023	Segmentation	nuScenes and SemanticKITTI	RangeViT	nuScenes: RangeViT-CS: mIoU: 75.2% SemanticKITTI: RangeViT-CS: mIoU: 64.00%	Need further enhancement in tokenization methods for LiDAR data and exploration of advanced techniques like FlexiViT or Perceiver IO, as well as consideration of tokenizing raw 3D data instead of 2D projections for improved 3D semantic segmentation in autonomous driving.
Zhang et al. [118]	2023	Detection	N/A	HCRNS-CT	HCRNS-CT reduces mapping time, enhancing efficiency with a 2D grid map in 3D terrain.	The construction of its proposed map takes longer than traditional maps, potentially impacting real-time mapping applications.
Acun et al. [119]	2023	Detection	CityScapes	D3NET	The detection performance of 92.55% is achieved at 237 fps on a Titan XP GPU and above 30 fps on a Jetson Nano module.	Potential loss of contextual information is observed when images are split into thin vertical strips, resulting in decreased visibility of objects such as trees.
Chase Jr et al. [120]	2023	Detection	SIM10k, CityScapes	YOCO	YOCO achieves a 3% increase in average precision at 0.5 IoU than YOLOv3.	Challenges may arise in extending the YOCO technique to diverse scenarios, and additional validation is required in real-world scenarios with varying terrains and lighting conditions.
Goh et al. [121]	2022	Segmentation	161 training images	–	Accuracy: 91.1%	Limited by a scarcity of labeled samples, hindering practical applicability, and lacks a discussion on the computational cost of contrastive pretraining, impacting scalability.
Sathyamoorthy et al. [122]	2022	Segmentation	RGB images of terrain surfaces	TerraPN	Success rates (up to 35.84% higher), Vibration cost of the trajectories (up to 21.52% lower), Slowing the robot on bumpy, deformable surfaces (up to 46.76% slower)	Unable to estimate navigability costs on completely non-traversable surfaces, susceptibility to adverse effects from significant lighting changes, and challenges in distinguishing subtle elevation changes owing to the absence of depth and elevation inputs.

(continued on next page)

Table 8 (*continued*).

Zürn et al. [123]	2020	Segmentation, Classification	Self-made dataset	SE, SE-R	Accuracy: SE: 94.07%, SE-R: 94.81%	Limitations include incomplete segmentation masks owing to the robot's limited traversal, challenges in handling class imbalance, and a need for more extensive generalization analysis across diverse terrains.
Zhang et al. [124]	2020	Segmentation	Cloud data	Decision tree, Maximum likelihood, SVM	Accuracy: 92.8%	Face control signal lag owing to long distances and obstruction by buildings and terrain, hindering real-time requirements for autonomous movement.
Mehrabi et al. [125]	2021	Segmentation	LiDAR Three-dimensional point cloud data	Gaussian Process	Average success rate: 93.5%	Dependent on a specifically labeled dataset, potential sensitivity to parameters, lack of detailed discussion on computational complexity, and limited insight into the algorithm's behavior with untruncated data.
Shamsolmoali et al. [126]	2020	Segmentation	Massachusetts	Adversarial Spatial Pyramid Network	Achieved IOU of 78.86% with 14.89M parameters and 86.78B FLOPs	Does not address the scalability of the proposed model across diverse datasets nor the impact of increasing computational costs with enhanced model complexity.
Zhang et al. [127]	2021	Segmentation	CasNet	RoadDA	IoU: 74.92%, F1 score: 85.81%	The limitation of GPU resources prevents direct training of the model using high-resolution remote sensing images.
Chang et al. [128]	2022	Segmentation	KITTI road	Uncertainty aware Symmetric Network	Real-time inference speed of 43+ FPS	Does not explore the network's performance under varying environmental conditions or address potential limitations in generalizing across diverse and more challenging datasets.
Milli et al. [129]	2023	Segmentation	KITTI road	(3MT) Road Segmentation	3MT-RoadSeg achieves a higher MaxF score than all RGB input methods and all RGB+LiDAR methods.	Does not address the potential challenges in data synchronization and calibration between LiDAR and camera sensors, nor does it evaluate the robustness of the method across varied environmental and lighting conditions.

D3NET on CityScapes at varying frame rates and observed a potential loss of contextual information. Chase Jr et al. [120] improved the precision in the SIM10k and CityScapes datasets with YOCO, emphasizing the need for further validation across different planetary environments.

These studies highlight rapid advancements in segmentation and detection technologies, employing novel datasets and models to achieve remarkable performance metrics. However, they also highlight significant limitations, such as data quality, algorithmic adaptability to diverse environments, and the need for computational efficiency. These insights underscore the progress in terrain detection and segmentation and map the challenges and opportunities for future research in enhancing autonomous vehicle navigation and safety.

5. Applications and case studies

The applications of terrain detection and segmentation are vast and varied, encompassing both structured and unstructured environments. This section addresses these applications by providing detailed case studies and examples from both the domains.

5.1. Structured environments

Structured environments, such as urban roads and highways, present unique challenges and opportunities for terrain detection and segmentation. In these environments, the primary focus is on accurately detecting and segmenting various types of terrain to enhance navigation safety and efficiency.

5.1.1. Urban roads

Urban roads are characterized by a complex network of lanes, intersections, pedestrian crossings, and various road signs. Terrain detection and segmentation in urban settings are crucial for developing advanced driver-assistance systems (ADAS) and autonomous vehicles. Technologies such as LIDAR, computer vision, and deep learning have been extensively used to segment road surfaces, detect obstacles, and recognize traffic signs. For instance, Horváth et al. [130] presented a method for detecting roads and sidewalks using LIDAR technology in urban environments. The proposed real-time solution (operating at 20 Hz) addresses challenges such as road imperfections and high data bandwidth by utilizing three parallel algorithms for sidewalk edge detection. This method ensures robustness and parameter insensitivity, making it suitable for local path planning in autonomous vehicles. The authors validate their approach using the KITTI dataset and made the source code and data publicly available for reproducibility and further research. In addition, Dong et al. [131] presented a novel method that leveraging Deep Convolutional Neural Networks (DCNNs) to achieve a balance between high accuracy and real-time processing for the semantic segmentation of urban street scenes. The proposed method introduces a Lightweight Baseline Network with Atrous convolution and Attention (LBN-AA) to efficiently generate dense feature maps. It incorporates Distinctive Atrous Spatial Pyramid Pooling (DASPP) for multi-scale object detection and a Spatial Detail-Preserving Network (SPN) to maintain high-resolution spatial information. Finally, an (FFN) combines deep semantic and shallow spatial features. Extensive experiments on Cityscapes and CamVid datasets demonstrate that the method achieves high accuracy (73.6% mIoU on Cityscapes) at impressive speeds (51 fps), making it suitable for real-time applications in intelligent transportation systems.

5.1.2. Highways

Although highways are, more straightforward than urban roads, still require precise terrain detection and segmentation for effective navigation. Key applications include lane detection, road boundary identification, and hazard detection. Kalpoma et al. [132] presented an automated system to measure road quality using satellite images, addressing the inefficiencies of manual inspection methods. The researchers developed a dataset using high-resolution images from Google Earth Pro, focusing on various national highways in Bangladesh. They employed nine deep learning models – Deeplabv3, DeeplabV3plus, MAnet, LinkNet, Unet, FPN, Unetplusplus, PAN, and PSPNet – to segment roads from these images. The DeeplabV3plus model outperformed the others, achieving over 97% accuracy across different training–validation ratios. Liu et al. [133] proposed a comprehensive framework to address the challenges of detecting and evaluating highway traffic congestion using advanced deep learning technologies. Traditional methods for traffic congestion detection are often slow, labor-intensive, and depend on extensive monitoring equipment. The proposed framework leverages deep learning to accurately detect, evaluate, and predict traffic congestion, thereby aiding traffic management in formulating effective strategies for mitigating congestion. The study introduced a self-coding deep learning model to classify traffic data and employed a SoftMax-based prediction model to forecast congestion with high accuracy. The research was validated using data from the Shanghai expressway network, demonstrating a detection accuracy of 98.6% and a prediction accuracy of 92% during peak hours.

5.2. Unstructured environments

Unstructured environments, including rural paths and off-road terrains, pose various challenges owing to their irregular and unpredictable nature. Terrain detection and segmentation in these settings are essential for applications in agriculture, military operations, and adventure sports.

5.2.1. Rural paths

Rural paths often lack clear markings and can vary significantly in terms of the surface type and condition. Terrain detection and segmentation can help in developing robust navigation systems for agricultural machinery, ensuring efficient and precise operations. For example, Barba-Guaman et al. [134] explored the use of an NVIDIA's Jetson Nano for object detection in complex rural environments. It evaluates various deep learning models, including ssd-mobilenet v1 and v2, ssd-inception v2, pednet, and multiped, to determine their accuracy and processing efficiency in this low-power embedded system. The study an that while pednet excels in pedestrian detection, ssd-mobilenet v2 and ssd-inception v2 are more effective for vehicle detection. Overall, Jetson Nano is capable of running multiple neural networks simultaneously, making it a viable option for real-time object detection in resource-constrained settings. The results highlight the potential for deploying advanced computer vision applications in rural areas, emphasizing the need for optimization to balance the accuracy and processing speed. Yang et al. [135] proposed a novel method for extracting road networks from high-resolution remote sensing images in rural areas. The method utilizes an Ensemble Wasserstein Generative Adversarial Network with Gradient Penalty (E-WGAN-GP), integrating U-Net and BiSeNet as generators to address class imbalance and enhance the robustness of road extraction. By incorporating a spatial penalty term in the loss function and leveraging an ensemble strategy, the E-WGAN-GP model significantly improves road extraction performance. Experimental results on the GaoFen-2 and DeepGlobe datasets demonstrate that the proposed method achieves superior precision and recall, outperforming existing state-of-the-art methods with an F1-score of 0.85 and IoU of 0.73.

5.2.2. Off-road terrains

Off-road terrain present the most significant challenge for terrain detection owing to its unpredictable nature, including rocks, mud, and water bodies. Applications in this domain include military reconnaissance, search and rescue operations, and recreational activities such as off-road racing. Advanced sensors and machine learning algorithms were employed to analyze and segment these challenging terrains. For instance, Shon et al. [136] present a neural network model using Transformer architecture to evaluate the drivability of off-road surfaces based solely on Controller Area Network (CAN)-bus signals from vehicles. The network categorizes the terrain into three types: sand, mud, and pebble, and assesses drivability as either drivable or challenging. By leveraging CAN-bus signals, the model circumvents the need for external sensors, enabling real-time, accurate terrain condition detection. To enhance reliability, a post-processing algorithm using an exponential moving average (EMA) was implemented, to filter out erroneous estimates and minimize frequent changes in detection. The study demonstrates high accuracy and robustness in various off-road environments, highlighting the practical applicability of the proposed method for improving safety and performance in off-road driving scenarios. The study by Fritz et al. [137] focuses on off-road terrain classification to improve Advanced Driver Assist Systems (ADAS) in environments typical of low- and middle-income countries, where poorly maintained roads often challenge current ADAS technologies. Using convolutional neural networks, model were developed to classify terrains based on road roughness according to the ISO8608:2016 standard. The team created an image database from forward- and downward-facing cameras capturing various terrains and trained two classification models: one developed from scratch and another utilizing transfer learning on a pre-trained model. The results demonstrated high accuracy in terrain classification, indicating potential improvements in vehicle safety and performance under off-road conditions. The study suggests further research with larger datasets under diverse conditions to enhance the robustness of these models.

6. Challenges and future directions

This section outlines the challenges of Terrain Detection and Segmentation for Autonomous Vehicle Navigation and discusses potential future work.

6.1. Challenges

Adaptive Navigation in Unstructured Terrain: Navigating unstructured environments presents distinct challenges compared with structured settings, primarily because the lack of clear path definitions and unpredictable elements. In unstructured environments such as rural paths and off-road terrains, autonomous vehicles must contend with irregular terrain features, varying surface textures, and potential obstacles such as rocks, mud, and vegetation, which are not typically found on urban roads and highways. These conditions require robust detection and segmentation technologies capable of dynamic adaptation. The absence of standard traffic signals and road markings further complicates the navigation process, demanding advanced sensing and machine-learning algorithms to interpret complex scenes and make real-time driving decisions. This stark contrast highlights the significant technological advancements required to enhance autonomy in unstructured settings, focusing on improving sensor accuracy, data fusion capabilities, and machine learning models to effectively handle environmental variability and uncertainty.

Variability in Terrain Types: The diversity of terrain types poses a significant challenge for robotic navigation systems, particularly for terrain recognition and segmentation [138]. This diversity includes landscapes such as flat surfaces, rocky terrain, slopes, forests, water surfaces, and urban environments. Accurately identifying and categorizing these terrain types in real-time is a complex task that requires robust sensor fusion techniques that combine data from LiDAR, cameras, radar, and IMUs. The dynamic nature of outdoor environments introduces uncertainties, including changing weather conditions and human-induced surface variations, which further complicates accurate terrain determination. In addition, the transition between different terrain types can be abrupt or gradual, making it challenging to delineate the boundaries accurately. This requires the development of sophisticated algorithms to detect subtle changes in terrain characteristics. In addition, topographic variability often leads to ambiguity and misclassification errors, requiring advanced machine learning and pattern recognition techniques to improve the classification accuracy and adaptability.

Complexities and Challenges in Sensor Fusion Integration: Sensor fusion in terrain detection and segmentation for robotic navigation involves integrating data from multiple sensors to understand the environment comprehensively. Sensor fusion for terrain recognition and segmentation for robot navigation faces several challenges, including integrating heterogeneous sensor data, including LiDAR, cameras, radar, and IMUs, to account for different resolutions, accuracies, and update rates which is a complex task [139]. In addition, ensuring accurate data synchronization and sensor calibration is essential to avoid inconsistencies and errors, and managing noise, measurement errors, and uncertainties in the environment is critical for reliable understanding of the terrain. Adapting fusion algorithms to deal with dynamic terrain features and updating terrain models in real time are additional challenges, especially given the computational complexity and need for efficient real-time processing. Moreover, the seamless integration of fused terrain data and navigation algorithms is crucial for efficient route planning, obstacle avoidance, and localization, which requires compatibility and consistency between the two components.

Dynamic Environments: In robotic navigation, terrain recognition and segmentation in dynamic environments with changing conditions such as moving objects, changing lighting, and unpredictable terrain changes can be challenging. Key hurdles include accurate object detection and tracking amidst clutter, adaptation to environmental variability, integration of data from different sensors amidst noise and occlusions, providing real-time processing for timely decision-making, semantic understanding of terrain types for efficient navigation, maintaining adaptability and robustness in the face of uncertainties, and prioritizing safety through collision-avoidance mechanisms and emergency protocols.

Obstacle Differentiation: Distinguishing between obstacles, recognizing terrain, and segmenting it presents many challenges for robotic navigation [140]. Relying on sensors such as lidar, radar, and cameras introduces limitations in the resolution, range, and field of view, making it difficult to accurately distinguish between obstacles and terrain features. In addition, navigation through complex terrain with different obstacles, such as rocks, vegetation, and bodies of water, is difficult, which is further complicated by the similarity of thesensor data characteristics. Dynamic environments further complicate the situation, because obstacles are not static and are prone to movement or change, thus requiring adaptive navigation strategies. Ambiguity in sensor data owing to noise or environmental factors, such as shadows and reflections, adds another layer of complexity, potentially leading to navigation errors. Adverse weather and changing light conditions can degrade sensor performance, thereby challenging the robustness of the system. In addition, the real-time processing requirements for timely decision-making can strain computing resources, especially in complex environments, and introduce latency, which is undesirable in safety-critical applications.

Low Visibility Conditions: Poor visibility conditions, such as fog, heavy rain, snowfall, or darkness, pose a significant challenge for terrain recognition and segmentation required for robotic navigation. These conditions limit the effective range of sensors, such as LiDAR, radar, and cameras, leading to incomplete or inaccurate data collection while causing interference and distortion in sensor readings [141]. The reduced contrast and blurred landmarks make it difficult for algorithms to accurately detect features, thereby increasing the risk of navigation errors and collisions. Dynamic environmental changes, such as puddles, flooding during heavy rainfall, or altered surfaces owing to snowfall, further complicate the navigation of robots that rely on static maps or models. Navigational uncertainty is increased because robots have difficulty locating and mapping their environment accurately without reliable environmental cues, posing a safety risk in dynamic environments, such as roads or construction sites.

Edge Cases and Uncommon Terrain: Edge Cases and Uncommon Terrain refer to atypical and infrequent environmental conditions that pose challenges for robotic navigation owing to their unpredictability and limited representation in training data. Terrain recognition and segmentation in robotic navigation face significant challenges owing to the diverse and unpredictable nature of real-world environments. Edge cases include a broad spectrum of unusual terrain types, from loose gravel to marshy areas and cliffs, which are often underrepresented in training data, hindering the ability of traditional algorithms to classify and navigate accurately. Machine learning models trained on common terrain types have difficulty generalizing to extreme cases, leading to navigation errors or failures. In addition, edge cases often involve terrain features that are difficult to detect with standard sensors, such as LiDAR, or cameras, such as transparent or highly reflective surfaces, such as ice or glass. Adaptability is critical, because robots must adjust their navigation strategy in real time to navigate edge situations, thereby requiring saferobust decision-making algorithms. Safety concerns are manifold, as navigating edge situations poses risks to the robot and its environment, especially in unstable or dangerous terrains. Furthermore, in remote or less-explored regions, data on edge situations and unusual terrain conditions are rare or unavailable, posing a challenge for the practical training of the models.

Computational Efficiency: Terrain recognition and segmentation in robot navigation are significant computational efficiency challenges, such as managing large amounts of sensor data from various sources such as LiDAR, cameras, and depth sensors in real-time while coping with the complexity of terrains ranging from flat surfaces to rugged landscapes with obstacles such as rocks and vegetation. Algorithms must accurately segment diverse terrain while ensuring real-time operation, which is essential for timely decision-making in dynamic environments. This must be achieved within limited computational resources, balancing the need for accuracy with available processing power and memory. In addition, energy-efficient algorithms are essential for extending the operational lifetimes of battery-powered robots. Sensor fusion further complicates this situation, requiring efficient algorithms to integrate data from multiple sensors without excessive computational overheads.

Semantic Understanding: Semantic Understanding in terrain detection and segmentation in robotic navigation refers to the ability of robots to interpret and categorize terrain types based on their visual or sensory characteristics for effective navigation. Semantic understanding of terrain recognition and segmentation for robotic navigation poses several challenges [142]. The complexity of natural terrain, which includes different types of grass, gravel, rocks, and water, requires specific detection and segmentation algorithms. Variability in terrain appearance owing to changes in lighting, weather, and seasonal variations leads to ambiguity, making accurate classification difficult. In addition, when relying on sensors, such as cameras and LiDAR, limitations in resolution, range, and accuracy must be considered, affecting semantic understanding. The mapping sensor data to specific terrain classes in noisy or dynamic environments requires sophisticated algorithms. Real-time decision-making is crucial because it requires efficient sensor data processing while ensuring accuracy. Adaptation of field models to new environments is essential for generalization. Integrating terrain recognition and segmentation with navigation tasks, such as route planning, requires seamless coordination for a safe and efficient passage.

Semantic Segmentation in 3D: Semantic segmentation in 3D terrain detection and segmentation for robotic navigation involves categorizing each point in a three-dimensional environment to understand its semantic meaning, which is crucial for safe and accurate navigation. Field data captured in 3D can be vast and complex, requiring significant computational resources for processing and analysis; however, the efficient management of these data remains challenging. Unbalanced distributions between terrain classes such as 'soil', 'vegetation', and 'rocks' can lead to biased models and misclassification, especially for minority classes. Natural environments exhibit a variety of terrain types, from flat surfaces to rough terrains, requiring the effective generalization of models for reliable navigation. Real-time adaptation to dynamic environmental changes such as weather or erosion requires robust and flexible segmentation algorithms. In addition, sensor noise and occlusions caused by obstacles such as vegetation are challenging, requiring efficient data management strategies for accurate segmentation. Achieving an advanced semantic understanding to distinguish different terrain features, such as grass, shrubs, and trees, is essential yet challenging. Moreover, computational efficiency is critical for real-time navigation, which requires algorithms that balance thecomplexity and performance to maintain high accuracy.

6.2. Future directions

Exploration of Multimodal Fusion: Multimodal fusion in terrain detection is a promising area of research with the potential to significantly improve the accuracy and robustness of autonomous vehicle navigation systems. Investigating the seamless integration of data from different sensors, including LiDAR, radar, and cameras, can provide a more comprehensive understanding of the environment. By leveraging the strengths of each sensor modality, researchers can compensate for the limitations of individual sensors, such as LiDAR's ability to provide detailed geometric information, radar's ability to penetrate adverse weather conditions, and cameras ability to capture rich visual information. Research opportunities in this area include to optimizing fusion algorithms, determining the most effective weightings for each sensor input, and developing sophisticated models that can dynamically adapt to different environmental conditions. Furthermore, it is crucial to explore how these fused modalities can improve perception in complex scenarios, such as urban environments or rugged terrain. Ultimately, multimodal fusion research has the potential to revolutionize terrain detection and provide autonomous vehicles with a more robust and adaptive perception system, thereby improving overall safety and reliability in real-world driving scenarios.

Real-time Dynamic Terrain Modeling: Exploring dynamic terrain modeling in real time opens exciting research opportunities with profound implications for autonomous vehicles. Building on existing terrain detection systems, it is necessary to develop responsive models that can adapt instantly to environmental changes. This research could include the development of algorithms that can process sensor data in real-time to dynamically update the terrain model and ensure that an autonomous vehicle receives accurate and updated information for effective navigation. In addition, researchers can explore the integration of machine learning techniques to improve the predictive capabilities of these models and enable them to anticipate changes in terrain conditions based on historical data and patterns. Investigating the optimal balance between the model complexity and computational efficiency is crucial for ensuring real-time applicability. In addition, including contextual information such as weather conditions and road infrastructure can contribute to a more holistic understanding of terrain dynamics. Developing such real-time dynamic terrain models is an important research direction that will provide autonomous vehicles with the flexibility to navigate rapidly changing environments while ensuring robust hazard avoidance.

Continual Learning for Adaptability: Continuous learning is a promising method for improving the adaptability and performance of terrain detection systems over time. Research opportunities in this area could focus on developing and implementing continuous learning algorithms that allow terrain detection models to adapt dynamically to changing environmental conditions, terrain, and infrastructure. Investigating methods to facilitate incremental learning,in which the system can continuously assimilate knowledge from new data without catastrophic forgetting. This includes exploring meta-learning, online learning, and transfer-learning techniques to utilize prior knowledge while effectively incorporating new terrain features. In addition, this research area can develop strategies for autonomous systems that self-monitor and evaluate their performance and trigger updates or retraining processes in response to performance degradation or environmental changes. In addition, it is essential to understand ethical implications and ensure robust safety mechanisms during continuous learning processes. By addressing the area of constant learning, researchers can contribute to the development adaptive terrain recognition systems that can maintain optimal performance under different and changing scenarios.

Human–robot collaboration in navigation: Human–robot collaboration in the context of terrain detection systems is an exciting topic for research in autonomous vehicle technology. Investigating how these systems can improve the interaction and communication between autonomous vehicles and human drivers or pedestrians is critical to ensuring safe and efficient navigation in shared spaces. This research area could include the development of intuitive and informative communication interfaces that convey the intentions and actions of autonomous vehicles to their human counterparts. This could include the development of visual cues, acoustic signals, or haptic feedback mechanisms that enable effective communication of vehicle intentions and promote a shared understanding of the environment. In addition, research on integrating predictive models and decision algorithms that consider human behavior patterns can contribute to predictive and cooperative navigation strategies. Research efforts can also focus on user-centered design principles to ensure that collaborative interfaces are user-friendly, minimize ambiguity, and increase trust in autonomous systems. This research can pave the way for safer and more harmonious interactions in mixed-use transportation scenarios by addressing the complex dynamics of human–robot collaboration.

Explainability and Trustworthiness: The Explainability and trustworthiness of terrain detection models are critical for promoting public acceptance and ensuring the safe use of autonomous vehicles. Future research in this area may investigate methods for improving the interpretability of complex machine learning models, particularly the deep learning architectures used in terrain detection systems. Research on techniques for generating human-understandable explanations for decisions made by these models is essential. This includes developing interpretable feature representations, exploring attentional mechanisms, and incorporating causal inference approaches to clarify the factors that influence a model's predictions. In addition, researchers could conduct user-centered studies to understand the specific requirements of confidence-building explanations, considering different stakeholders such as drivers, pedestrians, and regulators. Developing standardized frameworks and metrics to assess the explainability of terrain detection models could also be an essential contribution. In addition, exploring methods for incorporating user feedback into the model refinement process thus creating a feedback loop that increases trustworthiness is a relevant research direction. By addressing these aspects, researchers can play a crucial role in developing effective terrain detection systems and building trust in the decision-making processes.

7. Conclusions

This survey has meticulously charted the advancements and ongoing challenges in terrain detection and segmentation technologies crucial for autonomous driving, revealing an impressive trajectory from traditional methods to sophisticated machine learning and AI models.

Integrating these technologies into autonomous vehicle systems underscores a complex, multidisciplinary effort that combines sensor data fusion, advanced algorithmic interpretation, and real-time decision-making to enhance navigation in complex environments. Despite significant progress, the field faces challenges, such as strengthening algorithmic robustness across diverse conditions, improving real-time processing efficiencies, and ensuring system safety and reliability. Future research is needed to tackle these challenges by advancing algorithmic sophistication, refining sensors and data processing technologies, and exploring new learning paradigms to reduce reliance on extensively labeled datasets. The journey ahead for terrain detection and segmentation technologies is not only about refining autonomous vehicles' operational capabilities but also about contributing to safer, more efficient, and transformative transportation solutions, emphasizing the collective endeavor of the research and engineering communities to push the boundaries of what is possible in autonomous driving.

CRediT authorship contribution statement

Md Mohsin Kabir: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Jamin Rahman Jim:** Writing – original draft, Formal analysis, Data curation. **Zoltán Istenes:** Writing – review & editing, Supervision, Investigation.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgment

The authors thank Intelligent Field Robotic Systems and Eötvös Loránd University, Hungary, for their invaluable research support and supervision.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.inffus.2024.102644.

References

[1] Y. Zhang, M. Kamargianni, A review on the factors influencing the adoption of new mobility technologies and services: autonomous vehicle, drone, micromobility and mobility as a service, Transp. Rev. 43 (3) (2023) 407–429.

[2] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, K. Dietmayer, Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges, IEEE Trans. Intell. Transp. Syst. 22 (3) (2020) 1341–1360.

[3] Y. Mo, Y. Wu, X. Yang, F. Liu, Y. Liao, Review the state-of-the-art technologies of semantic segmentation based on deep learning, Neurocomputing 493 (2022) 626–646.

[4] B. Baheti, S. Innani, S. Gajre, S. Talbar, Eff-unet: A novel architecture for semantic segmentation in unstructured environment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 358–359.

[5] S. Hakak, T.R. Gadekallu, P.K.R. Maddikunta, S.P. Ramu, M. Parimala, C. De Alwis, M. Liyanage, Autonomous vehicles in 5G and beyond: A survey, Veh. Commun. 39 (2023) 100551.

[6] L. Chen, S. Teng, B. Li, X. Na, Y. Li, Z. Li, J. Wang, D. Cao, N. Zheng, F.-Y. Wang, Milestones in autonomous driving and intelligent vehicles—Part II: Perception and planning, IEEE Trans. Syst. Man Cybern. Syst. (2023).

[7] A. Mohan, A.K. Singh, B. Kumar, R. Dwivedi, Review on remote sensing methods for landslide detection using machine and deep learning, Trans. Emerg. Telecommun. Technol. 32 (7) (2021) e3998.

[8] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, A. Alamri, Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review, Remote Sens. 12 (9) (2020) 1444.

[9] W. Alkhadour, J. Zraqou, A. Al-Helali, S. Al-Ghananeem, Traffic accidents detection using geographic information systems (GIS), Int. J. Adv. Comput. Sci. Appl. 12 (4) (2021).

[10] M.H. Faykus, B. Selee, M. Smith, Utilizing Neural Networks for Semantic Segmentation on RGB/LiDAR Fused Data for Off-road Autonomous Military Vehicle Perception, Tech. Rep., SAE Technical Paper, 2023.

[11] Z. Pan, S.L. Lau, X. Yang, N. Guo, X. Wang, Automatic pavement crack segmentation using a generative adversarial network (GAN)-based convolutional neural network, Results Eng. 19 (2023) 101267.

[12] C. Wu, X. Bi, J. Pfrommer, A. Cebulla, S. Mangold, J. Beyerer, Sim2real transfer learning for point cloud segmentation: An industrial application case on autonomous disassembly, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 4531–4540.

[13] A. Gupta, A. Anpalagan, L. Guan, A.S. Khwaja, Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues, Array 10 (2021) 100057.

[14] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, W. Shi, Computing systems for autonomous driving: State of the art and challenges, IEEE Internet Things J. 8 (8) (2020) 6469–6486.

[15] M.H. Nampoothiri, B. Vinayakumar, Y. Sunny, R. Antony, Recent developments in terrain identification, classification, parameter estimation for the navigation of autonomous robots, SN Appl. Sci. 3 (2021) 1–14.

[16] K. Muhammad, T. Hussain, H. Ullah, J. Del Ser, M. Rezaei, N. Kumar, M. Hijji, P. Bellavista, V.H.C. de Albuquerque, Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks, IEEE Trans. Intell. Transp. Syst. (2022).

[17] X. Ma, W. Ouyang, A. Simonelli, E. Ricci, 3D object detection from images for autonomous driving: a survey, IEEE Trans. Pattern Anal. Mach. Intell. (2023).

[18] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, Int. J. Surg. 88 (2021) 105906.

[19] K. Li, K. Chen, H. Wang, L. Hong, C. Ye, J. Han, Y. Chen, W. Zhang, C. Xu, D.-Y. Yeung, et al., Coda: A real-world road corner case dataset for object detection in autonomous driving, in: European Conference on Computer Vision, Springer, 2022, pp. 406–423.

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

[21] J. Carpentier, G. Saurel, G. Buondonno, J. Mirabel, F. Lamiraux, O. Stasse, N. Mansard, The pinocchio C++ library – A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives, in: SII 2019 - International Symposium on System Integrations, Paris, France, 2019, p. 1, URL https://hal.laas.fr/hal-01866228.

[22] M. Wigness, S. Eum, J.G. Rogers, D. Han, H. Kwon, A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2019, pp. 5000–5007.

[23] A.H. Al-Dabbagh, R. Ronsse, 3D-point cloud dataset of various geometrical terrains in urban environments recorded during human locomotion, Open Data @ UCLouvain (2021) http://dx.doi.org/10.14428/DVN/S0LSIW.

[24] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, S. Savarese, Segcloud: Semantic segmentation of 3d point clouds, in: 2017 International Conference on 3D Vision, 3DV, IEEE, 2017, pp. 537–547.

[25] S. Xie, J. Gu, D. Guo, C.R. Qi, L. Guibas, O. Litany, Pointcontrast: Unsupervised pre-training for 3d point cloud understanding, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, Springer, 2020, pp. 574–591.

[26] P. Morasso, V. Mohan, Pinocchio: A language for action representation, Cogn. Robot. 2 (2022) 119–131.

[27] L. Montaut, Q. Le Lidec, A. Bambade, V. Petrik, J. Sivic, J. Carpentier, Differentiable collision detection: a randomized smoothing approach, in: 2023 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2023, pp. 3240–3246.

[28] P. Jiang, P. Osteen, M. Wigness, S. Saripalli, Rellis-3d dataset: Data, benchmarks and analysis, in: 2021 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2021, pp. 1110–1116.

[29] P. Ni, X. Li, W. Xu, D. Kong, Y. Hu, K. Wei, Robust 3D semantic segmentation based on multi-phase multi-modal fusion for intelligent vehicles, IEEE Trans. Intell. Veh. (2023).

[30] G. Chustz, S. Saripalli, Rooad: Rellis off-road odometry analysis dataset, in: 2022 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2022, pp. 1504–1510.

[31] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, A. Torralba, Scene parsing through ade20k dataset, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 633–641.

[32] B. Zhou, H. Zhao, X. Puig, T. Xiao, S. Fidler, A. Barriuso, A. Torralba, Semantic understanding of scenes through the ade20k dataset, Int. J. Comput. Vis. 127 (2019) 302–321.

[33] J. Uijlings, T. Mensink, V. Ferrari, The missing link: Finding label relations across datasets, in: European Conference on Computer Vision, Springer, 2022, pp. 540–556.

[34] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer, 2014, pp. 740–755.

[35] S. Rostianingsih, A. Setiawan, C.I. Halim, COCO (creating common object in context) dataset for chemistry apparatus, Procedia Comput. Sci. 171 (2020) 2445–2452.

[36] K. Gauen, R. Dailey, J. Laiman, Y. Zi, N. Asokan, Y.-H. Lu, G.K. Thiruvathukal, M.-L. Shyu, S.-C. Chen, Comparison of visual datasets for machine learning, in: 2017 IEEE International Conference on Information Reuse and Integration, IRI, IEEE, 2017, pp. 346–355.

[37] M. Rottmann, M. Reese, Automated detection of label errors in semantic segmentation datasets via deep learning and uncertainty quantification, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 3214–3223.

[38] J. Breitenstein, T. Fingscheidt, Amodal cityscapes: a new dataset, its generation, and an amodal semantic segmentation challenge baseline, in: 2022 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2022, pp. 1018–1025.

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

[40] Y. Pan, B. Gao, J. Mei, S. Geng, C. Li, H. Zhao, Semanticposs: A point cloud dataset with large quantity of dynamic instances, in: 2020 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2020, pp. 687–693.

[41] Y. Liao, J. Xie, A. Geiger, KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d, IEEE Trans. Pattern Anal. Mach. Intell. 45 (3) (2022) 3292–3310.

[42] R. Krajewski, J. Bock, L. Kloeker, L. Eckstein, The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems, in: 2018 21st International Conference on Intelligent Transportation Systems, ITSC, IEEE, 2018, pp. 2118–2125.

[43] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, L. Eckstein, The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections, in: 2020 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2020, pp. 1929–1934.

[44] A. Pierson, W. Schwarting, S. Karaman, D. Rus, Learning risk level set parameters from data sets for safer driving, in: 2019 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2019, pp. 273–280.

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

[46] Q.-H. Pham, P. Sevestre, R.S. Pahwa, H. Zhan, C.H. Pang, Y. Chen, A. Mustafa, V. Chandrasekhar, J. Lin, A 3D dataset: Towards autonomous driving in challenging environments, in: 2020 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2020, pp. 2267–2273.

[47] S. Xu, D. Zhou, J. Fang, J. Yin, Z. Bin, L. Zhang, Fusionpainting: Multimodal fusion with adaptive attention for 3d object detection, in: 2021 IEEE International Intelligent Transportation Systems Conference, ITSC, IEEE, 2021, pp. 3047–3054.

[48] G. Neuhold, T. Ollmann, S. Rota Bulo, P. Kontschieder, The mapillary vistas dataset for semantic understanding of street scenes, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4990–4999.

[49] K. Sakurada, M. Shibuya, W. Wang, Weakly supervised silhouette-based semantic scene change detection, in: 2020 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2020, pp. 6861–6867.

[50] P. Meletis, G. Dubbelman, Training of convolutional networks on multiple heterogeneous datasets for street scene semantic segmentation, in: 2018 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2018, pp. 1045–1050.

[51] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, R. Yang, The apolloscape open dataset for autonomous driving and its application, IEEE Trans. Pattern Anal. Mach. Intell. 42 (10) (2019) 2702–2719.

[52] A. Soni, P. Padamwar, K.R. Konda, Contextual road lane and symbol generation for autonomous driving, in: 2021 20th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2021, pp. 865–872.

[53] R. Chandra, T. Guan, S. Panuganti, T. Mittal, U. Bhattacharya, A. Bera, D. Manocha, Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms, IEEE Robot. Autom. Lett. 5 (3) (2020) 4882–4890.

[54] J. Fritsch, T. Kühnl, A. Geiger, A new performance measure and evaluation benchmark for road detection algorithms, in: 16th International IEEE Conference on Intelligent Transportation Systems, ITSC 2013, 2013, pp. 1693–1700, http://dx.doi.org/10.1109/ITSC.2013.6728473.

[55] Y. Chang, F. Xue, F. Sheng, W. Liang, A. Ming, Fast road segmentation via uncertainty-aware symmetric network, in: 2022 International Conference on Robotics and Automation, ICRA, IEEE, 2022, pp. 11124–11130.

[56] R. Fan, H. Wang, P. Cai, M. Liu, Sne-roadseg: Incorporating surface normal information into semantic segmentation for accurate freespace detection, in: European Conference on Computer Vision, Springer, 2020, pp. 340–356.

[57] K. Lis, K. Nakka, P. Fua, M. Salzmann, Detecting the unexpected via image resynthesis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2152–2161.

[58] C. Liang, W. Wang, J. Miao, Y. Yang, Gmmseg: Gaussian mixture based generative semantic segmentation models, Adv. Neural Inf. Process. Syst. 35 (2022) 31360–31375.

[59] Y. Liu, C. Ding, Y. Tian, G. Pang, V. Belagiannis, I. Reid, G. Carneiro, Residual pattern learning for pixel-wise out-of-distribution detection in semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 1151–1161.

[60] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, R. Raskar, Deepglobe 2018: A challenge to parse the earth through satellite images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 172–181.

[61] P. Shamsolmoali, M. Zareapoor, H. Zhou, R. Wang, J. Yang, Road segmentation for remote sensing images using adversarial spatial pyramid networks, IEEE Trans. Geosci. Remote Sens. 59 (6) (2021) 4673–4688, http://dx.doi.org/10.1109/TGRS.2020.3016086.

[62] L. Zhang, M. Lan, J. Zhang, D. Tao, Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–13, http://dx.doi.org/10.1109/TGRS.2021.3104032.

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

[64] Y. Chang, F. Xue, F. Sheng, W. Liang, A. Ming, Fast road segmentation via uncertainty-aware symmetric network, in: 2022 International Conference on Robotics and Automation, ICRA, 2022, pp. 11124–11130, http://dx.doi.org/10.1109/ICRA46639.2022.9812452.

[65] C. Liu, L.-C. Chen, F. Schroff, H. Adam, W. Hua, A.L. Yuille, L. Fei-Fei, Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 82–92.

[66] G.J. Brostow, J. Fauqueur, R. Cipolla, Semantic object classes in video: A high-definition ground truth database, Pattern Recognit. Lett. 30 (2) (2009) 88–97.

[67] J. Wang, C. Gou, Q. Wu, H. Feng, J. Han, E. Ding, J. Wang, Rtformer: Efficient design for real-time semantic segmentation with transformer, Adv. Neural Inf. Process. Syst. 35 (2022) 7423–7436.

[68] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, N. Sang, Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation, Int. J. Comput. Vis. 129 (2021) 3051–3068.

[69] E.S. Gedraite, M. Hadad, Investigation on the effect of a Gaussian blur in image filtering and segmentation, in: Proceedings ELMAR-2011, IEEE, 2011, pp. 393–396.

[70] S. Paris, P. Kornprobst, J. Tumblin, F. Durand, et al., Bilateral filtering: Theory and applications, Found. Trends® Comput. Graph. Vis. 4 (1) (2009) 1–73.

[71] X. Meng, N. Currit, K. Zhao, Ground filtering algorithms for airborne LiDAR data: A review of critical issues, Remote Sens. 2 (3) (2010) 833–860.

[72] R. Raguram, J.-M. Frahm, M. Pollefeys, A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus, in: Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II 10, Springer, 2008, pp. 500–513.

[73] M. Shah, Y. Xiao, N. Subbanna, S. Francis, D.L. Arnold, D.L. Collins, T. Arbel, Evaluating intensity normalization on MRIs of human brain with multiple sclerosis, Med. Image Anal. 15 (2) (2011) 267–282.

[74] S.-C. Huang, F.-C. Cheng, Y.-S. Chiu, Efficient contrast enhancement using adaptive gamma correction with weighting distribution, IEEE Trans. Image Process. 22 (3) (2012) 1032–1041.

[75] D.C. Lepcha, B. Goyal, A. Dogra, K.P. Sharma, D.N. Gupta, A deep journey into image enhancement: A survey of current and emerging trends, Inf. Fusion 93 (2023) 36–76.

[76] M.T. Rasheed, G. Guo, D. Shi, H. Khan, X. Cheng, An empirical study on retinex methods for low-light image enhancement, Remote Sens. 14 (18) (2022) 4608.

[77] A. Kiran, D. Vasumathi, Data mining: min–max normalization based data perturbation technique for privacy preservation, in: Proceedings of the Third International Conference on Computational Intelligence and Informatics: ICCII 2018, Springer, 2020, pp. 723–734.

[78] H. Henderi, T. Wahyuningsih, E. Rahwanto, Comparison of min-max normalization and Z-score normalization in the K-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer, Int. J. Inform. Inf. Syst. 4 (1) (2021) 13–20.

[79] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, K. Xu, Geometric transformer for fast and robust point cloud registration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11143–11152.

[80] S.K. Paul, P. Hoseini, M. Nicolescu, M. Nicolescu, Performance analysis of keypoint detectors and binary descriptors under varying degrees of photometric and geometric transformations, 2020, arXiv preprint arXiv:2012.04135.

[81] X.S. Poma, E. Riba, A. Sappa, Dense extreme inception network: Towards a robust cnn model for edge detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 1923–1932.

[82] N. Kaur, N. Nazir, et al., A review of local binary pattern based texture feature extraction, in: 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), IEEE, 2021, pp. 1–4.

[83] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, J. Saeed, A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction, J. Appl. Sci. Technol. Trends 1 (2) (2020) 56–70.

[84] C. Labrín, F. Urdinez, Principal component analysis, in: R for Political Data Science, Chapman and Hall/CRC, 2020, pp. 375–393.

[85] C. Ricciardi, A.S. Valente, K. Edmund, V. Cantoni, R. Green, A. Fiorillo, I. Picone, S. Santini, M. Cesarelli, Linear discriminant analysis and principal component analysis to predict coronary artery disease, Health Inform. J. 26 (3) (2020) 2181–2192.

[86] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

[87] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, 2016, pp. 21–37.

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

[89] C.R. Qi, L. Yi, H. Su, L.J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Process. Syst. 30 (2017).

[90] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, B. Chen, Pointcnn: Convolution on x-transformed points, Adv. Neural Inf. Process. Syst. 31 (2018).

[91] H. Xue, C. Liu, F. Wan, J. Jiao, X. Ji, Q. Ye, Danet: Divergent activation for weakly supervised object localization, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6589–6598.

[92] B. Baheti, S. Innani, S. Gajre, S. Talbar, Semantic scene segmentation in unstructured environment with modified DeepLabV3+, Pattern Recognit. Lett. 138 (2020) 223–229.

[93] J. Huang, G. Huang, Z. Zhu, Y. Ye, D. Du, Bevdet: High-performance multi-camera 3d object detection in bird-eye-view, 2021, arXiv preprint arXiv:2112.11790.

[94] Y. Wang, V.C. Guizilini, T. Zhang, Y. Wang, H. Zhao, J. Solomon, Detr3d: 3d object detection from multi-view images via 3d-to-2d queries, in: Conference on Robot Learning, PMLR, 2022, pp. 180–191.

[95] T. Hospedales, A. Antoniou, P. Micaelli, A. Storkey, Meta-learning in neural networks: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2021) 5149–5169.

[96] B. Cai, M. Wang, Y. Chen, Y. Hu, M. Liu, MFF-net: A multi-feature fusion network for community detection in complex network, Knowl.-Based Syst. 252 (2022) 109408.

[97] D. Xu, D. Anguelov, A. Jain, Pointfusion: Deep sensor fusion for 3d bounding box estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 244–253.

[98] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848–6856.

[99] Y. Xu, T. Fan, M. Xu, L. Zeng, Y. Qiao, Spidercnn: Deep learning on point sets with parameterized convolutional filters, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 87–102.

[100] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Bisenet: Bilateral segmentation network for real-time semantic segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 325–341.

[101] J. Zhou, M. Hao, D. Zhang, P. Zou, W. Zhang, Fusion PSPnet image segmentation based method for multi-focus image fusion, IEEE Photonics J. 11 (6) (2019) 1–12.

[102] T. Cortinhal, G. Tzelepis, E. Erdal Aksoy, Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds, in: Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part II 15, Springer, 2020, pp. 207–222.

[103] Y. Liu, Y. Wang, Y. Li, Q. Li, J. Wang, SETR-YOLOv5n: A lightweight low-light lane curvature detection method based on fractional-order fusion model, IEEE Access 10 (2022) 93003–93016.

[104] S. Abdallaoui, E.-H. Aglzim, A. Chaibet, A. Kribèche, Thorough review analysis of safe control of autonomous vehicles: path planning and navigation techniques, Energies 15 (4) (2022) 1358.

[105] U. Rosolia, S. De Bruyne, A.G. Alleyne, Autonomous vehicle control: A nonconvex approach for obstacle avoidance, IEEE Trans. Control Syst. Technol. 25 (2) (2016) 469–484.

[106] S. Chuprov, P. Belyaev, R. Gataullin, L. Reznik, E. Neverov, I. Viksnin, Robust autonomous vehicle computer-vision-based localization in challenging environmental conditions, Appl. Sci. 13 (9) (2023) 5735.

[107] J. Funke, M. Brown, S.M. Erlien, J.C. Gerdes, Collision avoidance and stabilization for autonomous vehicles in emergency scenarios, IEEE Trans. Control Syst. Technol. 25 (4) (2016) 1204–1216.

[108] R. Thakker, N. Alatur, D.D. Fan, J. Tordesillas, M. Paton, K. Otsu, O. Toupet, A.-a. Agha-mohammadi, Autonomous off-road navigation over extreme terrains with perceptually-challenging conditions, in: Experimental Robotics: The 17th International Symposium, Springer, 2021, pp. 161–173.

[109] P. Kim, J. Chen, J. Kim, Y.K. Cho, SLAM-driven intelligent autonomous mobile robot navigation for construction applications, in: Advanced Computing Strategies for Engineering: 25th EG-ICE International Workshop 2018, Lausanne, Switzerland, June 10-13, 2018, Proceedings, Part I 25, Springer, 2018, pp. 254–269.

[110] M.S.A. Mahmud, M.S.Z. Abidin, A.A. Emmanuel, H.S. Hasan, Robotics and automation in agriculture: present and future applications, Appl. Model. Simul. 4 (2020) 130–140.

[111] F. Thompson, R. Galeazzi, Robust mission planning for autonomous marine vehicle fleets, Robot. Auton. Syst. 124 (2020) 103404.

[112] A. Girma, N. Bahadori, M. Sarkar, T.G. Tadewos, M.R. Behnia, M.N. Mahmoud, A. Karimoddini, A. Homaifar, IoT-enabled autonomous system collaboration for disaster-area management, IEEE/CAA J. Autom. Sin. 7 (5) (2020) 1249–1262.

[113] M.G. Müller, M. Durner, W. Boerdijk, H. Blum, A. Gawel, W. Stürzl, R. Siegwart, R. Triebel, Uncertainty estimation for planetary robotic terrain segmentation, in: 2023 IEEE Aerospace Conference, IEEE, 2023, pp. 1–8.

[114] S.K. Panda, Y. Lee, M.K. Jawed, Agronav: Autonomous navigation framework for agricultural robots and vehicles using semantic segmentation and semantic line detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 6271–6280.

[115] E. Firkat, F. An, B. Peng, J. Zhang, T. Mijit, A. Ahat, J. Zhu, A. Hamdulla, FGSeg: Field-ground segmentation for agricultural robot based on LiDAR, Comput. Electron. Agric. 211 (2023) 107965.

[116] N. Steinke, D. Goehring, R. Rojas, Groundgrid: Lidar point cloud ground segmentation and terrain estimation, IEEE Robot. Autom. Lett. 9 (1) (2023) 420–426.

[117] A. Ando, S. Gidaris, A. Bursuc, G. Puy, A. Boulch, R. Marlet, RangeViT: Towards vision transformers for 3D semantic segmentation in autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 5240–5250.

[118] Y. Zhang, S. Wang, A robot navigation system in complex terrain based on statistical features of point clouds, IEEE Trans. Intell. Veh. (2023).

[119] O. Acun, A. Küçükmanisa, Y. Genç, O. Urhan, D3NET (divide and detect drivable area net): deep learning based drivable area detection and its embedded application, J. Real-Time Image Process. 20 (2) (2023) 16.

[120] T. Chase. Jr., C. Gnam, J. Crassidis, K. Dantu, You only crash once: Improved object detection for real-time, sim-to-real hazardous terrain detection and classification for autonomous planetary landings, 2023, arXiv preprint arXiv:2303.04891.

[121] E. Goh, J. Chen, B. Wilson, Mars terrain segmentation with less labels, in: 2022 IEEE Aerospace Conference, AERO, IEEE, 2022, pp. 1–10.

[122] A.J. Sathyamoorthy, K. Weerakoon, T. Guan, J. Liang, D. Manocha, Terrapn: Unstructured terrain navigation using online self-supervised learning, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2022, pp. 7197–7204.

[123] J. Zürn, W. Burgard, A. Valada, Self-supervised visual terrain classification from unsupervised acoustic feature learning, IEEE Trans. Robot. 37 (2) (2020) 466–481.

[124] P. Zhang, Y. Xiao, X. Wang, B. Duan, Semantic segmentation of point clouds of field obstacle-crossing terrain for multi-legged rescue equipment based on random forest, in: 2020 International Conference on Artificial Intelligence and Electromechanical Automation, AIEA, IEEE, 2020, pp. 147–153.

[125] P. Mehrabi, H.D. Taghirad, A gaussian process-based ground segmentation for sloped terrains, in: 2021 9th RSI International Conference on Robotics and Mechatronics, ICRoM, IEEE, 2021, pp. 371–377.

[126] P. Shamsolmoali, M. Zareapoor, H. Zhou, R. Wang, J. Yang, Road segmentation for remote sensing images using adversarial spatial pyramid networks, IEEE Trans. Geosci. Remote Sens. 59 (6) (2020) 4673–4688.

[127] L. Zhang, M. Lan, J. Zhang, D. Tao, Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–13.

[128] Y. Chang, F. Xue, F. Sheng, W. Liang, A. Ming, Fast road segmentation via uncertainty-aware symmetric network, in: 2022 International Conference on Robotics and Automation, ICRA, IEEE, 2022, pp. 11124–11130.

[129] E. Milli, Ö. Erkent, A.E. Yılmaz, Multi-modal multi-task (3MT) road segmentation, IEEE Robot. Autom. Lett. (2023).

[130] E. Horváth, C. Pozna, M. Unger, Real-time LiDAR-based urban road and sidewalk detection for autonomous vehicles, Sensors 22 (1) (2021) 194.

[131] G. Dong, Y. Yan, C. Shen, H. Wang, Real-time high-performance semantic image segmentation of urban street scenes, IEEE Trans. Intell. Transp. Syst. 22 (6) (2020) 3258–3274.

[132] K.A. Kalpoma, A.S. Aurgho, M.M.I. Shizan, F.H. Ani, A.R. Bondhon, Deep learning image segmentation for satellite images of national highways of bangladesh, in: IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2023, pp. 6894–6897.
[133] Y. Liu, Z. Cai, H. Dou, Highway traffic congestion detection and evaluation based on deep learning techniques, Soft Comput. 27 (17) (2023) 12249–12265.
[134] L. Barba-Guaman, J. Eugenio Naranjo, A. Ortiz, Deep learning framework for vehicle and pedestrian detection in rural roads on an embedded GPU, Electronics 9 (4) (2020) 589.
[135] C. Yang, Z. Wang, An ensemble wasserstein generative adversarial network method for road extraction from high resolution remote sensing images in rural areas, IEEE Access 8 (2020) 174317–174324.
[136] H. Shon, S. Choi, K. Huh, Real-time terrain condition detection for off-road driving based on transformer, IEEE Trans. Intell. Transp. Syst. (2024).
[137] L. Fritz, H.A. Hamersma, T.R. Botha, Off-road terrain classification, J. Terramech. 106 (2023) 1–11.
[138] L. Wijayathunga, A. Rassau, D. Chai, Challenges and solutions for autonomous ground robot scene understanding and navigation in unstructured outdoor environments: A review, Appl. Sci. 13 (17) (2023) 9877.
[139] Z. Wang, Y. Wu, Q. Niu, Multi-sensor fusion in automated driving: A survey, IEEE Access 8 (2019) 2847–2868.
[140] M. Wermelinger, P. Fankhauser, R. Diethelm, P. Krüsi, R. Siegwart, M. Hutter, Navigation planning for legged robots in challenging terrain, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2016, pp. 1184–1189.
[141] Y. Zhang, A. Carballo, H. Yang, K. Takeda, Perception and sensing for autonomous vehicles under adverse weather conditions: A survey, ISPRS J. Photogramm. Remote Sens. 196 (2023) 146–177.
[142] B. Kuang, C. Gu, Z.A. Rana, Y. Zhao, S. Sun, S.G. Nnabuife, Semantic terrain segmentation in the navigation vision of planetary rovers—a systematic literature review, Sensors 22 (21) (2022) 8393.