

Maps for Autonomous Driving: Full-process Survey and Frontiers

Pengxin Chen^{a,b}, Zhipeng Luo^{*c}, Xiaoqi Jiang^d, Zhangcai Yin^a, Jonathan Li, *Fellow IEEE*^{e,f}

^a*School of Resource and Environment Engineering, Wuhan University of Technology, Wuhan, China*

^b*Department of Smart Driving Product, IAS BU, Huawei Technologies, Shanghai, China*

^c*Key Laboratory of Data Science and Intelligence Application of Fujian Province University and School of Computer Science, Minnan Normal University, Zhangzhou, China*

^d*Global Technology Innovation Center, Chery Automobile Co., Ltd, Shanghai, China*

^e*Department of Geography and Environmental Management, University of Waterloo, Waterloo, Canada*

^f*Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada*

Abstract

Maps have always been an essential component of autonomous driving. With the advancement of autonomous driving technology, both the representation and production process of maps have evolved substantially. The article categorizes the evolution of maps into three stages: High-Definition (HD) maps, Lightweight (Lite) maps, and Implicit maps. For each stage, we provide a comprehensive review of the map production workflow, with highlighting technical challenges involved and summarizing relevant solutions proposed by the academic community. Furthermore, we discuss cutting-edge research advances in map representations and explore how these innovations can be integrated into end-to-end autonomous driving frameworks.

Keywords:

Autonomous Driving, HD Map, Lite Map, Implicit Map

1. Introduction

Maps have long been an essential component of driving, providing critical spatial information for navigation and decision-making. In human-driven scenarios, Standard-Definition (SD) maps, commonly referred to as navigation maps, are generally sufficient. These maps typically include road-level structure and traffic rules, offering guidance for basic routing without requiring detailed lane-level information. Human drivers depend primarily on their own perception and reasoning capabilities to navigate with the aid of SD maps.

However, in Autonomous Driving (AD), the requirements for map precision and richness increase substantially. Autonomous vehicles rely heavily on HD maps which provide rich, centimeter-level lane information, including road geometry, lane boundaries, traffic signals, semantic landmarks, and even road surface features. These HD maps act as a prior knowledge base, enhancing localization accuracy, supporting perception systems, informing prediction models, and enabling rule-compliant planning.

Despite their critical role, HD maps present considerable limitations. They are expensive to produce, often require manual annotations and specialized survey vehicles, and suffer from slow update cycles and high maintenance costs, thereby hinder

scalability. In response, Lite maps have emerged as a promising alternative. Lite maps retain essential lane-level information but leverage crowd-sourced data and automated generation techniques to significantly reduce production costs and support more frequent updates, offering a more scalable and practical solution for real-world AD deployment.

A more recent evolution is the development of implicit map representations, aligned with the trend toward end-to-end autonomous driving systems. Instead of storing explicit geometric and semantic elements, Implicit maps encode spatial and behavioral priors directly into neural network representations. These learned representations can be integrated into perception, prediction, and planning modules, enabling differentiable processing and backpropagation, which are essential in joint learning systems. The goal of this paradigm is to develop map representations that not only support driving tasks but also promote more human-like, context-aware decision-making in autonomous vehicles.

In this survey, we provide an comprehensive review of the maps used for autonomous driving. Fig. 1 provides an overview of our work. We begin by introducing the integration of maps in autonomous driving systems and the development of maps in industry. The main body is organized into four core sections: HD Map, Lite Map, Implicit Map, and a discussion on future trends. This arrangement is according to the chronological order of map development. In each development stage, we address the challenges encountered and review related works.

- 1) HD Map Section: Focuses on localization, multi-trip mapping, static perception, and topology generation.

*Corresponding author

Email addresses: pengxin.chen@connect.polyu.hk (Pengxin Chen), lzp2680@mnnu.edu.cn (Zhipeng Luo*), jiangxiaoqi1@mychery.com (Xiaoqi Jiang), yinzhangcai@whut.edu.cn (Zhangcai Yin), junli@uwaterloo.ca (Jonathan Li, *Fellow IEEE*)

- 2) **Lite Map Section:** Highlights online vectorization, change detection, map updates, and traffic-flow trajectory mining.
- 3) **Implicit Map Section:** Explores different implicit representations and world models.
- 4) **Future Trends Section:** Discusses emerging directions in the field, including fully automatic Lite maps, zero-shot map learning, vision-language-action models and foundation models.

Through this structured organization, we aim to provide a thorough and insightful overview of modern mapping technologies and their evolving role in autonomous driving.

1.1. Integration of Maps in AD Systems

Maps have consistently played a vital role in AD architectures. Fig. 2 illustrates how maps are integrated into convention and modern AD systems, respectively.

In conventional modular AD architectures, each processing step, including the localization, perception, prediction, and planning, interacts with a prebuilt map in a predefined manner. For instance, map matching can help realize accurate localization in GNSS-denied environments. The prediction module relies on map information, along with the historical trajectories of vehicles or pedestrians, to forecast their future movements. Similarly, the planning module depends on the map's topology and traffic rules to generate appropriate control strategies for vehicles.

In contrast, modern learning-based AD systems employ an end-to-end network architecture where multiple modules are optimized jointly. Within this framework, map representations are often integrated into a differentiable form, supporting gradient-based optimization and continuous feature learning across perception, prediction, and planning. This approach aims to improve coordination among functional components and ultimately achieve more human-like driving behavior. That said, such end-to-end systems have not yet reached the level of maturity required for large-scale commercial deployment.

Throughout the evolution of autonomous driving, maps themselves have also undergone significant transformation: moving from high-precision LiDAR-based surveying toward crowd-sourced data collection, from prebuilt offline maps to online map detection, and from explicit geometric representations to implicit neural embeddings. Regardless of these shifts, maps remain an indispensable element in AD systems—never absent, always evolving.

1.2. Map Development in Industry

Based on the authors' industrial experiences for the last decade, we divide the development of maps for AD into three stages as shown in Fig. 3.

- 1) **HD Map Stage (2015–2021):** TomTom pioneered the first commercial HD map in 2015 [174], establishing a critical foundation for map-dependent autonomous driving systems. During this phase, industry efforts concentrated on creating highly detailed and semantically rich

maps that included precise lane geometries, traffic signs, and road attributes. Major contributors included TomTom with Road DNA, HERE with HD Live Map, and Mobileye with RoadBook. Significant developments also came from Baidu's Apollo 2.5 and Waymo Driver. These systems relied extensively on pre-built HD maps to support high-accuracy localization and contextual decision-making. Because of the expensive production and maintenance pipeline, most HD map products only cover highways for cost considerations.

- 2) **Lite Map Stage (2021–Future):** Beginning around 2021, the industry transitioned toward "Lite map", namely the lightweight maps, which aimed to reduce the dependency on dense and static HD maps by introducing more scalable and adaptive solutions. Tesla pioneered this trend with its FSD auto-labeling solution [171], emphasizing real-time perception and automated annotation. This stage also saw the emergence of Mobileye REM [136], Huawei's Road-Code, NavInfo's HD Lite, Tencent's HD Air, and Baidu's MapAuto, all focusing on lightweight, more updateable maps. These maps balanced efficiency and functionality, enabling wider geographic coverage with less infrastructure overhead. It is worthwhile to note that Lite maps remain the mainstream of industry at present and is far from being replaced by Implicit maps. Because the maintenance costs have been greatly reduced, Lite maps can cover urban roads.
- 3) **Implicit Map Stage (2023–Future):** Since approximately 2023, a new trend has emerged toward Implicit maps, in which environmental knowledge is encoded implicitly within neural network parameters rather than through traditional explicit map formats. Mapping becomes an internalized process through AI large-scale learning. Rather than relying on explicit maps, vehicles understand their environment through learned models. Tesla and Wayve's World Models [10, 187], NVIDIA's CosMos [4] and Xpeng's WFM [199] represent this stage. These approaches aim to achieve highly generalizable and scalable autonomy by integrating map knowledge into end-to-end AD architectures.

1.3. Comparison to Related Surveys

This survey distinguishes itself from prior works in two key aspects.

Firstly, while existing surveys [51, 103, 11, 1, 121] offer limited emphasis on industrial perspectives, this review incorporates insights grounded in real-world applications. As autonomous driving technology rapidly evolves toward production vehicles, distinct challenges emerge at different developmental stages. Accordingly, we adopt a dynamic, industry-informed viewpoint and categorize map evolution into three milestone phases: 1) HD Map; 2) Lite Map; and 3) Implicit Map. This staging allows us to systematically examine the unique technical challenges and innovations characteristic of each period.

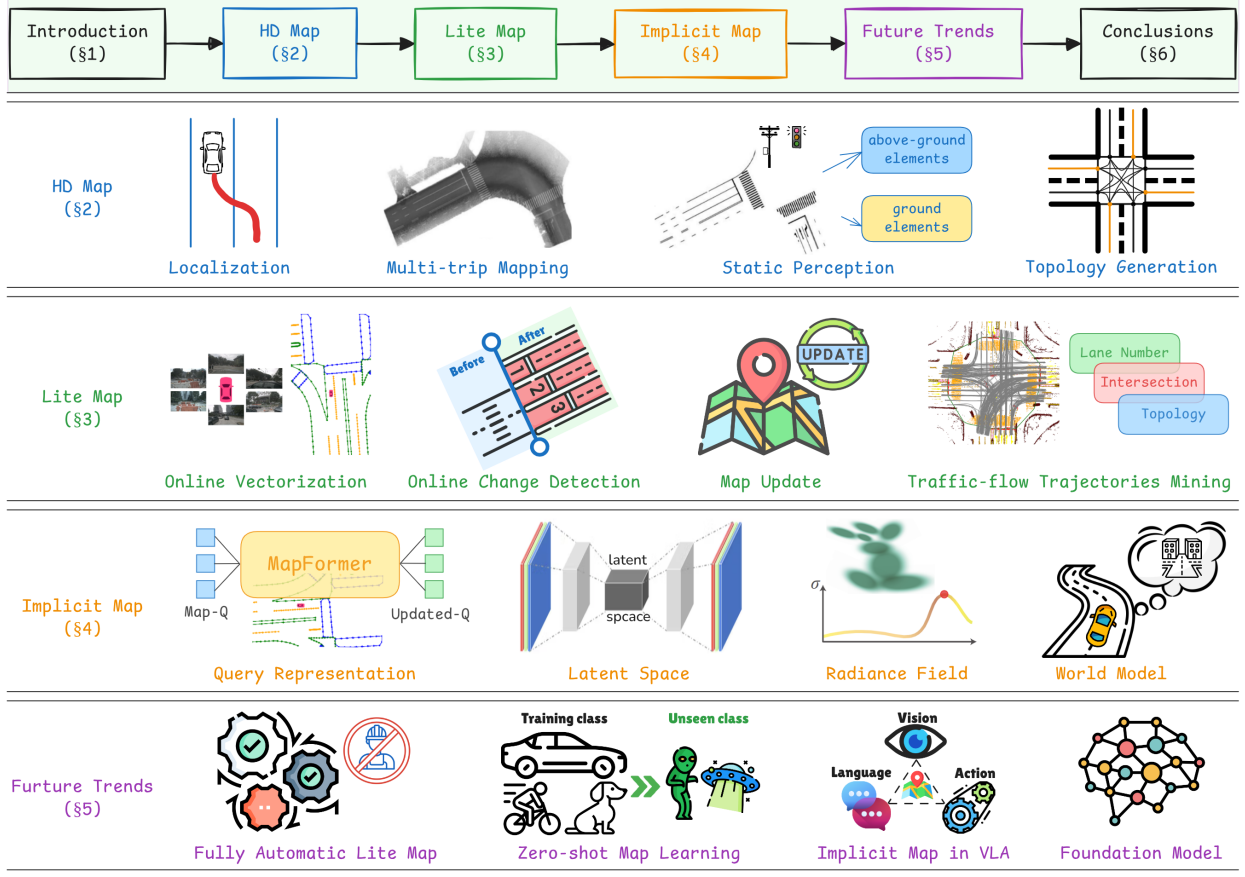


Fig. 1: Survey at A Glance. The introduction section provides a technical overview of the modern autonomous driving and the map development. The structure of main body is organized into HD Map, Lite Map and Implicit Map according to the order of their appearance. The paper ends with summarizing future trends and conclusions.

Secondly, with regard to scope, previous surveys often concentrate on isolated topics, such as localization & mapping [191, 31], object detection [126], lane topology reasoning [211], crowdsourcing update [65, 21] and end-to-end generation [98], or focus exclusively on specific HD map production pipelines [36, 127, 14, 169]. In contrast, our work provides a comprehensive full-process survey that spans the entire map production workflow, covering all essential modules from data collection to representation and application.

1.4. Contributions

The main contributions of our survey are as follows:

- 1) We present a comprehensive review of map representations for autonomous driving. To the best of our knowledge, this is the first survey that systematically covers the entire developmental pipeline across three key stages: HD maps, Lite maps, and Implicit maps.
- 2) We identify and analyze critical challenges faced at each stage from an industrial perspective. To support this analysis, we examine more than 200 relevant publications, offering insights grounded in both technical and practical contexts.

- 3) We summarize emerging trends in Implicit maps and their role in end-to-end autonomous driving systems. Although most of these approaches have not yet been widely deployed commercially, we emphasize their potential to shape future autonomous driving models and encourage the community to actively engage with these developments. It is our hope that this survey will serve as a valuable reference and provide a holistic outlook for researchers and practitioners in both academia and industry.

2. HD Map: The First Victory on Highways

HD maps are characterized by lane-level precision and centimeter-level accuracy, making them a foundational infrastructure for autonomous driving systems. However, HD maps also face significant challenges, including high production costs and low update frequency. *As a result, the early AD solutions which depends on HD maps could only cover highways.* This section elaborates on the HD map production pipeline and its key technical components.

2.1. Production Pipeline

We divide the production pipeline of HD map into three stages: Mapping, Perception and Map Compilation, as illus-

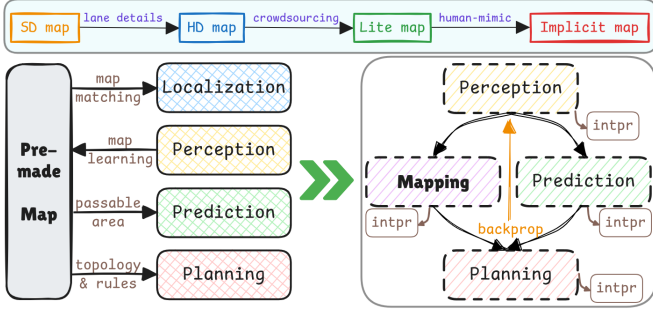


Fig. 2: An overview of the evolution and integration of maps in autonomous driving systems. The top row illustrates the progression from **SD map**, sufficient for human drivers, to **HD map** with detailed lane information, followed by **Lite map** enabled through crowdsourcing, and finally to **Implicit map** designed for human-like, end-to-end autonomous driving. The bottom half contrasts the **traditional modular architecture** (left) where pre-made maps independently support localization, perception, prediction, and planning with a **modern learning-based paradigm** (right). On the right side, a shared Mapping module is jointly trained with other components via interpretable representations and backpropagation (marked as “intpr” and “backprop”, respectively), allowing tighter integration and greater adaptability.

trated in Fig. 4.

- 1) **Mapping:** This stage aims to construct a globally consistent point cloud map by integrating data from multiple survey trips. For a survey vehicle, the location of each timestamp can be computed by integrated navigation or SLAM methods. After that, a point cloud map can be acquired by stitching laser scans. During this process, dynamic objects, such as pedestrians, cyclists and cars, are filtered out to retain only static structures. However, due to nonlinear errors and drift in localization, point cloud maps surveyed from different trips at the same place usually cannot be aligned well. To solve this problem, optimization-based algorithms are employed to align maps from different trips. Finally, the mapping process outputs a clean point cloud map as well as street-view images with absolute locations.
- 2) **Perception:** As the downstream process of mapping, this stage involves extracting map elements essential for autonomous driving from raw point cloud maps and images. Both entity elements (e.g., ground features and above-ground objects) and logical elements (e.g., lane topology and traffic light associations) are processed in parallel. A detailed taxonomy of these elements is provided in [Section. 2.2.2](#).
- 3) **Map Compilation:** The extracted elements require further processing to support downstream applications such as planning and control (PnC), simulation, and map rendering. Map platforms compile these elements into structured, query-efficient vector representations. For instance, road elements are vectorized and spatially partitioned to facilitate efficient storage and retrieval (see [Fig. 5](#)).

2.2. Map specifications

2.2.1. File management

Before discussing the elements in HD maps, we would like to introduce file management. As illustrated in [Fig. 5](#), map artifacts are usually stored and read by administrative divisions. Self-driving cars need to download the HD map portfolio belonging to their city. When a car starts navigating, both SD maps (i.e. navigation map) and HD maps are loaded into the memory. Since a HD map of the entire city is too large to be loaded to the memory all at once, the common practice is to only load the map data along the navigation route. Therefore, smaller scales of map division are required.

Tile is a classic strategy to manage HD artifacts hierarchically. [170] introduced a nested grid tile structure, known as geometry clip maps, for real-time terrain rendering. It efficiently supports high-resolution detail and smooth level-of-detail (LOD) transitions using a multilevel grid-based caching system. [123] applies hierarchical tile structures to planetary-scale terrain rendering. It introduces a “4–8” tile hierarchy, allowing efficient adaptive refinement of both geometry and texture using a pyramid-based tile organization. [219] proposes a method for efficiently building tile pyramids from large-scale raster datasets. The generated multi-resolution tiles significantly accelerate visualization and interaction with massive geospatial data. [181] focuses on vector tile management using a hierarchical tile pyramid structure combined with NoSQL storage (HBase). It enables scalable and high-performance spatial indexing and retrieval for large vector datasets. [203] describes a distributed approach for constructing vector tile pyramids using a Hadoop-based platform. It uses a quadtree hierarchy to support scalable spatial connectivity analysis for agricultural land.

Tiles are hierarchical, compact, rectangular, friendly for indexing and retrieval, but they also have a disadvantage. That is, the tile representation mixes multi-level roads within a tile. More specifically, in complex interchanges, roads within a single connected region may exhibit vertical overlaps in the 2D projection. For instance, in downtown area, there may be an elevated bridge and an underground tunnel at the same place, so the road network will be interlaced vertically. These overlapping scenarios pose significant challenges for the downstream PnC module because we all know that it operates on planar road networks.

To address this issue, it is necessary to further split a connected region into multiple patches according to the vertical stacking relationships of roads. The goal is to ensure that each resulting patch contains no overlapping structures in 2D, while satisfying the following constraints:

- 1) **Graph connectivity:** The road network within each patch must remain topologically connected.
- 2) **Production specifications:** Requirements such as maximum patch size, boundary completeness, and topological integrity must be maintained.

Patches are irregular polygons. They are partitioned according to the “no-interlacement” rule. For example, if an elevated

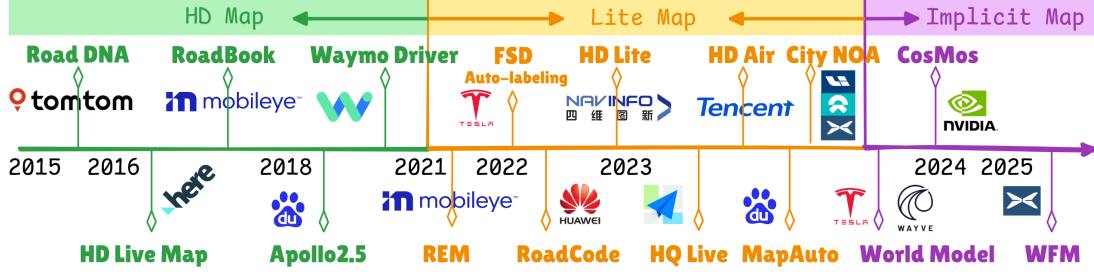


Fig. 3: Key companies, technologies, and milestones from 2015 to 2025, illustrating the industry’s transition from high-definition maps to lightweight maps, and now to implicit, AI-driven models.

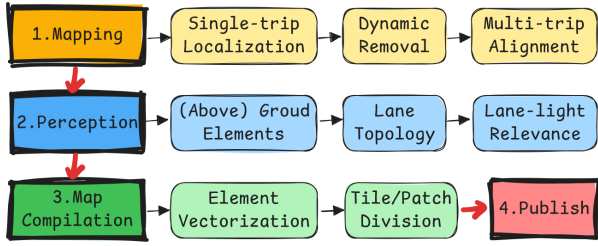


Fig. 4: Typical pipeline of HD map production

road is found above a main road, these two roads will be separated into two patches. The same rule applies to tunnels. According to the rule, the road network in a patch can be considered as a plane, thus facilitating downstream PnC. As the shapes of patches are irregular, they may overlap each other. A self-driving car needs to identify which patch it is in based on real-time locations and navigation routes.

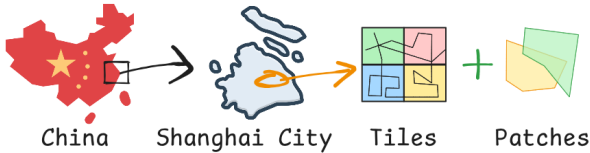


Fig. 5: Map artifacts are managed by administrative divisions

2.2.2. Elements

HD maps employ a structured, object-oriented schema to represent the complex relationships among diverse road elements. However, a unified standard has not yet been established across the industry, as different map providers often adopt varying specifications [139, 9, 145, 3, 64]. To facilitate a clearer and systematic understanding, we distill the principal elements of an HD map into a simplified conceptual framework, illustrated in Fig. 6. At the center lies the **Road**, which serves as the core entity and is linked to several sub-components.

Road: Each road is defined by a unique ID and its geometric representation (‘Geo/LineString’). It also contains higher-level topology information and serves as a container for related objects such as traffic signs, road markings, traffic lights, lanes, poles, curbs, fences, and crosswalks.

Lane: A road may contain multiple lanes. Each lane is described by its geometry, length, width, and functional attributes such as ‘LaneType’ and ‘TurnType’. The schema supports detailed lane-level connectivity through fields like ‘PrevLaneID’, ‘NextLaneID’, and lateral relationships (‘LeftLaneID’, ‘RightLaneID’). These refer to the IDs of neighboring lanes. The schema also includes indicators for stop lines, traffic light associations, and intersection connectivity.

Intersection: Intersections are modeled using polygonal geometries and contain references to incoming and outgoing lanes, as well as related crosswalks.

Traffic Light: Traffic lights are linked to lanes and defined by their spatial coordinates (‘Geo/Point’) and control information (‘LightInfo’), enabling accurate modeling of signalized intersections.

Traffic Sign and Road Marking: These are associated with roads and described by type and geometry. They provide semantic cues for vehicle behavior and navigation.

Pole/Curb/Fence: These roadside objects are modeled as linear geometries and categorized by type. They contribute to the environmental context and are relevant for perception and localization tasks.

Crosswalk: Crosswalks are defined as independent polygonal objects but are also referenced by roads and intersections, aiding pedestrian-aware planning and safety modules.

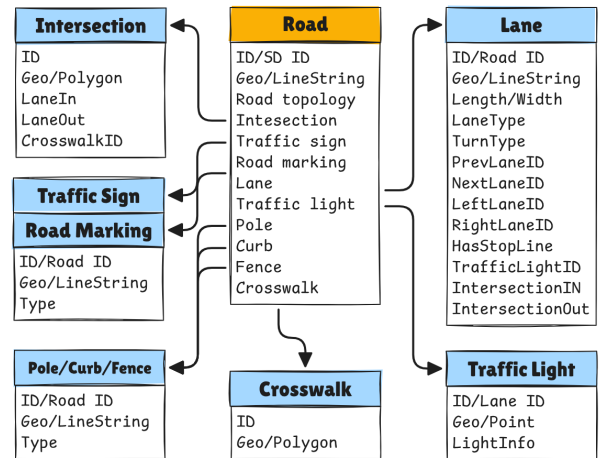


Fig. 6: Typical elements in HD map

2.3. Localization

Precise localization of survey vehicles is a fundamental component in the production of HD maps. These maps require centimeter-level precision to support Advanced Driver-Assistance Systems (ADAS) and autonomous driving. Survey vehicles equipped with high-precision GNSS, IMU, wheel odometer, LiDAR, and camera systems traverse road networks to collect rich environmental data. The effectiveness of this process hinges on the ability to precisely estimate the vehicle's position and orientation. Fig. 7 illustrates the hierarchical enhancement of localization strategies employed in HD map surveying vehicles.

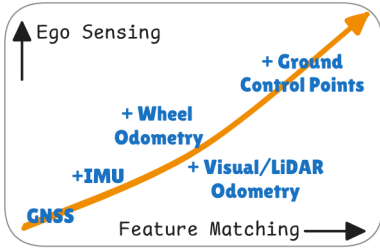


Fig. 7: A schema of localization methods. The horizontal axis represents the increasing reliance on feature matching techniques, while the vertical axis reflects the advancement in ego-motion sensing capabilities.

2.3.1. GNSS

As the core component of positioning solutions, GNSS can provide accurate position and speed in open-sky areas. [178] presents an early and influential analysis of the effect of correlator spacing on GPS receiver accuracy. It introduces the concept of “narrow correlator spacing”, which significantly improves pseudorange measurement accuracy, thereby enhancing both position and velocity estimation. The methodology laid a foundation for modern GPS receiver design. [155] explores techniques for integrity monitoring in GNSS-based navigation, particularly for aviation. It evaluates the accuracy and robustness of position and velocity estimates derived from GPS, focusing on safety-critical applications.

As the localization for HD map production does not require real-time processing, PPK (Post-Processed Kinematic) is preferred in many HD map manufactures such as NovAtel Inertial Explorer [141]. Compared to RTK (Real-Time Kinematic), PPK does not require real-time data transmission. Therefore, PPK is more robust to data loss and allows full processing optimization. [241] describes a robust framework for multi-GNSS PPK using double-difference carrier-phase processing. It improves ambiguity resolution by leveraging observations from multiple constellations, achieves centimeter-level precision, and analyzes ambiguity-fixing performance in post-processing mode. [242] proposes strategies to improve integer ambiguity resolution in PPK by using quality control and optimized hypothesis testing. It enhances the ambiguity-fixing rate and robustness of the post-processing kinematic solution.

Modern GNSS systems go beyond GPS and now include multiple constellations. Global GNSS systems with worldwide coverage include GPS (USA), GLONASS (Russia), Galileo (EU) and BeiDou (China). Regional systems include QZSS (Japan) and IRNSS/NavIC (India). Fig. 8 provides an overview of these systems.

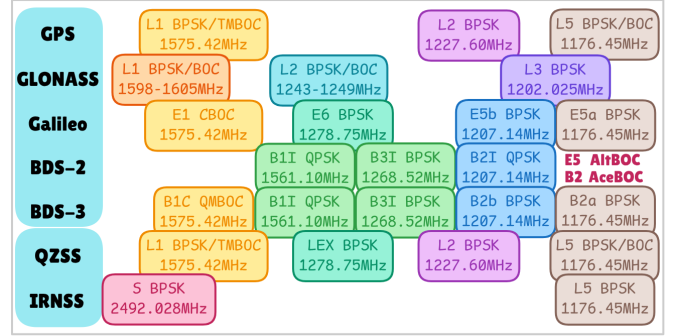


Fig. 8: GNSS Signal Frequencies and Modulation Schemes. While all GNSS systems operate in similar frequency bands, each employs distinct modulation schemes to balance compatibility, precision, and robustness.

2.3.2. IMU

IMU propagation can compensate for GNSS signal outages, enabling continuous positioning and smoothing the trajectory during GNSS-denied periods. [230] proposes a practical integration of a low-cost MEMS IMU with single-frequency GPS. It compares loose versus tight coupling strategies in dense urban scenarios. Using an Extended Kalman Filter (EKF), the authors analyze velocity/position error resilience during GNSS interruptions and highlighted drift behavior typical of MEMS-grade IMUs. [190] inspires a paradigm shift from extended Kalman filters to Factor Graph Optimization (FGO) in GNSS/INS navigation. The authors demonstrate FGO's superior accuracy during GNSS outages in urban canyons and discuss the influence of optimization window sizes.

2.3.3. Wheel Odometry

For vehicle localization, wheel odometry provides high-frequency motion updates that enhance short-term accuracy and stability. It is especially useful during GNSS outages or in environments where sensor visibility is limited, such as tunnels or urban canyons. [227] proposes a new navigation algorithm for all-wheel steering robots that combines GNSS, INS, odometry, and wheel angle data. It reduces positioning errors by using steering angles and speed to correct drift, even when GNSS is unavailable. Field tests show the method works well in both standard and all-wheel steering modes. [137] addresses key issues in low-cost GNSS/INS/odometer fusion, such as odometer scale drift, sensor misalignment, lever arm effects, and wheel slip. It proposes solutions including scale and misalignment estimation, lever arm compensation, and fault detection using a two-stage Kalman filter. Results show improved positioning accuracy, especially in GNSS-denied environments. [112]

presents a high-precision vehicle navigation system that tightly fuses GNSS PPP-RTK, MEMS IMU, wheel odometer, and vehicle motion constraints using a Kalman filter.

2.3.4. Visual And LiDAR Odometry

In extreme circumstances, such as a tunnel over 10KM long, the GNSS/IMU/Wheel Odometry fusion still suffers from the accumulation of drift errors. Feature matching provides inter-frame translation and attitude to solve this issue. In terms of sensor types, we classify the feature matching into 2 categories: **visual odometry (VO)** and **LiDAR odometry (LO)**. VO captures texture features [148, 27], while LO extracts geometry features [220, 38]. LO can acquire the real translation scale inherently, but VO needs to fuse the prior from IMU or binocular calibration to get it. In-P3VINS [109] proposes an invariant optimization-based system that fuses GNSS PPP, IMU, and Visual-SLAM for high-precision and consistent positioning. Built on an invariant factor graph framework, it models all raw measurements—including GNSS carrier phase with ambiguity states—for improved estimation. [108] proposes a tightly coupled GNSS PPP/INS/LiDAR integration system to meet the high-precision positioning demands of urban IoT applications like autonomous driving. It introduces a LiDAR sliding-window plane-feature tracking method to enhance both accuracy and efficiency. Field tests show the system achieves submeter-level accuracy in GNSS-challenging environments, significantly outperforming traditional GNSS/INS integration.

2.3.5. Ground Control Points

When neither texture nor geometry features are visible, both visual and LiDAR odometry will fail to localize itself. Ground control points (GCPs) provide fixed, precisely surveyed reference locations that anchor the HD map to real-world coordinates, ensuring accurate georeferencing. They can also help correct and reduce accumulated errors from GNSS, sensor drift, and mapping processes, thereby improving overall map reliability. They serve as ground truth for validating and calibrating HD map features, ensuring mapping quality meets stringent autonomous driving requirements. [177] aims to improve the efficiency of HD map production in Taiwan by using diverse surveying methods and multiple sources of ground control points (GCPs). The approach supports flexible and scalable map creation, allowing positioning within 0.5m and contributing to the evolution of HD map standards. [42] outlines a verification and validation procedure for HD maps, emphasizing the role of GCPs in enhancing the relative and absolute accuracy of point clouds used in autonomous driving systems. The study highlights how GCPs contribute to meeting the stringent accuracy requirements of HD maps, ensuring reliable localization and navigation for autonomous vehicles.

2.4. Multi-trip Mapping

The localization methods introduced in Section 2.3 are sufficient for the map construction of a single trip. Generally speaking, a single-trip map cannot cover all the lanes of a road due to sensors' range and environmental occlusion. Therefore, in

practice, multiple trips are needed to make a complete map. However, different maps surveyed at the same place normally do not align well due to the different localization errors of them, thus leading to a scientific problem: multi-trip mapping. The problem is to aggregate the maps from multiple trips and ensure the mapping error is minimized.

Taking Huawei RoadCode as an example, a cloud mapping server is established to aggregate mass data captured by multiple vehicles. In 2021, the on-cloud mapping module is used to merge scalar maps uploaded by production cars [149]. To save uploading bandwidth, the scalar map is designed in the form of a grid map whose resolution is $0.1 \times 0.1 \times 0.1m$. Every grid's information contains position, semantic labels, and counts of each semantic label. Semantic labels include ground, lane line, stop line, ground sign, and crosswalk. With the development of end-to-end online vectorized HD map generation, related work has shown decent speed and accuracy to be deployed onboard [122, 116, 216, 218]. The vectorized form of a single-trip map seems to be a better choice to upload, taking into consideration of transmission bandwidth, computational overhead and surveying regulations. Therefore, in 2024, the on-cloud mapping module turns to directly consuming visual vectorized elements provided by the on-board cognition deployed on multiple cars [37]. It then outputs global vectorized map which can be used for localization, planning and navigation. The new solution is more lightweight and efficient compared with its scalar-merging counterpart.

2.4.1. Collaborative SLAM

The multi-trip mapping task is quite similar to the collaborative SLAM defined in robotics literature. Taking self-driving cars as robots, the only difference between multi-trip mapping and collaborative SLAM might be that the former does not need real-time inter-robots communication. CoSLAM [245] is the first visual-only SLAM solution in a dynamic environment with multiple cameras moving independently. This work introduces inter-camera pose estimation and inter-camera mapping to handle dynamic objects during localization and mapping, while maintaining uncertainty estimates for each map point to improve robustness. It also dynamically clusters cameras into groups based on view overlap, allowing real-time group splitting and merging, which enhances scalability and adaptability in dynamic environments. However, CoSLAM relies on the assumption that all cameras are synchronized and observe the same scene at initialization. Together with the requirement for a GPU, they render CoSLAM impractical to run online onboard multiple robots. CVI-SLAM [89] proposes a centralized collaborative SLAM framework which is more efficient. In the framework, multiple agents equipped with visual-inertial sensors share data with a central server, enabling efficient and robust multi-agent localization and mapping. It offloads computationally intensive tasks like global map optimization to the central server, reducing the onboard processing load while maintaining agent autonomy through onboard visual-inertial odometry. As a centralized framework, it inherently relies on a single server for data association and optimization, suffering from a communication bottleneck between the robots and the server.

In contrast, decentralized solutions, such as DiSCo-SLAM [80] and Swarm-SLAM [235], rely only on occasional communication between the robots are better suited for large-scale deployment. DiSCo-SLAM uses the Scan Context descriptor [94] in multi-robot LiDAR SLAM for data-efficient communication. The descriptor is lightweight and resilient to unknown initial conditions. Swarm-SLAM, a recent opensource multi-robot SLAM system, presents a decentralized, scalable, and sparse collaborative SLAM framework specifically tailored for multi-robot (swarm) systems operating in GPS-denied environments. The system is modular and sensor-agnostic, supporting LiDAR, stereo, and RGB-D sensors, enhancing its flexibility across various robotic platforms. It also introduces a novel inter-robot loop closure prioritization strategy, which significantly reduces communication overhead and improves the speed of map convergence. Recently, there is a trend of using neural network to solve the collaborative SLAM problem. The CP-SLAM [72] is an example. It is the first dense collaborative neural SLAM framework based on a novel neural point based representation.

2.4.2. Loop Closure Detection

Loop closure detection and factor graph optimization are typically the back-end solvers used in multi-trip mapping.

Some work focuses on the loop closure detection of visual images. [7] proposes an online loop-closure detection approach by extending the bag-of-words image classification method with local shape and color features and incorporating Bayesian filtering to estimate loop-closure probability incrementally. The underlying idea is to compare a new image with all of the previous ones. If a loop closure is detected, it will be added to the back-end graph as an edge. Then, graph optimization is performed to share errors among edges according to customized weights. The size of the graph will gradually increase as the scene expands. If not controlled, the graph will be too large for real-time computation. To address this, a memory management mechanism, introduced by [99], is used to limit the data processed by global loop closure detection and graph optimization. Thus, online constraints are independent of the size of the environment. As the prevalence of using 3D Gaussian Splat (3DGS) [91] to represent 3D scenes, [244] presents LoopSplat, a dense RGB-D SLAM system that uses 3DGS submaps. The LoopSplat incorporates online loop closure through direct 3DGS registration and pose graph optimization for global consistency.

The LiDAR loop closure detection is also a significant task. [195] proposes FastLCD, a fast and compact loop-closure detection method that uses discriminative multimodal descriptors and a pre-trained machine learning classifier, followed by a double-deck verification strategy to improve accuracy. LCD-Net [30] is a deep learning-based loop closure detection network for LiDAR point clouds, which combines a shared encoder, a place recognition head, and a novel relative pose head based on differentiable unbalanced optimal transport for end-to-end training. [66] proposes a loop closure detection approach for LiDAR SLAM that uses bird's-eye view density images of locally generated maps, rather than individual scans, to create

robust and compact descriptors suitable for long sequences and varying sensor patterns. It enables real-time loop closure detection with smaller feature complexity and database size.

Actually, there is an important trick in autonomous driving to solve the problem of loop closure detection. The trick is to match a car's continuous trajectory with the SD road networks that can be acquired from navigation maps. This is a well-defined task called map matching in navigation literature. It will be discussed in Section. 2.4.3.

2.4.3. SD Map Matching

Given SD road networks and a vehicle's moving trajectory, the goal of map matching is to project each point of the trajectory onto the road networks. The purpose is to recover the driving route of a car given its historical trajectory. Low localization accuracy and complex road topology are two difficulties to in this task.

The study of SD map matching can be traced back to [16]. Now the solution paradigm has been basically determined. The paradigm uses Hidden Markov Model (HMM) to formulate the map matching problem and then applies Viterbi algorithm to solve the HMM. Most papers use this paradigm with some differences of formulating the distance from the given trajectory and road networks [22, 20, 140, 124, 61].

2.5. Static Perception for HD Map Construction

Static perception is a fundamental stage in HD map generation, focusing on the extraction of both ground elements (encompassing pavement, markings, and lane lines) and non-ground objects, such as traffic signs, guardrails, pole-like structures, and street trees, as shown in Fig. 9.

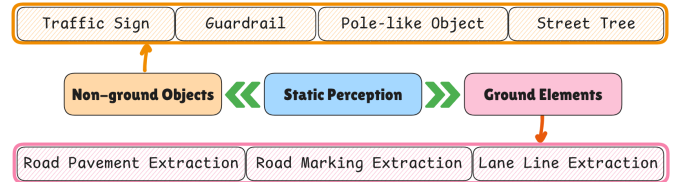


Fig. 9: Ground elements detection and non-ground objects detection for static perception

2.5.1. Ground Elements Extraction

Road Pavement Extraction: As shown in Fig. 10, early approaches primarily relied on heuristic ideas. Segmentation based techniques formulate the task as an image segmentation problem, frequently integrating geometric features [73], probabilistic Support Vector Machines (SVM), and graph cuts [41]. Conversely, tracking-based methods connect seed points using parabolic models [76], geodesic algorithms [134], or particle filtering [237]. However, these methods often face limitations in complex urban scenes due to their dependence on empirical features and seed point initialization. Deep learning (DL) approaches have become the predominant solutions. Encoder-decoder architectures significantly enhance multi-scale feature learning. To address occlusion challenges, post-processing

techniques including tensor voting [58] and Generative Adversarial Networks (GANs) [224] are employed to reconnect fragmented road regions. Despite improved accuracy, effectively modeling topological relationships in heavily occluded areas, such as those under dense vegetation, remains challenging.

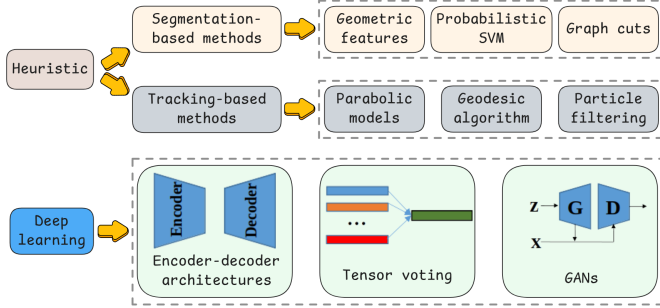


Fig. 10: Two kinds of road pavement extraction methods

Road Marking Extraction: As shown in Fig. 11, three primary strategies characterize this task. Traditional 2D vision-based methods exploit the inherent color contrast between markings and pavement, often employing edge detection or Hough transforms [213]. While shallow CNNs have been applied for marking classification [13], they exhibit significant performance degradation under varying illumination conditions and occlusions. Geo-Referenced Feature (GRF) Image-based Methods apply radiometric correction [63] and multi-threshold segmentation to LiDAR intensity data projected onto 2D images. However, this projection inevitably causes information loss, frequently resulting in incomplete extractions. In contrast, point-based methods directly process 3D point clouds using intensity thresholds and geometric filters [189], demonstrating superior robustness. Key limitations include difficulty in detecting worn markings with low reflectivity and susceptibility to environmental obstructions.

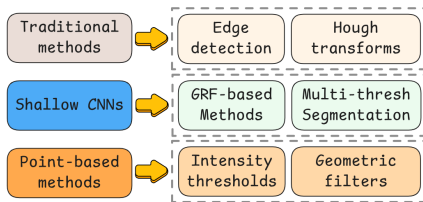


Fig. 11: Three kinds of methods for road marking extraction

Lane Line Extraction: Traditional pipelines encompass several approaches. The geometric model typically involves edge detection followed by polynomial fitting [132]. The Hough Transform (HT) detects lines, which are subsequently refined using B-splines [47]. The energy minimization technique uses conditional random fields to model multi-lane correlations [82]. Deep learning approaches currently dominate recent research. Segmentation-focused networks, including encoder-decoder CNNs [95] and Spatial CNNs [143], excel at capturing long-range contextual features. Weakly supervised methods have also emerged, leveraging multi-sensor data to re-

duce the substantial costs associated with manual annotation [24]. Future research may focus on integrating sequence modeling techniques, such as RNNs or Transformers, to further improve performance, particularly in challenging occlusion scenarios.

2.5.2. Non-ground Objects Detection

Fig. 12 gives an overview of non-ground objects detection.

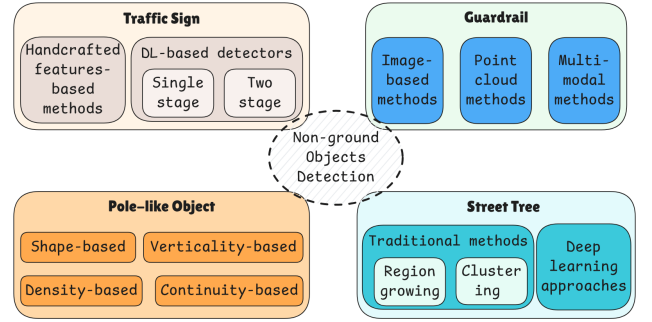


Fig. 12: Four common categories in non-ground objects detection

Traffic Sign Detection: Historically, image-based methods utilizing handcrafted features are prevalent. These include color/shape thresholding [46], Maximally Stable Extremal Regions (MSER) [158], and HOG-SVM classifiers [217]. DL-based detectors have largely supplanted these approaches. Two-stage frameworks like Faster R-CNN [154] and Mask R-CNN [69] achieve high accuracy but face efficiency challenges. Single-stage models, such as variants of YOLO [152], effectively balance speed and performance [12]. Point cloud-based methods employ techniques like supervoxel segmentation [207], Hough forests [182], and Bag-of-Visual-Phrases combined with deep Boltzmann machines [215]. However, sparse point distribution and sensitivity to noise pose critical challenges for detection relying solely on point clouds. Consequently, fusion strategies incorporating image data represent a promising direction for enhancing overall robustness.

Guardrail Detection: Image-based approaches, often utilizing stereo-vision systems with geometric constraints [159], exhibit limited reliability, particularly under nighttime conditions. Point cloud methods primarily rely on spatial clustering and feature-based filtering (e.g., corner features, height constraints) to detect both anti-fall and lane-separating guardrails [57], although their dependence on handcrafted features restricts generalization capability. Multi-modal fusion techniques, which combine image patches classified by CNNs with corresponding point clouds, demonstrate enhanced detection performance [131].

Pole-like Object Detection: Methods for detecting pole-like objects can be categorized into four main types, as shown in Fig. 12. Shape-based approaches, such as RANSAC or least-squares fitting, identify cylindrical structures [100, 79]. Verticality-based methods exploit the characteristic vertical continuity of these objects in 3D space [56, 204]. Density-based techniques project points onto 2D grids and apply clustering

algorithms like DBSCAN [205, 233]. Continuity-based methods utilize voxel analysis to extract vertically connected components [194]. While verticality/continuity methods generally offer higher accuracy, density-based approaches favor computational efficiency. A common limitation across all categories is their reliance on handcrafted features.

Street Tree Detection: Traditional methods focus on point cloud segmentation. Region growing algorithms utilize properties like curvature [17], normals [176], or voxel merging [180] to isolate trees. Clustering algorithms, such as Mean Shift [54] and Normalized Cuts [153], are also employed to separate individual trees. More recently, deep learning approaches, particularly R-CNN frameworks augmented with attention modules, have shown promise in improving detection accuracy under occlusion conditions [198]. A persistent challenge in segmentation arises from the spatial proximity of trees to streetlights and other urban structures [151].

2.5.3. Summary

The extraction of road surfaces is increasingly adopting DL-based segmentation techniques, yet significant challenges remain in preserving topological connectivity under heavy occlusion. Road marking extraction methods benefit from direct point cloud analysis but require enhanced robustness against degradation and environmental factors. Traffic sign detection leverages advanced 2D detectors effectively but necessitates further improvements in detecting small targets and developing robust multimodal fusion strategies. Detection of pole-like objects and street trees still relies heavily on traditional point cloud geometry analysis, warranting greater integration with advanced 3D deep learning paradigms. Guardrail detection demands the development of more adaptive feature learning frameworks. Two persistent overarching challenges remain across static perception tasks:

- 1) **Data Dependency:** The labor-intensive nature of annotating 3D point clouds for training and validation remains a significant bottleneck. Promising solutions involve exploring synthetic data generation [114] and transfer learning techniques to reduce annotation burdens.
- 2) **Unified Frameworks:** Current pipelines typically process different map elements separately. There is an emerging need for end-to-end models capable of jointly extracting both ground and non-ground features [133] within a unified framework to improve consistency and efficiency.

2.6. Topology Generation

As the "road genome" of autonomous driving systems, topological maps abstractly describe key nodes (e.g., intersections, lane points) and connection relationships (e.g., lane centerlines, traffic rules) in road networks. They serve as core elements for efficient path planning and navigation decision-making. Topological maps offer high storage efficiency, low computational burden, and easy dynamic updates, which makes them more suitable for real-time localization and environmental understanding in autonomous vehicles. With the advancement of

autonomous driving technology, topology generation methods have evolved from traditional geometric algorithms to data-driven approaches. Based on different technical routes and core ideas, as shown in Fig. 13, topology generation methods can be divided into four categories: graph search and geometric constraints methods, deep learning-based methods, hierarchical structures and progressive reasoning methods, and multimodal fusion and map prior-based methods. In this subsection, we analyze the core ideas, advantages, and limitations of each category, and provides prospects for future research directions.

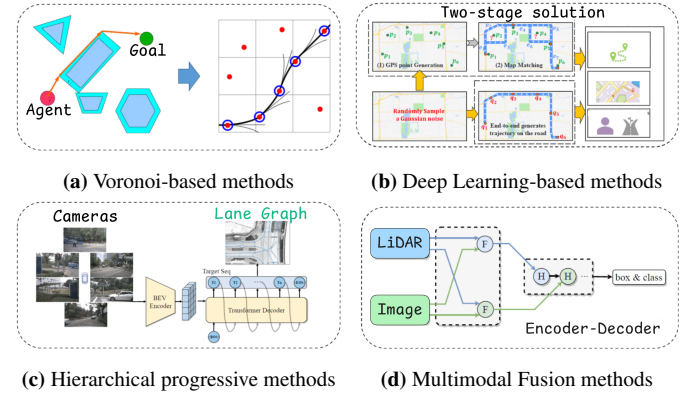


Fig. 13: Four categories of topologic generation methods

2.6.1. Graph Search and Geometric Constraints

This category is grounded in graph theory and computational geometry. It achieves path planning by constructing a Configuration space (C-space) and a Generalized Voronoi Diagram (GVD). The core lies in transforming physical spatial obstacles into topological constraints and employing heuristic search algorithms to identify optimal paths. [18] proposed a product structure algorithm for high-dimensional configuration spaces, reducing complexity through topological decomposition, significantly enhancing the planning efficiency of multi-robot systems. [160] introduces a Voronoi-guided hybrid A* algorithm, which integrates the topological separation characteristics of Voronoi diagrams with the Reed–Shepp kinematic model. [119] offers theoretical rigor and provable safety. Its typical framework consists of three levels: environmental topology modeling, path space decomposition, and conflict avoidance, making it suitable for multi-agent collaborative navigation in structured environments. [70] employs direction-constrained search and real-time replanning mechanisms to control the collision rate in dynamic scenarios, addressing the limitations of traditional methods in responding to environmental dynamics.

Though these methods have achieved satisfactory results in some scenes, they still face several challenges. Firstly, the accuracy of dynamic obstacle prediction depends on the completeness of environmental modeling. Secondly, the computational complexity of high-dimensional configuration spaces grows exponentially with the number of agents. Thirdly, generated paths require post-processing to meet vehicle dynamic constraints.

2.6.2. Generative Models and Deep Learning

Generative models learn the distribution patterns of topological structures in a data-driven manner, with their core advantage being the reduction of manual design costs and the generation of road networks that conform to real statistical characteristics. The typical workflow of these methods consists of three stages: data preprocessing, topological feature learning, and controllable generation, making them suitable for large-scale road network construction in autonomous driving simulation scenarios. Early approaches were represented by GANs, such as [142], while diffusion models have recently become mainstream, improving generation quality through latent space topological constraints. For example, structure-aware diffusion models [188] incorporate persistent homology theory to address the issue of fragmented generated trajectories, improving topological accuracy by 32%. Recently, language-guided traffic simulation models [236] fuse Large Language Models (LLMs) with diffusion models, enabling road network generation controlled by natural language instructions.

Despite the promising performance of these methods in certain scenarios, several challenges persist. Firstly, generation quality heavily relies on large-scale annotated data, leading to poor generalization in few-shot scenarios. In addition, the black-box nature of the diffusion process makes topological logic difficult to trace. Furthermore, generated results often violate actual traffic rules. Future work needs to develop few-shot learning paradigms, explore topology-aware interpretability visualization techniques (such as persistence diagrams), and establish generation mechanisms constrained by physical rules.

2.6.3. Hierarchical Structure and Progressive Reasoning

Hierarchical progressive methods simulate human cognitive processes by decomposing complex topological problems into parallelizable subtasks through hierarchical abstraction. The core concept is "local construction-global integration", and a typical framework consists of a feature extraction layer, a topology reasoning layer, and a path optimization layer. The polyline Transformer detection network [77] employs vector representations instead of pixel-level masks, increasing inference speed compared to traditional methods, validating the efficiency advantages of hierarchical representation. The multi-agent hierarchical control system [228] introduces mid-level language-driven commands to decouple high-level reasoning from low-level control. In addition, [197] proposes a lane graph autoregressive expansion framework, which achieves modeling of complex structures like roundabouts through incremental node addition and dynamic adjacency matrix updates.

For this kind of methods, the current bottlenecks lie in inter-level error propagation and real-time performance balancing. Autoregressive expansion models are prone to cumulative errors, and multi-agent collaboration requires solving communication latency issues. Future research should explore hierarchical error correction methods based on attention mechanisms, develop edge computing architectures to enhance real-time responsiveness, and integrate spatiotemporal hierarchical representations to improve adaptability in dynamic environments.

2.6.4. Multimodal Fusion and Map Priors

These methods construct robust environmental perception models by integrating heterogeneous sensor data with map prior knowledge. Its technical core lies in cross-modal attention mechanisms and a fusion architecture with prior constraints, which can compensate for performance degradation of single sensors in occluded and noisy scenarios. [212] proposes a multimodal fusion transformer that aligns camera semantics with LiDAR geometric features through cross-attention. [210] designs a modal interaction encoder with a dual-stream transformer architecture that preserves modality specificity while enabling directed information exchange. In [144], authors propose a cross-modal attention topology reasoning framework that fuses standard map data with satellite imagery, achieving a topology inference F1-score of 0.87 in GPS-deprived scenarios.

Although these methods offer feasible exploration directions, they suffer from several main challenges including modal alignment accuracy, prior bias, and computational complexity. Spatiotemporal registration errors in heterogeneous data can easily lead to topological misjudgments, while the timeliness of map priors affects adaptability in dynamic scenarios. In addition, the multimodal feature processing increases hardware computational demands. Future work needs to establish dynamic update models for prior knowledge, and design lightweight fusion architectures suitable for vehicular embedded platforms.

2.7. Summary And Limitations

In this section, we start from introducing the production pipeline and map specification of HD map. Through it, readers can have an overview of HD maps and have knowledge of what kind of map elements are required for autonomous driving. After, we elaborate on the following key technical points: localization, multi-trip mapping, static perception and topology generation, covering both classic and cutting-edge papers.

HD maps make autonomous driving truly possible for the first time. The industry has relied on HD maps to achieve L2 to L4 autonomous driving on highways. However, it is difficult to extend it to urban roads that are featured with more complex topology and more frequent construction or repair. This is because the high production cost, low update frequency, lack of quick response to reality change and significant data transmission of HD maps. These limitations discourage the AD industry from relying on HD maps on urban roads.

3. Lite Map: The Conquest of Urban Roads

By the end of 2021, most AD companies had completed commercial applications on highways, but the practice of AD on urban roads was still a challenge for them at that time. Both industry and academia started to put effort on finding an economically acceptable solution for urban roads. The concept of Lite maps started to appear and it might be most appropriate solution till today and even in the future.

3.1. Comparison of Lite and HD Map

The Lite map is an effective and acceptable solution for facilitating urban-road autonomous driving. Fig. 14 illustrates the differences between the Lite map and HD maps. More specifically, the differences lie in the following aspects:

- 1) **Vehicle for map collection:** Professional surveying vehicles are used for HD map collection. These vehicles are normally owned by map providers for commercial use. In contrast, Lite maps rely on the data collected by production cars which are mass-produced cars and offered for sale to the public.
- 2) **Sensors:** Surveying-grade sensors are used for HD map collection while consumer-grade sensors are used for Lite maps. The latter are relatively much cheaper.
- 3) **Production pipeline:** For HD maps, raw sensor data are transmitted directly via hard disks. Almost all mapping and perception algorithms are post-processed in a data center. In the pipeline of Lite map production, by contrast, data are collected through a crowdsourcing manner. Every car plays the role of a edge computer. The single-trip mapping and perception algorithms directly run on cars. Raw sensor data are discarded after on-board vectorization for user's privacy concern. Only vectorized elements are uploaded to a data server using cellular data network on condition that change detection reality changes are detected on board.
- 4) **Map elements:** Lite maps can be regarded as a simpler version of HD maps. The map elements of Lite maps are simpler and with lower accuracy, but are sufficient for the use of autonomous driving. For example, the mapping of traffic lights and signs does not need centimeter-level accuracy at all.
- 5) **Research focuses:** After inheriting the algorithms of HD maps, Lite maps are faced with new challenges. Research is focused on online vectorization, online change detection, Map differential update and traffic-flow trajectory mining. Some research points are newly encountered in the Lite map production. We will elaborate on them in this section.

Except for the differences between Lite map and HD map, there are also many things in common, including file management, element specifications, localization and some object detection algorithms. Since these have already been discussed in Section. 2, we will focus on these differences in this section.

3.2. Online Vectorization

3.2.1. Task Definition

The task of online vectorization or online map learning takes as input onboard sensor observations, which normally are surrounding images and LiDAR scans, and output vectorized road

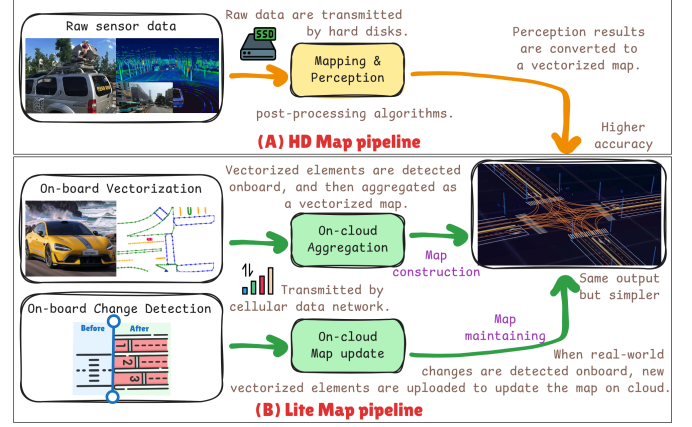


Fig. 14: Comparison of Lite and HD map in terms of production pipeline and research focus.

elements such as lanes, boundaries, crosswalk and intersections. This task is significant in modern autonomous driving systems. It not only facilitates the Lite map construction, but also helps real-time motion prediction and planning. Although many of recent papers still refer to online vectorization as HD map learning, we still argue that it should be called Lite map due to the fact that neither its mapping accuracy nor element abundance meets the criterion of HD maps. Here we briefly divide the online vectorization, a real research hotspot in recent years, into two categories: single-frame detection and long-sequence modeling.

3.2.2. Single-frame Detection

The single-frame detection consumes surrounding images and/or LiDAR scans at a single time t , and then directly generated vectorized elements. It is independent at any timestamps and does not rely on previous or feature observations.

HMapNet [106] proposes a novel projection module that converts features from perspective view to bird's-eye view when depth is missing. Then, a BEV decoder outputs semantic segmentation, instance embeddings and lane directions. Finally, a post-processing module is used to vectorize these predicted instances. The limitation, however, is that HMapNet relies on the rasterized map predictions, and its heuristic post-processing step restricts the model's scalability and performance.

The idea behind HMapNet is segmentation followed by a post-processing DBSCAN [52]. The whole framework is indirect. By contrast, VectorMapNet [122] solves the problem in an end-to-end manner. VectorMapNet is the first work designed for end-to-end vectorized map learning from onboard observations. It takes advantage of DETR [29] to directly formulate the map learning as a detection task. VectorMapNet uses map queries to represent road elements in a Transformer decoder. The map queries are similar to the object queries used in DETR. The architecture of VectorMapNet comprises a BEV feature extractor, a DETR-like map element detector and a polyline generator. Since the position and shape of map elements predicted

by the detector are not satisfied, the polyline generator then focuses on refining the detailed geometry of the map elements. This is actually a two-stage coarse-to-fine framework.

Unlike VectorMapNet, MapTR [116] proposes a single-stage framework. It designs a hierarchical query embedding scheme to flexibly encode structured map information and perform hierarchical bipartite matching for map element learning. The parallel and structured framework makes learning process higher efficiency. However, the limitation of MapTR is that it is designed to model map elements lacking physical directions, which means it does not account for map elements with practical directions such as centerlines. This will hinder the use of downstream motion prediction and planning. As an improved version, MapTRV2 [117] also builds on the permutation-equivalent modeling, but does not permute centerlines and directly chooses the given order as targeted permutation. Therefore, MapTRV2 is capable of detecting centerlines with directions.

The biggest limitation of single-frame detection is its small range of perception which is normally 60×30 meters. Worse still, in complex scenarios such as occlusions, the perception range is severely affected due to the loss of visual information.

3.2.3. Long-sequence Modeling

Different from the single-frame detection which only consumes images/scans at a single timestamp, the long-sequence modeling leverages historical observations in a long sequence to boost the range of perception. StreamMapNet [216] proposes a streaming strategy that encodes all historical information into the memory feature to save cost and build long-term association. This idea is borrowed from 3D objection detection works: StreamPETR [184], Sparse4Dv2 [118] and VideoBEV [67]. More specifically, the strategy selects the former foregoing top k queries based on confidence score as potential tracking queries, and then concatenates them with initialized queries. Note that the network is also DETR-like and query-based. As a result, the perception range of the local vectorized map is extended to 100×50 meters without compromising map quality.

MapTracker [35] argues that the aforementioned methods detect road elements anew in every frame, potentially guided by previous frames without consistency enforcement. Inspired by MeMOT [26] and MeMOTR [59] which design memory mechanisms in tracking Transformer to preserve long-term consistency, MapTracker formulates the map learning as a tracking task using memory-based raster and vector latents for temporally consistent reconstruction. It explicitly associates tracked map elements from historical frames to further enhance temporal consistency.

MapExpert [218] raises another concern. Traditional visual detection targets are normally cube-like objects such as vehicles, pedestrians and animals that can readily represented by bounding boxes. However, the road elements in map learning task varies significantly. For instance, lanes are often smooth curves but crosswalks are usually rectangular. Such variability poses a significant challenge for DETR-like decoders. To tackle the problem, MapExpert proposes an expert-based method that addresses the challenges of representing diverse non-cubic map

elements. It introduces sparse experts guided by routers, a novel auxiliary balance loss for expert load distribution, and a Learnable Weighted Moving Descent module for efficient temporal BEV fusion. MapExpert achieves state-of-the-art accuracy and efficiency on nuScenes [25] and Argoverse2 [193] benchmarks.

3.2.4. Overview And Limitations

Overall, the above-mentioned methods can also be summarized as segmentation-based, detection-based and tracking-based according to their formulation, as illustrated in Fig. 15. Additionally, we provide a table listing recent open-source work on online vectorization. Readers can refer to Appendix.A for quick access.

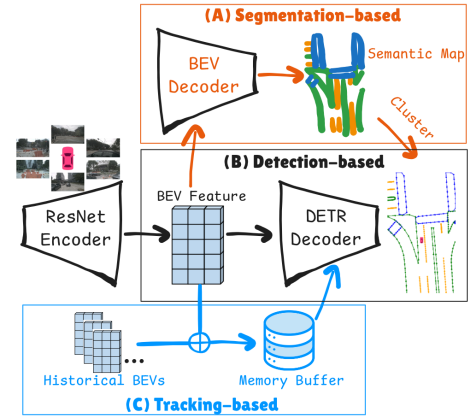


Fig. 15: An architecture overview of three categories of online vectorization. (A) The segmentation-based methods generate a semantic map through a BEV decoder and then clusters it into structured elements. (B) The detection-based methods utilize a DETR-like decoder to directly detect map elements from BEV features in an end-to-end manner. (C) The tracking-based methods are featured with a memory buffer that enhances temporal consistency by aggregating historical BEV features so as to support more stable and accurate map reconstruction.

Despite the increasing popularity on the online vectorization for the online Lite map learning, the state-of-the-art performance in terms of both mapping accuracy and element abundance is still far from HD map. Many of recent works can only generate lane (centerline), boundary and crosswalk which are obviously not enough to support autonomous driving. Therefore, the online vectorization is still an open problem. We advocate the proposal of datasets with more element annotations from industry. In addition, methods regarding zero-shot, few-shot and transfer learning should be considered to enable networks to detect more types of road elements.

3.3. Crowd-sourcing Map Maintenance

Map maintenance strongly correlates with two tasks: change detection and map update.

Change detection: The change detection aims to identify road structural changes that are misaligned with map database. We briefly classify this task into two types: post-processed change detection and online change detection. In

post-processed change detection, changes are detected after data (e.g., constructed maps or images) are transmitted to a data server. On the contrary, online change detection is run onboard and only when changes are detected beforehand will data be uploaded.

Map update: The map update is to replace outdated map data with new one. It can also be divided into two types: direct update and incremental update. In direct update, the map for a tile or patch (we have introduced in Section. 2.2.1) is re-surveyed and will replace an old tile or patch directly. The remaining work is topological connections with adjacent tiles and patches. By contrast, the incremental update operates on smaller unit, namely roads. It replaces a road with an updated one and fix its topology. Therefore, the incremental update is also called road-level update.

Post-processing change detection and direct update are generally used in traditional HD map maintenance. [191] provides a well overview on them. Related works usually involve semantic map comparison [87, 222] and images comparison [90, 88] of the same location at two different dates. However, such maintenance frameworks would be highly impracticable for modern crowdsourcing map maintenance. The reason is threefold:

- 1) Volume of data transmission. Since the transmission is done by consuming users' data plan subscribed to a mobile network operator, it is impractical to send large data packages (e.g., high-resolution images or point cloud maps) from onboard systems.
- 2) Users' privacy. Images and trajectories contain much sensitive information. Uploading such data is at a risk of violating users' privacy.
- 3) Map freshness. The change of road structure and topology should be quickly detected and updated to maintain map freshness.

As we discussed in Section. 2.7 that urban roads are characterized by more frequent construction and repair compared to highways, the Lite map is hence required to be daily updated for safe autonomous driving. Reliable online map learning together with online change detection seems to be the correct solution. Taking advantage of massive crowd-sourced data, many AD companies nowadays employ a multi-source updating system for dynamic map maintenance. A typical system framework is illustrated in Fig. 16, where three data sources contribute to map updates: change detection (Section. 3.3.1), reward route (Section. 3.3.2) and traffic flow mining (Section. 3.3.3).

3.3.1. Online Change Detection

The online change detection is designed to detect structural or topological road differences from the pre-downloaded map. During normal driving, data will not be uploaded to avoid excess data transmission and privacy violation. But when a change is detected, the local Lite maps learned by online vectorization will be uploaded for map updating. Sometimes, low-resolution images will be uploaded as well for manual quality assessment.

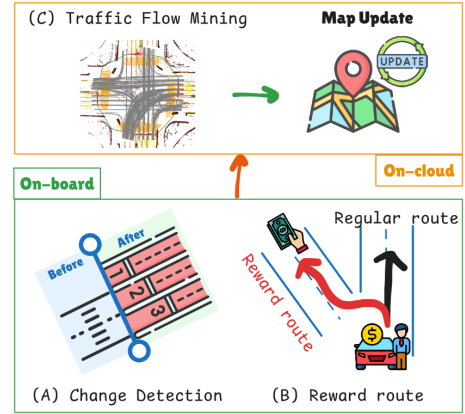


Fig. 16: A modern system for dynamic map maintenance through crowd-sourced vehicle data. It is divided into on-board and on-cloud components. (A) Change Detection: identifying road structural changes (e.g., lane modifications) by comparing before-and-after states, and (B) Reward Route: drivers are incentivized to take specific routes to gather valuable data. (C) Traffic Flow Mining: extracting road elements by analyzing traffic patterns. This collaborative framework ensures that Lite map remains accurate and up to date.

ExelMap [192] rethinks map update as element-based change detection. It proposes the first end-to-end explainable element-based map change detection architecture. The architecture is based on LaneSegNet [110] which consists of a BEV encoder, a Transformer decoder and a lane segment predictor. ExelMap further adds two distinct, independently operating heads with binary output for element-wise deletion and insertion detection. Both change detection heads are simply linear layers. The final output are twofold: an updated map representation and a novel change map with a change status: unchanged, inserted or deleted.

CleanMAP [163] introduces a Multimodal Large Language Model (MLLM)-based distillation framework that assigns quantitative confidence scores (0–10) to crowdsourced map data by evaluating lane visibility parameters (e.g., blur, lighting, occlusion). It employs a dynamic piecewise confidence-scoring function coupled with a confidence-driven local map fusion strategy to rank and fuse the most reliable local maps for updates. CleanMAP ensures that changes in local maps are accurately aligned with the confidence score-based map, maintaining consistency and accuracy in the map update.

RTMap [49] handles road change detection as part of its real-time, on-vehicle recursive mapping system. As the vehicle navigates, RTMap continuously compares new sensor observations against the existing pre-downloaded map. It uses a probabilistic model that considers positional uncertainty to determine whether structural changes have occurred in the road environment. Detected changes are treated as loop-edge constraints within RTMap's pose graph optimization framework. This enables the system to align submaps consistently and update the global map structure on the fly, supporting ongoing fusion of multiple agent observations to refine road element positions and update the prior HD or Lite map over time.

The open problems for online change detection are twofold:

- 1) **Rate of false alarm.** According to the author's observation in the industry, most companies' online change detection algorithms have a high false alarm rate. Many change notifications uploaded to the cloud are wrong due to various reasons. These reasons may include temporary road occupation, perception errors and sensor occlusions.
- 2) **Online map repair.** In current autonomous driving practices, if a road change is detected, the AD system will call the user to take over the vehicle. This will inevitably affect the user experience. Therefore, it is a very meaningful task to repair the car-side map in time according to the change detection results.

These interesting yet challenging problems are real demands from AD companies. We urge the academia to pay more attention on them.

3.3.2. Reward Route

For some newly built remote roads where few vehicles travel on them, it would be difficult to quickly build Lite map by spontaneous crowd-sourcing data alone. As a supplementary data source, the navigation software provides a reward route for nearby users. Users can choose to accept the reward route assignment or reject. As shown in Fig. 16(B), if accept, users will be navigated to make a detour for data collection. In return, users will get a cash reward. Many companies, such as Uber and Huawei, have adopted such reward mechanism to improve crowd-sourcing update speed. The key research question is how to reasonably distribute reward routes and identify potential users.

3.3.3. Traffic-flow Trajectory Mining

Massive users provide anonymous vehicular trajectory data. The task of traffic-flow trajectory mining is to extract useful information from these data, contributing to on-cloud Lite map maintenance. Here, we summarize the possible road elements that can be mined with some representative works as follows.

- 1) **SD road network.** [60] proposes a geospatial big data-driven framework for automatic urban road geometry extraction. The test is conducted on Didi Chuxing GAIA [48] which is a publicly available crowd-sourcing taxis dataset. The framework consists of three main steps: trajectory compression to reduce data noise and redundancy, clustering to identify road centerlines, and vectorization to generate structured road map elements. [102] presents DF-DRUNet, a decoder fusion model that combines remote sensing images and GPS trajectories for road network extraction. It uses two parallel dilated Res-U-Net to effectively learn complementary features.
- 2) **Intersection topology.** [229] proposes CITT, a three-phase framework for calibrating road intersection topology using vehicle trajectory data. It includes trajectory quality improvement, core zone detection, and topology

calibration within the intersection influence zone, identifying incorrect or missing turning paths by analyzing unmatched trajectories. CITT effectively handles noisy data and diverse intersection shapes, achieving high-accuracy intersection topology updates. [147] proposes a fully automatic and scalable framework for generating topological maps of complex intersections by leveraging crowd-sourced semantic data and traffic flow information. The system infers intersection topology directly from observed vehicle movements without manual annotation.

- 3) **Intersection geometry.** [104] proposes a novel intersection extraction method based on GNSS vehicle trajectories. It applies mean-shift clustering to trace road segments and estimate accurate road orientations. Dominant orientations around a location are then used to determine intersection positions, ensuring both geometric and topological correctness. [168] detects intersection boundaries via turning angle clustering, identifies entrances and exits using a novel approach, and reconstructs geometric structures with assigned turning rules. [32] enhances intersection detection from noisy, low-sampling GNSS data by proposing a three-step method. The steps include (a) stay-point detection: to filter out false turn-points unrelated to intersections, improving the reliability of turning data; (b) turn-point compensation: to estimate accurate turning locations by analyzing geometric patterns and hotspot areas, using a density-based turning angle model to refine sparse trajectory data; and (c) post-classification: to refine intersection candidates by clustering and classification using geometric features.
- 4) **Interchange networks.** [86] introduces a novel method for generating accurate road interchange networks using a forward and reverse tracking mechanism based on long-term trajectory continuity. The process includes (a) pre-processing: to filter out non-interchange and abnormal trajectories; (b) subgraph extraction: to detect potential transition nodes (where roads diverge or converge); (c) forward and reverse Tracking: to identify divergence nodes through forward tracking and convergence nodes; and (d) two-stage fusion: to merge forward and reverse tracking results.
- 5) **Centerline.** [33] extracts centerlines from low-frequency GPS trajectory data using a divide-and-conquer, intersection-first approach. This approach effectively balances precision and recall under low-sampling conditions, producing smoother and more connected road centerlines than traditional KDE-based approaches. [44] employs morphological analysis to generate elongated polygons that represent road ribbons, followed by skeletonization to reconstruct the road centerlines and the overall road network. The method is based on raw trajectory polylines, avoiding raster resampling.
- 6) **Road boundary.** [209] proposes a road boundary extraction approach using Delaunay triangulation (DT) on

crowdsourced GPS vehicle traces. It constructs DT and Voronoi diagrams over the interpolated trajectories to derive geometric descriptors (e.g., Voronoi cell area, triangle edge length). Then, a boundary detection model followed by a region-growing algorithm is applied to extract complete road boundaries.

- 7) **Lane** [8] proposes a three-step automatic method for constructing lane-level routable digital maps from trajectory data. The steps include a QuickBundles clustering for road node detection, a Fréchet distance-based dissimilarity matrix, and a Gaussian Mixture Model (GMM) for estimating lane positions as well as lane numbers.

There are many other applications of traffic-flow data such as crash rate analysis, real-time traffic dispatching and travel time estimation (ETA). They will not be discussed further in this paper because we are more focused on extracting static map elements using traffic-flow.

3.4. Summary And Limitations

In this section, we first compare the similarities and differences between HD maps and Lite maps, highlighting the new challenges encountered during the Lite map stage. Then, we conduct an in-depth discussion on the research hotspots in the Lite map, including online vectorization, online change detection and traffic-flow trajectory mining. Through this section, readers should be able to understand why Lite maps are more advanced than HD maps.

Lite maps have helped autonomous driving companies expand from highways to urban areas, realizing the true commercialization of self-driving cars. The explicit map representations connect modular tasks (i.e., perception, prediction and planning), but they also make AD systems suffer from accumulative modular errors and deficient task coordination. This makes us to think whether the Lite map is the ultimate form for AD systems? The answer is still unclear now, but researchers are working on more human-like end-to-end AD systems where maps are implicitly represented in neural networks.

4. Implicit Map: March Towards End-to-End AD

As shown in Fig. 17, this section aims to systematically review the Implicit map in autonomous driving, structured into four kinds of technologies. First, query-based representations are examined through three paradigms: generative goal-directed synthesis, context-conditioned formulation, and unified cross-task architectures. Subsequently, latent space methods integrate spatial structuring and temporal dynamics to compress high-dimensional sensory data into structured latent spaces. Next, Neural Radiance Fields (NeRF)-based approaches focus on 3D scene reconstruction, semantic understanding, and efficiency optimization techniques. Finally, world models encompass environmental representation, dynamic evolution modeling, and model enhancement technologies for predictive simulation. Subsequent sections will critically analyze strengths, limitations, and trade-offs including efficiency-

expressiveness balance and safety verification to guide future research directions. Fig. 18 provides the chronological overview of several representative implicit map methods.

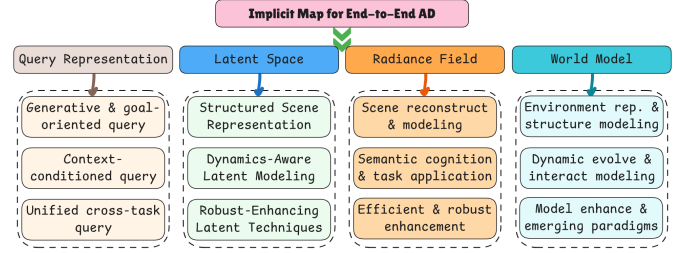


Fig. 17: Implicit map-based methods for autonomous driving

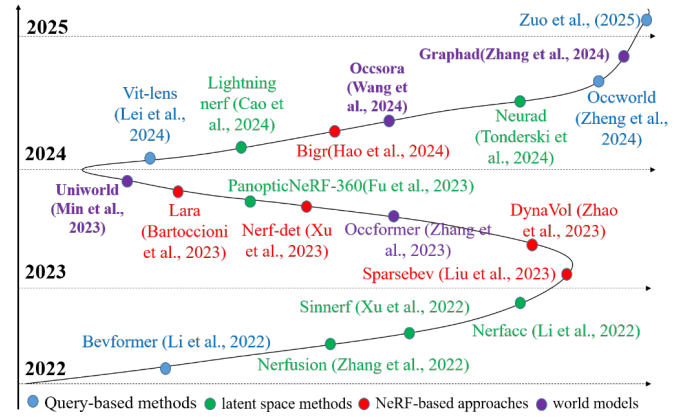


Fig. 18: Chronological overview of several representative implicit map methods

4.1. Query Representation Methods

Query-based representations have revolutionized end-to-end autonomous driving by enabling structured interaction between perception, prediction, and planning modules. As shown in Fig. 19, existing works can be divided into three methodological lenses: generative and goal-oriented query synthesis, context-conditioned query formulation, and unified cross-task query architectures. We critically examine how these approaches balance expressiveness, efficiency, and interpretability while identifying fundamental limitations in temporal coherence, causal reasoning, and safety verification that impede real-world deployment.

4.1.1. Generative & Goal-Oriented Query Synthesis

This paradigm leverages goal-directed mechanisms to produce diverse, multimodal trajectory hypotheses. Potential-based diffusion motion planning [125] pioneers diffusion-based query generation for probabilistic trajectory sampling, capturing complex maneuver distributions through composable potential functions. Trajectory Prediction using LSTM with Attention mechanism [5] injects intent awareness through goal-conditioned queries, enabling human-like anticipation of agent

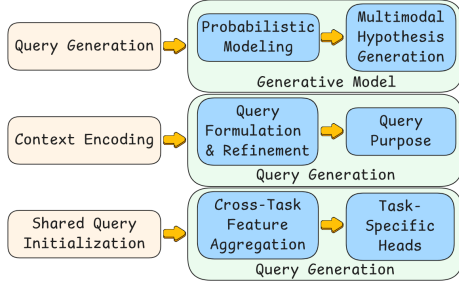


Fig. 19: Three methodological lenses of query-based representations

objectives by weighting historical trajectory features. GaussianWorld [246] extends continuity by modeling trajectories as queryable spatiotemporal Gaussian fields, while Tracking dynamic flow [172] integrates optical flow priors to anchor predictions in observed motion dynamics. Point Cloud Forecasting as a Proxy for 4D Occupancy Forecasting [93] further generalizes this concept to 4D occupancy spaces by rendering point clouds from predicted occupancy.

These methods demonstrate their distinct strengths. The robust probabilistic modeling capability enables the capture of multimodal trajectory distributions in complex scenarios. Besides, the generative framework inherently supports uncertainty quantification, providing a theoretical basis for risk assessment. Despite their representational power, these methods exhibit several fundamental limitations:

- 1) The trade-off between the computational efficiency and representational power: Probabilistic modeling (e.g., diffusion processes) requires hundreds of iterative refinements, conflicting with real-time planning constraints.
- 2) Input quality dependency: Goal-conditioned approaches assume perfect intent recognition and environmental perception, which degrade under noisy or ambiguous inputs.
- 3) Safety mechanism deficiency: None of the generative frameworks incorporate formal verification of safety boundaries, risking physically invalid trajectory proposals.

4.1.2. Context-Conditioned Query Formulation

These techniques contextualize queries using linguistic, temporal, or structural cues to align with human cognition and environmental constraints. DAGGER-Q [138] early combined query mechanisms with imitation learning, using confidence and risk queries to dynamically trigger expert interventions. This reduced data annotation costs and provided an early framework for context-driven efficient learning. A Language Agent for Autonomous Driving [130] and OccWorld [234] fuse natural language commands with motion/occupancy queries for human-aligned planning, while QA-Drive [92] uses question-answer pairs to generate explainable planner justifications. DP-Bridge [83] ensures temporal consistency through decoupled past/future queries, and IC-Mapper [243] encodes road topology priors for map-consistent perception. Think-on-Graph [165] introduces traffic rule verification via knowledge graph

queries, and MAPO [39] employs LLMs to abstract high-level decisions into symbolic queries.

These approaches naturally integrate human instructions and common-sense knowledge, enhancing the intuitiveness of human-machine interaction, and dynamically adjusting query strategies to adapt to scene changes, improving robustness in complex environments. However, they suffer from several core challenges:

- 1) Ambiguity in cross-modal grounding: Natural language commands often contain spatial vagueness (e.g., "near the tall building") which resists precise geometric interpretation.
- 2) Causal explainability gap: QA-based justification frameworks correlate with planning decisions but fail to establish causal relationships between query features and output actions.
- 3) Temporal robustness limitations: Historical context accumulation introduces drift in long-horizon planning, particularly in dynamic environments with occluded agents.

4.1.3. Unified Cross-Task Query Architectures

This category unifies disparate driving tasks through shared query representations that bridge perception-planning gaps. Neural Motion Planner with Query-based Occupancy [45] first proposed planning-driven sparse spatiotemporal queries, enabling direct perception-planning interaction. BEVFormer [115] efficiently aggregated spatiotemporal features via dynamic BEV queries, with its sparse query mechanism widely adopted by subsequent frameworks like UniAD [78]. Meta-Query [186] enables few-shot adaptation to novel scenarios through learnable query initialization.

Unified cross-task query architectures break down traditional modular barriers, enabling end-to-end optimization from perception to planning. They efficiently share multimodal features, significantly reducing information redundancy; and enhance complex task coordination capabilities, supporting multi-objective optimization in dynamic scenarios. However, they encounter multi-task optimization conflicts, where perception accuracy and planning safety objectives compete in gradient descent. Additionally, they face the exponential growth in system complexity with task scale, failing to meet scalability requirements for urban-level scenarios.

4.1.4. Summary

Generative and goal-oriented methods capture multimodal trajectory distributions through probabilistic modeling, providing uncertainty quantification for complex scenario decision-making, yet face inherent conflicts between computational efficiency and real-time performance. Context-conditioned approaches achieve deep integration of human intent and environmental features, enhancing system adaptability and explainability, while struggling with contextual ambiguity resolution and causal reasoning bottlenecks. Unified cross-task architectures break traditional modular barriers, enabling end-to-end perception-planning optimization via shared latent spaces, but

are constrained by multi-objective conflicts and safety verification complexity. These three categories collectively advance query-driven autonomous systems, yet none have fully resolved the robustness-efficiency-safety trilemma in dynamic open environments.

4.2. Latent Space Representation Methods

Latent space representations have emerged as a critical enabler for end-to-end autonomous driving systems, compressing high-dimensional sensor data into compact, structured feature spaces for efficient perception, prediction, and planning. As shown in Fig. 20, we categorize the existing methods into three core methodological paradigms: Structured Scene Representation, Dynamics-Aware Latent Modeling, and Robustness-Enhancing Latent Techniques.

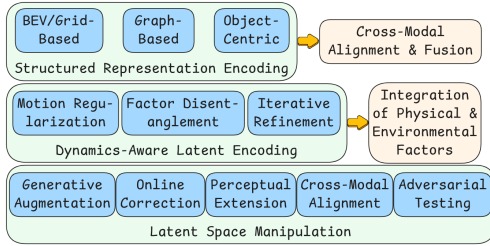


Fig. 20: Three methodological lenses of latent space representation methods

4.2.1. Structured Scene Representation

This paradigm focuses on geometrically consistent and topologically meaningful latent organization of driving scenes. SparseBEV [120] optimizes efficiency through sparsified convolutional latent grids, addressing computational bottlenecks in real-time applications with its fully sparse architecture for panoptic occupancy prediction. Object-centric organization is pioneered by DynaVol [231], which allocates dedicated latent slots to dynamic objects via 3D voxelization, and ViT-Lens [101], which employs hierarchical transformers for multi-task feature sharing across omni-modal inputs. Cross-modal alignment is achieved in LaRa [15] through radar-augmented BEV fusion and COLA [157] via contrastive learning for sensor-invariant representations.

These methods excel in efficient spatial information compression by transforming high-dimensional sensor data into compact latent spaces through techniques such as BEV projection and sparse convolutional grids, enabling real-time scene perception as demonstrated in BEVFormer and SparseBEV. Furthermore, they exhibit strong multi-modal fusion capabilities, facilitating cross-sensor alignment (e.g., radar-augmented BEV fusion in LaRa and contrastive learning for sensor-invariant representations in COLA). However, they face inherent limitations: BEV projections suffer from depth estimation errors and occlusion artifacts, leading to feature degradation in distant or occluded regions; graph-based approaches like VectorMapNet rely on predefined topological templates, limiting adaptability to novel road layouts (e.g., temporary construction

zones); and object-centric methods depend heavily on upstream detection accuracy, resulting in cascading errors for small or ambiguous objects.

4.2.2. Dynamics-Aware Latent Modeling

These techniques embed temporal dynamics and motion priors directly into latent spaces for accurate trajectory forecasting and planning. SceneFlow [196] integrates optical flow as motion regularization, enabling smooth latent space transitions across frames through self-supervised learning of rigid flow consistency. DriveLatent [240] disentangles static and dynamic scene factors via a variational autoencoder, allowing separate manipulation of environmental elements and agent motions. LatentDiffuser [111] leverages diffusion models within low-dimensional latent spaces to refine planning trajectories iteratively, bridging perception and control with energy-guided sampling. Physical environment interactions are encoded in Multi-Body Neural Scene Flow [179], which embeds terrain roughness features for off-road planning via isometry-regularized flow estimation, and HADNet [162], which uses holistic attention to fuse agent-road-traffic light dynamics.

These methods effectively embed temporal dynamics into latent spaces through different techniques, enhancing trajectory prediction accuracy for complex multi-agent interactions. They support multimodal future trajectory generation via diffusion models (LatentDiffuser) and adaptive motion priors, accommodating diverse traffic scenarios. Additionally, these methods integrate physical environmental cues such as terrain roughness and road-traffic light dynamics, improving adaptability to unstructured environments. Nevertheless, critical challenges persist. Firstly, computational complexity remains high due to iterative optimization in diffusion-based planners and multi-body motion decomposition, hindering real-time performance for high-speed driving. Secondly, nonlinear motion modeling (e.g., emergency maneuvers) remains inadequate, leading to flow field distortions in extreme scenarios. Finally, vulnerability to data sparsity limits generalization to rare events like accidents or sudden obstacles, increasing overfitting risks.

4.2.3. Robustness-Enhancing Latent Techniques

This category addresses domain adaptation, sensor degradation, and scenario generalization challenges through latent space manipulation. GELA [173] employs diffusion-based latent augmentation to synthesize rare scenarios, enhancing robustness to long-tail events via conditional diffusion models on dynamic graphs. LAC [68] corrects calibration errors through online affine transformations in binary latent codes, maintaining performance under sensor misalignment. [128] generates latent maps beyond sensor range using contrastive learning-equipped GANs, extending perceptual horizons with 14.6% higher MAP than traditional methods. Adversarial robustness is addressed by GANav [164], which tests planners against latent-space perturbations, while [84] aligns visual and linguistic latent spaces for improved command interpretation via contrastive language-image pre-training.

These methods strengthen model generalization through generative data augmentation (e.g., GELA’s diffusion-based syn-

thesis) and adversarial training (e.g., GANav’s latent space perturbations), addressing long-tail events such as adverse weather and sensor failures. They achieve cross-domain adaptability via contrastive learning (COLA) and vision-language alignment (CLIP-Drive), enabling sensor-invariant and scene-agnostic representations. Computational efficiency is optimized through binary latent coding (EcoLane) and sparse feature enhancement (LAC), making them suitable for onboard embedded systems. However, these techniques exhibit notable limitations: 1) Generative augmentation heavily depends on high-quality real-world data distributions, risking physically implausible samples in data-scarce scenarios (e.g., rural roads); 2) adversarial training and domain adaptation lack quantifiable robustness boundaries, leading to performance degradation under unseen sensor degradation modes (e.g., lens contamination); and 3) safety validation mechanisms for augmented samples are absent, potentially introducing misleading features such as unsafe driving patterns in extreme scenario generation.

4.2.4. Summary

These methods advance latent space representation from spatial structuring, temporal dynamics, and robustness perspectives, respectively, yet all face fundamental trade-offs between physical consistency and computational efficiency. Future research should prioritize integrating geometric-dynamic joint modeling to mitigate cascading errors from BEV projection and motion estimation, embedding physical constraints into generative augmentation to ensure safety and realism of synthetic samples, and developing lightweight architectures to balance model complexity with real-time inference requirements for autonomous driving systems.

4.3. Neural Radiance Fields based methods

With the groundbreaking progress of Neural Radiance Fields (NeRF) in 3D scene representation, their application in autonomous driving is undergoing a critical transition from theoretical exploration to system integration. As shown in Fig. 21, we categorize the existing works into three kinds: the scene reconstruction and modeling methods, semantic understanding and task applications methods, and efficiency and robustness enhancements methods.

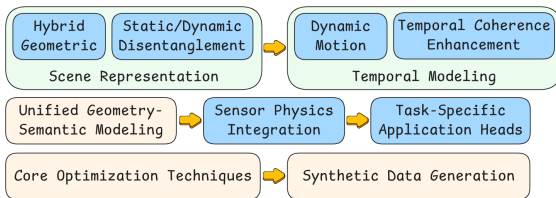


Fig. 21: Three methodological lenses of neural radiance fields-based methods

4.3.1. Scene Reconstruction and Modeling

Recent advancements in Neural Radiance Fields (NeRF) for scene reconstruction have focused on balancing accuracy,

speed, and dynamic environment handling. Hybrid approaches like Street Gaussians [206] combine 3D Gaussians with point clouds to achieve 135 fps rendering for dynamic urban scenes, prioritizing real-time performance through explicit geometric representations. In contrast, S-NeRF++ [40] employs a dual-branch architecture to disentangle static backgrounds and dynamic vehicles, improving reconstruction fidelity but increasing computational complexity. Dynamic modeling techniques such as Neural Radiance Flow [50] introduce differentiable scene flow for unsupervised motion estimation, while TimeNeRF [81] enhances temporal coherence through recurrent rendering. Occlusion handling solutions include GANeRF [156] with adversarial completion, reducing LPIPS by 50% on Tanks and Temples. Lightning-NeRF [28] integrates LiDAR for illumination decomposition, achieving 5× faster training through hybrid representation.

These methods demonstrate notable strengths in balancing efficiency and fidelity, as hybrid approaches like Lightning-NeRF and explicit representations such as Street Gaussians achieve 5-10× speedups while maintaining state-of-the-art reconstruction quality. These techniques excel in dynamic handling through dual-branch architectures like S-NeRF++ and 4D radiance flows such as Neural Radiance Flow, enabling robust modeling of moving objects and temporal changes. Despite these advancements, several limitations persist, including artifacts from transient dynamics where heuristic handling of sudden motion (e.g., pedestrians, fast vehicles) causes rendering distortions, as observed in EmerNeRF’s limited improvement for extreme transience. In addition, View-dependent ray sampling strategies remain a bottleneck, limiting real-time performance in open-road scenarios with rapid viewpoint changes. Furthermore, illumination decomposition methods exhibit LiDAR dependency, degrading performance in LiDAR-sparse regions and restricting applicability in sensor-limited setups.

4.3.2. Semantic Understanding and Task Applications

Semantic NeRF advancements enable unified geometry-semantic modeling and task-specific optimizations for autonomous driving. PanopticNeRF-360 [55] generates panoramic urban scene annotations with view-consistent panoptic labels, reducing manual labeling costs by 1.5 hours per scene. GSNeRF [43] fuses semantic and geometric features through a two-stage pipeline, and improves 3D scene understanding via depth-guided rendering. For 3D detection, NeRF-Det [200] avoids point cloud conversion losses. It obtains a significant improvement on ScanNet by leveraging geometric-aware volumetric representations. NeRF-Loc [166] achieves cm-level localization precision through Transformer-based feature alignment, outperforming traditional methods on Cambridge Landmarks. NeuRAD [175] supports closed-loop simulation by modeling sensor physics (rolling shutter, LiDAR beam divergence), enabling safety-critical scenario testing.

Semantic understanding methods offer unified representation capabilities, as seen in PanopticNeRF-360 and GSNeRF, which enable simultaneous geometry-semantic modeling and improve label propagation accuracy. Task-specific optimizations deliver significant gains, with NeRF-Det eliminating point cloud

conversion losses and NeRF-Loc achieving sub-centimeter localization precision critical for autonomous navigation. However, query latency remains problematic, with NeRF-Det and NeRF-Loc requiring 30-50ms per query, exceeding the real-time threshold for autonomous driving. Besides, physical inconsistencies also persist, as EditNeRF’s latent manipulation overlooks light transport physics, causing unrealistic shadows and reflections in edited scenes.

4.3.3. Efficiency and Robustness Enhancements

These methods aim to address deployment barriers for real-world autonomous systems. NeRFacc [107] achieves 1.5-20× training speedup via hash encoding and CUDA fusion, enabling real-time rendering. For environmental robustness, ClimateNeRF [113] employs degradation-aware ray sampling for rain/snow conditions, while PC-NeRF [74] uses sparse LiDAR regularization to improve reconstruction accuracy under 67% data loss. Sensor fusion methods like NeRFusion [223] integrate Camera-LiDAR-radar via cross-attention. Additionally, SinnNeRF [201] generates annotated synthetic data to mitigate real-data scarcity.

These efficiency enhancement methods deliver substantial computational gains, with NeRFacc and Lightning-NeRF achieving 10× speedups that make real-time operation feasible for autonomous driving platforms. Environmental adaptability is strengthened through methods like ClimateNeRF and PC-NeRF, which maintain performance in adverse conditions and sensor data loss, showing an improvement in rainy scenes. Despite these advancements, they suffer several deployment challenges. Calibration sensitivity is a key issue. When sensor synchronization delays exceeding 10ms, it will cause fusion failures and degrade radar-camera consistency. In addition, the sim-real gap also persists, as material property discrepancies in SimNeRF-generated data lead to 10% lower detection accuracy when transferred to real-world scenarios.

4.3.4. Summary

Neural Radiance Fields (NeRF) have transitioned from theoretical models to practical autonomous driving components, with advancements in dynamic modeling, semantic integration, and efficiency. Key achievements include 10× speedups, cm-level localization, and multi-sensor fusion robustness. However, critical gaps remain. Transient dynamics handling causes artifacts, while semantic resolution limits small object recognition, and calibration errors hinder sensor fusion. Future research should focus on spatiotemporal-consistent joint modeling, task-driven differentiable rendering, and neuro-symbolic integration to reduce annotation reliance. Addressing these barriers may enable NeRF to serve as the foundation for next-generation “environmental universal models” in autonomous driving.

4.4. World Model

World models constitute a critical component in autonomous driving systems, enabling the prediction of dynamic environment evolution to inform safe planning and decision-making.

We categorize the existing methods into three primary kinds based on their core objectives and technical approaches: Environment Representation & Structured Modeling, Dynamic Evolution & Interaction Modeling, and Model Enhancement & Emerging Paradigms.

4.4.1. Environment Representation and Structured Modeling Methods

This paradigm focuses on developing efficient, robust, and structured representations of complex driving scenes, forming the foundational state upon which predictions are made. OccFormer [226] exemplifies advancements in dense scene understanding by proposing an instance-aware 4D occupancy prediction framework that unifies geometric structure and dynamic evolution through agent-scene interaction. Similarly, GAIA-1 [71] adopts an object-centric perspective, explicitly modeling key scene entities (e.g., vehicles, pedestrians) with integrated geometric constraints, aligning more closely with human scene cognition. GraphAD [225] abstract scene elements as nodes within a graph structure, utilizing Graph Neural Networks to learn complex spatial and interactive dependencies, thereby providing a powerful tool for relational reasoning. QuAD [19] enhances practical utility by designing a world model capable of on-demand queries (e.g., collision checking), significantly improving planning efficiency.

Despite these significant strides in scene comprehension and structured expression, substantial challenges persist. Dense representations like occupancy grids incur prohibitive computational and memory overhead, hindering real-time deployment, while object-centric approaches are critically dependent on the accuracy of upstream perception modules (detection/tracking), making them vulnerable to error propagation. In addition, the effectiveness of graph-based models is contingent upon the completeness and appropriateness of the defined relationships between nodes. Furthermore, vision-only models exhibit limited robustness under adverse conditions such as severe weather, heavy occlusion, or extreme lighting variations.

4.4.2. Dynamic Evolution and Interaction Modeling Methods

This category focuses on forecasting the future trajectory of scene states, emphasizing physical laws, multi-agent interactions, and the inherent uncertainty of predictions. Physics-based modeling is prominently featured in UniWorld [135], which employs coupled differential equations to jointly model agent kinematics and environmental physical effects (e.g., friction, aerodynamics), demonstrating deep integration of physics and agent behavior. [53] addresses the specific demands of off-road driving by incorporating soil mechanics into its dynamic predictions. For modeling interactions, InterSim [167] introduces game-theoretic frameworks to explicitly capture strategic interactions (e.g., competition, cooperation) among multiple agents (vehicles, pedestrians), providing a principled approach to complex behavioral prediction. [6] directly predicts joint trajectory sets for all interacting agents, mitigating the cascading errors inherent in sequential “predict-then-react” pipelines. Generative approaches are represented by DriveDreamer [185],

which harnesses diffusion models to synthesize high-quality future scene states, and OccSora [183], which utilizes variational auto-encoders and latent space regularization to encourage diverse predictions and alleviate mode collapse. Recognizing the criticality of uncertainty for safety, [105] implements Bayesian Neural Networks to propagate uncertainty through the prediction horizon, forming a foundation for risk-aware decision-making. [75] enforces bidirectional temporal consistency losses to ensure the coherence of long-horizon predictions, indirectly enhancing prediction reliability.

Nevertheless, significant limitations remain. Physics-based models often entail high computational costs incompatible with real-time onboard requirements. Game-theoretic approaches suffer from high computational complexity and often employ simplified models of agent intent and rationality. Generative models (e.g., diffusion, VAE) are hampered by slow inference speeds, and ensuring the physical plausibility and controllability of generated outcomes is challenging. Bayesian methods for uncertainty estimation are computationally intensive and struggle to reliably disentangle epistemic (model) uncertainty from aleatoric (data) uncertainty.

4.4.3. Model Enhancement and Emerging Paradigm Methods

This final category encompasses efforts to augment specific critical capabilities of world models or integrate disruptive novel AI paradigms to address complex challenges and enhance applicability. [62] pioneers the integration of formal verification techniques to provide mathematical guarantees on prediction safety boundaries, representing a crucial step towards certifiably safe systems. [161] employs meta-learning strategies to enable rapid few-shot adaptation of world models to entirely new urban layouts, significantly boosting generalization and deployment efficiency. Explorations of emerging paradigms feature DriveGPT4 [202], which innovatively utilizes Large Language Models (LLMs) to capture state transition rules and semantic scene knowledge, leveraging the powerful reasoning and knowledge representation capabilities of LLMs for complex scenario understanding. [221] addresses data scarcity and long-tail challenges by focusing on generating high-fidelity synthetic scenarios for effective world model pre-training.

While these approaches represent cutting-edge explorations, they face distinct shortcomings. The scope of formal verification in safety-focused models is often limited and operates at high levels of abstraction, struggling to encompass the full complexity of real-world driving. Training differentiable models like NeuralDriving can be complex and prone to instability. The efficacy of meta-learning approaches is heavily dependent on the diversity of meta-training data and the inherent generalization capability of the base model. LLM-based models contend with inherent challenges, including significant inference latency, real-time performance limitations, potential "hallucinations," and difficulties in effectively grounding domain-specific driving knowledge. In addition, synthetic data generation grapples with the persistent "reality gap" between synthetic and real-world data, and ensuring the generated data distribution effectively covers critical scenarios and variations is challenging.

4.4.4. Summary

The world model increasingly integrates physical laws and interactive priors to improve prediction realism, while advancing structured representations (e.g., graph-based, object-centric). It balances generative models (e.g., diffusion) for diverse prediction and probabilistic techniques for uncertainty quantification. Recent research also focuses on verification, causal reasoning, and adaptability. Key challenges include computational efficiency, complex interaction modeling, robust uncertainty estimation, sim-to-real transfer, multi-modal sensor fusion, interpretability, and generalization. Addressing these requires interdisciplinary collaboration and foundation models to achieve verifiable and scalable autonomous driving.

4.5. Conclusion for implicit map

Collectively, these implicit map-based methods establish the theoretical foundation for modern autonomous driving representation learning, reflecting the field's evolution from modular perception systems to integrated end-to-end architectures. Cross-cutting trends include the convergence of computer vision and robotics principles, the increasing integration of physical constraints into data-driven models, and the shift toward unified frameworks that bridge perception, prediction, and planning. Fundamental challenges persist in reconciling representational expressiveness with real-time computational demands, developing formal safety verification methods for learned representations, and achieving robust generalization across diverse environmental conditions. Future research directions should prioritize hybrid modeling approaches that combine structured priors with neural networks, explainable representation learning techniques for critical scenario analysis, and standardized evaluation metrics that account for both performance and safety requirements. These advances will be pivotal in transitioning from research prototypes to deployable autonomous systems capable of operating reliably in complex real-world environments. Additionally, we list recent open-source work on implicit maps in Appendix.A.

5. Future Trends

Considering the aforementioned challenges and opportunities, we list some critical directions for future research as follows.

5.1. Fully Automatic Lite Map

End-to-end autonomous driving has not yet matured to the point where it can be commercially deployed at scale. As a result, explicitly designed lightweight maps remain the mainstream approach in the industry today. The number of related academic papers are also showing an increasing trend year by year, as summarized in Appendix.A. Looking ahead, a major research direction will continue to build upon the Lite map concept. Due to the fact that the automation rate in industry has always been low, the focus will be on maintaining high-quality map generation while achieving fully automated pipelines. The

ultimate goal is to create a closed-loop system, where data collection, processing, and updates are seamlessly integrated without the need for any human intervention, enabling scalable and cost-effective map maintenance to support autonomous driving applications.

5.2. Zero-shot Map Learning

Zero-shot learning (ZSL) is a machine learning approach where a model is able to recognize objects without having seen any labeled examples during training [146, 34]. Zero-shot learning is promising to complement online Lite map learning by enabling semantic flexibility, generalization to unseen environments, and reduced reliance on labeled data. Interesting research points might include:

- 1) Zero-shot Map Element Recognition: Develop models that can identify rare or novel road features (e.g., new sign types, local lane conventions) purely through zero-shot transfer from pretrained vision–language models.
- 2) Reducing Label Dependency: Online Lite map learning often struggles with the need for high-quality, human-labeled data for training or validation. Zero-shot capabilities allow models to generalize to novel road structures or layouts without needing supervised fine-tuning.
- 3) Uncertainty-aware ZSL for Mapping: Combine ZSL with uncertainty estimation to help determine when a newly observed feature is outside the model’s generalization range, improving safety and map reliability.

5.3. Implicit Map in VLA

Vision-Language-Action (VLA) Model is an emerging paradigm in autonomous driving and embodied AI that aims to unify perception, reasoning, and control within a single model [23, 232, 96]. A VLA model takes visual inputs (e.g., camera images), language inputs (e.g., route commands, traffic rules, or instructions), and outputs corresponding driving actions (e.g., steering, throttle, braking). This architecture is inspired by the way humans interpret their surroundings, understand instructions, and interact with the environment in a goal-driven manner [239, 238, 85].

Another concept in VLA is “memory knowledge base” or “experience layer”. The purpose of this concept is to leverage the map experience stored in the cloud to assist the vehicle in better understanding driving scenarios. For example, when the vehicle approaches an intersection, can it retrieve similar intersections from the knowledge base to help predict possible routes, even under obstructed conditions?

VLA models naturally support the idea of implicit mapping: instead of relying on structured maps, they learn to associate visual cues and linguistic context with action policies. For instance, a VLA model might implicitly understand where an intersection or crosswalk is, or the layout of a parking lot, based solely on its multimodal training data, without ever using a formalized map structure. This makes implicit maps a byproduct or internal representation within a VLA model, offering a flexible and potentially scalable alternative to explicit maps.

5.4. Foundation Model

Foundation models are large-scale, pre-trained models that serve as a general-purpose backbone for a wide range of downstream tasks. Trained on massive and diverse datasets using self-supervised or weakly supervised learning, these models, such as GPT [2, 129], CLIP [150, 214], or ViT [97, 208], exhibit strong generalization, transferability, and the ability to adapt to new tasks with minimal additional data or fine-tuning. Foundation models can naturally support the formation and use of Implicit maps in multiple ways:

- 1) Representation Learning: Foundation models learn latent spatial structures and semantic context from large-scale visual (and possibly LiDAR or multimodal) data, forming internal maps without the need for manual annotations or explicit formats.
- 2) Multimodal Integration: Foundation models can combine vision, language, and temporal context, enabling them to reason about traffic layouts, navigational goals, or obstacles based on implicit cues, much like how humans understand an environment from limited inputs.
- 3) Memory and Attention: Transformer-based foundation models use attention mechanisms to dynamically retrieve spatially and semantically relevant information, which can serve as a form of implicit localization and mapping.

6. Conclusions

With the gradual adoption of autonomous driving, traditional HD map production methods have become inadequate to meet the frequency demands of urban road updates. Significant changes have taken place in maps. This article categorizes this evolution into three stages: HD maps, Lite maps, and Implicit maps. We provide a detailed elaboration on the characteristics of these three stages, the map production workflows, the technical challenges encountered, and relevant academic research.

Despite the fact that many automakers used “mapless autonomous driving” as a marketing gimmick in recent years, time has proven that their so-called “mapless” approach actually refers to driving without HD maps. Maps, however, continue to exist in other forms within the autonomous driving system and have never truly been absent. Lite maps and Implicit maps stand as the clearest evidence of this.

In the future, the development trend of maps will lean toward greater lightweight design. Cloud-based map databases will assist vehicle-side real-time perception, ultimately paving the way for end-to-end autonomous driving.

Acknowledgements

The authors would like to acknowledge the support of the National Natural Science Foundation of China (grant No.: 42301503, 42171415).

References

- [1] Abdelkader, G., Alghamdi, T., Elgazzar, K., Khamis, A., 2023. Hd maps for connected and automated vehicles: Enabling technologies and future directions, in: 2023 IEEE International Conference on Smart Mobility (SM), IEEE. pp. 91–97.
- [2] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al., 2023. Gpt-4 technical report. arXiv preprint arXiv:2303.08774 .
- [3] AG, D., GmbH, V., 2014. Opencrg: Open curved regular grid (crg) road surface data format. URL: <https://www.opencrg.org/>. version 1.0.6 released 9 April 2014; open-source C-API and MATLAB tool suite for detailed road-surface representation :contentReference[oaicite:2]index=2.
- [4] Agarwal, N., Ali, A., Bala, M., Balaji, Y., Barker, E., Cai, T., Chattopadhyay, P., Chen, Y., Cui, Y., Ding, Y., Dworakowski, D., ..., Zolkowski, A., 2025. Cosmos world foundation model platform for physical ai. arXiv preprint arXiv:2501.03575 .
- [5] Ahmadi, A.M.S., Semnani, S.H., 2023. Human trajectory prediction using lstm with attention mechanism. arXiv preprint arXiv:2309.00331 .
- [6] Ahmadi, E., Mercurius, R., Alizadeh, S., Rezaee, K., Rasouli, A., 2024. Curb your attention: Causal attention gating for robust trajectory prediction in autonomous driving. arXiv preprint arXiv:2410.07191 .
- [7] Angeli, A., Filliat, D., Doncieux, S., Meyer, J.A., 2008. Fast and incremental method for loop-closure detection using bags of visual words. IEEE transactions on robotics 24, 1027–1037.
- [8] Arman, M.A., Tampère, C.M., 2021. Lane-level routable digital map reconstruction for motorway networks using low-precision gps data. Transportation research part C: emerging technologies 129, 103234.
- [9] ASAM e.V., 2023. ASAM OpenDRIVE Standard. <https://www.asam.net/standards/detail/opendrive>. Latest version 1.8 released on 22 November 2023; XML-based format for road network description used in driving/traffic simulation :contentReference[oaicite:1]index=1.
- [10] Ashok Elluswamy (Tesla), 2023. Building vision foundation models for autonomous driving. https://www.youtube.com/watch?v=6x-Xb_uT7ts.
- [11] Asrat, K.T., Cho, H.J., 2024. A comprehensive survey on high-definition map generation and maintenance. ISPRS International Journal of Geo-Information 13, 232.
- [12] Avramović, A., Sluga, D., Tabernik, D., Skočaj, D., Stojnić, V., Ilc, N., 2020. Neural-network-based traffic sign detection and recognition in high-definition images using region focusing and parallelization. IEEE Access 8, 189855–189868.
- [13] Bailo, O., Lee, S., Rameau, F., Yoon, J.S., Kweon, I.S., 2017. Robust road marking detection and recognition using density-based grouping and machine learning techniques, in: 2017 IEEE winter conference on applications of computer vision (WACV), IEEE. pp. 760–768.
- [14] Bao, Z., Hossain, S., Lang, H., Lin, X., 2022. High-definition map generation technologies for autonomous driving. arXiv preprint arXiv:2206.05400 .
- [15] Bartoccioni, F., Zablocki, É., Bursuc, A., Pérez, P., Cord, M., Alahari, K., 2023. Lara: Latents and rays for multi-camera bird’s-eye-view semantic segmentation, in: Conference on robot learning, PMLR. pp. 1663–1672.
- [16] Bernstein, D., Kornhauser, A., et al., 1996. An introduction to map matching for personal navigation assistants .
- [17] Besl, P.J., Jain, R.C., 2002. Segmentation through variable-order surface fitting. IEEE Transactions on pattern analysis and machine intelligence 10, 167–192.
- [18] Bhattacharya, S., Likhachev, M., Kumar, V., 2012. Topological constraints in search-based robot path planning. Autonomous Robots 33, 273–290.
- [19] Biswas, S., Casas, S., Sykora, Q., Agro, B., Sadat, A., Urtasun, R., 2024. Quad: Query-based interpretable neural motion planning for autonomous driving, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 14236–14243.
- [20] Bloit, J., Rodet, X., 2008. Short-time viterbi for online hmm decoding: Evaluation on a real-time phone recognition task, in: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE. pp. 2121–2124.
- [21] Boubakri, A., Gammar, S.M., Brahim, M.B., Filali, F., 2022. High definition map update for autonomous and connected vehicles: A survey, in: 2022 International Wireless Communications and Mobile Computing (IWCMC), IEEE. pp. 1148–1153.
- [22] Brakatsoulas, S., Pfoser, D., Salas, R., Wenk, C., 2005. On map-matching vehicle tracking data, in: Proceedings of the 31st international conference on Very large data bases, pp. 853–864.
- [23] Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., et al., 2023. Rt-2: Vision-language-action models transfer web knowledge to robotic control. arXiv preprint arXiv:2307.15818 .
- [24] Bruls, T., Maddern, W., Morye, A.A., Newman, P., 2018. Mark yourself: Road marking segmentation via weakly-supervised annotations from multimodal data, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 1863–1870.
- [25] Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O., 2020. nuscenes: A multi-modal dataset for autonomous driving, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11621–11631.
- [26] Cai, J., Xu, M., Li, W., Xiong, Y., Xia, W., Tu, Z., Soatto, S., 2022. Memot: Multi-object tracking with memory, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8090–8100.
- [27] Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M., Tardós, J.D., 2021. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. IEEE transactions on robotics 37, 1874–1890.
- [28] Cao, J., Li, Z., Wang, N., Ma, C., 2024. Lightning NeRF: Efficient hybrid scene representation for autonomous driving, in: IEEE International Conference on Robotics and Automation (ICRA).
- [29] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers, in: European conference on computer vision, Springer. pp. 213–229.
- [30] Cattaneo, D., Vaghi, M., Valada, A., 2022. Lcdnet: Deep loop closure detection and point cloud registration for lidar slam. IEEE Transactions on Robotics 38, 2074–2093.
- [31] Charroud, A., El Moutaouakil, K., Palade, V., Yahyaoui, A., Onyekpe, U., Eyo, E.U., 2024. Localization and mapping for self-driving vehicles: a survey. Machines 12, 118.
- [32] Chen, B., Ding, C., Ren, W., Xu, G., 2020. Extended classification course improves road intersection detection from low-frequency gps trajectory data. ISPRS International Journal of Geo-Information 9, 181.
- [33] Chen, B., Ding, C., Ren, W., Xu, G., 2021a. Automatically tracking road centerlines from low-frequency gps trajectory data. ISPRS International Journal of Geo-Information 10, 122.
- [34] Chen, J., Geng, Y., Chen, Z., Pan, J.Z., He, Y., Zhang, W., Horrocks, I., Chen, H., 2023a. Zero-shot and few-shot learning with knowledge graphs: A comprehensive survey. Proceedings of the IEEE 111, 653–685.
- [35] Chen, J., Wu, Y., Tan, J., Ma, H., Furukawa, Y., 2024a. Maptracker: Tracking with strided memory fusion for consistent vector hd mapping, in: European Conference on Computer Vision, Springer. pp. 90–107.
- [36] Chen, L., Li, Y., Huang, C., Xing, Y., Tian, D., Li, L., Hu, Z., Teng, S., Lv, C., Wang, J., et al., 2023b. Milestones in autonomous driving and intelligent vehicles—part i: Control, computing system design, communication, hd map, testing, and human behaviors. IEEE Transactions on Systems, Man, and Cybernetics: Systems 53, 5831–5847.
- [37] Chen, P., Jiang, X., Zhang, Y., Tan, J., Jiang, R., 2024b. Mapcvv: On-cloud map construction using crowdsourcing visual vectorized elements towards autonomous driving. IEEE Robotics and Automation Letters .
- [38] Chen, P., Shi, W., Bao, S., Wang, M., Fan, W., Xiang, H., 2021b. Low-drift odometry, mapping and ground segmentation using a backpack LiDAR system. IEEE Robotics and Automation Letters 6, 7285–7292. doi:10.1109/LRA.2021.3097060.
- [39] Chen, Y., Wen, Z., Fan, G., Chen, Z., Wu, W., Liu, D., Li, Z., Liu, B., Xiao, Y., 2024c. Mapo: Boosting large language model performance with model-adaptive prompt optimization. arXiv preprint arXiv:2407.04118 .
- [40] Chen, Y., Zhang, J., Xie, Z., Li, W., Zhang, F., Lu, J., Zhang, L., 2025. S-nerf++: Autonomous driving simulation via neural reconstruction and

- generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [41] Cheng, G., Wang, Y., Gong, Y., Zhu, F., Pan, C., 2014. Urban road extraction via graph cuts based probability propagation, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 5072–5076.
 - [42] Chiang, K.W., Wang, C.K., Hong, J.H., Li, P.L., Yang, C.S., Tsai, M.L., Lee, J., Lin, S., 2022. Verification and validation procedure for high-definition maps in taiwan. *Urban Informatics* 1, 18.
 - [43] Chou, Z.T., Huang, S.Y., Liu, I., Wang, Y.C.F., et al., 2024. Gsnerf: Generalizable semantic neural radiance fields with enhanced 3d scene understanding, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20806–20815.
 - [44] Dal Poz, A.P., Martins, E., Zanin, R.B., 2022. Road network extraction using gps trajectories based on morphological and skeletonization algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 43, 239–245.
 - [45] Dalal, M., Yang, J., Mendonca, R., Khaky, Y., Salakhutdinov, R., Pathak, D., 2024. Neural mp: A generalist neural motion planner. *arXiv preprint arXiv:2409.05864*.
 - [46] De La Escalera, A., Moreno, L.E., Salichs, M.A., Armingol, J.M., 1997. Road traffic sign detection and classification. *IEEE transactions on industrial electronics* 44, 848–859.
 - [47] Deng, J., Han, Y., 2013. A real-time system of lane detection and tracking based on optimized ransac b-spline fitting, in: Proceedings of the 2013 research in adaptive and convergent systems, pp. 157–164.
 - [48] Didi Chuxing, 2019. Didi chuxing gaia initiative open dataset. <https://gaia.didichuxing.com>. Accessed: 2025-07-03.
 - [49] Du, Y., Yang, S., Wang, L., Hou, Z., Cai, C., Tan, Z., Chen, M., Huang, S.S., Li, Q., 2025. Rtmap: Real-time recursive mapping with change detection and localization [arXiv:2507.00980](https://arxiv.org/abs/2507.00980).
 - [50] Du, Y., Zhang, Y., Yu, H.X., Tenenbaum, J.B., Wu, J., 2021. Neural radiance flow for 4d view synthesis and video processing, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE Computer Society. pp. 14304–14314.
 - [51] Elghazaly, G., Frank, R., Harvey, S., Safko, S., 2023. High-definition maps: Comprehensive survey, challenges, and future perspectives. *IEEE Open Journal of Intelligent Transportation Systems* 4, 527–550.
 - [52] Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise, in: *kdd*, pp. 226–231.
 - [53] Fan, J., Fan, L., Ni, Q., Wang, J., Liu, Y., Li, R., Wang, Y., Wang, S., 2024. Perception and planning of intelligent vehicles based on bev in extreme off-road scenarios. *IEEE Transactions on Intelligent Vehicles* 9, 4568–4572.
 - [54] Ferraz, A., Bretar, F., Jacquemoud, S., Gonçalves, G., Pereira, L., Tomé, M., Soares, P., 2012. 3-d mapping of a multi-layered mediterranean forest using als data. *Remote Sensing of Environment* 121, 210–223.
 - [55] Fu, X., Zhang, S., Chen, T., Lu, Y., Zhou, X., Geiger, A., Liao, Y., 2023. Panopticnerf-360: Panoramic 3d-to-2d label transfer in urban scenes. *arXiv preprint arXiv:2309.10815*.
 - [56] Fukano, K., Masuda, H., 2015. Detection and classification of pole-like objects from mobile mapping data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2, 57–64.
 - [57] Gao, J., Chen, Y., Junior, J.M., Wang, C., Li, J., 2020. Rapid extraction of urban road guardrails from mobile lidar point clouds. *IEEE Transactions on Intelligent Transportation Systems* 23, 1572–1577.
 - [58] Gao, L., Song, W., Dai, J., Chen, Y., 2019. Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network. *Remote Sensing* 11, 552.
 - [59] Gao, R., Wang, L., 2023. Memotr: Long-term memory-augmented transformer for multi-object tracking, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9901–9910.
 - [60] Gao, S., Li, M., Rao, J., Mai, G., Prestby, T., Marks, J., Hu, Y., 2021. Automatic urban road network extraction from massive gps trajectories of taxis, in: *Handbook of Big Geospatial Data*. Springer, pp. 261–283.
 - [61] Goh, C.Y., Dauwels, J., Mitrovic, N., Asif, M.T., Oran, A., Jaillet, P., 2012. Online map-matching based on hidden markov model for real-time traffic sensing applications, in: 2012 15th International IEEE Conference on Intelligent Transportation Systems, IEEE. pp. 776–781.
 - [62] Greve, E., Büchner, M., Vödisch, N., Burgard, W., Valada, A., 2024. Collaborative dynamic 3d scene graphs for automated driving, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 11118–11124.
 - [63] Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M., Yang, B., 2014. Using mobile laser scanning data for automated extraction of road markings. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 93–107.
 - [64] Guo, Y., Zhou, J., Dong, Q., Li, B., Xiao, J., Li, Z., 2025. Refined high-definition map model for roadside rest area. *Transportation Research Part A: Policy and Practice* 195, 104463.
 - [65] Guo, Y., Zhou, J., Li, X., Tang, Y., Lv, Z., 2024. A review of crowdsourcing update methods for high-definition maps. *ISPRS International Journal of Geo-Information* 13, 104.
 - [66] Gupta, S., Guadagnino, T., Mersch, B., Vizzo, I., Stachniss, C., 2024. Effectively detecting loop closures using point cloud density maps, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 10260–10266.
 - [67] Han, C., Yang, J., Sun, J., Ge, Z., Dong, R., Zhou, H., Mao, W., Peng, Y., Zhang, X., 2024. Exploring recurrent long-term temporal fusion for multi-view 3d perception. *IEEE Robotics and Automation Letters*.
 - [68] Hao, S., Liu, X., Qi, X., Zhao, S., Zi, B., Xiao, R., Han, K., Wong, K.Y.K., 2024. Bigr: Harnessing binary latent codes for image generation and improved visual representation capabilities. *arXiv preprint arXiv:2410.14672*.
 - [69] He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.
 - [70] Ho, S.L., Lin, J.K., Chou, K.Y., Chen, Y.P., 2022. Voronoi diagram based collision-free a* algorithm for mobile vehicle in complex dynamic environment, in: 2022 IEEE International Conference on Consumer Electronics-Taiwan, IEEE. pp. 429–430.
 - [71] Hu, A., Russell, L., Yeo, H., Murez, Z., Fedoseev, G., Kendall, A., Shotton, J., Corrado, G., 2023a. Gaia-1: A generative world model for autonomous driving. *arXiv preprint arXiv:2309.17080*.
 - [72] Hu, J., Mao, M., Bao, H., Zhang, G., Cui, Z., 2023b. Cp-slam: Collaborative neural point-based slam system. *Advances in Neural Information Processing Systems* 36, 39429–39442.
 - [73] Hu, J., Razdan, A., Femiani, J.C., Cui, M., Wonka, P., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Transactions on Geoscience and Remote Sensing* 45, 4144–4157.
 - [74] Hu, X., Xiong, G., Zang, Z., Jia, P., Han, Y., Ma, J., 2024a. Pcnrf: Parent-child neural radiance fields using sparse lidar frames in autonomous driving environments. *IEEE Transactions on Intelligent Vehicles*.
 - [75] Hu, X., Zhang, L., Meng, D., Han, Y., Yuan, L., 2024b. Gitsr: Graph interaction transformer-based scene representation for multi vehicle collaborative decision-making. *arXiv preprint arXiv:2411.01608*.
 - [76] Hu, X., Zhang, Z., Tao, C.V., 2004. A robust method for semi-automatic extraction of road centerlines using a piecewise parabolic model and least square template matching. *Photogrammetric Engineering & Remote Sensing* 70, 1393–1398.
 - [77] Hu, Y., Wang, Z., Huang, Z., Liu, Y., 2023c. Polyroad: polyline transformer for topological road-boundary detection. *IEEE Transactions on Geoscience and Remote Sensing* 62, 1–12.
 - [78] Hu, Y., Yang, J., Chen, L., Li, K., Sima, C., Zhu, X., Chai, S., Du, S., Lin, T., Wang, W., et al., 2023d. Planning-oriented autonomous driving, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 17853–17862.
 - [79] Huang, J., You, S., 2015. Pole-like object detection and classification from urban point clouds, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 3032–3038.
 - [80] Huang, Y., Shan, T., Chen, F., Englot, B., 2021. Disco-slam: Distributed scan context-enabled multi-robot lidar slam with two-stage global-local graph optimization. *IEEE Robotics and Automation Letters* 7, 1150–1157.
 - [81] Hung, H.H., Do, H.P., Li, Y.H., Huang, C.C., 2024. Timenerf: Building generalizable neural radiance fields across time from few-shot input views, in: Proceedings of the 32nd ACM International Conference on Multimedia, pp. 253–262.
 - [82] Hur, J., Kang, S.N., Seo, S.W., 2013. Multi-lane detection in urban driving environments using conditional random fields, in: 2013 IEEE

- Intelligent vehicles symposium (IV), IEEE. pp. 1297–1302.
- [83] Ji, H., Lin, T., Li, H., 2024. Dpbridge: Latent diffusion bridge for dense prediction. *arXiv preprint arXiv:2412.20506*.
- [84] Jia, C., Yang, Y., Xia, Y., Chen, Y.T., Parekh, Z., Pham, H., Le, Q., Sung, Y.H., Li, Z., Duerig, T., 2021. Scaling up visual and vision-language representation learning with noisy text supervision, in: *International conference on machine learning*, PMLR. pp. 4904–4916.
- [85] Jiang, A., Gao, Y., Sun, Z., Wang, Y., Wang, J., Chai, J., Cao, Q., Heng, Y., Jiang, H., Zhang, Z., et al., 2025. Diffvla: Vision-language guided diffusion planning for autonomous driving. *arXiv preprint arXiv:2505.19381*.
- [86] Jiao, F., Xiang, L., Deng, Y., 2025. Automatic extraction of road interchange networks from crowdsourced trajectory data: A forward and reverse tracking approach. *ISPRS International Journal of Geo-Information* 14, 234.
- [87] Jo, K., Kim, C., Sunwoo, M., 2018. Simultaneous localization and map change update for the high definition map-based autonomous driving car. *Sensors* 18, 3145.
- [88] Kannan, S.S., Min, B.C., 2024. Zeroscd: Zero-shot street scene change detection. *arXiv preprint arXiv:2409.15255*.
- [89] Karrer, M., Schmuck, P., Chli, M., 2018. Cvi-slam—collaborative visual-inertial slam. *IEEE Robotics and Automation Letters* 3, 2762–2769.
- [90] Kataoka, H., Shirakabe, S., Miyashita, Y., Nakamura, A., Iwata, K., Satoh, Y., 2016. Semantic change detection with hypermaps. *arXiv preprint arXiv:1604.07513* 2.
- [91] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42, 139–1.
- [92] Khan, Z., BG, V.K., Schultze, S., Chandraker, M., Fu, Y., 2023. Exploring question decomposition for zero-shot vqa. *Advances in Neural Information Processing Systems* 36, 56615–56627.
- [93] Khurana, T., Hu, P., Held, D., Ramanan, D., 2023. Point cloud forecasting as a proxy for 4d occupancy forecasting, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1116–1124.
- [94] Kim, G., Kim, A., 2018. Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map, in: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE. pp. 4802–4809.
- [95] Kim, J., Park, C., 2017. End-to-end ego lane estimation based on sequential transfer learning for self-driving cars, in: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 30–38.
- [96] Kim, M.J., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., Rafailov, R., Foster, E., Lam, G., Sanketi, P., et al., 2024. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*.
- [97] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al., 2023. Segment anything, in: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026.
- [98] Kwag, J., Toth, C., 2024. A review on end-to-end high-definition map generation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48, 187–194.
- [99] Labbe, M., Michaud, F., 2014. Online global loop closure detection for large-scale multi-session graph-based slam, in: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE. pp. 2661–2666.
- [100] Lam, J., Kusevic, K., Mrstik, P., Harrap, R., Greenspan, M., 2010. Urban scene extraction from mobile ground based lidar data, in: *Proceedings of 3DPVT*, pp. 1–8.
- [101] Lei, W., Ge, Y., Yi, K., Zhang, J., Gao, D., Sun, D., Ge, Y., Shan, Y., Shou, M.Z., 2024. Vit-lens: Towards omni-modal representations, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26647–26657.
- [102] Li, B., Gao, J., Chen, S., Lim, S., Jiang, H., 2024. Df-drnet: A decoder fusion model for automatic road extraction leveraging remote sensing images and gps trajectory data. *International Journal of Applied Earth Observation and Geoinformation* 127, 103632.
- [103] LI, B., GUO, Y., ZHOU, J., TANG, Y., DONG, Q., LI, Z., 2024. Development and prospects of high definition map for intelligent vehicle. *Geomatics and Information Science of Wuhan University* 49, 491–505.
- [104] Li, L., Li, D., Xing, X., Yang, F., Rong, W., Zhu, H., 2017. Extraction of road intersections from gps traces based on the dominant orientations of roads. *ISPRS International Journal of Geo-Information* 6, 403.
- [105] Li, P., Grana, D., Liu, M., 2024. Bayesian neural network and bayesian physics-informed neural network via variational inference for seismic petrophysical inversion. *Geophysics* 89, M185–M196.
- [106] Li, Q., Wang, Y., Wang, Y., Zhao, H., 2022a. Hdmapnet: An online hd map construction and evaluation framework, in: *2022 International Conference on Robotics and Automation (ICRA)*, IEEE. pp. 4628–4634.
- [107] Li, R., Gao, H., Tancik, M., Kanazawa, A., 2023a. Nerfacc: Efficient sampling accelerates nerfs. *arXiv preprint arXiv:2305.04966*.
- [108] Li, S., Wang, S., Zhou, Y., Shen, Z., Li, X., 2022b. Tightly coupled integration of gnss, ins, and lidar for vehicle navigation in urban environments. *IEEE Internet of Things Journal* 9, 24721–24735.
- [109] Li, T., Hua, T., Fu, M., Zhang, W.A., Pei, L., Yu, W., Truong, T.K., 2025. In-p 3 vins: Tightly-coupled ppp/ins/visual slam based on invariant optimization approach. *IEEE Transactions on Intelligent Transportation Systems*.
- [110] Li, T., Jia, P., Wang, B., Chen, L., Jiang, K., Yan, J., Li, H., 2023b. Laneseget: Map learning with lane segment perception for autonomous driving. *arXiv preprint arXiv:2312.16108*.
- [111] Li, W., 2023. Efficient planning with latent diffusion. *arXiv preprint arXiv:2310.00311*.
- [112] Li, X., Qin, Z., Shen, Z., Li, X., Zhou, Y., Song, B., 2022c. A high-precision vehicle navigation system based on tightly coupled ppp-rtk/ins/odometer integration. *IEEE Transactions on Intelligent Transportation Systems* 24, 1855–1866.
- [113] Li, Y., Lin, Z.H., Forsyth, D., Huang, J.B., Wang, S., 2023c. Climated-erf: Extreme weather synthesis in neural radiance field, in: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3227–3238.
- [114] Li, Y., Ma, L., Zhong, Z., Liu, F., Chapman, M.A., Cao, D., Li, J., 2020. Deep learning for lidar point clouds in autonomous driving: A review. *IEEE Transactions on Neural Networks and Learning Systems* 32, 3412–3432.
- [115] Li, Z., Wang, W., Li, H., Xie, E., Sima, C., Lu, T., Qiao, Y., Dai, J., 2022d. Bevformer: Learning bird’s-eye-view representation from multi-camera images via spatiotemporal transformers. *arXiv preprint arXiv:2203.17270*.
- [116] Liao, B., Chen, S., Wang, X., Cheng, T., Zhang, Q., Liu, W., Huang, C., 2023. Maptr: Structured modeling and learning for online vectorized hd map construction, in: *International Conference on Learning Representations*.
- [117] Liao, B., Chen, S., Zhang, Y., Jiang, B., Zhang, Q., Liu, W., Huang, C., Wang, X., 2024. Maptrv2: An end-to-end framework for online vectorized hd map construction. *International Journal of Computer Vision*, 1–23.
- [118] Lin, X., Lin, T., Pei, Z., Huang, L., Su, Z., 2023. Sparse4d v2: Recurrent temporal fusion with sparse model. *arXiv preprint arXiv:2305.14018*.
- [119] Liu, D., Wang, J., 2022. Path planning algorithm for ground unmanned vehicles based on topological constraints in unstructured environments. *Control Decis* 37, 3103–3114.
- [120] Liu, H., Teng, Y., Lu, T., Wang, H., Wang, L., 2023a. Sparsebev: High-performance sparse 3d object detection from multi-camera videos, in: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 18580–18590.
- [121] Liu, R., Wang, J., Zhang, B., 2020. High definition map for automated driving: Overview and analysis. *The Journal of Navigation* 73, 324–341.
- [122] Liu, Y., Yuan, T., Wang, Y., Wang, Y., Zhao, H., 2023b. Vectormapnet: End-to-end vectorized hd map learning, in: *International Conference on Machine Learning*, PMLR. pp. 22352–22369.
- [123] Lok, M.H., Duchaineau, M., Joy, K.I., 2005. Real-Time Optimal Adaptation for Planetary Geometry and Texture: 4–8 Tile Hierarchies. *IEEE Transactions on Visualization and Computer Graphics* 11, 432–444. doi:10.1109/TVCG.2005.39.
- [124] Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., Huang, Y., 2009. Map-matching for low-sampling-rate gps trajectories, in: *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 352–361.

- [125] Luo, Y., Sun, C., Tenenbaum, J.B., Du, Y., 2024. Potential based diffusion motion planning, in: International Conference on Machine Learning, PMLR. pp. 33486–33510.
- [126] Luo, Z., Gao, L., Xiang, H., Li, J., 2023. Road object detection for hd map: Full-element survey, analysis and perspectives. *ISPRS Journal of Photogrammetry and Remote Sensing* 197, 122–144.
- [127] Lyu, H., Perez, J.S.B., Huang, Y., Li, K., Shan, M., Worrall, S., 2024. Online high-definition map construction for autonomous vehicles: A comprehensive survey.
- [128] Mahara, A., Rishe, N.D., 2024. Generative adversarial model equipped with contrastive learning in map synthesis, in: Proceedings of the 2024 6th International Conference on Image Processing and Machine Vision, pp. 107–114.
- [129] Mao, J., Qian, Y., Ye, J., Zhao, H., Wang, Y., 2023a. Gpt-driver: Learning to drive with gpt. arXiv preprint arXiv:2310.01415.
- [130] Mao, J., Ye, J., Qian, Y., Pavone, M., Wang, Y., 2023b. A language agent for autonomous driving. arXiv preprint arXiv:2311.10813.
- [131] Matsumoto, H., Mori, Y., Masuda, H., 2019. Extraction and shape reconstruction of guardrails using mobile mapping data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42, 1061–1068.
- [132] McCall, J.C., Trivedi, M.M., 2006. Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. *IEEE transactions on intelligent transportation systems* 7, 20–37.
- [133] Mi, L., Zhao, H., Nash, C., Jin, X., Gao, J., Sun, C., Schmid, C., Shavit, N., Chai, Y., Anguelov, D., 2021. Hdmagen: A hierarchical graph generative model of high definition maps, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4227–4236.
- [134] Miao, Z., Wang, B., Shi, W., Zhang, H., 2014. A semi-automatic method for road centerline extraction from vhr images. *IEEE Geoscience and Remote Sensing Letters* 11, 1856–1860.
- [135] Min, C., Zhao, D., Xiao, L., Nie, Y., Dai, B., 2023. Uniworld: Autonomous driving pre-training via world models. arXiv preprint arXiv:2308.07234.
- [136] Mobileye, 2021. RemTM hd maps for continuous adas/autonomous driving. <https://www.mobileye.com/news/press-releases/2021/03/18>.
- [137] Mu, M., Zhao, L., 2021. A data fusion algorithm of gnss/ins/odometer integrated system in consideration of total odometer errors, in: 2021 21st International Conference on Control, Automation and Systems (IC-CAS), IEEE. pp. 1093–1098.
- [138] Mullins, G.E., Dress, A.G., Stankiewicz, P.G., Appler, J.D., Gupta, S.K., 2018. Accelerated testing and evaluation of autonomous vehicles via imitation learning, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE. pp. 5636–5642.
- [139] NDS e.V., 2018. Navigation Data Standard (NDS). <https://www.nds-association.org/>. Standard version 2.5.4 released on 12 December 2018, automotive-grade, binary map database format supporting incremental updates and DRM :contentReference[oaicite:0]index=0.
- [140] Newson, P., Krumm, J., 2009. Hidden markov map matching through noise and sparseness, in: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 336–343.
- [141] Novatel Inc., 2024. Inertial explorer - waypoint post-processing software. <https://novatel.com/products/waypoint-post-processing-software/inertial-explorer>. Accessed on June 23, 2025.
- [142] Owaki, T., Machida, T., 2020. Roadnetgan: Generating road networks in planar graph representation, in: International Conference on Neural Information Processing, Springer. pp. 535–543.
- [143] Pan, X., Shi, J., Luo, P., Wang, X., Tang, X., 2018. Spatial as deep: Spatial cnn for traffic scene understanding, in: Proceedings of the AAAI conference on artificial intelligence.
- [144] Pei, M., Shan, J., Li, P., Shi, J., Huo, J., Gao, Y., Shen, S., 2025. Sept: Standard-definition map enhanced scene perception and topology reasoning for autonomous driving. *IEEE Robotics and Automation Letters*.
- [145] Poggenhans, F., Pauls, J.H., Janosovits, J., Orf, S., Naumann, M., Kuhnt, F., Mayr, M., 2018. Lanelet2: A high-definition map framework for the future of automated driving, in: Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Hawaii, USA. URL: <http://www.mrt.kit.edu/z/publ/download/2018/Poggenhans2018Lanelet2.pdf>.
- [146] Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C.P., Wang, X.Z., Wu, Q.J., 2022. A review of generalized zero-shot learning methods. *IEEE transactions on pattern analysis and machine intelligence* 45, 4051–4070.
- [147] Qin, T., Huang, H., Wang, Z., Chen, T., Ding, W., 2023. Traffic flow-based crowdsourced mapping in complex urban scenario. *IEEE Robotics and Automation Letters* 8, 5077–5083.
- [148] Qin, T., Li, P., Shen, S., 2018. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE transactions on robotics* 34, 1004–1020.
- [149] Qin, T., Zheng, Y., Chen, T., Chen, Y., Su, Q., 2021. A light-weight semantic map for visual localization towards autonomous driving, in: 2021 IEEE international conference on robotics and automation (ICRA), IEEE. pp. 11248–11254.
- [150] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision, in: International conference on machine learning, PmLR. pp. 8748–8763.
- [151] Rastiveis, H., Shams, A., Sarasua, W.A., Li, J., 2020. Automated extraction of lane markings from mobile lidar point clouds based on fuzzy inference. *ISPRS Journal of Photogrammetry and Remote Sensing* 160, 149–166.
- [152] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788.
- [153] Reitberger, J., Schnörr, C., Krzystek, P., Stilla, U., 2009. 3d segmentation of single trees exploiting full waveform lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing* 64, 561–574.
- [154] Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence* 39, 1137–1149.
- [155] Rife, J., Phelts, R.E., Bhatti, J.A., Walter, T., 2004. Navigation integrity monitoring for aircraft approaches using gps. *GPS Solutions* 8, 193–203.
- [156] Roessle, B., Müller, N., Porzi, L., Buló, S.R., Kotschieder, P., Nießner, M., 2023. Ganerf: Leveraging discriminators to optimize neural radiance fields. *ACM Transactions on Graphics (TOG)* 42, 1–14.
- [157] Saeed, A., Grangier, D., Zeghidour, N., 2021. Contrastive learning of general-purpose audio representations, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 3875–3879.
- [158] Salti, S., Petrelli, A., Tombari, F., Fioraio, N., Di Stefano, L., 2015. Traffic sign detection via interest region extraction. *Pattern recognition* 48, 1039–1049.
- [159] Scharwächter, T., Schuler, M., Franke, U., 2014. Visual guard rail detection for advanced highway assistance systems, in: 2014 IEEE Intelligent Vehicles Symposium Proceedings, IEEE. pp. 900–905.
- [160] Sedighi, S., Nguyen, D.V., Kapsalas, P., Kuhnert, K.D., 2019. Implementing voronoi-based guided hybrid a* in global path planning for autonomous vehicles, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE. pp. 3845–3852.
- [161] Seo, J., Kim, T., Ahn, S., Kwak, K., 2024. Metaverse: Meta-learning traversability cost map for off-road navigation, in: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE. pp. 13190–13197.
- [162] Shan, Y., Lu, H., Lou, W., 2023. A hybrid attention and dilated convolution framework for entity and relation extraction and mining. *Scientific Reports* 13, 17062.
- [163] Shaw, A.K., Jiang, K., Wen, T., Sah, C.K., Shi, Y., Yang, M., Yang, D., Lian, X., 2025. Cleanmap: Distilling multimodal llms for confidence-driven crowdsourced hd map updates, in: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 3798–3807.
- [164] Shukla, N., Banerjee, S., 2023. Generating adversarial attacks in the latent space, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 730–739.
- [165] Sun, J., Xu, C., Tang, L., Wang, S., Lin, C., Gong, Y., Ni, L.M., Shum, H.Y., Guo, J., 2023a. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. arXiv preprint arXiv:2307.07697.

- [166] Sun, J., Xu, Y., Ding, M., Yi, H., Wang, C., Wang, J., Zhang, L., Schwager, M., 2023b. Nerf-loc: Transformer-based object localization within neural radiance fields. *IEEE Robotics and Automation Letters* 8, 5244–5250.
- [167] Sun, Q., Huang, X., Williams, B.C., Zhao, H., 2022. Intersim: Interactive traffic simulation via explicit relation modeling, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 11416–11423.
- [168] Tang, J., Deng, M., Huang, J., Liu, H., 2019. A novel method for road intersection construction from vehicle trajectory data. *IEEE Access* 7, 95065–95074.
- [169] Tang, X., Jiang, K., Yang, M., Liu, Z., Jia, P., Wijaya, B., Wen, T., Cui, L., Yang, D., 2023. High-definition maps construction based on visual sensor: A comprehensive survey. *IEEE Transactions on Intelligent Vehicles*.
- [170] Tanner, C., Jones, T., Ritschard, W., 2004. Geometry clipmaps: Terrain visualization with nested grids, in: ACM SIGGRAPH, pp. 237–245. doi:10.1145/1186562.1015738.
- [171] Tesla, I., 2022. Tesla ai day 2022: Fsd auto labeling. https://www.youtube.com/watch?v=0DSJsvid_SU. Tesla presents its progress on AI and self-driving technologies.
- [172] Tian, B., Zhang, S., Chen, S., Zhang, Y., Peng, K., Zhang, H., Wang, D., 2024a. Tracking dynamic flow: Decoding flow fluctuations through performance in a fine motor control task. *IEEE Transactions on Affective Computing*.
- [173] Tian, Y., Qi, Y., Jiang, A., Huang, Q., Guo, J., 2024b. Latent conditional diffusion-based data augmentation for continuous-time dynamic graph model. *arXiv preprint arXiv:2407.08500*.
- [174] TomTom, 2015. Tomtom hd map with roaddna - product sheet. <https://download.tomtom.com/open/banners/HD-Map-with-RoadDNA-Product-Sheet.pdf>.
- [175] Tonderski, A., Lindström, C., Hess, G., Ljungbergh, W., Svensson, L., Petersson, C., 2024. NeuRAD: Neural rendering for autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14895–14904.
- [176] Tóvári, D., Pfeifer, N., 2005. Segmentation based robust interpolation—a new approach to laser data filtering. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36, 79–84.
- [177] Tsai, M.L., Chiang, K.W., Hsieh, C.Y., Lin, S., Srinara, S., Chiu, Y.T., 2023. Alternative gcp sources for accurate hd map production. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48, 855–860.
- [178] Van Dierendonck, A.J., Fenton, P., Ford, T., 1992. Theory and performance of narrow correlator spacing in a gps receiver, in: Proceedings of IEEE Position Location and Navigation Symposium (PLANS), IEEE, pp. 115–122.
- [179] Vidanapathirana, K., Chng, S.F., Li, X., Lucey, S., 2023. Multi-body neural scene flow. *arXiv preprint arXiv:2310.10301*.
- [180] Vo, A.V., Truong-Hong, L., Laefer, D.F., Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing* 104, 88–100.
- [181] Wan, L., Huang, Z., Peng, X., 2016. An effective nosql-based vector map tile management approach. *ISPRS International Journal of Geo-Information* 5, 215. doi:10.3390/ijgi5110215.
- [182] Wang, H., Wang, C., Luo, H., Li, P., Cheng, M., Wen, C., Li, J., 2014. Object detection in terrestrial laser scanning point clouds based on hough forest. *IEEE Geoscience and Remote Sensing Letters* 11, 1807–1811.
- [183] Wang, L., Zheng, W., Ren, Y., Jiang, H., Cui, Z., Yu, H., Lu, J., 2024. Occsora: 4d occupancy generation models as world simulators for autonomous driving. *arXiv preprint arXiv:2405.20337*.
- [184] Wang, S., Liu, Y., Wang, T., Li, Y., Zhang, X., 2023a. Exploring object-centric temporal modeling for efficient multi-view 3d object detection, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 3621–3631.
- [185] Wang, X., Zhu, Z., Huang, G., Chen, X., Zhu, J., Lu, J., 2023b. Drive-dreamer: Towards real-world-driven world models for autonomous driving. *arXiv preprint arXiv:2309.09777*.
- [186] Wang, Z., Wang, J., Jiang, C., 2022. Unified multimodal model with unlikelihood training for visual dialog, in: Proceedings of the 30th ACM International Conference on Multimedia, pp. 4625–4634.
- [187] Wayve, 2023. Introducing gaia-1: A cutting-edge generative ai model for autonomy. <https://wayve.ai/thinking/scaling-gaia-1/>.
- [188] Wei, T., Lin, Y., Guo, S., Lin, Y., Huang, Y., Xiang, C., Bai, Y., Wan, H., 2024. Diff-rntraj: A structure-aware diffusion model for road network-constrained trajectory generation. *IEEE Transactions on Knowledge and Data Engineering*.
- [189] Wen, C., Sun, X., Li, J., Wang, C., Guo, Y., Habib, A., 2019. A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. *ISPRS journal of photogrammetry and remote sensing* 147, 178–192.
- [190] Wen, W., Pfeifer, T., Bai, X., Hsu, L.T., 2020. It is time for factor graph optimization for gns/ins integration: Comparison between fgo and ekf. *arXiv preprint arXiv:2004.10572*.
- [191] Wijaya, B., Jiang, K., Yang, M., Wen, T., Wang, Y., Tang, X., Fu, Z., Zhou, T., Yang, D., 2024. High definition map mapping and update: A general overview and future directions. *arXiv preprint arXiv:2409.09726*.
- [192] Wild, L., Ericson, L., Valencia, R., Jensfelt, P., 2024. Exelmap: Explainable element-based hd-map change detection and update. *arXiv preprint arXiv:2409.10178*.
- [193] Wilson, B., Qi, W., Agarwal, T., Lambert, J., Singh, J., Khandelwal, S., Pan, B., Kumar, R., Hartnett, A., Pontes, J.K., et al., 2023. Argoverse 2: Next generation datasets for self-driving perception and forecasting. *arXiv preprint arXiv:2301.00493*.
- [194] Wu, B., Yu, B., Yue, W., Shu, S., Tan, W., Hu, C., Huang, Y., Wu, J., Liu, H., 2013. A voxel-based method for automated identification and morphological parameters estimation of individual street trees from mobile laser scanning data. *Remote Sensing* 5, 584–611.
- [195] Xiang, H., Shi, W., Fan, W., Chen, P., Bao, S., Nie, M., 2021. Fastlcd: A fast and compact loop closure detection approach using 3d point cloud for indoor mobile mapping. *International Journal of Applied Earth Observation and Geoinformation* 102, 102430.
- [196] Xiang, X., Abdein, R., Lv, N., Yang, J., 2023. Self-supervised learning of scene flow with occlusion handling through feature masking. *Pattern Recognition* 139, 109487.
- [197] Xie, M., Zeng, S., Chang, X., Liu, X., Pan, Z., Xu, M., Wei, X., 2025. Seqgrowgraph: Learning lane topology as a chain of graph expansions. *arXiv preprint arXiv:2507.04822*.
- [198] Xie, Q., Li, D., Yu, Z., Zhou, J., Wang, J., 2019. Detecting trees in street images via deep learning with attention module. *IEEE Transactions on Instrumentation and Measurement* 69, 5395–5406.
- [199] Xpeng Motors, 2025. Scaling up autonomous driving via large foundation models.
- [200] Xu, C., Wu, B., Hou, J., Tsai, S., Li, R., Wang, J., Zhan, W., He, Z., Vajda, P., Keutzer, K., et al., 2023. Nerf-det: Learning geometry-aware volumetric representation for multi-view 3d object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 23320–23330.
- [201] Xu, D., Jiang, Y., Wang, P., Fan, Z., Shi, H., Wang, Z., 2022. Sinnerf: Training neural radiance fields on complex scenes from a single image, in: European Conference on Computer Vision, Springer, pp. 736–753.
- [202] Xu, Z., Zhang, Y., Xie, E., Zhao, Z., Guo, Y., Wong, K.Y.K., Li, Z., Zhao, H., 2024. Drivept4: Interpretable end-to-end autonomous driving via large language model. *IEEE Robotics and Automation Letters*.
- [203] Yan, e.a., 2023. Map vector tile construction for arable land spatial connectivity analysis based on the hadoop cloud platform. *Frontiers in Earth Science* 11, 1234732. doi:10.3389/feart.2023.1234732.
- [204] Yan, L., Li, Z., Liu, H., Tan, J., Zhao, S., Chen, C., 2017. Detection and classification of pole-like road objects from mobile lidar data in motorway environment. *Optics & Laser Technology* 97, 272–283.
- [205] Yan, W.Y., Morsy, S., Shaker, A., Tulloch, M., 2016. Automatic extraction of highway light poles and towers from mobile lidar data. *Optics & Laser Technology* 77, 162–168.
- [206] Yan, Y., Lin, H., Zhou, C., Wang, W., Sun, H., Zhan, K., Lang, X., Zhou, X., Peng, S., 2024. Street gaussians: Modeling dynamic urban scenes with gaussian splatting, in: European Conference on Computer Vision, Springer, pp. 156–173.
- [207] Yang, B., Dong, Z., Zhao, G., Dai, W., 2015. Hierarchical extraction of urban objects from mobile laser scanning data. *ISPRS Journal of*

- Photogrammetry and Remote Sensing 99, 45–57.
- [208] Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth anything: Unleashing the power of large-scale unlabeled data, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10371–10381.
 - [209] Yang, W., Ai, T., Lu, W., 2018. A method for extracting road boundary information from crowdsourcing vehicle gps trajectories. *Sensors* 18, 1261.
 - [210] Yang, Z., Song, N., Li, W., Zhu, X., Zhang, L., Torr, P.H., 2025. Deepinteraction++: Multi-modality interaction for autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
 - [211] Yao, Y., Fan, M., Xu, S., Xiong, H., Liu, X., Hu, W., Huang, W., 2025. A concise survey on lane topology reasoning for hd mapping. *arXiv preprint arXiv:2504.01989*.
 - [212] Ye, T., Jing, W., Hu, C., Huang, S., Gao, L., Li, F., Wang, J., Guo, K., Xiao, W., Mao, W., et al., 2023. Fusionad: Multi-modality fusion for prediction and planning tasks of autonomous driving. *arXiv preprint arXiv:2308.01006*.
 - [213] Ying, Z., Li, G., 2016. Robust lane marking detection using boundary-based inverse perspective mapping, in: 2016 IEEE international conference on acoustics, speech and signal processing (icassp), IEEE. pp. 1921–1925.
 - [214] Yu, J., Wang, Z., Vasudevan, V., Yeung, L., Seyedhosseini, M., Wu, Y., 2022. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917*.
 - [215] Yu, Y., Li, J., Wen, C., Guan, H., Luo, H., Wang, C., 2016. Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data. *ISPRS journal of photogrammetry and remote sensing* 113, 106–123.
 - [216] Yuan, T., Liu, Y., Wang, Y., Wang, Y., Zhao, H., 2024. Streammapnet: Streaming mapping network for vectorized online hd map construction, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 7356–7365.
 - [217] Zaklouta, F., Stanculescu, B., 2014. Real-time traffic sign recognition in three stages. *Robotics and autonomous systems* 62, 16–24.
 - [218] Zhang, D., Chen, D., Zhi, P., Chen, Y., Yuan, Z., Li, C., Zhou, R., Zhou, Q., et al., 2025. Mapexpert: Online hd map construction with simple and efficient sparse map element expert, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 14745–14753.
 - [219] Zhang, e.a., 2016. An efficient tile-pyramids building method for fast visualization of massive geospatial raster datasets. *ISPRS International Journal of Geo-Information* 5, 215. doi:[10.3390/ijgi5110215](https://doi.org/10.3390/ijgi5110215).
 - [220] Zhang, J., Singh, S., 2017. Low-drift and real-time LiDAR odometry and mapping. *Autonomous Robots* 41, 401–416.
 - [221] Zhang, J., Xing, Z., Gu, F., Wang, Y., Wang, T., Chen, J., 2024a. Changes in the microflora on the seed surface and seed vigor of maize (zea mays) under different conditions. *PloS one* 19, e0311258.
 - [222] Zhang, P., Zhang, M., Liu, J., 2021a. Real-time hd map change detection for crowdsourcing update based on mid-to-high-end sensors. *Sensors* 21, 2477.
 - [223] Zhang, X., Bi, S., Sunkavalli, K., Su, H., Xu, Z., 2022. Nerfusion: Fusing radiance fields for large-scale scene reconstruction, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5449–5458.
 - [224] Zhang, X., Han, X., Li, C., Tang, X., Zhou, H., Jiao, L., 2019. Aerial image road extraction based on an improved generative adversarial network. *Remote Sensing* 11, 930.
 - [225] Zhang, Y., Qian, D., Li, D., Pan, Y., Chen, Y., Liang, Z., Zhang, Z., Zhang, S., Li, H., Fu, M., et al., 2024b. Graphad: Interaction scene graph for end-to-end autonomous driving. *arXiv preprint arXiv:2403.19098*.
 - [226] Zhang, Y., Zhu, Z., Du, D., 2023. Occformer: Dual-path transformer for vision-based 3d semantic occupancy prediction. *arXiv preprint arXiv:2304.05316*.
 - [227] Zhang, Z., Niu, X., Tang, H., Chen, Q., Zhang, T., 2021b. Gns/ins/odo/wheel angle integrated navigation algorithm for an all-wheel steering robot. *Measurement Science and Technology* 32, 115122.
 - [228] Zhang, Z., Tang, S., Zhang, Y., Fu, T., Wang, Y., Liu, Y., Wang, D., Shao, J., Wang, L., Lu, H., 2024c. Ad-h: Autonomous driving with hierarchical agents. *arXiv preprint arXiv:2406.03474*.
 - [229] Zhao, L., Mao, J., Pu, M., Liu, G., Jin, C., Qian, W., Zhou, A., Wen, X., Hu, R., Chai, H., 2020. Automatic calibration of road intersection topology using trajectories, in: 2020 IEEE 36th International Conference on Data Engineering (ICDE), IEEE. pp. 1633–1644.
 - [230] Zhao, S., Chen, Y., Farrell, J.A., 2016. High-precision vehicle navigation in urban environments using an mems imu and single-frequency gps receiver. *IEEE Transactions on Intelligent Transportation Systems* 17, 2854–2867. doi:[10.1109/TITS.2016.2543664](https://doi.org/10.1109/TITS.2016.2543664).
 - [231] Zhao, Y., Gao, S., Wang, Y., Yang, X., 2023. Dynavol: Unsupervised learning for dynamic scenes through object-centric voxelization. *arXiv preprint arXiv:2305.00393*.
 - [232] Zhen, H., Qiu, X., Chen, P., Yang, J., Yan, X., Du, Y., Hong, Y., Gan, C., 2024. 3d-vla: A 3d vision-language-action generative world model. *arXiv preprint arXiv:2403.09631*.
 - [233] Zheng, H., Tan, F., Wang, R., 2016. Pole-like object extraction from mobile lidar data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41, 729–734.
 - [234] Zheng, W., Chen, W., Huang, Y., Zhang, B., Duan, Y., Lu, J., 2023. Oc-cworld: Learning a 3d occupancy world model for autonomous driving. *arXiv preprint arXiv: 2311.16038*.
 - [235] Zhong, S., Qi, Y., Chen, Z., Wu, J., Chen, H., Liu, M., 2023a. Dcl-slam: A distributed collaborative lidar slam framework for a robotic swarm. *IEEE sensors journal* 24, 4786–4797.
 - [236] Zhong, Z., Rempe, D., Chen, Y., Ivanovic, B., Cao, Y., Xu, D., Pavone, M., Ray, B., 2023b. Language-guided traffic simulation via scene-level diffusion, in: Conference on robot learning, PMLR. pp. 144–177.
 - [237] Zhou, J., Bischof, W.F., Caelli, T., 2006. Road tracking in aerial images based on human-computer interaction and bayesian filtering. *ISPRS journal of photogrammetry and remote sensing* 61, 108–124.
 - [238] Zhou, X., Han, X., Yang, F., Ma, Y., Knoll, A.C., 2025a. Opendrivevla: Towards end-to-end autonomous driving with large vision language action model. *arXiv preprint arXiv:2503.23463*.
 - [239] Zhou, Z., Cai, T., Zhao, S.Z., Zhang, Y., Huang, Z., Zhou, B., Ma, J., 2025b. Autovla: A vision-language-action model for end-to-end autonomous driving with adaptive reasoning and reinforcement fine-tuning. *arXiv preprint arXiv:2506.13757*.
 - [240] Zhu, D., Munderloh, M., Rosenhahn, B., Stückler, J., 2019. Learning to disentangle latent physical factors for video prediction, in: German Conference on Pattern Recognition, Springer. pp. 595–608.
 - [241] Zhu, F., et al., 2020. A multi-gnss differential phase kinematic post-processing method. *Remote Sensing* 12, 2727. doi:[10.3390/rs12172727](https://doi.org/10.3390/rs12172727).
 - [242] Zhu, F., Zhang, H., 2020. A method of improving ambiguity fixing rate for post-processing kinematic gnss data. *Satellite Navigation* 1, 12. doi:[10.1186/s43020-020-00022-y](https://doi.org/10.1186/s43020-020-00022-y).
 - [243] Zhu, J., Yang, Z., Shi, Y., Fang, J., Xue, J., 2024a. Ic-mapper: Instance-centric spatio-temporal modeling for online vectorized map construction, in: Proceedings of the 32nd ACM International Conference on Multimedia, pp. 9961–9969.
 - [244] Zhu, L., Li, Y., Sandström, E., Huang, S., Schindler, K., Armeni, I., 2024b. Loopsplat: Loop closure by registering 3d gaussian splats. *arXiv preprint arXiv:2408.10154*.
 - [245] Zou, D., Tan, P., 2012. Coslam: Collaborative visual slam in dynamic environments. *IEEE transactions on pattern analysis and machine intelligence* 35, 354–366.
 - [246] Zuo, S., Zheng, W., Huang, Y., Zhou, J., Lu, J., 2025. Gaussianworld: Gaussian world model for streaming 3d occupancy prediction, in: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 6772–6781.

Appendix A. Related Opensourced Work

Table A.1: Part of opensourced work for implicit map

Year	Methods	Opensource Code
2019	DisentPhys	github.com/TsuTikgiau/DisentPhys4VidPredict/
2021	COLA	github.com/google-research/google-research/tree/master/cola
2022	BEVFormer NerfAcc NeRFusion SinNeRF	github.com/HFAiLab/BEVFormer github.com/nerfstudio-project/nerfacc github.com/jetd1/NeRFusion github.com/VITA-Group/SinNeRF
2023	ExpQD SparseBEV DynaVol LaRa PanNeRF360 NeRFDet OccFormer	github.com/codezakh/decomposition-0shot-vqa github.com/MCG-NJU/SparseBEV github.com/zyp123494/DynaVol github.com/valeoai/LaRa github.com/fuxiao0719/PanopticNeRF/tree/panopticnerf360 github.com/facebookresearch/NeRF-Det github.com/zhangyp15/OccFormer
2024	PotentialMP OccWorld IC-Mapper Neural MP Vit-Lens MBNSF BiGR LightningNeRF NeuRAD GraphAD DriveDreamer OccSora CurbSG	github.com/devinluo27/potential-motion-plan-release github.com/wzzheng/OccWorld github.com/Brickzhuantou/IC-Mapper github.com/mihdalal/neuralmotionplanner github.com/TencentARC/ViT-Lens github.com/kavisha725/MBNSF github.com/haoosz/BiGR github.com/VISION-SJTU/Lightning-NeRF github.com/georghess/neurad-studio github.com/zhangyp15/GraphAD github.com/JeffWang987/DriveDreamer github.com/wzzheng/OccSora github.com/robot-learning-freiburg/CURB-SG
2025	GaussianWorld	github.com/zuosc19/GaussianWorld

* Updated by June 2025.

Table A.2: Part of opensourced work for online vectorization

Year	Methods	Opensource Code
2022	HDMaNet	github.com/Tsinghua-MARS-Lab/HDMaNet
2023	CenterLineDet VectorMapNet MapTR NMP PivotNet BeMapNet MapVR TopoNet ScalableMap	github.com/TonyXuQAQ/CenterLineDet github.com/Mrmoore98/VectorMapNet_code github.com/hustvl/MapTR github.com/Tsinghua-MARS-Lab/neural_map_prior github.com/wenjie710/PivotNet github.com/er-muyue/BeMapNet github.com/ZhangGongjie/MapVR github.com/OpenDriveLab/TopoNet github.com/jingyiyu/ScalableMap
2024	MapTRV2 SMERF GeMap SatforHDMaNet LaneGAP Topo2D TopoLogic LaneGraph2Seq ADMap CGNet StreamMapNet P-MapNet MapQR SQD-MapNet MapTracker MGMap GenMapping HRMapNet	github.com/hustvl/MapTR github.com/NVlabs/SMERF github.com/cnzzx/GeMap github.com/xjtu-cs-gao/SatforHDMaNet github.com/hustvl/LaneGAP github.com/homothetic/Topo2D github.com/Franpin/TopoLogic github.com/fudan-zvg/RoadNet github.com/hht1996ok/ADMap github.com/XiaoMi/CGNet github.com/yuantianyan01/StreamMapNet github.com/jike5/P-MapNet github.com/HXMap/MapQR github.com/shuowang666/SQD-MapNet github.com/woodfrog/maptracker github.com/xiaolul2/MGMap github.com/lynn-yu/GenMapping github.com/HXMap/HRMapNet
2025	Uni-PrevPredMap PrevPredMap FastMap MapFM SDTagNet SuperMapNet UIGenMap MapDR InstaGraM Chameleon	github.com/pnnnnnnn/Uni-PrevPredMap github.com/pnnnnnnn/PrevPredMap github.com/hht1996ok/FastMap github.com/LIvanoff/MapFM github.com/immelf/SDTagNet github.com/zhouruqin/SuperMapNet github.com/xiaolul2/UIGenMap github.com/MIV-XJTU/MapDR github.com/juyebshin/InstaGraM github.com/XR-Lee/neural-symbolic

* Updated by June 2025.