Cooperative Perception for Automated Driving: A Survey of Algorithms, Applications, and Future Directions

Chuheng Wei*, *Student Member, IEEE*, Guoyuan Wu, *Senior Member, IEEE*, Matthew J. Barth, *Fellow, IEEE*

***Abstract*—Cooperative Perception (CP) has emerged as an important research topic in connected and automated transportation systems, significantly enhancing situational awareness in challenging conditions such as limited communication bandwidth, harsh weather, and poor lighting. By sharing and aggregating data from multiple sources—including vehicles, roadside units, and other infrastructure—CP addresses critical issues like occlusion and long-range perception in complex environments. This paper provides a comprehensive review of existing CP methodologies, evaluating their effectiveness across diverse scenarios and fusion layers. We investigate approaches ranging from early to late collaboration strategies, highlighting how multi-modal sensing at the network level can improve detection accuracy, system robustness, and adaptability. Additionally, we review publicly available datasets, identifying key gaps in their coverage of heterogeneous sensors, adverse conditions, and large-scale spatial scenarios. Finally, we propose actionable future research directions, including advanced algorithms for adverse environments, adaptive fusion strategies for heterogeneous sensors, and the integration of emerging technologies such as HD maps, large language models, and end-to-end frameworks. These advancements aim to ensure reliability, interoperability, and scalability for real-world connected and automated vehicle applications.**

***Index Terms*—Cooperative Perception, Sensor Fusion, Adverse Conditions, Vehicle-to-Everything, Dataset, Survey**

I. Introduction

As Connected and Automated Vehicle (CAV) technologies continue to evolve, Cooperative Perception (CP) has become quite valuable for enhancing safety and situational awareness in complex driving environments. By leveraging shared information from multiple sources (i.e., nodes), such as vehicles, roadside infrastructure, and other connected entities, CP augments each participant's perception capabilities through data exchange [1]. This sharing process typically relies on communication frameworks such as Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and Vehicle-to-Everything (V2X), enabling individual vehicles to benefit from a collective perception network [2]. Through such collaboration, CP extends sensing range and addresses challenges like occlusion [3], making it particularly effective in scenarios such as dense urban intersections [4] and high-speed highways [5].

In real-world deployments, however, achieving robust CP under adverse conditions remains challenging. Issues such as communication constraints, sensor degradation, hardware failures, and unfavorable weather often complicate the deployment of reliable perception for connected and automated vehicles. In particular, limited bandwidth and unpredictable delays in V2X scenarios can hinder the timely exchange of sensing data, diminishing the benefits of collective perception [6]. Additionally, environmental factors like fog, rain, and low-light conditions introduce noise and reduce detection accuracy, while hardware problems—such as low-resolution sensors or physical damage—further degrade perception quality [1], [7].

To address these challenges, a variety of CP approaches have been explored. These methods span multiple fusion layers—early, intermediate, late, and hybrid—and employ flexible data-sharing schemes, from raw sensor data to feature-level and object-level representations. With increased emphasis on V2X-enabled perception, CP strategies that share and aggregate multi-modal data from multiple nodes show great promise in mitigating occlusion, improving long-range perception, and maximizing system resilience to adverse environments.

We further assess state-of-the-art datasets, identifying critical gaps, such as limited sensor heterogeneity, inadequate coverage of adverse conditions, and constrained spatial diversity. While existing datasets have driven progress in perception tasks, they often lack the complexity required to fully evaluate CP system robustness under real-world scenarios. To address this, we discuss emerging technologies, including HD maps, large language models (LLMs), and 5G/6G communication, which introduce new opportunities to advance CP systems and overcome these limitations.

In addition, this paper addresses non-technical challenges, such as the lack of standardized protocols, trustworthiness (i.e., integrity) concerns , and ethical considerations, all of which hinder large-scale deployments. By combining a comprehensive evaluation of methodologies, datasets, and emerging technologies, this review sets the stage for advancing CP systems, ensuring their scalability, robustness, and adaptability in real-world automated driving applications.

The contributions of this paper are as follows:

1) **Comprehensive Method Evaluation:** This paper con-

Chuheng Wei, Dr. Guoyuan Wu, and Dr. Matthew J. Barth are with the College of Engineering, Center for Environmental Research and Technology, University of California at Riverside, Riverside, CA, 92507.

* Corresponding author. E-mail: chuheng.wei@email.ucr.edu

Manuscript received May 19, 2025; revised February 14, 2025.

ducts a rigorous and systematic evaluation of multi-agent cooperative perception (CP) methodologies. We emphasize their theoretical underpinnings, practical applications, and performance trade-offs, categorizing approaches across different collaboration layers (e.g., early, intermediate, and late) and operational contexts. By examining these diverse methods, we provide a holistic understanding of CP's capabilities and limitations, highlighting how various strategies address real-world challenges such as communication constraints, adverse environments, and heterogeneous sensor setups.

2) **Dataset Insights:** Through an in-depth analysis of both virtual and real-world datasets, this study identifies critical shortcomings in their diversity, multimodal coverage, and scalability. The findings underscore the importance of creating datasets that reflect the complexity of real-world CP applications, fostering advancements in robust and adaptable CP systems.
3) **Emerging Trends and Technologies:** This paper explores the opportunities and advantages for CP development arising from advancements in high-definition maps, large language models (LLMs), next-generation communication technologies (5G/6G), and end-to-end automated driving. These emerging innovations present new venues to address persistent challenges in CP, such as communication efficiency, multimodal integration, and operational reliability.
4) **Advancing Real-World Applicability Through Cross-Disciplinary Synergies:** The paper examines how CP systems can be refined to tackle real-world challenges, including dynamic environmental conditions, heterogeneous hardware configurations, and communication constraints. By integrating interdisciplinary insights, the study outlines strategies to improve the reliability, scalability, and adaptability of CP technologies for practical automated driving deployments.
5) **Comprehensive Gap Analysis and Future Research Directions:** This paper identifies fundamental gaps across the dimensions of technology, non-technical, and datasets in CP research, as well as providing actionable research directions. These include developing perception algorithms resilient to adverse conditions, adaptive fusion strategies for heterogeneous sensors, and scalable frameworks designed for real-world operational diversity.

II. System Architecture Framework for Cooperative Perception

A. Sensors in Automated Driving Systems

Sensors are fundamental to the perception capabilities of automated driving systems, enabling vehicles to understand, interpret, and interact with their surroundings. These sensors form the backbone of Advanced Driver Assistance Systems (ADAS) and higher levels of automation, where precise and reliable environmental data are critical for safety and operational efficiency. The integration of various sensing modalities ensures robust perception by leveraging complementary data sources, mitigating the limitations of individual sensors. This multi-sensor framework not only enhances situational awareness but also builds resilience against adverse conditions and unexpected scenarios. Fig. 1 illustrates a typical sensor configuration on a modern automated vehicle. These include LiDAR for 3D spatial mapping, cameras for visual context, radar for velocity and range measurement, thermal cameras for low-light detection, GNSS for positioning, IMU/INS for motion estimation, and ultrasonic sensors for near-field obstacle detection. Below, we outline the characteristics and applications of key sensors used in automated vehicles, emphasizing their specific advantages and constraints.

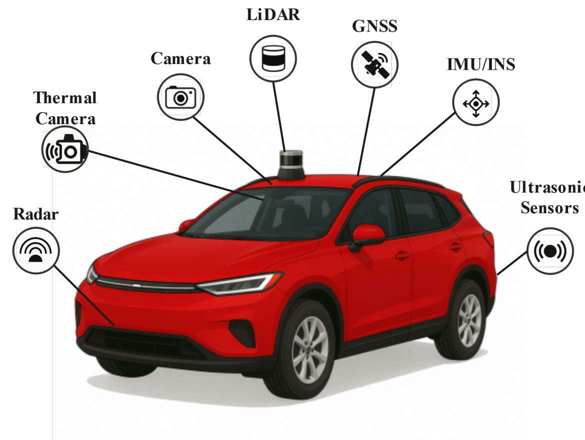

Fig. 1. An overview of typical sensors mounted on automated vehicles

a) Camera: Cameras function by capturing light reflected from objects in the environment and converting it into digital images for subsequent analysis [8]. As one of the most prevalent sensor types in automated vehicles, cameras provide high-resolution visual data that closely resembles human visual perception. They are particularly effective in tasks such as object detection [9], lane recognition [10], and traffic sign identification [11], facilitated by advancements in computer vision and deep learning methodologies. Specialized camera configurations, such as fisheye cameras with ultra-wide-angle lenses, can capture 360-degree fields of view for comprehensive situational awareness, though they require distortion correction for accurate data processing [12], [13]. Despite these strengths, cameras face inherent challenges in accurately estimating 3D spatial information [14] and velocities [15], especially in scenarios involving dense traffic and occlusions [16]. Furthermore, their performance is highly susceptible to degradation under suboptimal lighting conditions or adverse weather [17].

b) Radar: Radar (Radio Detection and Ranging) uses radio waves to detect objects by measuring the time delay and frequency shift of the returned signals [18]. This mechanism makes radar well-suited for detecting motion and measuring object velocity. Radar sensors are essential for velocity measurement and detecting objects under adverse weather conditions, such as rain or fog [19]. They are particularly effective in applications requiring motion analysis, such as collision avoidance and parking assistance. However, their lower resolution compared to cameras and LiDAR makes them less suitable for distinguishing objects with fine detail [20]. Despite these challenges, radar's robustness to environmental disturbances makes it indispensable for automated driving.

c) LiDAR: LiDAR (Light Detection and Ranging) works by emitting laser pulses and measuring the time it takes for the reflected signals to return, enabling precise distance calculations [21]. This allows LiDAR to generate detailed 3D point clouds that represent the surrounding environment. Their ability to create detailed point clouds makes them highly effective in detecting objects and mapping environments with variable features [22]. LiDAR is particularly valuable in tasks requiring fine-grained spatial information, such as obstacle avoidance [23] and path planning [24]. However, its sparse data output and high cost remain significant limitations, making it less optimal for certain real-time applications [25]. Recently, Frequency-Modulated Continuous-Wave (FMCW) LiDAR has emerged as a promising technology that combines the advantages of LiDAR and radar, offering both precise 3D spatial information and direct velocity measurements [21].

d) Thermal Camera: Thermal cameras detect infrared radiation emitted by objects, allowing them to visualize heat signatures instead of visible light [26]. This capability makes thermal cameras effective for detecting pedestrians, animals, or other heat-emitting objects, especially in low-light or nighttime conditions [27]. Their ability to adapt to changes in lighting conditions enhances their usability in scenarios where conventional cameras may fail. However, thermal imaging typically provides lower resolution than standard cameras and requires specialized algorithms for accurate interpretation.

e) Ultrasonic Sensors: Ultrasonic sensors function by emitting high-frequency sound waves and measuring the time it takes for the echoes to return after bouncing off nearby objects [28]. This mechanism enables them to detect obstacles at close range, making them ideal for applications like parking assistance and low-speed navigation [28]. Their simplicity and low cost make them popular for proximity detection, but their limited range and susceptibility to environmental noise restrict their broader applicability in high-speed or long-range scenarios.

f) GNSS: GNSS (Global Navigation Satellite System) receivers provides location data by receiving signals from satellites to calculate the vehicle's absolute position [29]. It is crucial for navigation and localization, especially in mapping and long-distance driving applications. Augmented GNSS techniques, such as Real-Time Kinematic (RTK), significantly improve accuracy, down to the decimeter level [30]. However, GNSS signals can be disrupted in urban canyons, tunnels, or areas with dense foliage, which limits its reliability in certain environments [31].

g) IMU/INS: Inertial Measurement Units (IMU) and Inertial Navigation Systems (INS) measure vehicle motion by detecting acceleration, angular velocity, and magnetic field using accelerometers, gyroscopes, and magnetometers [32]. These sensors provide critical information about vehicle dynamics, orientation, and short-term positioning, making them essential for motion estimation and sensor fusion applications. IMU/INS data are particularly valuable for bridging gaps in GNSS coverage, such as in tunnels or urban canyons, and for providing high-frequency motion updates that complement lower-rate positioning systems [33]. However, IMU sensors are prone to drift over time due to bias accumulation, requiring periodic calibration and fusion with absolute positioning systems like GNSS to maintain accuracy [34].

While automated vehicles employ diverse sensor modalities as described above, current cooperative perception algorithms primarily focus on camera and LiDAR integration. This concentration reflects both the maturity of perception algorithms for these sensors and their proven effectiveness in object detection and environmental mapping tasks. However, recent cooperative perception datasets have begun incorporating diverse sensor types such as radar [35] and thermal cameras [36] to address this limitation.

Other sensor modalities, such as GNSS and IMU/INS, while essential for cooperative perception systems, primarily provide positioning, motion, and proximity information rather than direct environmental perception data. These sensors enable critical functions such as spatial alignment between cooperative agents, coordinate transformation, and vehicle state estimation. However, since their primary role is to support rather than perform perception tasks, they are not typically considered core perception sensors in fusion algorithm design. Therefore, the subsequent sections on cooperative perception methods focus primarily on camera and LiDAR-based approaches, which constitute the main perception sensing modalities in current research.

B. Systematic Architecture of Cooperative Perception

Cooperative perception systems require a robust architecture not only for performance, scalability, and robustness, but also for adaptability to various operational challenges, particularly under adverse circumstances such as rain, fog, nighttime driving, and occlusions in congested urban areas [37]. It is imperative for these systems to integrate data from multiple sources—such as vehicles, roadside infrastructure (e.g., RSUs), and cellular base stations—in order to gain a comprehensive understanding of the surroundings [38], [39]. These systems must address issues such as data heterogeneity, communication constraints, and dynamic environmental changes while ensuring low latency and high reliability [40].

Fig. 2 depicts an example of how cooperative perception can be implemented in a city environment. The map highlights overlapping coverage zones of roadside units (RSUs) and cellular base stations, illustrating how vehicles leverage either Dedicated Short-Range Communications (DSRC) or Cellular Vehicle-to-Everything (C-V2X) links for data exchange. Within each coverage area, vehicles and RSUs communicate sensor information in real time, enabling collaborative awareness of critical events such as sudden braking, traffic congestion, or obstacles in blind spots. Simultaneously, cellular infrastructure can relay aggregated data to cloud services for large-scale analytics and further decision support.

System efficiency and adaptability are greatly influenced by the choice of architecture. Centralized models consolidate global data in the cloud or at powerful edge nodes, offering

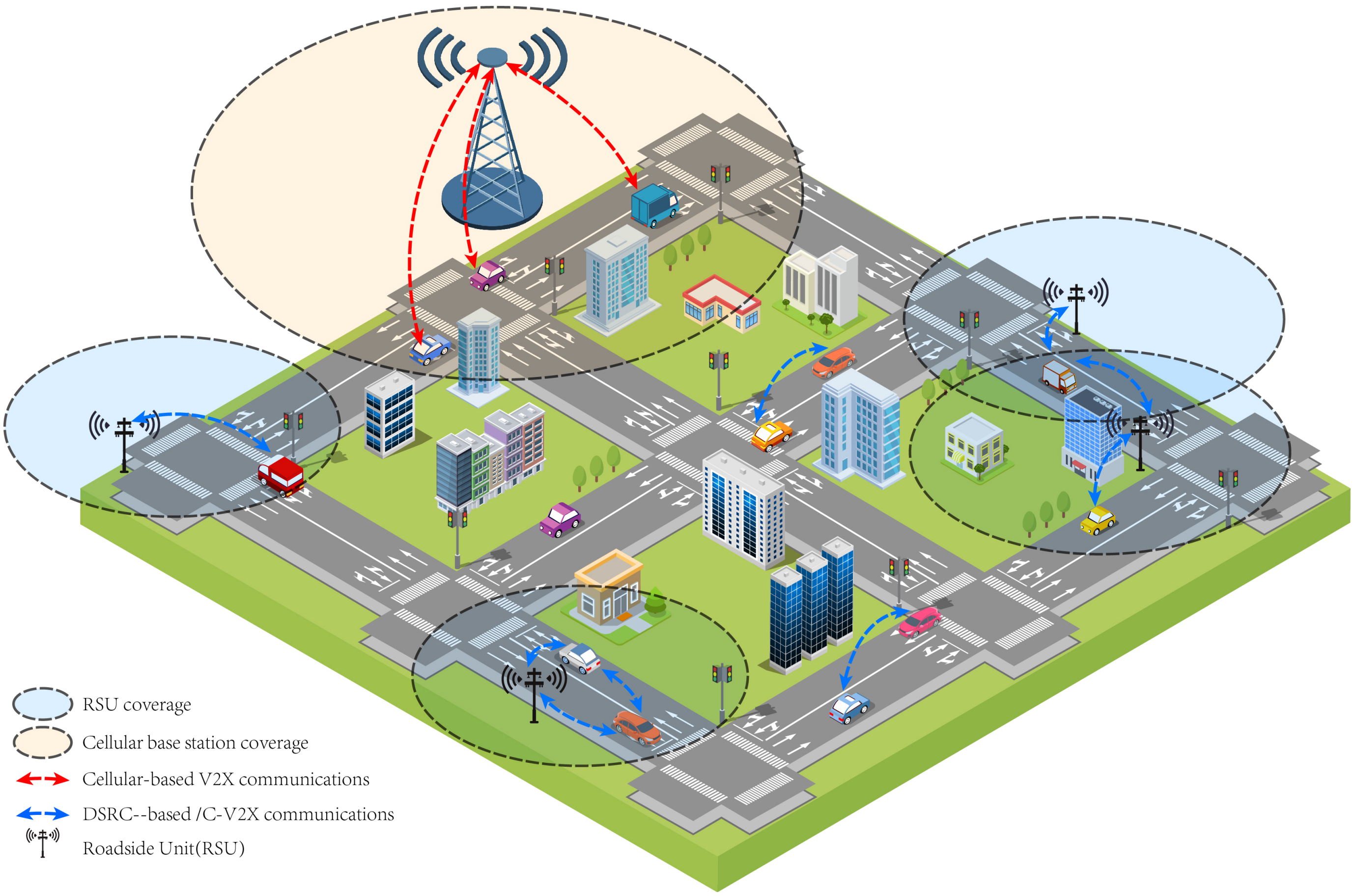

Fig. 2. An urban scenario illustrating cooperative perception among multiple vehicles and infrastructure. The figure shows Vehicle-to-Everything (V2X) communication networks including Dedicated Short-Range Communications (DSRC) and Cellular Vehicle-to-Everything (C-V2X) data exchange protocols enabling communication between connected vehicles and roadside units (RSUs).

high computational power but relying heavily on stable connectivity. In contrast, distributed models enable local, on-board processing to provide flexibility and resilience. Hybrid models seek to balance these trade-offs by combining the benefits of centralized data aggregation with localized autonomy [20], [37]. In any of these architectural setups, the communication backbone—whether DSRC, C-V2X, or cellular—plays a pivotal role in ensuring low-latency, reliable data sharing for cooperative perception.

1) Centralized Architectures:

a) Cloud Computing: Cloud computing represents one of the most established centralized architectures in cooperative perception. In this approach, raw data from vehicle and infrastructure sensors are transmitted to cloud servers, where powerful computational resources are used for tasks such as data fusion, 3D mapping, and object detection [41]. The ability to perform sophisticated analyses and leverage historical data stored in the cloud provides significant advantages in scalability and long-term system learning [42]. Cloud servers can also integrate data from multiple regions to build large-scale, high-accuracy maps [43]. However, this approach is heavily reliant on high-bandwidth, low-latency communication networks. In scenarios with poor connectivity, such as remote or rural areas, the performance of cloud-based systems can degrade significantly [44], [45]. Additionally, the latency introduced by data transmission to and from the cloud may be unacceptable for time-critical applications, such as collision avoidance and emergency braking.

b) Infrastructure-Based Computing: Infrastructure-based computing shifts the computational workload from centralized cloud servers to localized roadside units (RSUs) or edge servers [46]. RSUs are strategically placed within the transportation network to collect and process data from vehicles within their coverage area [47]. This architecture reduces reliance on long-distance communication links and enables real-time processing for applications such as traffic flow management and dynamic obstacle detection [2]. By aggregating data locally, infrastructure-based computing supports efficient data sharing among nearby vehicles and enhances system responsiveness. However, the deployment of RSUs and associated edge computing infrastructure involves substantial upfront investment and maintenance costs [48]. Scalability remains a challenge in urban environments, where high vehicle densities may overwhelm the processing capabilities of individual RSUs [49].

2) Distributed Architectures:

a) Vehicle-Based Computing: Vehicle-based computing decentralizes perception by enabling individual vehicles to perform onboard data processing. Equipped with high-performance processors and specialized hardware, vehicles can handle tasks such as sensor fusion, object detection, and trajectory planning independently [8], [50], [51]. This approach minimizes the dependency on external infrastructure and reduces communication overhead, making it particularly robust in scenarios with unreliable network coverage [20], [52]. However, variations in the computational capabilities of different vehicles and inconsistencies in sensor quality can lead to disparities in perception accuracy. Furthermore, the lack of global context in purely vehicle-based systems may limit their ability to address complex cooperative scenarios, such as multi-vehicle trajectory prediction in dense traffic [53].

b) Edge Computing: Edge computing distributes computational tasks across devices located at the network's edge, including vehicles, RSUs, and localized servers [54], [55]. By processing data closer to its source, edge computing significantly reduces latency and mitigates network congestion. It supports real-time cooperative perception by enabling rapid data exchange and localized decision-making, particularly in applications such as collision avoidance and intersection management [56]. This architecture is highly adaptable, capable of dynamically allocating computational resources based on demand [57]. However, challenges include ensuring consistency across distributed nodes, managing resource constraints, and addressing security vulnerabilities, particularly when multiple stakeholders operate the edge devices [58].

3) Hybrid Architectures:

Hybrid architectures integrate features from both centralized and distributed models to optimize performance across diverse conditions. In a typical hybrid setup, vehicles and edge devices handle time-sensitive tasks locally, such as object detection and path planning, while cloud servers perform computationally intensive tasks like large-scale data aggregation and model training [59], [60]. This layered approach reduces the communication burden on the cloud and enhances system scalability. By leveraging the cloud's ability to store and analyze historical data, hybrid architectures also enable continuous learning and adaptation to evolving traffic patterns and environmental conditions [61]. Nevertheless, designing effective hybrid systems requires careful coordination to ensure seamless integration of local and global data streams. This includes aligning data formats and timestamps across nodes, maintaining consistency between real-time decisions made locally and aggregated insights from the cloud, and handling network interruptions through data buffering and synchronization mechanisms. Improvements can be achieved by developing unified data schemas, adaptive synchronization strategies, and hierarchical decision frameworks that balance immediacy and global context.

Each architectural model offers unique trade-offs in latency, scalability, computational requirements, and robustness [8]. Centralized architectures excel in global data aggregation and resource-intensive processing but are hindered by communication delays and reliance on stable connectivity. Distributed architectures enhance flexibility and responsiveness, particularly in dynamic environments, but face challenges in maintaining consistency and scalability. Hybrid models attempt to balance these trade-offs, leveraging the strengths of both centralized and distributed approaches to provide adaptable solutions for cooperative perception. Understanding the interplay of these architectures is critical for developing systems that maintain reliable performance under adverse conditions.

C. Adverse Conditions Affecting Perception Systems

There are various adverse conditions that can substantially impair perception capabilities of automated driving systems that operate in diverse and dynamic environments. The challenges stem from environmental factors and hardware limitations, each of which presents a unique obstacle to acquiring, processing, and making reliable decisions. Here, we discuss these conditions in detail and examine how cooperative perception can mitigate these negative effects.

1) Environmental Factors:

Adverse weather conditions, such as rain, fog and snow, pose significant challenges to automated driving systems [62], [63]. Rain introduces additional noise to images and LiDAR data, as water droplets on sensors or road surfaces can create reflections and distort measurements [64]. Fog reduces visibility and attenuates the strength of signals from both cameras and LiDAR, limiting the effective sensing range [65]. Low-light conditions, such as nighttime driving, exacerbate the limitations of cameras by reducing their ability to capture high-quality images and detect objects accurately [66].

Detection, tracking, and localization performance are affected by these environmental factors directly, resulting in poor accuracy and reliability. If traffic signs or lane markings are obscured by rain or fog, they are more likely to be misclassified or missed [67], [68]. The effectiveness of visual-based sensors is further diminished in low-light conditions when visibility is low, making it difficult to see pedestrians, vehicles, and road boundaries [69]. Especially in complex or high-risk scenarios, automated systems are unable to make safe and reliable decisions because of these limitations.

By sharing complementary sensor data, cooperative perception offers a promising solution to these challenges. To support individual vehicles, infrastructure-based sensors, which may not be affected by adverse weather conditions, can provide additional perspectives or redundant information [70]. Feature-level or object-level data sharing allows automated systems to aggregate and integrate diverse sensor inputs, improving robustness under conditions where individual sensors might fail [71]. Furthermore, CP enables dynamic adaptation to environmental challenges by leveraging real-time data from multiple nodes to extend the effective sensing range and reduce blind spots.

2) Hardware Limitations:

Individual automated vehicles and their hardware components, including sensors and computing units, are prone to limitations that can affect perception accuracy. Low-resolution cameras and aged ISP (Image Signal Processing) systems may fail to capture fine details in complex scenes, reducing object detection and classification reliability [72], [73]. Low-resolution LiDAR sensors generate sparse point clouds, which may lack the density required for precise 3D perception in scenarios such as urban intersections or dense traffic [74]. Physical damage, such as cracked lenses [75] or degraded sensor materials [76], further compounds these issues, leading to inconsistent or inaccurate sensor outputs.

These hardware limitations constrain the ability of automated systems to operate effectively, particularly in scenarios requiring high precision and reliability. For example, sparse LiDAR data may result in incomplete object contours, making it challenging to detect small or partially obscured objects [77]. Similarly, low-resolution images or damaged camera lenses can hinder the identification of critical visual cues, such as traffic signs or road surface markings, increasing the risk of errors in navigation or decision-making [72].

Cooperative perception mitigates these hardware limitations by enabling vehicles to leverage sensor data from other nodes within the network. Vehicles equipped with higher-quality sensors can share processed information, such as object positions or feature maps, to support those with less capable hardware. Infrastructure-based sensors, which are often more robust and less susceptible to physical damage, can serve as reliable sources of supplementary data. This collaborative approach enhances overall system performance by compensating for individual sensor deficiencies and enabling consistent perception across heterogeneous platforms, demonstrating how CP systems turn hardware diversity into a strength rather than a weakness.

D. Representative Applications and Benefits of Cooperative Perception

Cooperative Perception is not only a technical framework for multi-agent sensor fusion, but also a critical enabler for real-world automated driving applications. In this subsection, we summarize several key use cases that demonstrate the practical value of CP and outline how it contributes to improved safety, robustness, and perception performance in complex scenarios. These applications focus on perception-level improvements and have been widely studied in both academic research and practical deployments.

1) Intersection Safety and Management:

Intersections represent one of the most critical applications for cooperative perception, where occlusions and limited field of view create significant safety challenges. By integrating vehicle-mounted sensors with roadside infrastructure, CP systems create comprehensive situational awareness that extends far beyond individual vehicle capabilities.

Real-world deployments demonstrate substantial improvements in intersection scenarios. The DAIR-V2X dataset [78] shows remarkable performance gains through cooperative perception: the VIC-Sync algorithm achieves 120% improvement in Average Precision for 3D object detection (AP_{3D}) compared to vehicle-only vision-based perception and 33% improvement over roadside-only vision systems. For point cloud-based detection, cooperative approaches demonstrate 60% improvement over single-vehicle systems and 183% improvement compared to roadside-only perception. Infrastructure-vehicle cooperation proves to be particularly effective for the detection of vulnerable road users, showing significant improvements in pedestrian detection accuracy in complex intersection scenarios with heavy occlusions.

2) Highway and High-Speed Scenarios:

Highway applications focus on long-range detection and coordination during high-speed maneuvers, where extended sensing range is critical for safe lane changes, merging, and collision avoidance. Highways present unique perception challenges due to high-speed vehicle movements that require extended detection ranges for enough reaction times, frequent occlusions caused by large vehicles (trucks, buses), and the need for precise long-range object tracking across multiple lanes. Cooperative perception addresses the fundamental limitation of individual vehicle sensors in providing sufficient early warning for high-speed decision-making.

Highway studies demonstrate substantial improvements through cooperative perception systems. Research shows continuous improvement in detection accuracy as participation rates increase, with optimal performance achieved at 70-80% participation rates in dense traffic scenarios [79]. Real-world testing reveals that cooperative systems maintain high-quality tracking performance under typical communication delays ($\sim$100ms), successfully eliminating false negatives from occlusions with track mismatch rates as low as 0.017%. Additionally, cooperative localization achieves Root Mean Square (RMS) positioning errors of 0.074m compared to 0.127m for single-vehicle systems, providing the precision necessary for safe high-speed automated driving operations [80].

3) Urban Dense Traffic Environment:

Urban environments present unique challenges with high object density, frequent occlusions, and complex traffic patterns. Cooperative perception leverages multiple viewpoints to maintain consistent object detection and tracking in these challenging conditions.

Research in dense urban scenarios demonstrates substantial improvements through cooperative perception. In roundabout scenarios with building-constrained line-of-sight conditions, cooperative systems show optimal performance with participation rates around 50%, with significant perception improvements observed even at 15% participation rates [79]. The spatial diversity gain from multiple sensors proves particularly effective, with early fusion schemes using two sensors outperforming the best single sensor by 20-85% in complex intersection scenarios. This improvement stems from increased point cloud density and reduced false negatives caused by occlusions, enabling better object detection quality and more accurate bounding box estimation in areas with overlapping sensor coverage [81].

4) Adverse Weather and Challenging Conditions:

One of the most promising applications of cooperative perception lies in adverse weather conditions, where individual vehicle sensors often experience significant performance

TABLE I
OVERVIEW OF MULTI-SENSOR FUSION METHODS AND THEIR APPLICATIONS(FOR INTERMEDIATE FUSION, THE CATEGORIES ARE: TRAD: TRADITIONAL FUSION, ATTN: ATTENTION-BASED FUSION, TOPO: TOPOLOGY-BASED FUSION)

Method	Year	Scenario	Collaborate Level					Sensors	Task
			Early	Intermediate			Late		
				Trad	Attn	Topo			
Cooper [83]	2019	V2V	✓					Lidar	Object Detection
V2X-PC [84]	2024	V2V	✓					Lidar	Object Detection
F-Cooper [85]	2019	V2V		✓				Lidar	Object Detection
[86]	2020	V2I		✓				Camera	Pedestrain Re-Identification
PillarGrid [87]	2022	V2I		✓				Lidar	Object Detection
[88]	2022	V2V		✓				Lidar	Object Detection
SiCP [89]	2023	V2I		✓				Lidar	Object Detection
CoCa3D [16]	2023	V2I, V2U		✓				Camera	Object Detection
CoBEVFusion [90]	2023	V2V		✓				Camera, Lidar	Object Detection, Semantic Segmentation
CooPercept [91]	2024	V2V		✓				Lidar	Object Detection
CoopDet3D [92]	2024	V2X		✓				Camera, Lidar	Semantic Segmentation
UniV2X [93]	2024	V2I		✓				Camera	Object Detection &Tracking, Occupancy
Attentive Fusion [94]	2022	V2V			✓			Lidar	Object Detection
VIC-Sync [78]	2022	V2I			✓			Camera or Lidar	Object Detection
COOPERNAUT [95]	2022	V2V			✓			Lidar	Object Detection
CoBEVT [96]	2022	V2V			✓			Lidar	BEV Semantic Segmentation
V2VFormer [97]	2023	V2V			✓			Lidar	Object Detection
V2VFormer++ [98]	2023	V2V			✓			Camera, Lidar	Object Detection
MKD-Cooper [99]	2023	V2V			✓			Lidar	Object Detection
VIMI [100]	2023	V2I			✓			Camera	Object Detection
HM-ViT [101]	2023	V2V			✓			Camera or Lidar	Object Detection
QUEST [102]	2024	V2I			✓			Camera	Object Detection
HP3D-V2V [103]	2024	V2V			✓			Lidar	Object Detection
V2VNet [104]	2020	V2V				✓		Lidar	Object Detection
V2X-ViT [105]	2022	V2X			✓	✓		Lidar	Object Detection
HYDRO-3D [106]	2023	V2X			✓	✓		Lidar	Object Detection &Tracking
[107]	2024	V2V			✓	✓		Lidar	Object Detection
V2X-ViTv2 [108]	2024	V2X			✓	✓		Lidar	Object Detection
CollabGAT [109]	2024	V2V			✓	✓		Lidar	Object Detection
[110]	2017	V2I					✓	Camera	Lane Detection
Poster [111]	2020	V2V					✓	Camera	Object Detection
[112]	2020	V2I					✓	Camera	Object Detection
VIPS [113]	2022	V2I					✓	Lidar	Object Detection
CooperFuse [114]	2024	V2X, I2I					✓	Lidar	Object Detection
[81]	2020	I2I	✓				✓	Lidar	Object Detection
DiscoNet [115]	2021	V2V	✓			✓		Lidar	Object Detection
ML-Cooper [116]	2022	V2V	✓	✓			✓	Lidar	Object Detection
[117]	2023	V2V			✓		✓	Lidar	Object Detection

degradation. Cooperative systems leverage data redundancy and complementary sensing capabilities to maintain perception performance when individual sensors fail.

Research demonstrates substantial improvements in adverse weather scenarios through cooperative perception. In highway scenarios with heavy rain (90 mm/h), cooperative perception with 20% equipment rate achieves 15% higher detection rates than local perception, while 40% equipment rate shows 21.8% improvement [82]. The Adver-City dataset provides additional evidence, revealing performance gaps of 12.59 AP@50 points between optimal and most challenging conditions for models not trained on adverse weather. When trained specifically on adverse weather data, cooperative systems demonstrate improved robustness across diverse conditions including rain, fog, and challenging lighting scenarios, highlighting the critical importance of weather-aware training for reliable cooperative perception deployment [70].

III. METHOD OF COOPERATIVE PERCEPTION

Cooperative perception (CP) methods can be categorized into three main fusion strategies based on the stage at which collaboration occurs: early fusion, intermediate fusion, and late fusion. These correspond to data-level fusion, feature-level fusion, and fusion of individual agents' independent outputs, respectively. As shown in Fig.3, these stages reflect the progressive integration of information in CP systems.

Early research in CP predominantly focused on data-level alignment in early fusion or the combination and filtering of outputs from individual nodes in late fusion. However, as researchers recognized the computational power and adaptability of neural networks, intermediate fusion methods have emerged as the mainstream approach. These methods leverage various feature fusion techniques to enhance performance [71], [118]. An overview of these algorithms is presented in Table I.

A. Early Fusion

Early fusion involves collaboration at the input data level, as illustrated in Fig. 3-(a). In this approach, raw data from

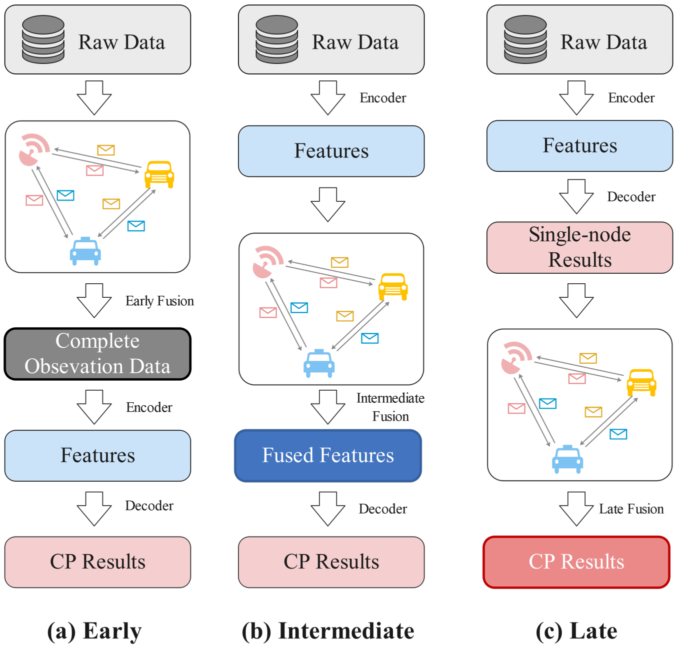

Fig. 3. Overview of different collaboration strategy for Connected Automated Vehicles: (a) Early fusion; (b) Intermediate fusion; (c) Late fusion

multiple agents are shared and aggregated to provide a more comprehensive view of the environment. This enables broader situational awareness and long-range perception, facilitating the completion of perception tasks from both holistic and extended perspectives.

One representative method, Cooper [83], integrates data from ego vehicles and other connected vehicles in a vehicular network to enhance perception capabilities. By leveraging lidar point clouds, Cooper fuses raw sensor data from multiple nodes at the data level. It proposes a 3D object detection framework to process aligned point cloud data, demonstrating the feasibility of performing cooperative perception through point cloud data transmission using existing vehicular network technologies. This method significantly improves detection accuracy and driving safety.

Raw sensor data, such as high-resolution images (typically 1-10 MB per frame) and dense LiDAR point clouds (1-7 MB per frame), require substantial bandwidth when transmitted at the frequencies needed for real-time perception (10-20 Hz), potentially consuming 100+ Mbps per vehicle [119]. As a result, sharing raw data typically requires substantial communication bandwidth and may lead to network congestion due to high data loads. This limitation has prompted recent algorithms to explore preprocessing techniques that reduce transmission volume. For instance, V2X-PC [84] adopts a novel approach to early fusion by defining a new message unit called the point cluster. Instead of directly transmitting raw point clouds, V2X-PC preprocesses the data into clusters and uses a Point Cluster Aggregation (PCA) module to match data from the same object across agents. This processed data is then utilized for cooperative perception, significantly alleviating communication overhead while maintaining perception accuracy.

B. Intermediate Fusion

Features extracted by neural networks are significantly more compact than raw sensor data because they represent high-level semantic information rather than pixel-level or point-level details. For example, while a raw LiDAR frame might contain hundreds of thousands of 3D points, its corresponding feature representation captures the essential spatial and semantic patterns in a much smaller dimensional space, typically reducing data volume by orders of magnitude. Intermediate fusion is now the most common approach in CP because it reduces the high communication demands and bandwidth constraints of early fusion methods. By sharing and fusing features instead of raw data, it significantly decreases data volume and network congestion, making CP more practical across multiple nodes. The structure is shown in Fig. 3-(b). Based on fusion principles, intermediate fusion methods can be divided into traditional fusion, topology-based fusion, and attention-based fusion [71], [118].

1) Traditional Intermediate Fusion Methods:

Traditional feature fusion methods involve basic feature aggregation techniques such as summation, max pooling, and average pooling. These straightforward operations allow for efficient feature sharing and simplification.

Building on the Cooper framework [83], Chen et al. [85] proposed F-Cooper, a cooperative perception framework that employs feature-based fusion during the training phase. This method capitalizes on the compact nature of features to enable real-time edge computing, effectively mitigating the bandwidth limitations of vehicular networks and avoiding network congestion. By focusing on feature-level sharing, F-Cooper ensures efficient perception without the excessive communication overhead associated with raw data sharing.

Subsequent research has refined traditional methods by focusing on the feature extraction and fusion of lidar data. For example, PillarGrid [87] and SiCP [89] achieve collaborative object detection by integrating features from both vehicle-side and roadside units. These methods leverage complementary perspectives to enhance detection accuracy. Zhang et al. proposed CooPercept [91], which utilizes voxel features processed by multiple agents. By applying a maxout function, the framework extracts the most salient features for semantic segmentation tasks, ensuring robust performance and efficient feature utilization across agents. Moreover, Wang et al. [88] proposed a method that utilizes Octomap-based point cloud compression to preprocess and reduce the size of lidar data before fusion. This approach specifically targets the irregular characteristics of Pillar data, using Octomap's hierarchical tree structure to efficiently compress and represent the data. Traditional fusion strategies are also applied in vision-based cooperative perception. Hu et al. introduced CoCa3D [16], a 3D object detection framework that improves robustness by sharing feature information among multiple agents. This approach expands the field of view of individual cameras and addresses depth ambiguity issues inherent to single-node systems. Additionally, Zeng et al. [86] developed a vehicle-road cooperative pedestrian recognition system that extracts pedestrian image features using convolutional neural networks

(CNNs). By adjusting feature dimensions based on wireless channel conditions, this system reduces communication overhead while maintaining real-time pedestrian detection reliability.

Furthermore, with automated vehicles increasingly equipped with multi-modal sensors, some methods leverage diverse sensor data for cooperative perception. CoBEVFusion, proposed by Qiao and Zulkernine [90], uses a Dual Window-based Cross-Attention (DWCA) module to integrate camera and lidar information from each vehicle into a unified BEV (Bird's Eye View) representation. By sharing BEV features among connected vehicles, CoBEVFusion enables semantic segmentation and object detection in the BEV domain. Zimmer et al. introduced CoopDet3D [92], which combines BEVFusion [120] and PillarGrid to encode sensor data into features. These features are aggregated using max pooling at the feature level for 3D object detection tasks, further demonstrating the versatility and effectiveness of traditional intermediate fusion methods.

2) Attention-based Intermediate Fusion Methods:

With the introduction of attention mechanisms [121], Transformers have emerged as a powerful alternative to traditional neural networks [122]. In cooperative perception (CP), feature-level fusion leveraging attention mechanisms enables efficient integration of multi-sensor information through dynamic weight assignment and global modeling capabilities. This not only enhances the robustness of perception systems by mitigating the influence of noisy data but also ensures reliability under real-world conditions [123]. Furthermore, the ability to optimally match multimodal features and process them in real-time makes CP more accurate, adaptable, and transparent in complex traffic scenarios.

Xu et al. introduced AttentionFusion [94], which incorporates a single-head self-attention fusion module. This module employs attention mechanisms at precise spatial locations to capture interactions within specific regions of the feature map. Chen et al. [103] proposed HP3D-V2V, a feature extraction network that processes lidar voxel point sets. This method utilizes max and mean reduction operators as an adaptive feature fusion module focused solely on spatial relationships. Meanwhile, MKD-Cooper [99] introduces a collaborative attention fusion (CAF) module that applies attention mechanisms across both spatial and channel dimensions, enabling deep feature aggregation among connected vehicles.

Beyond spatial alignment, attention mechanisms are also utilized for adaptive information fusion. COOPERNAUT [95] incorporates Point Transformer [124] methodologies for encoding and aggregating features, retaining spatial information through a point encoder and representation aggregator. V2VFormer employs processed 3D point cloud features and recalibrates feature importance based on positional relevance using a Spatial-Channel Transformer (SAT), while adaptively enabling semantic feature interaction with a Channel-Wise Transformer (CWT). Building on this, V2VFormer++ [98] integrates image information with a dynamic channel fusion (DCF) module, adaptively merging image and point cloud features for enhanced perception.

Attention mechanisms have also shown promise in 2D image-based cooperative perception. Xu et al. proposed Cobevt [96], which utilizes a fused axial attention module (FAX) to capture local and spatial interactions between views and agents. Fan et al.'s QUEST framework [102] employs a cross-agent query interaction module to achieve query fusion and complementation. Wang et al.'s VIMI [100] adopts a multi-scale cross-attention module, extracting features by generating attention weights for each scale through cross-attention operations, resulting in robust 3D object detection outcomes from images.

Despite these advances in point cloud and image-based features, most studies focus on fusing data from identical sensor types. However, in practical scenarios, connected and automated vehicles (CAVs) and roadside units often deploy heterogeneous sensors, necessitating fusion across different sensor types. To address this gap, Xiang et al. introduced HM-ViT, the first unified cross-modal V2V cooperative perception framework [101]. HM-ViT fuses BEV features from multiple agents using both local window-based attention and sparse global grid-based attention to capture local and global cues. This framework successfully enables information interaction and fusion across heterogeneous sensors.

3) Topology-based Intermediate Fusion Methods:

As cooperative perception inherently involves multiple nodes (agents), it is natural to model these nodes and their communication relationships as a topological graph. Each agent can be represented as a fixed or dynamic node, while the distances, communication conditions, or other relationships between nodes are represented as edge weights. Through message passing and information aggregation, the states of each node are iteratively updated. Consequently, many studies have incorporated topological graphs into CP models using methods such as graph neural networks (GNN) [125] and graph attention networks (GAT) [126] to address challenges in collaboration, complementarity, and efficient information exchange between nodes.

Wang et al. were among the first to apply GNN models to CP. They proposed a novel push-and-pull (P&P) framework, V2VNet [104]. This model first extracts lidar point cloud features and processes them into BEV (Bird's Eye View) representations. These features are then compressed and broadcast to other vehicles, where a spatially aware GNN [125] is used to achieve cross-vehicle feature fusion. This approach highlights the potential of graph-based models in integrating data across agents for improved perception.

Ahmed et al. [107] introduced a method that incorporates graph attention networks (GAT) into CP. By modeling agents as nodes in a topological graph, their relationships are encoded as edge weights determined by attention coefficients. This adaptive approach allows the model to learn the significance of each perception point dynamically, focusing on information-rich regions that contribute most to perception. Similarly, CollabGAT [109] builds on this structure by incorporating more complex attention mechanisms at both channel and spatial levels, further improving the model's ability to extract and utilize meaningful information.

Xu et al. proposed the V2X-ViT model [105], which alternates between heterogeneous multi-agent self-attention (HMSA) and multi-scale window self-attention (MSwin) layers. This architecture adaptively fuses features while capturing spatial interaction relationships among agents. The updated version, V2X-ViTv2 [108], enhances multi-scale perception capabilities, making it even more effective in handling heterogeneous multi-agent interactions. Building on these models, Meng et al. introduced HYDRO-3D [106], which incorporates a spatiotemporal deep neural network. HYDRO-3D integrates global object tracking operations with low-level local operations, achieving improved object detection and tracking performance within a topological graph framework.

C. Late Fusion

Late fusion methods, as shown in Fig. 3-(c), primarily integrate detection results using traditional data fusion techniques rather than relying on the perception capabilities of neural networks. These methods were more common during the early stages of CP development, as they assign confidence levels to different modules and combine their outputs for final detection.

Zhao et al. [110] proposed a method where vehicle-mounted cameras detected lane markings while roadside cameras identified vehicles. By sharing perception information between vehicle and roadside units, the system used Dempster-Shafer (DS) theory [127], [128] to update detection confidence, improving lane detection accuracy. Khalifa et al. [112] designed a multi-view cooperative framework for intersections, leveraging geometric transformations to map detection results between vehicle-mounted and roadside cameras. Pereira et al. developed the Poster platform [111], which integrates data from multiple sensors and GNSS modules to encode, decode, and fuse detection results efficiently.

To address the limitations of these early-stage methods, more recent approaches have focused on lightweight and efficient frameworks. Shi et al. [113] introduced VIPS, a system that encodes simplified representations of objects and their relationships for efficient post-detection matching. By using roadside lidar to track the ego vehicle's position, VIPS achieves reliable detection with minimal communication overhead. Similarly, Zheng et al. proposed CooperFuse [114], a late fusion algorithm that combines object detection and tracking results from individual vehicles. CooperFuse calculates motion and scale consistency by analyzing kinematic and dynamic properties across frames, integrating confidence scores, and assessing inter-frame scale changes to achieve robust bounding box fusion.

D. Hybrid Fusion Methods

In addition to standalone fusion strategies, hybrid fusion methods combine features from different stages of fusion, leveraging their respective strengths for enhanced cooperative perception. These methods often integrate the benefits of early and late fusion while incorporating advanced intermediate techniques for specific tasks, providing flexibility and adaptability across various scenarios.

In [81], Arnold et al. proposed two approaches for cooperative 3D object detection that exemplify the complementary strengths of hybrid fusion. Their early fusion method aggregates raw point cloud data from spatially distributed sensors, while the late fusion method integrates independently detected bounding boxes from these sensors. Experimental results demonstrated that combining these methods improved detection performance compared to using either approach alone.

Liu et al. [117] extended the hybrid concept by introducing a region-based fusion framework. For areas with overlapping sensor coverage, they utilized cross-agent attention and global context attention modules for intermediate fusion, achieving robust feature integration. For regions covered by a single sensor, late fusion methods based on Gaussian Mixture Models (GMM) were employed to ensure reliable result sharing, demonstrating the flexibility of hybrid approaches in handling diverse sensor configurations.

Other hybrid methods focus on adaptive and dynamic information utilization. Xie et al. [116] proposed ML-Cooper, a multi-layer cooperative perception framework that simultaneously transmits data, features, and results. To optimize performance under varying communication bandwidth constraints, the framework employs reinforcement learning to dynamically adjust the proportion of each type of information utilized. This adaptive strategy highlights the potential of hybrid fusion methods in real-world, bandwidth-limited scenarios.

Li et al. further explored hybrid fusion with their DiscoNet model [115], integrating knowledge distillation to align early and intermediate fusion strategies. In DiscoNet, the early fusion model serves as a teacher, providing feature maps that act as target outputs for the intermediate fusion student model. By enforcing consistency through knowledge distillation, DiscoNet achieves enhanced feature learning. Additionally, DiscoNet maps vehicle features to BEV space and employs an edge encoder to establish topological correlations among features from different vehicles. The edge encoder uses convolutional layers for downsampling and normalization, producing efficient feature representations that facilitate robust hybrid fusion.

By combining the strengths of different fusion stages and adapting to diverse conditions, hybrid fusion methods offer a comprehensive solution for cooperative perception challenges.

IV. Datasets of Cooperative Perception

Datasets play an essential role in the development of cooperative perception algorithms. Through the integration of multi-sensor and multi-view data, cooperative perception is better equipped to address key limitations faced by single-view datasets, such as occlusions, restricted field of view (FOV), and sparse point cloud density. These datasets enable the training and evaluation of models that rely on data from multiple agents or infrastructure sensors to enhance environmental awareness. Table II provides a detailed comparison of common cooperative perception datasets, categorizing them into simulated and real-world datasets. These datasets span various scenarios, environments, and tasks critical to advancing cooperative perception systems in automated and connected vehicles.

TABLE II
A DETAILED COMPARISON OF AUTOMATED DRIVING DATASETS FOR COOPERATIVE PERCEPTION.

Simulated Dataset	**Year***	**View**	**Images**	**Point Cloud**	**Objects**	**Classes**	**Sensors†**	**Platform**
CoopInf [81]	2020	V2V	20k	20k	N/A	3	C, L	CARLA
V2V-Sim [104]	2020	V2V	–	51k	N/A	1	L	LiDARsim
CODD [129]	2021	V2V	–	10k	26.6k	2	L, G	CARLA
COMAP [130]	2021	V2V	–	8k	226k	2	L	CARLA , SUMO
V2X-Sim [131]	2022	V2X	60k	10k	26k	2	C, L, G, I,	CARLA, SUMO
OPV2V [94]	2022	V2V	44k	11k	233k	1	C, L, G, I	OpenCDA, CARLA, SUMO
DOLPHINS [132]	2022	V2X	42k	42k	292k	2	C,L	CARLA
V2XSet [105]	2022	V2X	-	11k	-	1	L	OpenCDA, CARLA, SUMO
IRV2V [133]	2023	V2V	34k	8k	1.5M	1	C, L, G, I	CARLA
CARTI [134]	2023	V2I	–	8k	346k	2	L	CARLA
DeepAccident [135]	2023	V2X	57k	57k	285k	2	C, L	CARLA
Scope [136]	2024	V2X	17k	17k	575M	5	C, L	CARLA, SUMO
Adver-City [70]	2024	V2I	24k	24k	889k	6	C, L, G, I	OpenCDA, CARLA

Real-world Dataset	**Year***	**View**	**Images**	**Point Cloud**	**Objects**	**Classes**	**Sensors†**	**Region**
T&J [83]	2019	V2V	N/A	N/A	N/A	1	C, L, R, F, G, I	USA
IPS300+ [137]	2021	V2I	14k	14k	4.5M	7	C, L, R, G	China
DAIR-V2X [78]	2022	V2I	71k	71k	1.2M	10	C, L, R, G, I	China
V2V4Real [138]	2023	V2V	40k	20k	240k	5	C, L, G, I	USA
UGVT [139]	2023	I2U	204k	-	N/A	21	C	China
LUCOOP [140]	2023	V2X	-	54k	34k	5	L, G, I	Germany
V2X-Seq [141]	2023	V2I	15k	15k	10k	9	C, L	China
TUMTraf-V2X [92]	2024	V2I	2k	5k	29k	8	C, L, G, I	Germany
RCooper [142]	2024	I2I	50k	30k	N/A	10	C, L	China
InScope [143]	2024	I2I	-	21k	188k	4	L	China
V2XReal [144]	2024	V2X	171k	33k	1.2M	10	C, L, G, I	China
V2AIX [145]	2024	V2I	N/A	N/A	N/A	N/A	C, L, G	Germany
HoloVIC [146]	2024	V2I	100k	100k	11.5M	3	C, L, F, G	China
V2X-Radar [35]	2024	V2I	40k	20k	350k	5	C, L, R, G, I	China
ISC [36]	2024	I2I	N/A	N/A	N/A	18	C, L, T	USA

* The publication year refers to the year the dataset was introduced in the respective paper.
† Sensor abbreviations: C: Camera, L: LiDAR, R: Radar, F: Fisheye Camera, T: Thermal Camera, G: GPS/GNSS receiver, I: IMU (based on sensors explicitly mentioned in the papers or existing in the datasets).
”–” indicates data that is not included in the dataset.
”N/A” indicates information not provided in publicly available paper.

Fig. 4 illustrates representative samples from both simulated and real-world cooperative perception datasets spanning clear and challenging visibility conditions. The visual comparison demonstrates the diversity and complementary nature of these datasets, highlighting how simulated environments can provide controlled testing scenarios while real-world data captures authentic driving complexities. This diversity in data sources and environmental conditions is crucial for developing robust cooperative perception algorithms that can adapt to varying operational scenarios, from ideal conditions to challenging environments with reduced visibility and complex traffic interactions.

A. Simulated Datasets

Simulated datasets are often created using powerful simulation tools designed to generate comprehensive data under a wide variety of scenarios, including diverse lighting [135] and weather conditions [70], [135]. Simulators such as CARLA, SUMO, and OpenCDA provide extensive customizability and precise control over environment parameters, enabling the realistic re-creation of different driving environments such as urban streets, rural roads, and highways [147]. CARLA [148], for instance, allows for realistic rendering of cities, suburbs, and rural landscapes with dynamic elements, including traffic, pedestrians, and even adverse weather like rain and fog. Vehicles within CARLA can be equipped with a variety of different sensor models, including RGB cameras, depth sensors, LiDAR, and radar, to emulate real-world data collection. However, limitations exist, such as the absence of accurate intensity and deactivation calculations for simulated LiDAR sensors.

SUMO is a microscopic road traffic simulator that models vehicular movement and traffic behaviors, often integrated with CARLA to manage vehicle routes, speeds, and interactions for realistic multi-agent traffic flows. SUMO [149] also enables the modeling of vulnerable road users (VRUs) like pedestrians and cyclists, providing more comprehensive simulations. OpenCDA [150] builds upon these capabilities, particularly focused on cooperative driving applications, facilitating scenarios that require complex perception, planning, and interaction between connected and automated vehicles (CAVs). By using these simulators together, datasets can be generated with realistic multi-vehicle traffic patterns, sensor data, and communication delays, providing valuable data for developing cooperative perception algorithms in challenging, high-stakes environments [151].

a) CoopInf: The CoopInf dataset [81], created by Arnold et al. using the CARLA simulator, is designed for cooperative 3D object detection in dynamic driving scenarios. It includes T-junction and roundabout setups covered by fixed roadside

(a) Simulated : Clear Day (b) Simulated : Nighttime (c) Real-World : Clear Day (d) Real-World : Nighttime

Fig. 4. Example front-view camera images from four representative datasets used in cooperative perception research: (a) Simulated dataset under clear weather conditions from OPV2V [94], (b) Simulated dataset under nighttime conditions from Adver-City [70], (c) Real-world dataset under clear weather conditions from DAIR-V2X [78], (d) Real-world dataset under nighttime conditions from TUMTraf-V2X [92].(Some of these images cropped to uniform display dimensions while maintaining original aspect ratios.)

cameras capturing RGB and depth images at 400×300 resolution. Detection areas span 80×40 m (T-junction) and a 96 m square (roundabout). Depth maps are converted to point clouds with approximately 200,000 points per frame, and additive Gaussian noise is applied for realistic simulation. CoopInf's diverse configurations and object types make it well-suited for evaluating cooperative perception in complex traffic scenarios.

b) V2V-Sim: The V2V-Sim dataset [104] was released by Wang et al., specifically focusing on cooperative vehicle-to-vehicle (V2V) perception. This dataset is unique in that it uses LiDARsim to generate synthetic LiDAR data, enabling more accurate point cloud representations than CARLA alone. LiDARsim leverages real-world datasets as its base and generates point clouds through ray tracing, followed by neural network processing, to simulate realistic LiDAR reflections and occlusions. In V2V-Sim, real-world scenes from the ATG4D dataset are reconstructed in the simulator, with each sample containing an average of 10 CAVs, up to a maximum of 63 CAVs.

c) CODD: CODD [129], proposed by Arnold et al., is another synthetic dataset aimed at cooperative perception and Simultaneous Localization and Mapping (SLAM) tasks. Generated using CARLA, CODD comprises 108 diverse scenes across urban, suburban, and rural settings, involving 4 to 16 CAVs equipped with 64-layer 360° LiDAR sensors with a range of 100 m and a recording rate of 5 Hz. The dataset provides annotated 3D bounding boxes for cars and pedestrians, totaling 204,250 3D boxes. Additionally, CODD offers ground-truth data for point cloud registration, enabling training for 3D object detection under low-frequency data and challenging pedestrian-rich environments.

d) COMAP: The COMAP dataset [130], introduced by Yuan et al., is a CARLA-SUMO collaborative dataset focused on V2V and infrastructure-based cooperative perception. Recorded on CARLA's Town05 map, it includes 21 intersection scenes, each containing 2 to 20 CAVs equipped with 32-layer LiDAR and RGB cameras, both operating at 10 Hz. This dataset is structured to support joint LiDAR-camera 3D object detection and semantic segmentation across vehicle categories. COMAP emphasizes scenarios with multiple non-cooperative agents to test models' robustness in dense, complex urban intersections.

e) V2X-Sim: The V2X-Sim dataset [131], released by Li et al., was initially created using CARLA and SUMO to simulate 100 intersection scenarios on CARLA's Town05 map, with 2 to 5 vehicles per scene acting as CAVs. Each CAV is equipped with a 32-layer 360° LiDAR sensor, with data recorded at 5 Hz. The main task is LiDAR-based 3D object detection. V2X-Sim 2.0, released in 2022, expanded on the initial version with additional scenarios, sensors, and tasks, including six high-resolution RGB cameras for each CAV and RSUs with matching sensors. This version supports BEV object detection, semantic segmentation, and tracking across various urban and suburban scenes, though its scene and sensor diversity remain limited.

f) OPV2V: OPV2V [94], created by Xu et al., is a cooperative perception dataset generated using OpenCDA, CARLA, and SUMO. It includes 73 diverse scenarios across urban, rural, and highway settings. Each scenario has 2 to 7 CAVs equipped with 64-layer LiDAR and four RGB cameras (800×600 resolution), capturing data at 10 Hz. The dataset supports cooperative 3D object detection, BEV segmentation, tracking, and prediction tasks. OPV2V is suitable for evaluating LiDAR-camera fusion methods but lacks non-vehicle road users, and its average CAV count per scene is relatively low.

g) DOLPHINS: The DOLPHINS dataset [132], introduced by Mao et al., is generated using CARLA and includes six diverse scenarios: four intersection scenes, a highway scene, and a rural road. Labeled in KITTI [152] format, it supports both V2V and V2X perception by including RSU views alongside CAVs. Vehicles and RSUs are equipped with 1920×1080 RGB cameras and a 64-layer LiDAR, capturing a range of weather effects like rain and fog. DOLPHINS is primarily used for 3D object detection, making it valuable for evaluating perception algorithms under adverse weather, though CARLA's LiDAR remains unaffected by these conditions. Its open-source code allows for easy reproduction and extension.

h) V2XSet: V2XSet [105], introduced by Xu et al., is generated using CARLA and OpenCDA. It includes 55 scenarios on 8 CARLA maps. Scenarios cover various road types, with intersections comprising around 50% of scenes. Each scenario has 2 to 7 CAVs, equipped with 32-layer LiDAR and four 360° RGB cameras, and some include an RSU. The primary task is cooperative 3D object detection, with realistic communication noise modeled for enhanced evaluation. However, the dataset includes only vehicles, lacks VRUs, and has some positioning inconsistencies in highway scenarios.

i) IRV2V: IRV2V [133], introduced by Wei et al., is the first cooperative perception dataset to incorporate asynchronous time frames, focusing on the impact of temporal desynchronization. Generated using CARLA, each scenario includes 2 to 5 CAVs equipped with RGB cameras and 32-layer LiDAR sensors. The dataset features dense traffic scenarios with over 1.5 million labeled objects, emphasizing BEV perception under asynchronous conditions, making it valuable for evaluating perception algorithms in dynamic and temporally misaligned environments.

j) CARTI: The CARTI dataset [134], built on the CARLA simulator, is designed for cooperative perception research, formatted for compatibility with open-source platforms like OpenMMLab [153] using KITTI [152]'s format. The dataset includes 3D point clouds with 2 to 7 perception nodes per scenario. Objects with insufficient LiDAR reflections are filtered out to improve data quality. The dataset utilizes different LiDAR configurations for onboard and roadside sensors, varying in height and field of view (FOV), capturing a diverse range of cooperative perception scenarios.

k) DeepAccident: The DeepAccident [135] dataset, created by Wang et al. using CARLA, simulates 12 accident scenarios, including rural roads, highways, and intersections, under diverse weather and lighting. Each scenario involves an ego vehicle, three CAVs, and infrastructure sensors across vehicle and VRU categories. The primary tasks are cooperative motion and accident prediction, along with 3D object detection, allowing testing of algorithms under adverse conditions. The consistent number of CAVs limits its application for studies involving varied multi-vehicle cooperation rates.

l) SCOPE: The SCOPE [136] dataset is the first synthetic multimodal cooperative perception dataset, designed for collective perception research. It offers over 40 scenarios with up to 24 collaborative agents. The dataset includes realistic weather effects and sensor models, along with recordings from digital twin maps of Karlsruhe and Tübingen. SCOPE emphasizes diverse environmental conditions and supports benchmarking for cooperative perception algorithms.

m) Adver-City: Adver-City [70] is the first open-source synthetic collaborative perception (CP) dataset designed for adverse weather conditions, bridging a key gap in CP research. Simulated in CARLA with OpenCDA, it includes 110 scenarios across six weather types: clear, rain (soft/heavy), fog, foggy rain, and glare. Combining data from vehicles and roadside units (LiDAR, cameras, GNSS, and IMUs), the dataset reflects real-world crash-inspired scenarios with varied object densities. It provides a valuable resource for improving perception models under challenging conditions.

B. Real-World Datasets

Real-world datasets play a crucial role in the development of cooperative perception for automated driving, as they capture authentic driving conditions and interactions that simulated datasets may overlook. These datasets, collected in diverse locations like intersections [78], [92], [137], [142], [146], highways [138], [145], and urban areas [140], [145], introduce high variability and realistic scenarios that are essential for evaluating the robustness of perception algorithms. Unlike simulated datasets, real-world data contain natural noise and environmental factors, which challenge models and reveal their limitations, allowing for more effective algorithm improvement and error analysis [147]. Real-world datasets are particularly valuable for testing perception under extreme weather conditions, unique traffic situations, and uncommon events, making them indispensable for advancing automated vehicle safety and reliability.

Currently, most real-world cooperative perception datasets are recorded in countries with established research infrastructures, such as China [35], [78], [132], [137], [139], [141]–[144], [146], the United States [83], [138], and Germany [92], [140], [145]. Each dataset reflects the unique characteristics of its location, contributing distinct features, environmental factors, and traffic behaviors. In these real-world settings, varying traffic densities, road configurations, and environmental conditions contribute to a richer and more challenging data landscape, enabling algorithms to better generalize across different contexts.

a) T&J: The T&J dataset [83] by Chen et al. is the first V2V cooperative perception dataset, primarily focusing on cooperative 3D object detection. It comprises data captured in a parking lot using two golf carts equipped with 16-layer 360° LiDAR sensors, high-resolution cameras, radar, GPS, and IMU sensors. Although limited by the lack of ground-truth object labels, this dataset laid foundational work for V2V cooperative perception, but it remains insufficient for neural network training due to the small sample size and lack of label details.

b) IPS300+: The IPS300+ [137] dataset by Wang et al. is recorded at an intersection in Beijing, China. It uses two sensor units, each containing two 1920×1080 resolution RGB cameras and a 360° LiDAR sensor with 80 layers and a 230 m range. Annotated at a 5 Hz frequency, the dataset features high density of 3D labels across categories such as cars, bicycles, pedestrians, tricycles, buses, trucks, and engineering vehicles. The primary task is 3D object detection.

c) DAIR-V2X: DAIR-V2X [78], proposed by Yu et al., is a large-scale real-world cooperative perception dataset collected across multiple intersections in Beijing, China. The dataset includes data from both CAV and RSU sensors across 10 categories. Each intersection is equipped with two 300-layer 100° LiDARs and 1920×1080 resolution RGB cameras, recorded at 10 Hz and 25 Hz, respectively. The DAIR-V2X dataset is divided into subsets, with DAIR-V2X-C providing cooperative CAV and RSU data, supporting multi-view cooperative 3D object detection with camera and LiDAR fusion benchmarks.

d) V2V4Real: The V2V4Real [138] dataset by Xu et al. is tailored for V2V cooperative perception, capturing multi-modal data on highways and urban roads across 347 km of highway and 63 km of city streets. Two vehicles are equipped with single cameras (with varying resolutions) and 32-layer 360° LiDARs with a 200 m range, along with GPS and IMU systems. The dataset covers five vehicle categories and is primarily designed for 3D object detection and tracking, providing a benchmark for perception in both unobstructed and dense highway environments.

e) UGVT: The UGVT dataset [139] introduces a novel computer vision task: I2U (Infrastructure-to-UAV) collaborative visual tracking, combining roadside and UAV perspectives for enhanced object tracking. This approach leverages the complementary strengths of ground-level views (detailed appearance) and aerial perspectives (global motion), enabling more robust and comprehensive tracking systems. UGVT includes 210 high-resolution video pairs (ground and UAV views), designed to evaluate tracking methods effectively.

f) LUCOOP: The LUCOOP dataset [140], released by Axmann et al., was recorded in Hanover, Germany, and is one of the few real-world datasets involving three CAVs. Each CAV is equipped with distinct sensors, including combinations of 16, 32, and 64-layer LiDARs. GNSS, IMU, and UWB sensors provide high-precision localization, allowing the dataset to support navigation, SLAM, and cooperative 3D object detection. LUCOOP's main focus is on urban navigation, with highly accurate CAV trajectories, enabling advanced localization and path planning studies. Although less suitable for general object detection due to repeated scenes, LUCOOP's diverse sensor setups provide valuable insights into early fusion methods and sensor domain adaptation across heterogeneous CAV configurations.

g) V2X-Seq: Built upon the DAIR-V2X dataset, V2X-Seq [141] was released by Yu et al. to support sequential perception and trajectory forecasting tasks. V2X-Seq introduces tracking IDs for moving objects, enabling cooperative tracking across scenarios. The dataset includes 95 scenarios for sequential perception, while the trajectory forecasting dataset includes 336 hours of CAV and RSU data. Objects are annotated with 8 categories and unique tracking information, making it particularly valuable for cooperative object tracking and motion prediction tasks, especially within V2I applications.

h) TUMTraf-V2X: The TUMTraf-V2X dataset [92] is designed to enhance cooperative perception for automated driving. Collected at a busy urban intersection in Germany, it captures complex traffic scenarios such as dense traffic, near-miss events, and traffic violations across eight object categories, offering a multi-sensor perspective from both vehicles and roadside infrastructure. The authors also provide an annotation tool specifically tailored for V2X datasets.

i) RCooper: The RCooper dataset [142], introduced by Hao et al. from Tsinghua University, is the first large-scale real-world dataset dedicated to roadside cooperative perception. Designed for common traffic scenes like intersections and corridors, it supports tasks like detection and tracking by enabling cross-infrastructure cooperation to extend sensing range and reduce blind spots in complex environments. The dataset, along with an annotation tool for V2X applications, is publicly available to facilitate research in advanced roadside perception.

j) InScope: The InScope dataset [143] is a real-world 3D infrastructure-side collaborative perception dataset designed to address occlusion issues in open traffic scenarios. Captured in Guangdong, China, this dataset uses multiple LiDAR systems positioned strategically to mitigate blind spots and improve perception under occluded conditions. InScope provides benchmarks for cooperative 3D object detection, multisource data fusion, domain adaptation, and 3D multi-object tracking tasks. A novel occlusion evaluation metric is introduced to quantify the impact of occlusions on detection performance, supporting V2X research and enhancing the capabilities of perception systems in tracking distant and obscured objects.

k) V2X-Real: The V2X-Real [144] dataset bridges the gap in real-world cooperative perception by integrating vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) collaboration. It features multimodal data collected using two connected automated vehicles and two smart infrastructure nodes, equipped with LiDAR and multi-view cameras. The dataset covers 10 categories and supports vehicle-centric, infrastructure-centric, and cross-agent cooperative tasks, providing benchmarks for state-of-the-art methods.

l) V2AIX: The V2AIX dataset [145] is a pioneering multi-modal dataset capturing European Telecommunications Standards Institute (ETSI) [154] ITS V2X messages in real-world urban and highway settings around Aachen, Cologne, and Düsseldorf, Germany. It includes V2X messages from vehicles and roadside units, along with complementary sensor data from LiDAR and cameras, collected only in relevant proximity to V2X-enabled vehicles. This dataset offers an invaluable resource for studying the penetration of V2X technology, analyzing real-world localization accuracy, and examining message adherence to ETSI standards across diverse traffic environments.

m) HoloVIC: HoloVIC [146] focuses on holographic vehicle-infrastructure cooperation (VIC) for intersections with complex layouts. It includes synchronized data from 6 to 18 sensors, featuring cameras, LiDARs, and fisheye cameras, across diverse intersection configurations. With over 100,000 frames and annotated 3D bounding boxes, HoloVIC supports tracking and detection tasks. The dataset highlights its ability to address occlusions and enhance roadside perception.

n) V2X-Radar: The V2X-Radar dataset [35] represents a pioneering multimodal dataset designed to enhance cooperative perception by incorporating 4D Radar alongside LiDAR and camera sensors. Captured under diverse weather and lighting conditions with a focus on challenging scenarios like obstructed intersections, the dataset supports cooperative, roadside, and single-vehicle perception tasks. It ensures precise temporal and spatial alignment across modalities, facilitating robust benchmarking of perception algorithms. By incorporating 4D Radar, the dataset addresses gaps in existing real-world datasets, enhancing performance under adverse conditions and enabling comprehensive exploration of cooperative perception strategies for automated driving.

o) ISC: The Intersection Safety Challenge (ISC) dataset [36] is a real-world I2I dataset collected to support research on intersection safety, with a strong focus on vulnerable road user (VRU) safety. It was recorded at a four-way signalized intersection using 20 roadside sensors: 8 CCTV cameras, 5 thermal cameras, 2 LiDAR sensors, and 4 radar sensors, alongside traffic signal phase and timing data. The dataset includes 18 distinct categories, comprising 2 vehicle classes and 16 VRU classes, such as pedestrians, cyclists, and other non-motorized road users. The diverse scenarios cover both daytime and nighttime conditions, with a range of potential and non-conflict situations, making the ISC dataset particularly valuable for studying VRU safety and improving intersection perception systems.

V. Emerging Trends in Cooperative Perception

Cooperative perception (CP) has witnessed substantial advancements in the last few years, with the development of numerous innovative frameworks and methodologies. These approaches have significantly enhanced the capabilities of automated systems by improving information fusion, decision-making, and adaptability. However, CP remains a rapidly evolving field, offering ample opportunities for integration with emerging technologies and novel research directions. Recent advancements in areas such as high-definition map utilization, large language models (LLMs), next-generation communication technologies (5G and 6G), and end-to-end automated driving frameworks have opened new possibilities for enhancing CP systems. These interdisciplinary integrations not only address existing challenges in perception accuracy and system robustness but also pave the way for innovative solutions that cater to increasingly complex and dynamic environments. As CP continues to mature, exploring these synergies will be instrumental in realizing its full potential and advancing the future of automated mobility.

A. Map Information in Cooperative Perception Systems

High-definition (HD) maps play a crucial role in enhancing perception tasks within automated driving systems by providing rich, contextual information about the environment [155], [156]. These maps contain detailed road geometry [157], traffic regulations [158], lane-level accuracy [159], and information about static objects such as buildings [160], traffic lights [161], and road signs [162]. By integrating this map information, vehicles can achieve more accurate detection and prediction by comparing real-time sensor data against pre-mapped features. In complex environments, such as urban areas with dense traffic and numerous obstacles, HD maps enable more precise localization, improving decision-making processes like path planning and obstacle avoidance [156], [159]. Furthermore, real-time map updates, facilitated by vehicle-to-everything (V2X) communication, allow vehicles to react promptly to dynamic changes, such as construction zones or temporary road closures [163]. The integration of map information enhances perception tasks by providing a consistent and detailed environmental context [164].

HD maps can act as references for static elements, improving the estimation accuracy of perception tasks through predefined location information [165]. Some studies have integrated HD map data to enhance vehicle localization accuracy. For example, HDNET [166] and MapFusion [167] leverage the geometric and semantic features of HD maps, incorporating them into feature-level processing of Lidar point clouds in deep neural networks. These encoded map features assist in verifying and improving 3D object detection. Endo et al. [168] proposed modules based on HD map-projected images, including an attention module and a transformed feature map module, demonstrating improved object detection accuracy in low-light conditions. While the direct application of HD maps in cooperative perception has not been widely explored, Jiang et al. [169] addressed this gap by proposing a map-based cooperative perception framework called Map Container. This framework uses maps as platforms to convert all information into a unified map coordinate space, enabling distributed fusion of diverse data sources. By integrating GNSS signals and matching sensor features with map features, the framework enhances the vehicle's estimation of environmental states.

The integration of map information into cooperative perception systems offers significant advantages by enhancing the precision and scope of environmental representation. HD maps provide detailed and accurate descriptions of road geometry and static objects, supporting advanced tasks like lane-level localization [170] and path planning [171]. Their structured data formats, such as vectorized or binary map representations, simplify complex road features into computationally efficient inputs for perception algorithms [172], [173]. HD maps also facilitate seamless integration with cooperative perception frameworks by offering a consistent reference that aligns sensor data from multiple nodes. This alignment is especially beneficial for cooperative tasks that require high precision, such as multi-vehicle trajectory planning or collaborative obstacle detection.

Cooperative perception, however, does not solely rely on the availability of HD maps. When maps are unavailable or outdated, CP systems can adapt by leveraging real-time data sharing among nodes [174]. This capability enables CP systems to construct a dynamic and context-aware understanding of the environment, effectively compensating for the absence of pre-existing maps. In such scenarios, CP proves particularly valuable for applications in dynamic or unstructured environments, such as construction zones or off-road conditions [175].

A robust CP framework must demonstrate adaptability, ensuring reliable performance irrespective of the availability of HD maps. While HD maps significantly enhance the accuracy and contextual understanding of perception tasks, CP systems should maintain consistent functionality in their absence by harnessing the collective sensing and data fusion capabilities of the network. This ability to dynamically adapt to varying levels of environmental information ensures that CP remains a scalable and effective solution, capable of addressing the complex and diverse operational scenarios encountered in automated driving, whether or not detailed map information is available.

B. Opportunities Brought by 5G and 6G Communication to Cooperative Perception

The development of 5G and the anticipated arrival of 6G networks represent significant advancements for cooperative perception (CP) systems in automated driving. These cutting-edge communication technologies provide remarkable improvements in data transfer speed, minimal latency, and the capacity to connect a vast number of devices, addressing key limitations of current CP frameworks [176], [177]. By enabling faster and more dependable communication between vehicles, roadside infrastructure, and cloud services, 5G and 6G enhance data sharing capabilities, fostering improved situational awareness and more efficient decision-making [178].

5G technology, with data rates reaching up to 10 Gbps and latency as low as one millisecond, enhances the responsiveness of CP systems [179]. These characteristics enable the near-instantaneous transmission of sensor data, such as object detection outputs or updated high-definition maps, ensuring real-time performance. Furthermore, 5G's substantial bandwidth accommodates the exchange of large and complex datasets, including high-resolution imagery and detailed Lidar point clouds, supporting comprehensive multimodal perception [180]. This capability is critical for scenarios requiring rapid coordination, such as vehicle interactions in urban traffic or merging maneuvers on highways.

Looking towards the future, 6G networks are expected to redefine the landscape of CP by achieving data rates surpassing 1 Tbps and reducing latency to fractions of a millisecond [181]. With the capacity to connect a higher density of devices, 6G will seamlessly integrate edge computing and cloud processing, enhancing collaboration within CP systems [182]. Additionally, 6G is envisioned to introduce AI-powered network functionalities and embedded sensing capabilities, which can further optimize the flow of information [183]. For instance, CP systems supported by 6G could proactively share critical data based on predictive analytics, improving safety and efficiency by preempting potential hazards or traffic disruptions.

Beyond improvements in communication performance, 5G and 6G also enable new CP architectures, such as hybrid edge-cloud frameworks and decentralized intelligence [184]. These models balance the need for immediate, localized processing with the advantages of global data aggregation through centralized resources. Moreover, the extensive connectivity provided by these networks facilitates the inclusion of additional data sources, such as IoT sensors and smart city infrastructure, enriching the overall perception capabilities of automated systems.

In summary, 5G and 6G technologies hold transformative potential for cooperative perception, addressing critical challenges such as high latency, limited bandwidth, and insufficient connectivity [185], [186]. By enabling smarter and more reliable data sharing, these networks are poised to revolutionize CP systems, ensuring they are adaptable, scalable, and well-suited to the increasing complexity of automated driving environments. The integration of these advanced communication technologies is a key enabler for future breakthroughs in automated mobility.

C. The Application of Large Language Models in Cooperative Perception Systems

Large Language Models (LLMs), such as GPT [187] and BERT-based architectures [188], have significantly advanced various domains by showcasing outstanding capabilities in natural language understanding, reasoning, and knowledge representation. While originally developed for text-based tasks, the inherent versatility of LLMs makes them suitable for applications that involve integrating multimodal data [189], including cooperative perception (CP) systems. Their strength lies in processing complex contextual information, learning hierarchical representations, and making data-driven predictions, which are essential for CP tasks [190].

In cooperative perception, LLMs can serve as effective tools for fusing information and supporting decision-making across heterogeneous data sources. CP systems integrate diverse sensor inputs, such as Lidar, radar, cameras, and map information, often enriched with metadata about environmental or communication conditions [191]. LLMs excel at capturing interdependencies between structured and unstructured data, enabling seamless interpretation of multimodal inputs [192]. Trained on domain-specific datasets, these models can encode semantic relationships across sensor modalities to enhance perception accuracy, even in complex environments.

Another notable application of LLMs in CP is their ability to enable cooperative decision-making while personalizing perception strategies [193]. By interpreting shared information, such as object-level features or real-time map updates, LLMs can infer high-level insights to facilitate collaborative tasks like multi-vehicle trajectory prediction or adaptive path planning. Additionally, their generative capabilities allow for the simulation of potential scenarios, enhancing system preparedness for unpredictable conditions, such as occlusions or dynamic hazards. By leveraging historical data and vehicle-specific behaviors, LLMs can also align CP strategies with individual node characteristics, optimizing communication and resource allocation in bandwidth-constrained scenarios.

Despite these promising applications, significant challenges remain in adopting LLMs for CP systems. These include high computational requirements [194], [195], the need for domain-specific fine-tuning [196], and ensuring real-time responsiveness [197], which is critical for safety-sensitive applications. Tackling these obstacles demands innovative approaches to model design, resource optimization, and system integration.

In conclusion, LLMs present transformative potential for cooperative perception by advancing multimodal data fusion, enhancing collaborative decision-making, and supporting context-aware and personalized strategies. Their capability to handle complex and heterogeneous data interactions positions them as a pivotal component in the evolution of CP systems, paving the way for more adaptive and intelligent automated driving solutions.

D. Opportunities for Cooperative Perception in End-to-End automated Driving

End-to-end automated driving, which relies on deep learning models to directly map sensory inputs to control outputs, has gained significant traction, especially with the introduction of advanced methods like UniAD [198]. While traditional automated systems decompose driving tasks into modular subtasks—such as perception, prediction, and planning—end-to-end frameworks seek to unify these processes into a seamless pipeline, leveraging large-scale data and sophisticated neural architectures [199].

One of the key challenges in end-to-end automated driving is ensuring reliable performance across diverse and dynamic environments [200]. The integration of CP addresses this issue by aggregating information from multiple nodes, such as vehicles and infrastructure, to create a more comprehensive understanding of the surroundings. Current research has explored the use of different types of sensors in end-to-end models, including camera-based [201], Lidar-based [202], and multimodal inputs [203]. For example, Li et al. proposed ICOP [204], a collaborative perception model for automated driving that incorporates images shared among multiple vehicles. ICOP employs an encoder to transform each vehicle's images into a comprehensive bird's-eye-view (BEV) representation and a decoder to generate driving commands from the aggregated feature vectors. This approach exemplifies how CP can enhance end-to-end models by improving their ability to handle occlusions, long-range perception, and complex scenarios like crowded intersections or adverse weather conditions.

Another critical opportunity lies in the ability of CP to enable collaborative decision-making in real-time. End-to-end systems often struggle to generalize to rare or unexpected situations due to the limitations of isolated sensory data. CP mitigates this issue by fostering shared situational awareness across multiple nodes, allowing models to capture subtle yet vital environmental cues that may otherwise be overlooked. For instance, Yu et al. [93] introduced UniV2X, an end-to-end framework for vehicle-infrastructure cooperative automated driving (VICAD). This framework integrates dynamic object perception, online mapping, and grid occupancy modules within a unified network. By employing a sparse-dense hybrid transmission strategy and cross-view data interaction, UniV2X enhances planning performance while ensuring data reliability and transmission efficiency. Its design highlights the potential of leveraging CP to overcome common limitations of end-to-end models, particularly in scenarios requiring collaborative and adaptive decision-making.

Despite these advances, most end-to-end driving solutions remain vehicle-centric, focusing primarily on onboard sensor data for perception and control. Incorporating infrastructure-side information—such as roadside LiDAR or high-mounted cameras—represents a promising yet less explored direction. By fusing infrastructure streams with onboard sensing within an end-to-end neural architecture, systems can gain extended coverage of occluded areas (e.g., intersections and blind corners) and enhance situational awareness beyond a single vehicle's field of view. Future research could design modules that efficiently encode infrastructure observations into a unified bird's-eye-view or spatiotemporal representation before generating control outputs. This approach may significantly improve long-range perception, safety, and adaptability under complex conditions, paving the way for robust connected and automated driving scenarios.

In conclusion, cooperative perception offers substantial opportunities to enhance end-to-end automated driving by addressing critical challenges such as limited sensory input, poor generalization, and adaptability in dynamic environments. By harnessing the collaborative and data-rich nature of CP, end-to-end systems can achieve improved situational awareness, robust decision-making, and greater overall reliability. As both CP and end-to-end approaches continue to evolve, their integration represents a critical step toward realizing fully automated and intelligent driving solutions.

VI. Future Directions

A. Technological Gaps

Despite considerable progress in sensor fusion and cooperative perception (CP), several technological hurdles remain unresolved, limiting their broader adoption and effectiveness in practical applications. These challenges include constraints in real-time processing, synchronization of multi-sensor data, perception accuracy, and system reliability under challenging conditions. Addressing these limitations is vital for enhancing the dependability and scalability of CP systems.

a) Communication Constraints: Communication remains a critical bottleneck in cooperative perception (CP) systems. Hardware limitations, such as restricted bandwidth and computational capacity, constrain the volume and frequency of data exchange among vehicles, or between vehicles and infrastructure [20]. Environmental factors like buildings [205], tunnels [206], and adverse weather [207] can further degrade communication reliability, especially in urban or remote regions. These constraints can introduce latency, limit the amount of data that can be shared, and force the system to select only the most critical information for transmission, potentially discarding other useful data. This selective transmission may result in suboptimal decisions and reduced perception accuracy. Although feature-level and result-level sharing [1], adaptive communication strategies [208], and edge-computing techniques offer promising mitigation [20], real-time, robust, and scalable communication under diverse conditions remains an open challenge.

b) Performance in Adverse Environments: One major limitation of current CP methods is their reduced effectiveness in adverse conditions—precisely the scenarios where multi-vehicle cooperation is most critical. While existing approaches excel at addressing issues like occlusion and long-range perception, their reliance on stable communication channels becomes problematic in scenarios with degraded transmission quality, such as interference, bandwidth constraints, or environmental noise. Adverse weather conditions exacerbate these communication challenges by directly affecting wireless signal propagation: rain and atmospheric moisture cause signal attenuation and scattering, while fog and heavy precipitation increase path loss in radio frequency transmissions, effectively reducing available bandwidth for V2X communications. Although many solutions focus on minimizing data transmission to reduce latency, they often overlook the dynamic nature of communication deterioration, which can severely impact the performance of cooperative perception systems in real-world deployments.

c) Lack of Adaptive Fusion Techniques: Another challenge lies in the adaptability of early and intermediate fusion techniques, specifically their ability to adjust fusion strategies based on varying input conditions, such as changes in sensor availability, resolution, reliability, or environmental context.

Traditional fusion methods typically assume fixed data formats and synchronization schemes, which limit their generalizability across diverse datasets and deployment scenarios. For instance, in real-world conditions where certain sensors may temporarily fail or produce noisy outputs [209], a rigid fusion pipeline cannot reweight or reconfigure the contribution of each modality accordingly. Adaptive fusion techniques, by contrast, would dynamically select or weigh sensor inputs based on real-time quality assessments or contextual cues, enabling robust performance under sensor dropout, occlusion, or adverse weather. The absence of such mechanisms currently restricts the scalability and resilience of cooperative perception systems in unpredictable environments.

d) Challenges in Sensor Heterogeneity: A further issue arises from the variability of sensor configurations across different nodes within a CP system. In real-world deployments, vehicles and infrastructure nodes are likely to have differing sensor capabilities, resolutions, and configurations. Current CP frameworks provide limited solutions for effectively integrating data from heterogeneous sensors. This gap becomes especially apparent when high-resolution data from advanced sensors must be combined with lower-resolution data from legacy or cost-effective hardware, creating a pressing need for new approaches to robust sensor fusion.

e) Limited Integration of Emerging Technologies: Emerging technologies such as HD maps, large language models (LLMs), and end-to-end learning frameworks offer significant potential for CP systems but remain underutilized. These technologies could substantially improve perception and decision-making tasks, yet their integration into CP frameworks is still in its early stages. For instance, the ability of LLMs to process heterogeneous and unstructured data or the potential of end-to-end frameworks to streamline perception and decision-making across nodes presents promising directions for future research.

f) Limited Sensor Modality Integration: Despite the availability of diverse sensor technologies, cooperative perception research remains heavily concentrated on camera and LiDAR-based approaches. While these sensors have proven effective for object detection and environmental mapping, the limited exploration of other modalities represents a significant technological gap. Radar sensors, which offer robust performance under adverse weather conditions, and thermal cameras, which excel in low-light scenarios, have received minimal attention in cooperative perception frameworks. This narrow focus limits the potential for developing truly robust multi-weather and multi-condition cooperative systems.

In summary, cooperative perception faces key technological gaps related to adverse conditions, adaptive fusion, sensor heterogeneity, and the incorporation of emerging technologies. Addressing these challenges will be pivotal for advancing CP systems, making them more reliable, resilient, and scalable in practical automated driving environments.

B. Non-Technical Gaps

a) Lack of Standardized Protocols: A notable non-technical gap in CP systems is the absence of universally accepted standards for data formats, communication protocols, and safety requirements. Without such standardization, ensuring interoperability across various manufacturers, Original Equipment Manufacturers (OEMs), and systems becomes challenging, ultimately hindering large-scale deployment. Existing frameworks, such as IEEE 802.11p [210] (DSRC)and C-V2X for vehicle-to-everything (V2X) communication, provide initial guidelines for wireless interfaces but do not fully address CP-specific requirements like multi-sensor data fusion, semantic consistency, or real-time bandwidth adaptation. Some recent SAE standards—including SAE J3216 [211] (taxonomy and definitions for cooperative and automated driving systems), SAE J3224 [212] (nomenclature and definitions for cooperative automated driving system functions), and SAE J3256 [213] (anticipated guidelines for cooperative automated driving)—begin to outline terminology and operational scenarios relevant to cooperative systems. However, these documents primarily focus on basic connectivity and functional descriptions rather than detailed technical specifications for CP data exchange. Similarly, European standards under ETSI ITS-G5 and C-ITS initiatives (e.g., ETSI EN 302 637 series [214]–[216], TS 103 324 [217]) facilitate interoperability for connected vehicles but have yet to establish comprehensive guidelines on robust cooperative perception frameworks, such as how to standardize feature-level fusion or manage dynamic data exchange policies under varying communication conditions. In short, while these standards represent critical steps toward more coordinated connected environments, they fall short of covering the full breadth of CP's complexities. Issues such as defining unified data formats for multi-modal perception, protocols for real-time sharing of sensor or feature maps, and safety requirements under adverse communication conditions remain largely unaddressed. This underscores the pressing need for enhanced and more detailed standards tailored specifically to the requirements of CP—standards that can ensure interoperability, data integrity, and reliability at scale.

b) Trustworthiness and Data Security Concerns: Ensuring the trustworthiness or integrity of CP systems is critical, as the safety of automated driving depends on the accuracy and integrity of shared data. Data exchanged between vehicles and infrastructure must be reliable, timely, and secure from tampering to prevent accidents and malicious interference. Existing frameworks, such as ETSI's ITS-Secure standards, address some security aspects but fail to comprehensively counter advanced threats like spoofing or data injection in CP scenarios. Additionally, robust trust management systems are essential to ensure reliable operation even when some nodes or networks are compromised.

c) Ethical and Privacy Challenges: Ethical and privacy concerns pose significant obstacles to the widespread adoption of CP systems. Sharing sensor data, such as camera feeds or lidar point clouds, raises issues regarding the protection of sensitive information, such as license plates or pedestrian identities. While regulations like Europe's GDPR offer guidelines for data protection, they do not fully address the unique challenges associated with CP systems. Clear, universally accepted guidelines for data sharing and usage are essential to balance the benefits of CP with the protection of individual privacy.

d) Collaboration Across Manufacturers and Stakeholders: Collaboration among manufacturers, OEMs, and other stakeholders is another barrier to CP adoption. Differences in hardware, software ecosystems, and proprietary technologies create silos that hinder seamless cooperation. For instance, while some manufacturers focus on cost-effective solutions with minimal sensor capabilities, others prioritize high-end systems with advanced sensors, leading to disparities in shared data quality. Initiatives like the 5G Automotive Association (5GAA) promote collaboration, but broader efforts are needed to address challenges such as cross-platform compatibility and shared innovation in perception technologies.

e) Societal Acceptance and Deployment Challenges: Public trust and acceptance remain critical challenges in deploying CP systems. Concerns about the safety and reliability of automated vehicles can delay adoption, especially in high-stakes scenarios such as urban traffic management. Moreover, regulatory and policy frameworks often lag behind technological advances, creating uncertainty for developers and manufacturers. While some regions have introduced testing frameworks for automated vehicles, comprehensive policies addressing multi-vehicle interactions and real-time decision-making in CP systems are still lacking. Overcoming these barriers requires aligning technological progress with public trust, clear regulations, and transparent communication about the benefits and limitations of CP.

In summary, addressing non-technical gaps is equally crucial for advancing CP systems. Establishing standardized protocols, ensuring data security, resolving ethical challenges, fostering collaboration, and improving societal acceptance are all essential for unlocking the full potential of cooperative perception in automated driving.

C. Need for Datasets

a) Homogeneity in Sensor Configurations: One significant limitation in current cooperative perception (CP) datasets is their reliance on homogeneous sensor configurations. Most datasets are constructed using sensors of the same type, model, and resolution, resulting in overly uniform detection scenarios. For example, datasets frequently feature identical Lidar and camera setups across all nodes, which simplifies data fusion but fails to reflect the diversity of real-world deployments, where vehicles and infrastructure often employ heterogeneous sensors with varying specifications. This homogeneity limits the generalizability of CP methods, as these systems may struggle to adapt to scenarios with mixed-resolution sensors or varying hardware capabilities.

b) Lack of Adverse Conditions: Another critical gap is the lack of representation for non-ideal conditions in existing real-world datasets. Adverse weather scenarios, such as rain, fog, and low light, are often underrepresented, as are datasets featuring non-ideal sensor outputs like high-noise Lidar point clouds or heavily blurred images. These conditions are precisely where CP systems have the potential to excel, yet their absence in datasets hampers progress toward developing robust perception models that can operate reliably across diverse and challenging environments. This gap also limits the ability to evaluate CP systems in scenarios where single-vehicle perception is most vulnerable, reducing the applicability of CP solutions in real-world deployments.

c) Limited Spatial and Node Coverage: The spatial and node coverage in most existing datasets is highly constrained. Current datasets often focus on small, specific areas, such as a single intersection, and involve an ideal number of nodes with well-optimized layouts for perception. While these setups provide controlled environments for testing CP methods, they fail to capture the complexities of larger-scale deployments. For instance, datasets rarely include scenarios with sparse or irregularly distributed agents, distant roadside units, or large-scale environments with varied road networks. This lack of diversity in spatial coverage and node configurations restricts the evaluation of CP systems' scalability and effectiveness in broader operational contexts.

d) Need for Diverse and Challenging Scenarios: The absence of datasets that combine multimodal fusion and cross-device cooperative perception scenarios further limits progress. Many current datasets emphasize single-modality perception tasks, such as camera-based object detection or Lidar-based tracking, without incorporating the full range of sensor data available in CP systems. Additionally, datasets often fail to include complex cooperative scenarios, such as interactions between heterogeneous nodes or agents with overlapping but incomplete fields of view. Expanding the scope of datasets to include these challenging scenarios would enable the development and benchmarking of CP systems that are more versatile and robust.

e) Future Dataset Development: To address these gaps, future datasets should prioritize diversity in sensor configurations, environmental conditions, and spatial layouts. Including heterogeneous sensors with varying resolutions and models would reflect real-world conditions more accurately. Datasets should also capture a broader range of adverse scenarios, such as extreme weather or degraded sensor outputs, to test the resilience of CP systems. Furthermore, expanding spatial coverage to include larger and less optimized environments, along with scenarios featuring sparse agents or irregular layouts, would enable the evaluation of CP systems' scalability and adaptability. By incorporating these elements, future datasets can significantly advance the development and application of cooperative perception systems.

D. Future Research Directions

While the previous subsections identify and analyze existing limitations across technological, non-technical, and dataset dimensions, this section synthesizes these findings into actionable research roadmaps for advancing cooperative perception (CP). Rather than simply reiterating the gaps, we provide concrete strategies, specific algorithmic approaches, and integrated solutions that address multiple challenge categories simultaneously. These research directions are designed to guide the community toward practical implementations that can overcome the identified barriers and unlock the full potential of CP systems in real-world automated driving applications.

a) Enhancing Perception in Adverse Environments: Improving perception reliability under challenging conditions, such as rain, snow, fog, and low-light or nighttime scenarios, is essential for ensuring CP systems' effectiveness. Future research should explore algorithms capable of processing degraded sensor outputs, such as blurred images or noisy Lidar data, while adapting to varying environmental factors. Techniques such as generative models or domain adaptation could mitigate the impacts of adverse weather by reconstructing or enhancing affected data. Additionally, multimodal data integration—leveraging radar or thermal cameras in conjunction with traditional sensors—offers a pathway to achieve robustness in low-visibility scenarios. These conditions are precisely where CP systems have the potential to excel, yet their absence in datasets hampers progress toward developing robust perception models that can operate reliably across diverse and challenging environments. Unlike hardware failures and communication problems, which can often be artificially simulated—such as reducing sensor resolution or selectively disabling LiDAR beams to emulate sensor degradation, or introducing latency and packet loss to mimic communication disruptions—authentic adverse weather conditions remain difficult to reproduce realistically in simulation environments. This is because weather phenomena like rain, fog, and snow introduce complex physical effects such as light scattering, signal attenuation, and sensor occlusion, which affect each sensing modality differently and are challenging to model with high fidelity [70]. As a result, datasets that capture real-world adverse weather conditions are particularly valuable for evaluating the robustness of cooperative perception systems, yet they are also difficult and costly to collect under genuine, naturally occurring scenarios.

b) Fusion Strategies for Heterogeneous Sensors Across Nodes: As CP systems scale to include diverse nodes, managing heterogeneous sensor configurations becomes a critical challenge. Research should develop fusion strategies that effectively integrate data from various modalities, such as cameras, Lidar, or a combination of both, while addressing differences in resolution, noise characteristics, and field of view. Advanced methods, such as graph neural networks (GNNs) or attention mechanisms, could facilitate dynamic and context-aware integration of disparate sensor inputs. Adaptive frameworks that adjust fusion strategies based on real-time sensor availability and quality could further enhance CP systems' flexibility and reliability in operational environments.

c) Real-World Validation and Comprehensive Dataset Development: Bridging the gap between experimental results and practical deployment necessitates rigorous real-world validation of CP systems. Future efforts should emphasize testing in diverse environments, including urban intersections, rural roads, and highways, with conditions such as adverse weather, uneven agent layouts, and fluctuating communication bandwidth. Simulated scenarios should also reflect real-world challenges, such as packet loss and limited connectivity. Concurrently, the development of rich, diverse datasets is essential to support these validations. Future datasets should include multimodal data, heterogeneous sensors, and large-scale environments that accurately reflect practical challenges. Establishing open-access platforms for dataset sharing can also promote collaboration and innovation across the research community.

d) Establishing Cross-Platform Standards: The absence of standardization in CP represents a significant bottleneck for scalability. Research should prioritize the development of universal standards for data formats, communication protocols, and system safety requirements. These standards must support interoperability among manufacturers, OEMs, and infrastructure systems while accommodating diverse hardware configurations and sensor resolutions. Incorporating privacy-preserving protocols and trust-validation mechanisms into these standards will further address ethical and security concerns, ensuring reliable and safe CP system implementations.

e) Integrating Emerging Technologies: Emerging technologies, including HD maps, large language models (LLMs), and end-to-end learning frameworks, present significant opportunities for advancing CP. HD maps can enhance situational awareness by providing contextual environmental details, while LLMs can interpret multimodal data and support collaborative decision-making. End-to-end learning frameworks, meanwhile, offer the potential to streamline the perception-to-planning pipeline, increasing system efficiency and adaptability. Research should investigate the integration of these technologies with CP to overcome challenges such as limited bandwidth, harsh weather conditions, hardware failures, and low-resolution sensors, ultimately improving detection accuracy and system resilience.

In conclusion, addressing these research directions will drive significant advancements in cooperative perception systems, ensuring they are robust, scalable, and adaptable to the diverse and dynamic demands of real-world automated driving applications.

VII. Conclusion

This paper provides a comprehensive review of cooperative perception (CP) methods, highlighting their pivotal role in connected and automated transportation. By systematically examining multi-agent CP approaches across different collaboration strategies, we reveal both the strengths and limitations of existing methodologies in diverse operational contexts. Our analysis also underscores the potential of emerging technologies—such as high-definition maps, next-generation communication networks (5G/6G), large language models (LLMs), and end-to-end frameworks—to further advance CP systems.

Despite considerable progress, our findings point to key gaps that impede the scalability and real-world deployment of CP. Technological challenges, including limited adaptability to adverse conditions and difficulties in handling heterogeneous sensor setups, persist. Non-technical barriers—such as the lack of standardized protocols, data integrity, and privacy issues—further hamper widespread adoption. Additionally, current datasets often lack the diversity in sensor configurations, adverse scenarios, and spatial coverage needed to develop robust, adaptable CP solutions.

To address these challenges, we propose a forward-looking research agenda focusing on three crucial aspects: (1) enhancing perception algorithms for adverse and complex environments, (2) developing adaptive collaborative strategies for heterogeneous sensors, and (3) establishing scalable and standardized CP frameworks. Collectively, these efforts aim to improve system reliability, adaptability, and efficiency under real-world conditions.

Furthermore, our study highlights the value of interdisciplinary innovations in bridging the gap between theoretical progress and practical implementation. By integrating next-generation communication technologies, multimodal data processing, and robust system architectures, future research can tackle pressing concerns such as communication constraints and reliability. This promises transformative strides in connected autonomous transportation, unlocking safer and more efficient mobility.

In conclusion, this work provides a foundational roadmap for advancing CP. By addressing the identified technical and non-technical gaps, while capitalizing on emerging technological opportunities, we can ensure that CP systems are not only theoretically sound but also practically scalable, adaptable, and capable of meeting the evolving demands of modern transportation ecosystems.

References

[1] A. Caillot, S. Ouerghi, P. Vasseur, R. Boutteau, and Y. Dupuis, "Survey on cooperative perception in an automotive context," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 204–14 223, 2022.

[2] Z. Bai, G. Wu, X. Qi, Y. Liu, K. Oguchi, and M. J. Barth, "Infrastructure-based object detection and tracking for cooperative driving automation: A survey," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 1366–1373.

[3] Z. Xiao, J. Shu, H. Jiang, G. Min, H. Chen, and Z. Han, "Overcoming occlusions: Perception task-oriented information sharing in connected and autonomous vehicles," *IEEE Network*, vol. 37, no. 4, pp. 224–229, 2023.

[4] X. Gao, X. Zhang, Y. Lu, Y. Huang, L. Yang, Y. Xiong, and P. Liu, "A survey of collaborative perception in intelligent vehicles at intersections," *IEEE Transactions on Intelligent Vehicles*, 2024.

[5] S. Mozaffari, E. Arnold, M. Dianati, and S. Fallah, "A comparative study of ego-centric and cooperative perception for lane change prediction in highway driving scenarios," in *Proceedings of the 2nd International Conference on Robotics, Computer Vision and Intelligent Systems*, vol. 1. SCITEPRESS-Science and Technology Publications, 2021, pp. 113–121.

[6] B. Anand, V. Barsaiyan, M. Senapati, and P. Rajalakshmi, "Real time lidar point cloud compression and transmission for intelligent transportation system," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*. IEEE, 2019, pp. 1–5.

[7] A. S. Mohammed, A. Amamou, F. K. Ayevide, S. Kelouwani, K. Agbossou, and N. Zioui, "The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review," *Sensors*, vol. 20, no. 22, p. 6532, 2020.

[8] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, and W. Shi, "Computing systems for autonomous driving: State of the art and challenges," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6469–6486, 2020.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[10] W. Farag, "Real-time detection of road lane-lines for autonomous driving," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 13, no. 2, pp. 265–274, 2020.

[11] J. Zhang, Z. Xie, J. Sun, X. Zou, and J. Wang, "A cascaded r-cnn with multiscale attention and imbalanced samples for traffic sign detection," *IEEE access*, vol. 8, pp. 29 742–29 754, 2020.

[12] R. W. Wood, "Xxiii. fish-eye views, and vision under water," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 12, no. 68, pp. 159–162, 1906.

[13] V. R. Kumar, C. Eising, C. Witt, and S. K. Yogamani, "Surround-view fisheye camera perception for automated driving: Overview, survey & challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 3638–3659, 2023.

[14] J. Mao, S. Shi, X. Wang, and H. Li, "3d object detection for autonomous driving: A comprehensive survey," *International Journal of Computer Vision*, vol. 131, no. 8, pp. 1909–1963, 2023.

[15] D. Fernández Llorca, A. Hernández Martínez, and I. Garcia Daza, "Vision-based vehicle speed estimation: A survey," *IET Intelligent Transport Systems*, vol. 15, no. 8, pp. 987–1005, 2021.

[16] Y. Hu, Y. Lu, R. Xu, W. Xie, S. Chen, and Y. Wang, "Collaboration helps camera overtake lidar in 3d detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9243–9252.

[17] C. Wei, G. Wu, and M. J. Barth, "Feature corrective transfer learning: End-to-end solutions to object detection in non-ideal visual conditions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23–32.

[18] M. A. Richards, J. Scheer, W. A. Holm, and W. L. Melvin, "Principles of modern radar," 2010.

[19] S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, and M. A. Kaafar, "The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car," *IEEE vehicular technology magazine*, vol. 14, no. 2, pp. 103–111, 2019.

[20] Z. Bai, G. Wu, M. J. Barth, Y. Liu, E. A. Sisbot, K. Oguchi, and Z. Huang, "A survey and framework of cooperative perception: From heterogeneous singleton to hierarchical cooperation," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[21] B. Behroozpour, P. A. Sandborn, M. C. Wu, and B. E. Boser, "Lidar system architectures and circuits," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 135–142, 2017.

[22] Y. Li, L. Ma, Z. Zhong, F. Liu, M. A. Chapman, D. Cao, and J. Li, "Deep learning for lidar point clouds in autonomous driving: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3412–3432, 2020.

[23] D. Hutabarat, M. Rivai, D. Purwanto, and H. Hutomo, "Lidar-based obstacle avoidance for the autonomous mobile robot," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*. IEEE, 2019, pp. 197–202.

[24] P. Teleweck and B. Chandrasekaran, "Path planning algorithms and their use in robotic navigation systems," in *Journal of Physics: Conference Series*, vol. 1207, no. 1. IOP Publishing, 2019, p. 012018.

[25] Z. Yan, L. Sun, T. Krajník, and Y. Ruichek, "Eu long-term dataset with multiple sensors for autonomous driving," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 697–10 704.

[26] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine vision and applications*, vol. 25, pp. 245–262, 2014.

[27] N. Bustos, M. Mashhadi, S. K. Lai-Yuen, S. Sarkar, and T. K. Das, "A systematic literature review on object detection using near infrared and thermal images," *Neurocomputing*, vol. 560, p. 126804, 2023.

[28] W. Xu, C. Yan, W. Jia, X. Ji, and J. Liu, "Analyzing and enhancing the security of ultrasonic sensors for autonomous vehicles," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 5015–5029, 2018.

[29] E. D. Kaplan and C. Hegarty, *Understanding GPS/GNSS: principles and applications*. Artech house, 2017.

[30] T. Takasu and A. Yasuda, "Development of the low-cost rtk-gps receiver with an open source program package rtklib," in *International symposium on GPS/GNSS*, vol. 1. International Convention Center Jeju Korea Seogwipo-si, Republic of Korea, 2009, pp. 1–6.

[31] S. Van der Spek, J. Van Schaick, P. De Bois, and R. De Haan, "Sensing human activity: Gps tracking," *Sensors*, vol. 9, no. 4, pp. 3033–3055, 2009.

[32] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory, Tech. Rep., 2007.

[33] H. Jing, Y. Gao, S. Shahbeigi, and M. Dianati, "Integrity monitoring of gnss/ins based positioning systems for autonomous vehicles: State-of-the-art and open challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 166–14 187, 2022.

[34] G. Hu, B. Gao, Y. Zhong, and C. Gu, "Unscented kalman filter with process noise covariance estimation for vehicular ins/gps integration system," *Information Fusion*, vol. 64, pp. 194–204, 2020.

[35] L. Yang, X. Zhang, J. Li, C. Wang, Z. Song, T. Zhao, Z. Song, L. Wang, M. Zhou, Y. Shen *et al.*, "V2x-radar: A multi-modal dataset with 4d radar for cooperative perception," *arXiv preprint arXiv:2411.10962*, 2024.

[36] U.S. DOT ITS, "Intersection safety challenge," 2024.

[37] S. Liu, J. Tang, Z. Zhang, and J.-L. Gaudiot, "Computer architectures for autonomous driving," *Computer*, vol. 50, no. 8, pp. 18–25, 2017.

[38] Z. Xiao, J. Shu, H. Jiang, G. Min, H. Chen, and Z. Han, "Perception task offloading with collaborative computation for autonomous driving," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 2, pp. 457–473, 2022.

[39] G. Cui, W. Zhang, Y. Xiao, L. Yao, and Z. Fang, "Cooperative perception technology of autonomous driving in the internet of vehicles environment: A review," *Sensors*, vol. 22, no. 15, p. 5535, 2022.

[40] E. Y. Bejarbaneh, H. Du, and F. Naghdy, "Exploring shared perception and control in cooperative vehicle-intersection systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[41] J. A. Guerrero-Ibanez, S. Zeadally, and J. Contreras-Castillo, "Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 122–128, 2015.

[42] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya, "A survey on vehicular cloud computing," *Journal of Network and Computer applications*, vol. 40, pp. 325–344, 2014.

[43] J. Lee, K. Lee, A. Yoo, and C. Moon, "Design and implementation of edge-fog-cloud system through hd map generation from lidar data of autonomous vehicles," *Electronics*, vol. 9, no. 12, p. 2084, 2020.

[44] K. Behravan, N. Farzaneh, M. Jahanshahi, and S. A. H. Seno, "A comprehensive survey on using fog computing in vehicular networks," *Vehicular Communications*, vol. 42, p. 100604, 2023.

[45] H. Du, S. Leng, F. Wu, X. Chen, and S. Mao, "A new vehicular fog computing architecture for cooperative sensing of autonomous driving," *IEEE Access*, vol. 8, pp. 10 997–11 006, 2020.

[46] H. Yang, Z. Wei, Z. Feng, X. Chen, Y. Li, and P. Zhang, "Intelligent computation offloading for mec-based cooperative vehicle infrastructure system: A deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7665–7679, 2022.

[47] C. M. Silva, B. M. Masini, G. Ferrari, and I. Thibault, "A survey on infrastructure-based vehicular networks," *Mobile Information Systems*, vol. 2017, no. 1, p. 6123868, 2017.

[48] R. Cai, Y. Feng, D. He, Y. Xu, Y. Zhang, and W. Xie, "A combined cable-connected rsu and uav-assisted rsu deployment strategy in v2i communication," in *ICC 2020-2020 IEEE international conference on communications (ICC)*. IEEE, 2020, pp. 1–6.

[49] S.-Y. Lin, C.-M. Huang, and T.-Y. Wu, "Multi-access edge computing-based vehicle-vehicle-rsu data offloading over the multi-rsu-overlapped environment," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 7–32, 2022.

[50] A. Taheri Tajar, A. Ramazani, and M. Mansoorizadeh, "A lightweight tiny-yolov3 vehicle detection approach," *Journal of Real-Time Image Processing*, vol. 18, no. 6, pp. 2389–2401, 2021.

[51] L. Chen, Y. Zhang, B. Tian, Y. Ai, D. Cao, and F.-Y. Wang, "Parallel driving os: A ubiquitous operating system for autonomous driving in cpss," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 886–895, 2022.

[52] L. Barba-Guaman, J. Eugenio Naranjo, and A. Ortiz, "Deep learning framework for vehicle and pedestrian detection in rural roads on an embedded gpu," *Electronics*, vol. 9, no. 4, p. 589, 2020.

[53] S. Lu and W. Shi, "Vehicle computing: Vision and challenges," *Journal of Information and Intelligence*, vol. 1, no. 1, pp. 23–35, 2023.

[54] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1697–1716, 2019.

[55] L. Liu, C. Chen, Q. Pei, S. Maharjan, and Y. Zhang, "Vehicular edge computing and networking: A survey," *Mobile networks and applications*, vol. 26, pp. 1145–1168, 2021.

[56] H. Xiao, J. Zhao, Q. Pei, J. Feng, L. Liu, and W. Shi, "Vehicle selection and resource optimization for federated learning in vehicular edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11 073–11 087, 2021.

[57] D. Grewe, M. Wagner, M. Arumaithurai, I. Psaras, and D. Kutscher, "Information-centric mobile edge computing for connected vehicle environments: Challenges and research directions," in *Proceedings of the Workshop on Mobile Edge Communications*, 2017, pp. 7–12.

[58] H.-H. Nguyen, D. N.-N. Tran, and J. W. Jeon, "Towards real-time vehicle detection on edge devices with nvidia jetson tx2," in *2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*. IEEE, 2020, pp. 1–4.

[59] Z. Ning, K. Zhang, X. Wang, L. Guo, X. Hu, J. Huang, B. Hu, and R. Y. Kwok, "Intelligent edge computing in internet of vehicles: A joint computation offloading and caching solution," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2212–2225, 2020.

[60] H. Shukur, S. R. Zeebaree, A. J. Ahmed, R. R. Zebari, O. Ahmed, B. S. A. Tahir, and M. A. Sadeeq, "A state of art: Survey for concurrent computation and clustering of parallel computing for distributed systems," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 148–154, 2020.

[61] M. Nijim and H. Albataineh, "Secure-stor: A novel hybrid storage system architecture to enhance security and performance in edge computing," *IEEE access*, vol. 9, pp. 92 446–92 459, 2021.

[62] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 146–177, 2023.

[63] J. Vargas, S. Alsweiss, O. Toker, R. Razdan, and J. Santos, "An overview of autonomous vehicles sensors and their vulnerability to weather conditions," *Sensors*, vol. 21, no. 16, p. 5397, 2021.

[64] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3937–3946.

[65] J. Li, R. Xu, J. Ma, Q. Zou, J. Ma, and H. Yu, "Domain adaptive object detection for autonomous driving under foggy weather," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 612–622.

[66] M. Schutera, M. Hussein, J. Abhau, R. Mikut, and M. Reischl, "Night-to-day: Online image-to-image translation for object detection within autonomous driving by night," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 3, pp. 480–489, 2020.

[67] Y. Xiao, A. Jiang, J. Ye, and M.-W. Wang, "Making of night vision: Object detection under low-illumination," *IEEE Access*, vol. 8, pp. 123 075–123 086, 2020.

[68] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive yolo for object detection in adverse weather conditions," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 2, 2022, pp. 1792–1800.

[69] X. Yin, Z. Yu, Z. Fei, W. Lv, and X. Gao, "Pe-yolo: Pyramid enhancement network for dark object detection," in *International Conference on Artificial Neural Networks*. Springer, 2023, pp. 163–174.

[70] M. Karvat and S. Givigi, "Adver-city: Open-source multi-modal dataset for collaborative perception under adverse weather conditions," *arXiv preprint arXiv:2410.06380*, 2024.

[71] Y. Han, H. Zhang, H. Li, Y. Jin, C. Lang, and Y. Li, "Collaborative perception in autonomous driving: Methods, datasets, and challenges," *IEEE Intelligent Transportation Systems Magazine*, 2023.

[72] M. Haris, G. Shakhnarovich, and N. Ukita, "Task-driven super resolution: Object detection in low-resolution images," in *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part V 28*. Springer, 2021, pp. 387–395.

[73] C. Wei, G. Wu, M. Barth, P. H. Chan, V. Donzella, and A. Huggett, "Enhanced object detection by integrating camera parameters into raw image-based faster r-cnn," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 4473–4478.

[74] I. del Pino, V. Vaquero, B. Masini, J. Sola, F. Moreno-Noguer, A. Sanfeliu, and J. Andrade-Cetto, "Low resolution lidar-based multi-object tracking for driving applications," in *Iberian Robotics conference*. Springer, 2017, pp. 287–298.

[75] K.-H. Chow, L. Liu, M. E. Gursoy, S. Truex, W. Wei, and Y. Wu, "Understanding object detection through an adversarial lens," in *Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part II 25*. Springer, 2020, pp. 460–481.

[76] Y. Zhao, H. Sun, and S. Wang, "Small object detection in medium–low-resolution remote sensing images based on degradation reconstruction," *Remote Sensing*, vol. 16, no. 14, p. 2645, 2024.

[77] C. Liu, C. Gao, F. Liu, J. Liu, D. Meng, and X. Gao, "Ss3d: Sparsely-supervised 3d object detection from point cloud," in *Proceedings of*

the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 8428–8437.

[78] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan *et al.*, "Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 361–21 370.

[79] D. D. Yoon, B. Ayalew, and G. M. N. Ali, "Performance of decentralized cooperative perception in v2v connected traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6850–6863, 2021.

[80] A. Miller, K. Rim, P. Chopra, P. Kelkar, and M. Likhachev, "Cooperative perception and localization for cooperative driving," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1256–1262.

[81] E. Arnold, M. Dianati, R. de Temple, and S. Fallah, "Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1852–1864, 2020.

[82] G. Volk, A. von Bernuth, and O. Bringmann, "Environment-aware development of robust vision-based cooperative perception systems," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 126–133.

[83] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 514–524.

[84] S. Liu, Z. Ding, J. Fu, H. Li, S. Chen, S. Zhang, and X. Zhou, "V2x-pc: Vehicle-to-everything collaborative perception via point cluster," *arXiv preprint arXiv:2403.16635*, 2024.

[85] Q. Chen, X. Ma, S. Tang, J. Guo, Q. Yang, and S. Fu, "F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds," in *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, 2019, pp. 88–100.

[86] Y. Zeng, C. Sun, L. Ding, J. Hu, Z. Chen, Y. Tian, and Y. Wang, "Pedestrian re-identification with adaptive feature dimension reduction for vehicle-road cooperative perception," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.

[87] Z. Bai, G. Wu, M. J. Barth, Y. Liu, E. A. Sisbot, and K. Oguchi, "Pillargrid: Deep learning-based cooperative perception for 3d object detection from onboard-roadside lidar," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 1743–1749.

[88] J. Wang, X. Guo, H. Wang, P. Jiang, T. Chen, and Z. Sun, "Pillar-based cooperative perception from point clouds for 6g-enabled cooperative autonomous vehicles," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, p. 3646272, 2022.

[89] D. Qu, Q. Chen, T. Bai, H. Lu, H. Fan, H. Zhang, S. Fu, and Q. Yang, "Sicp: Simultaneous individual and cooperative perception for 3d object detection in connected and automated vehicles," *arXiv preprint arXiv:2312.04822*, 2023.

[90] D. Qiao and F. Zulkernine, "Cobevfusion: Cooperative perception with lidar-camera bird's-eye view fusion," *arXiv preprint arXiv:2310.06008*, 2023.

[91] Y. Zhang, B. Chen, J. Qin, F. Hu, and J. Hao, "Coopercept: Cooperative perception for 3d object detection of autonomous vehicles," *Drones*, vol. 8, no. 6, p. 228, 2024.

[92] W. Zimmer, G. A. Wardana, S. Sritharan, X. Zhou, R. Song, and A. C. Knoll, "Tumtraf v2x cooperative perception dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 22 668–22 677.

[93] H. Yu, W. Yang, J. Zhong, Z. Yang, S. Fan, P. Luo, and Z. Nie, "End-to-end autonomous driving through v2x cooperation," in *https://arxiv.org/abs/2404.00717*, 2024.

[94] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, "Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2583–2589.

[95] J. Cui, H. Qiu, D. Chen, P. Stone, and Y. Zhu, "Coopernaut: End-to-end driving with cooperative perception for networked vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 252–17 262.

[96] R. Xu, Z. Tu, H. Xiang, W. Shao, B. Zhou, and J. Ma, "Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers," in *Conference on Robot Learning*. PMLR, 2023, pp. 989–1000.

[97] C. Lin, D. Tian, X. Duan, J. Zhou, D. Zhao, and D. Cao, "V2vformer: Vehicle-to-vehicle cooperative perception with spatial-channel transformer," *IEEE Transactions on Intelligent Vehicles*, 2024.

[98] H. Yin, D. Tian, C. Lin, X. Duan, J. Zhou, D. Zhao, and D. Cao, "V2vformer ++: Multi-modal vehicle-to-vehicle cooperative perception via global-local transformer," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[99] Z. Li, H. Liang, H. Wang, M. Zhao, J. Wang, and X. Zheng, "Mkd-cooper: Cooperative 3d object detection for autonomous driving via multi-teacher knowledge distillation," *IEEE Transactions on Intelligent Vehicles*, 2023.

[100] Z. Wang, S. Fan, X. Huo, T. Xu, Y. Wang, J. Liu, Y. Chen, and Y.-Q. Zhang, "Vimi: Vehicle-infrastructure multi-view intermediate fusion for camera-based 3d object detection," *arXiv preprint arXiv:2303.10975*, 2023.

[101] H. Xiang, R. Xu, and J. Ma, "Hm-vit: Hetero-modal vehicle-to-vehicle cooperative perception with vision transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 284–295.

[102] S. Fan, H. Yu, W. Yang, J. Yuan, and Z. Nie, "Quest: Query stream for practical cooperative perception," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 18 436–18 442.

[103] H. Chen, H. Wang, Z. Liu, D. Gu, and W. Ye, "Hp3d-v2v: High-precision 3d object detection vehicle-to-vehicle cooperative perception algorithm," *Sensors*, vol. 24, no. 7, p. 2170, 2024.

[104] T.-H. Wang, S. Manivasagam, M. Liang, B. Yang, W. Zeng, and R. Urtasun, "V2vnet: Vehicle-to-vehicle communication for joint perception and prediction," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 605–621.

[105] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2x-vit: Vehicle-to-everything cooperative perception with vision transformer," in *European conference on computer vision*. Springer, 2022, pp. 107–124.

[106] Z. Meng, X. Xia, R. Xu, W. Liu, and J. Ma, "Hydro-3d: Hybrid object detection and tracking for cooperative perception using 3d lidar," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 8, pp. 4069–4080, 2023.

[107] A. N. Ahmed, S. Mercelis, and A. Anwar, "Graph attention based feature fusion for collaborative perception," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2024, pp. 2317–2324.

[108] R. Xu, C.-J. Chen, Z. Tu, and M.-H. Yang, "V2x-vitv2: Improved vision transformers for vehicle-to-everything cooperative perception," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[109] A. N. Ahmed, S. Mercelis, and A. Anwar, "Collabgat: Collaborative perception using graph attention network," *IEEE Access*, 2024.

[110] X. Zhao, K. Mu, F. Hui, and C. Prehofer, "A cooperative vehicle-infrastructure based urban driving environment perception method using a ds theory-based credibility map," *Optik*, vol. 138, pp. 407–415, 2017.

[111] G. Pereira, P. M. d'Orey, and A. Aguiar, "Poster: Cooperative perception platform for intelligent transportation systems," in *2020 IEEE Vehicular Networking Conference (VNC)*. IEEE, 2020, pp. 1–2.

[112] A. B. Khalifa, I. Alouani, M. A. Mahjoub, and A. Rivenq, "A novel multi-view pedestrian detection database for collaborative intelligent transportation systems," *Future generation computer systems*, vol. 113, pp. 506–527, 2020.

[113] S. Shi, J. Cui, Z. Jiang, Z. Yan, G. Xing, J. Niu, and Z. Ouyang, "Vips: Real-time perception fusion for infrastructure-assisted autonomous driving," in *Proceedings of the 28th annual international conference on mobile computing and networking*, 2022, pp. 133–146.

[114] Z. Zheng, X. Xia, L. Gao, H. Xiang, and J. Ma, "Cooperfuse: A real-time cooperative perception fusion framework," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2024, pp. 533–538.

[115] Y. Li, S. Ren, P. Wu, S. Chen, C. Feng, and W. Zhang, "Learning distilled collaboration graph for multi-agent perception," *Advances in Neural Information Processing Systems*, vol. 34, pp. 29 541–29 552, 2021.

[116] Q. Xie, X. Zhou, T. Qiu, Q. Zhang, and W. Qu, "Soft actor–critic-based multilevel cooperative perception for connected autonomous vehicles," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 370–21 381, 2022.

[117] P. Liu, Z. Wang, G. Yu, B. Zhou, and P. Chen, "Region-based hybrid collaborative perception for connected autonomous vehicles," *IEEE Transactions on Vehicular Technology*, 2023.

[118] T. Huang, J. Liu, X. Zhou, D. C. Nguyen, M. R. Azghadi, Y. Xia, Q.-L. Han, and S. Sun, "V2x cooperative perception for autonomous driving:

Recent advances and challenges," *arXiv preprint arXiv:2310.03525*, 2023.

[119] X. Dong and M. L. Cappuccio, "Applications of computer vision in autonomous vehicles: Methods, challenges and future directions," *arXiv preprint arXiv:2311.09093*, 2023.

[120] Z. Liu, H. Tang, A. Amini, X. Yang, H. Mao, D. L. Rus, and S. Han, "Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation," in *2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023, pp. 2774–2781.

[121] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[122] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu *et al.*, "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.

[123] R. Gao, M. Li, S.-J. Yang, and K. Cho, "Reflective noise filtering of large-scale point cloud using transformer," *Remote Sensing*, vol. 14, no. 3, p. 577, 2022.

[124] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 16 259–16 268.

[125] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.

[126] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *International Conference on Learning Representations*, 2018.

[127] A. P. Dempster, "A generalization of bayesian inference," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 30, no. 2, pp. 205–232, 1968.

[128] ——, "Upper and lower probabilities induced by a multivalued mapping," in *Classic works of the Dempster-Shafer theory of belief functions*. Springer, 2008, pp. 57–72.

[129] E. Arnold, S. Mozaffari, and M. Dianati, "Fast and robust registration of partially overlapping point clouds," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1502–1509, 2021.

[130] Y. Yuan and M. Sester, "Comap: A synthetic dataset for collective multi-agent perception of autonomous driving," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 255–263, 2021.

[131] Y. Li, D. Ma, Z. An, Z. Wang, Y. Zhong, S. Chen, and C. Feng, "V2x-sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 914–10 921, 2022.

[132] R. Mao, J. Guo, Y. Jia, Y. Sun, S. Zhou, and Z. Niu, "Dolphins: Dataset for collaborative perception enabled harmonious and interconnected self-driving," in *Proceedings of the Asian Conference on Computer Vision*, 2022, pp. 4361–4377.

[133] S. Wei, Y. Wei, Y. Hu, Y. Lu, Y. Zhong, S. Chen, and Y. Zhang, "Asynchrony-robust collaborative perception via bird's eye view flow," *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[134] Z. Bai, G. Wu, M. J. Barth, Y. Liu, E. A. Sisbot, and K. Oguchi, "Vinet: Lightweight, scalable, and heterogeneous cooperative perception for 3d object detection," *Mechanical Systems and Signal Processing*, vol. 204, p. 110723, 2023.

[135] T. Wang, S. Kim, J. Wenxuan, E. Xie, C. Ge, J. Chen, Z. Li, and P. Luo, "Deepaccident: A motion and accident prediction benchmark for v2x autonomous driving," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 6, 2024, pp. 5599–5606.

[136] J. Gamerdinger, S. Teufel, P. Schulz, S. Amann, J.-P. Kirchner, and O. Bringmann, "Scope: A synthetic multi-modal dataset for collective perception including physical-correct weather conditions," *arXiv preprint arXiv:2408.03065*, 2024.

[137] H. Wang, X. Zhang, Z. Li, J. Li, K. Wang, Z. Lei, and R. Haibing, "Ips300+: a challenging multi-modal data sets for intersection perception system," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2539–2545.

[138] R. Xu, X. Xia, J. Li, H. Li, S. Zhang, Z. Tu, Z. Meng, H. Xiang, X. Dong, R. Song *et al.*, "V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 712–13 722.

[139] D. Sun, L. Cheng, S. Chen, C. Li, Y. Xiao, and B. Luo, "Uav-ground visual tracking: A unified dataset and collaborative learning approach," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[140] J. Axmann, R. Moftizadeh, J. Su, B. Tennstedt, Q. Zou, Y. Yuan, D. Ernst, H. Alkhatib, C. Brenner, and S. Schön, "Lucoop: Leibniz university cooperative perception and urban navigation dataset," in *2023 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2023, pp. 1–8.

[141] H. Yu, W. Yang, H. Ruan, Z. Yang, Y. Tang, X. Gao, X. Hao, Y. Shi, Y. Pan, N. Sun *et al.*, "V2x-seq: A large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5486–5495.

[142] R. Hao, S. Fan, Y. Dai, Z. Zhang, C. Li, Y. Wang, H. Yu, W. Yang, J. Yuan, and Z. Nie, "Rcooper: A real-world large-scale dataset for roadside cooperative perception," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 22 347–22 357.

[143] X. Zhang, Y. Li, J. Wang, X. Qin, Y. Shen, Z. Fan, and X. Tan, "Inscope: A new real-world 3d infrastructure-side collaborative perception dataset for open traffic scenarios," *arXiv preprint arXiv:2407.21581*, 2024.

[144] H. Xiang, Z. Zheng, X. Xia, R. Xu, L. Gao, Z. Zhou, X. Han, X. Ji, M. Li, Z. Meng *et al.*, "V2x-real: a largs-scale dataset for vehicle-to-everything cooperative perception," *arXiv preprint arXiv:2403.16034*, 2024.

[145] G. Kueppers, J.-P. Busch, L. Reiher, and L. Eckstein, "V2aix: A multi-modal real-world dataset of etsi its v2x messages in public road traffic," 2024.

[146] C. Ma, L. Qiao, C. Zhu, K. Liu, Z. Kong, Q. Li, X. Zhou, Y. Kan, and W. Wu, "Holovic: Large-scale dataset and benchmark for multi-sensor holographic intersection and vehicle-infrastructure cooperative," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 22 129–22 138.

[147] M. Yazgan, M. V. Akkanapragada, and J. M. Zöllner, "Collaborative perception datasets in autonomous driving: A survey," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2024, pp. 2269–2276.

[148] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[149] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE, 2018, pp. 2575–2582.

[150] R. Xu, Y. Guo, X. Han, X. Xia, H. Xiang, and J. Ma, "Opencda: an open cooperative driving automation framework integrated with co-simulation," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 1155–1162.

[151] S. Teufel, J. Gamerdinger, J.-P. Kirchner, G. Volk, and O. Bringmann, "Collective perception datasets for autonomous driving: A comprehensive review," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, 2024, pp. 1548–1555.

[152] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[153] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu *et al.*, "Mmdetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.

[154] ETSI. Abstract syntax notation definitions for tc its. [Online]. Available: forge.etsi.org/rep/ITS/asn1

[155] R. Liu, J. Wang, and B. Zhang, "High definition map for automated driving: Overview and analysis," *The Journal of Navigation*, vol. 73, no. 2, pp. 324–341, 2020.

[156] K. Kim, S. Cho, and W. Chung, "Hd map update for autonomous driving with crowdsourced data," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1895–1901, 2021.

[157] Y. Liu, T. Yuan, Y. Wang, Y. Wang, and H. Zhao, "Vectormapnet: End-to-end vectorized hd map learning," in *International Conference on Machine Learning*. PMLR, 2023, pp. 22 352–22 369.

[158] W. Xiao, N. Mehdipour, A. Collin, A. Y. Bin-Nun, E. Frazzoli, R. D. Tebbens, and C. Belta, "Rule-based optimal control for autonomous driving," in *Proceedings of the ACM/IEEE 12th International Conference on Cyber-Physical Systems*, 2021, pp. 143–154.

[159] Y. Zhou, Y. Takeda, M. Tomizuka, and W. Zhan, "Automatic construction of lane-level hd maps for urban scenes," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 6649–6656.

[160] V. Rheinstädter, "Digital maps in large, complex buildings: Effects on spatial performance and learning, influence of ability and environment," Ph.D. dissertation, ETH Zurich, 2019.

[161] M. Hirabayashi, A. Sujiwo, A. Monrroy, S. Kato, and M. Edahiro, "Traffic light recognition using high-definition map features," *Robotics and Autonomous Systems*, vol. 111, pp. 62–72, 2019.

[162] F. Ghallabi, G. El-Haj-Shhade, M.-A. Mittet, and F. Nashashibi, "Lidar-based road signs detection for vehicle localization in an hd map," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 1484–1490.

[163] Y. He, C. Bian, J. Xia, S. Shi, Z. Yan, Q. Song, and G. Xing, "Vi-map: Infrastructure-assisted real-time hd mapping for autonomous driving," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, 2023, pp. 1–15.

[164] E. Héry, P. Xu, and P. Bonnifait, "Consistent decentralized cooperative localization for autonomous vehicles using lidar, gnss, and hd maps," *Journal of Field Robotics*, vol. 38, no. 4, pp. 552–571, 2021.

[165] J. Jeong, Y. Cho, and A. Kim, "Hdmi-loc: Exploiting high definition map image for precise localization via bitwise particle filter," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6310–6317, 2020.

[166] B. Yang, M. Liang, and R. Urtasun, "Hdnet: Exploiting hd maps for 3d object detection," in *Conference on Robot Learning*. PMLR, 2018, pp. 146–155.

[167] J. Fang, D. Zhou, X. Song, and L. Zhang, "Mapfusion: A general framework for 3d object detection with hdmaps," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3406–3413.

[168] Y. Endo, E. Javanmardi, Y. Gu, and S. Kamijo, "High definition map aided object detection for autonomous driving in urban areas," *International journal of automotive technology*, vol. 24, no. 2, pp. 469–481, 2023.

[169] K. Jiang, Y. Shi, B. Wijaya, M. Yang, T. Wen, Z. Xiao, and D. Yang, "Map container: A map-based framework for cooperative perception," *arXiv preprint arXiv:2208.13226*, 2022.

[170] R. Matthaei, G. Bagschik, and M. Maurer, "Map-relative localization in lane-level maps for adas and autonomous driving," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE, 2014, pp. 49–55.

[171] Y. Ou, Y. Fan, X. Zhang, Y. Lin, and W. Yang, "Improved a* path planning method based on the grid map," *Sensors*, vol. 22, no. 16, p. 6198, 2022.

[172] Z. Zhang, Y. Zhang, X. Ding, F. Jin, and X. Yue, "Online vectorized hd map construction using geometry," in *European Conference on Computer Vision*. Springer, 2025, pp. 73–90.

[173] R. Kmet, Z. Dvorak, and M. Kvet, "Map of traffic accidents," *Transportation research procedia*, vol. 40, pp. 1418–1425, 2019.

[174] T. Ort, L. Paull, and D. Rus, "Autonomous vehicle navigation in rural environments without detailed prior maps," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 2040–2047.

[175] J. Gong, C. Gong, Y. Lin, Z. Li, and C. Lü, "Review on machine learning methods for motion planning and control policy of intelligent vehicles," *Transactions of Beijing institute of Technology*, vol. 42, no. 7, pp. 665–674, 2022.

[176] S.-W. Kim, Z. J. Chong, B. Qin, X. Shen, Z. Cheng, W. Liu, and M. H. Ang, "Cooperative perception for autonomous vehicle control on the road: Motivation and experimental results," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 5059–5066.

[177] C. Pilz, A. Ulbel, and G. Steinbauer-Wagner, "The components of cooperative perception-a proposal for future works," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 7–14.

[178] Z. Zhu, Q. Du, Z. Wang, and G. Li, "A survey of multi-agent cross domain cooperative perception," *Electronics*, vol. 11, no. 7, p. 1091, 2022.

[179] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5g be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[180] A. Ogawa, S. Kuroda, K. Ushida, R. Kudo, K. Tateishi, H. Yamashita, and T. Kantou, "Field experiments on sensor data transmission for 5g-based vehicle-infrastructure cooperation," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2018, pp. 1–5.

[181] A. I. Salameh and M. El Tarhuni, "From 5g to 6g—challenges, technologies, and applications," *Future Internet*, vol. 14, no. 4, p. 117, 2022.

[182] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, "Toward the 6g network era: Opportunities and challenges," *IT Professional*, vol. 22, no. 1, pp. 34–38, 2020.

[183] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6g networks," *IEEE network*, vol. 34, no. 6, pp. 272–280, 2020.

[184] B. Yang, X. Cao, K. Xiong, C. Yuen, Y. L. Guan, S. Leng, L. Qian, and Z. Han, "Edge intelligence for autonomous driving in 6g wireless system: Design challenges and solutions," *IEEE Wireless Communications*, vol. 28, no. 2, pp. 40–47, 2021.

[185] A. Jouini, H. Dalila, and C. Adnen, "Adas and 5g/6g-v2x implementation for sustainable and intelligent transport," in *2023 22nd Mediterranean Microwave Symposium (MMS)*. IEEE, 2023, pp. 1–6.

[186] E. Coronado, R. Behravesh, T. Subramanya, A. Fernàndez-Fernàndez, M. S. Siddiqui, X. Costa-Pérez, and R. Riggio, "Zero touch management: A survey of network automation solutions for 5g and 6g networks," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2535–2578, 2022.

[187] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[188] G. Jawahar, B. Sagot, and D. Seddah, "What does bert learn about the structure of language?" in *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics*, 2019.

[189] C. Cui, Y. Ma, X. Cao, W. Ye, Y. Zhou, K. Liang, J. Chen, J. Lu, Z. Yang, K.-D. Liao *et al.*, "A survey on multimodal large language models for autonomous driving," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 958–979.

[190] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang *et al.*, "A survey on evaluation of large language models," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 3, pp. 1–45, 2024.

[191] C. Xiang, C. Feng, X. Xie, B. Shi, H. Lu, Y. Lv, M. Yang, and Z. Niu, "Multi-sensor fusion and cooperative perception for autonomous driving: A review," *IEEE Intelligent Transportation Systems Magazine*, 2023.

[192] S. Yin, C. Fu, S. Zhao, K. Li, X. Sun, T. Xu, and E. Chen, "A survey on multimodal large language models," *National Science Review*, p. nwae403, 2024.

[193] J. Chen, Z. Liu, X. Huang, C. Wu, Q. Liu, G. Jiang, Y. Pu, Y. Lei, X. Chen, X. Wang *et al.*, "When large language models meet personalization: Perspectives of challenges and opportunities," *World Wide Web*, vol. 27, no. 4, p. 42, 2024.

[194] O. Friha, M. A. Ferrag, B. Kantarci, B. Cakmak, A. Ozgun, and N. Ghoualmi-Zine, "Llm-based edge intelligence: A comprehensive survey on architectures, applications, security and trustworthiness," *IEEE Open Journal of the Communications Society*, 2024.

[195] M. Zhang, B. Yuan, H. Li, and K. Xu, "Llm-cloud complete: Leveraging cloud computing for efficient large language model-based code completion," *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, vol. 5, no. 1, pp. 295–326, 2024.

[196] X. Lin, W. Wang, Y. Li, S. Yang, F. Feng, Y. Wei, and T.-S. Chua, "Data-efficient fine-tuning for llm-based recommendation," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024, pp. 365–374.

[197] F. De La Torre, C. M. Fang, H. Huang, A. Banburski-Fahey, J. Amores Fernandez, and J. Lanier, "Llmr: Real-time prompting of interactive worlds using large language models," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–22.

[198] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang *et al.*, "Planning-oriented autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 853–17 862.

[199] P. S. Chib and P. Singh, "Recent advancements in end-to-end autonomous driving using deep learning: A survey," *IEEE Transactions on Intelligent Vehicles*, 2023.

[200] L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, "End-to-end autonomous driving: Challenges and frontiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[201] K. Chitta, A. Prakash, and A. Geiger, "Neat: Neural attention fields for end-to-end autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 793–15 803.

[202] S. Hu, L. Chen, P. Wu, H. Li, J. Yan, and D. Tao, "St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning,"

in *European Conference on Computer Vision*. Springer, 2022, pp. 533–549.

[203] K. Chitta, A. Prakash, B. Jaeger, Z. Yu, K. Renz, and A. Geiger, "Transfuser: Imitation with transformer-based sensor fusion for autonomous driving," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 12 878–12 895, 2022.

[204] L. Li, Y. Cheng, C. Sun, and W. Zhang, "Icop: Image-based cooperative perception for end-to-end autonomous driving," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2024, pp. 2367–2374.

[205] R. Liang, L. Zhao, and P. Wang, "Performance evaluations of lora wireless communication in building environments," *Sensors*, vol. 20, no. 14, p. 3828, 2020.

[206] J.-M. Molina-Garcia-Pardo, M. Lienard, and P. Degauque, "Wireless communication in tunnels," in *Digital communication*. InTech, 2012.

[207] R. Uijlenhoet, A. Overeem, and H. Leijnse, "Opportunistic remote sensing of rainfall using microwave links from cellular communication networks," *Wiley Interdisciplinary Reviews: Water*, vol. 5, no. 4, p. e1289, 2018.

[208] Y. Zhu, C. Zhao, and Y. Du, "Latency impact analysis of point cloud fusion modes for cooperative perception in autonomous driving," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 1024–1029.

[209] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, p. 2140, 2021.

[210] Task Group p, "Ieee p802. 11p: Wireless access in vehicular environments (wave)," 2006.

[211] Automation, Cooperative Driving, "Taxonomy and definitions for terms related to cooperative driving automation for on-road motor vehicles," *SAE J3216*, 2020.

[212] SAE Advanced Applications Technical Committee, "V2x sensor-sharing for cooperative & automated driving," *SAE J3224*, 2022.

[213] SAE Advanced Applications Technical Committee, "Cooperative driving automation (cda) feature - infrastructure-based prescriptive cooperative merge," *SAE J3256*, 2024.

[214] ETSI, TCITS, "Intelligent transport systems (its); vehicular communications; basic set of applications; part 1: Functional requirements," 2011.

[215] ETSI, TCITS, "Intelligent transport systems (its); vehicular communications; basic set of applications; part 2: Specification of cooperative awareness basic service," 2011.

[216] ETSI, TCITS, "Intelligent transport systems (its); vehicular communications; basic set of applications; part 3: Specifications of decentralized environmental notification basic service," 2014.

[217] ETSI, ITS, "Intelligent transport system (its);vehicular communications; basic set of applications; collective perception service; release 2," 2023.