Bringing Different Views Together: A Hybrid Cooperative Perception Framework for Connected Autonomous Vehicles

Dominic Carrillo*, Michael Nutt*, Maarten Meijer†, Junaid Khan†, Song Fu* and Qing Yang*
*Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA
†PACCAR Technical Center, Mount Vernon, WA, USA

***Abstract*—Cooperative perception will be essential for connected autonomous vehicles to enhance object recognition and optimize path planning by sending data information about the surrounding environment. However, an inherent challenge in existing systems is the high bandwidth cost of transmitting information in real-time, which restricts cooperative perception's practicality. This work presents a hybrid cooperative perception fusion framework aimed at mitigating this issue by optimizing data transmission according to available bandwidth or through data reduction techniques. Our methods ensure that vehicles can rapidly transmit high-confidence data without overwhelming the network. Experimental results indicate that our methodology substantially diminishes data transmission sizes while maintaining object detection accuracy. For cooperative perception in autonomous vehicle systems, our approach provides a scalable and effective way to get past the bandwidth barrier.**

I. Introduction: Fusion Models Explained

In the field of autonomous vehicles, the integration of data from various modalities—such as cameras, LiDAR, and radar—has become a critical focus of research and innovation. This multimodal fusion facilitates a more thorough understanding of the vehicle's environment, which is vital for improving the accuracy of perception systems and ensuring their robustness. Examples of multi-modality fusion are Graph R-CNN [14] or MVX-Net [13], which effectively fuses data from LiDAR and Camera features to leverage the performance from both sensors.

Fusion between sensors is not the only innovation moving towards data fusion to improve the accuracy and precision of object detection systems. There have been advancements in utilizing data at different portions of the pipeline in the model to enrich features in the current inference. The concept of data-level fusion in autonomous vehicles can broadly be categorized into three types:

- **Early Fusion:** The raw data is fused at the start or before the processing pipeline, providing unprocessed and raw data early in the decision-making process [7].
- **Deep Fusion:** During the intermediate stage, such as the attention matrix of a transformer or the layers of a Convolution Neural Network (CNN), the features extracted from various sensors are fused [13].
- **Late Fusion:** Fused finalized output bounding boxes, regarding an object, at the final stage of the process, providing a common agreement from multiple independent evaluations [10].

These research works demonstrated the increasing use of complex fusion methods to address perception challenges faced in the autonomous vehicles field. The mix of different research where some models focus on multi-modality data fusion, and others concentrate on multi-level data fusion continues to make marginal improvements.

The object detection model's reliability and safety have been brought into question by recent incidents involving autonomous vehicles, attracting major public and government criticism [3]. CBS News reported an accident of a tragic collision in Texas with a Ford electric SUV that was utilizing the Blue Cruise system [9]. GM Cruise has openly published the findings of a third-party investigation into one specific incident, demonstrating their commitment to being upfront and making advancements in the safety of autonomous vehicles [6]. These incidents emphasize an immediate demand for robust safety protocols and reliability in automated driving systems.

Tesla has a major recall from incidents dating back to 2021, due to concerns regarding the system's effectiveness in detecting emergency vehicles [1]. An investigation by the National Highway Traffic Safety Administration on Tesla's Autopilot system demonstrates that government agencies are paying attention to proper safety features to warn and monitor drivers [2]. This investigation demonstrates increased concerns about safety testing and regulatory compliance to comprehend complex driving environments and potential hazard mitigation if the systems fail to operate as anticipated.

While manufacturers will adhere to stringent safety standards to ensure the safety and reliability of autonomous vehicles, there is significant potential to further enhance perception capabilities by shifting from single-agent to multi-agent fusion approaches. This transition leverages information from multiple vehicles, creating a more comprehensive view of the environment. Such cooperative perception not only improves detection accuracy but also enhances the overall safety and effectiveness of autonomous driving systems.

In 2019, an initial breakthrough into cooperative perception was the Cooper [4] approach, where a transmitting vehicle shared its raw point cloud data with the ego vehicle, which fused the information before the detection pipeline. From their

insight, the ego vehicle has a notable 10% improvement in object detection 80% of the time. This enhancement showcases that data sharing and fusion between vehicles is valuable in enriching object detection results. In correspondence, cooperative fusion has become a research topic to this day with its own three categories in early, deep, and late fusion [5, 11, 12].

Cooperative perception systems made innovations to exchange information between vehicles, resulting in road safety and vehicle efficiency. However, there are limitations in data mitigation and alignment that have not yet been explored thoroughly. The primary challenge is the bandwidth constraint during real-time data processing, which can limit the amount of data that can be communicated between vehicles. Our solution for this issue is to introduce a hybrid cooperative perception fusion framework. The motivation is to explore and refine the hybrid fusion framework that can significantly enhance the capabilities of cooperative perception systems. Optimizing the data transmission is the goal, by adapting to the corresponding data level that is available in the network bandwidth and only providing significant contextual data. Therefore, this approach aims to improve the effectiveness of data integration but also establish a robust framework that can adapt to varying network conditions without compromising the system's performance.

II. Research Challenges on the Road

Cooperative perception for autonomous vehicles continually evolve becoming complex, and there have been numerous technical challenges and discussions on how to address them before deployment. Notably in data fusion and alignment, these complexities present major obstacles directly influencing the safety and effectiveness of the systems. Data fusion is crucial for single and multi-agent systems as it involves the merging of data from many sensor inputs or between vehicles to create a unified data input. Before data fusion, the system must undergo alignment for precise synchronization and positioning of the data. The objective is to guarantee the alignment of data from different sources before the merging procedure. Both are necessary for enhancing vehicle perception in object detection capabilities and addressing their associated challenges is important.

A. Let Synchronize Across Vehicles: Alignment

There are several components in the alignment of data for cooperative perception in autonomous vehicles including synchronization, vehicle placement within coordinate systems, and sensor drift management. Data synchronization ensures the utility of the transmitter data to another vehicle which is essential for real-time cooperative decision-making. However, the shared data is in one perspective and must go through a process to map the transmitter's coordinates into the receiver's coordinate systems, a critical step for maintaining data integrity and relevance.

The bottom of Figure 1, illustrates this process where GPS/GNSS data are used to align the receiver's data for a common global reference, ensuring that the data is temporally and spatially synchronized. This step is crucial for the shared data to maintain its perspective and accuracy across different vehicle systems at different data levels. For example, taking the LiDAR's point cloud data from the transmitter would have to be transformed from a LiDAR coordinate system to a unified global coordinate system, being the GPS location, ensuring spatial accuracy across data from different sources. The transformation from the transmitter's location to the receiver's location is performed under the global coordinate system, therefore aligning the positions. Then, the third transformation is required back from global to the sensors coordinate system, aligning the two-point clouds similar seen in point cloud registration or SLAM research. As seen on the bottom of Figure 2 shows a combined point cloud, the alignment for feature maps is the same. An additional challenge that is out of scope in this discussion is the management of sensor drift, which can be mitigated through calibration and filtering techniques to maintain data accuracy.

B. Data Fusion

Once the data is aligned, the next significant procedure is to fuse them to generate one data input which would be a comprehensive understanding of the environment for the object detection system to process. This idea is not novel when it comes to single-agent fusion due to multiple techniques researched into sensor and temporal fusion. Extending those ideas into multi-agent fusion is to incorporate data from multiple vehicles to have a broader view of the environment. There are many different fusion techniques from CNN-based to transformer-based methods in their effectiveness in the merging of data. Each will influence the system's ability to make decisions accurately for complex scenarios like highways, city, or rural driving.

CNN-based and transformer-based algorithms are widely recognized for their effectiveness in fusing data from sensor-to-sensor or vehicle-to-vehicle. When it comes to CNN-based approaches, utilizing maxout or average pooling within the network provides increased performance in spatial data integration. This is accomplished by simplifying the input while emphasizing the most relevant features, which assist in image processing. In transformer-based approaches, the methods of self-attention and cross-attention manage complex features from data enhancing the model's efficiency. Researchers can dynamically merge visual cues and features from sensors or other vehicles that prioritize significant input information. These advanced techniques provide a powerful toolkit for achieving accurate and efficient data fusion for object detection systems.

III. Hybrid Cooperative Perception Fusion

Following the trends in the cooperative perception field, our research looks to the future of a novel approach to hybrid cooperative perception fusion and designing a framework to leverage the strengths of the different fusion techniques while addressing the limitations of bandwidth and data processing limitations. To overcome these limitations, we propose a

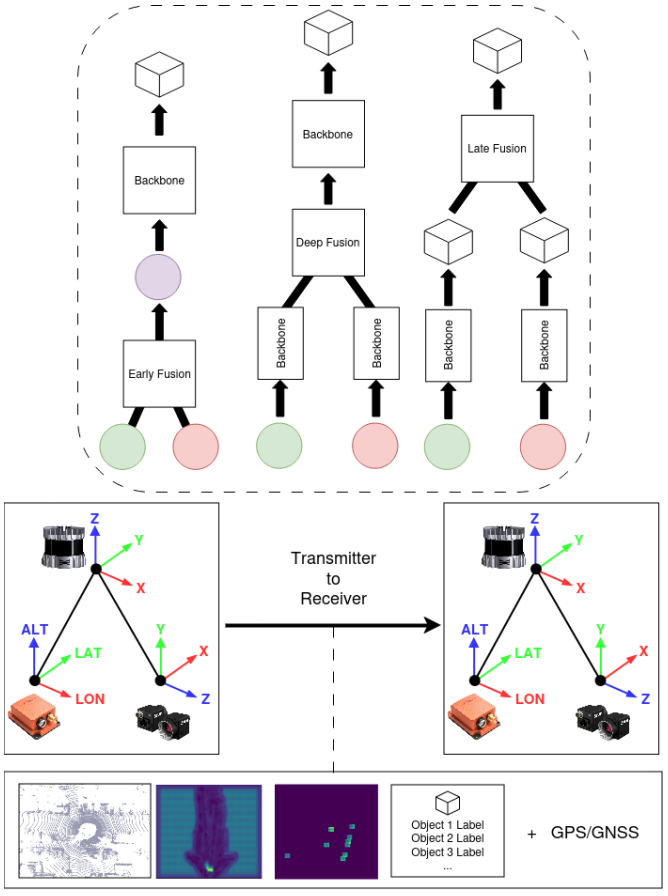

Fig. 1. The top section illustrates fusion frameworks for cooperative perception, where the proposed framework switches between early, deep, and late fusion to enhance object detection accuracy, computational workload, and real-time data communication. Early fusion merges raw sensor inputs, deep fusion merges network features, and late fusion merges detection outputs. The bottom section demonstrates transmitter-receiver interactions, highlighting data sharing and alignment across coordinate systems.

layered hybrid approach to data fusion, providing methods that help to overcome the deficiencies found in each fusion method by combining the best features of early and late fusion in one scenario and deep and late fusion in another. In this section, we will discuss the details of different strategies we developed: an adaptive fusion framework in section III-A, a hybrid fusion approach using LiDAR with predicted bounding boxes in section III-B, and using feature map with predicted bounding boxes in section III-C.

A. Adaptive Fusion Framework

The ability of networks to adapt to varying network conditions is essential for preserving system performance and efficiency in the rapidly evolving field of autonomous vehicles. The *adaptive fusion framework* we have developed is specifically designed to seamlessly switch between various levels of data fusion, namely early, deep, and late, depending on real-time evaluations of network bandwidth and traffic conditions, which is seen in Figure 1. This approach ensures optimal usage of available bandwidth by adjusting fusion strategies based

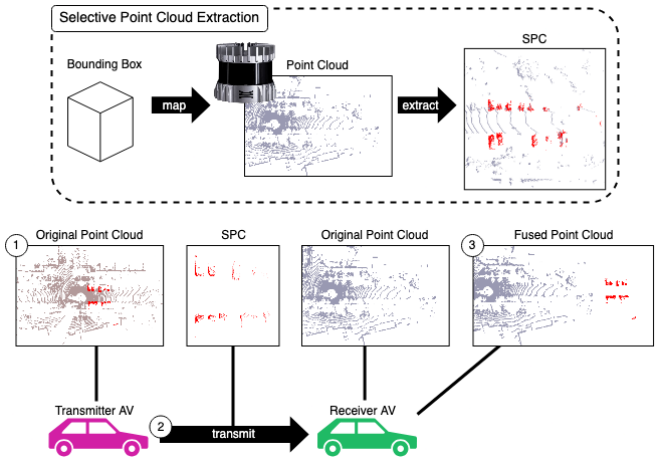

Fig. 2. The LiDAR-bounding box fusion process involves obtaining bounding boxes from point cloud inputs, mapping them back to the coordinate systems, to be extracted into a SPC which are transmitted, and fused with receiver's point cloud.

on real-time network conditions while maintaining the high accuracy of the perception system's output.

The implementation strategy for the adaptive fusion framework involves the following mechanisms: bandwidth monitoring, data prioritization, and dynamic fusion switch for both high and low bandwidth conditions. The system performs continuous bandwidth monitoring of the network and examines the rate of data traffic across the vehicle network. The monitoring process can be simplified by the use of integrated software that can estimate fluctuations in the bandwidth by analyzing past data and current network usage patterns. Regarding dynamic fusion switching, during periods of high bandwidth, when the network is not fully utilized, the system chooses to employ early or deep fusion. This allows for the transmission of larger message sizes, such as raw sensor data or feature maps, which contain more valuable information and can be done without compromising system latency. During periods of low bandwidth, when there is limited availability of data transmission capacity, the system can automatically switch to late fusion. This prioritizes the transmission of concise, high-level information such as bounding boxes or decision outputs, which have less bandwidth demands and are crucial for immediate vehicle responses. Furthermore, the framework can make use of a data prioritization mechanism that assesses the importance of data according to the driving context. This can be done in conjunction with bandwidth-based switching. Critical data that has a significant impact on vehicle decisions is given priority for transmission, guaranteeing that it is not delayed even when there are limitations on available bandwidth.

B. Hybrid LiDAR-Bounding Box Fusion

One distinctive feature of the adaptive fusion framework is its approach to switching between data levels, which can be adjusted depending on the network traffic. Despite this, inundating the network with a substantial amount of messages

poses a challenge. We aim to focus on resolving or mitigating the amount of data transmitted on the network. The hybrid LiDAR-bounding box fusion offers a solution for scenarios that have limited bandwidth by combining the accuracy of raw point cloud with the smaller size of the bounding boxes generated by the model.

Our approach is the ego vehicle will provide non-ego vehicles with precise data objects that it can see while reducing the size of the data needing to be sent. This is done by processing point clouds to generate new ones by extracting points from regions associated with object detection results. Utilizing the point cloud, from the ego's LiDAR sensor, of the surrounding environment which is inputted into an object detection model to obtain bounding boxes. These bounding boxes highlight the positions of detected objects including any objects that have low confidence value possibly due to being partially obscured. The initial process is to obtain the detected objects, bounding boxes, in a given point cloud, shown in Figure 2:(1). Would go through an alignment process to transform the bounding boxes to point cloud space before the extraction process. This transformation is critical and requires precise calibration of the LiDAR sensor.

Following the alignment is the extraction process, LiDAR points that fall inside the bounding boxes are extracted to generate a new point cloud, called Selected Point Cloud (SPC). This selective extraction significantly reduces the size of the point cloud data, as it concentrates only on regions where objects have been detected, thereby filtering out undesired data. The SPC is then transmitted to other vehicles for further processing, shown in Figure 2:(2). Once the point clouds are aligned, then they can be simply merged and fed into the object detection system to obtain similar or enhance the overall fidelity of the fused point cloud to obtain higher object detection results, shown in Figure 2:(3).

An additional option is to have the object detection model generate the detection results, and then iterate through the bounding box results to assess the accuracy of the detection using the confidence value. By utilizing the confidence value, one can determine whether it is advantageous to transmit the bounding box or the points for each detected object. If the level of confidence in the detection exceeds a certain threshold, the system will transmit only the bounding box information for the object to the receiving vehicle. Alternatively, it will identify the points contained within the bounding box and transmit those instead. As a demonstration, when the confidence level of the detection result is 80% or above, the system combines the bounding box features (confidence, coordinates, height, width, and length) and transmits them to other vehicles. If the confidence level is below 80%, the system utilizes the bounding box coordinates and dimensions to go through a point extraction procedure, resulting in the generation of a new point cloud that can be transmitted to other vehicles. By employing this method, we are minimizing the overall bandwidth needed for transmitting the data and guaranteeing that the receiving vehicle can utilize the information with a high level of certainty in the outcomes.

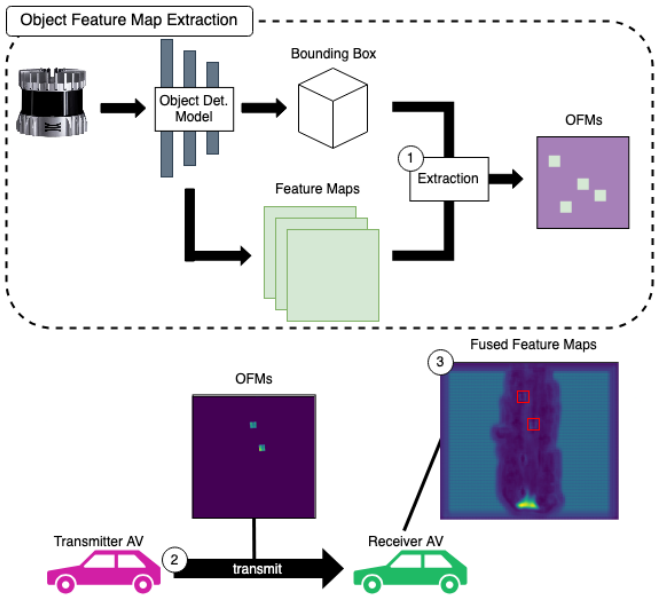

Fig. 3. The feature map-bounding box fusion process uses bounding boxes from the object detection model which are projected onto the feature map coordinate system to extract OFMs, which are transmitted and fused with the receiver's features. Fused features are shown in red boxes.

C. Hybrid Feature Map-Bounding Box Fusion

We would like to expand the technique using feature maps taken from object detection models, named hybrid feature map-bounding box fusion, in line with the hybrid LiDAR-bounding box fusion approach. This approach utilizes the contextual information found in feature maps to enhance the object detection process by extracting regions associated with objects. This approach aims to minimize the data transmission between vehicles while maintaining or enhancing detection evaluations.

By using feature maps rather than direct LiDAR point clouds, which are extracted during the object detection process, the hybrid feature map-bounding box fusion method differentiates itself. During the evaluations to obtain bounding boxes, feature maps are generated which is a representation that defines the features or attributes of the input data of the model. The bounding boxes tell us the locations and dimensions of detected objects, including vehicles, pedestrians, or other defined labels by the dataset. Thus, we can identify the relevant regions within feature maps by taking bounding boxes' locations and mapping them to the feature maps' locations. Accurate mapping of bounding boxes to extract regions from the feature map is crucial however knowing the the resolution size of the input point cloud we can transform location from one to the other.

In Figure 3:(1), the extracted process selects square regions within the feature map that correspond to the center point location of the bounding boxes. We chose this method over the alternatives such as exact dimension boxes that were rotated or circular regions in feature maps. Using a square region maintains the rich information around an object while simplifying

the extraction process along with the fusion process in later steps.

Once the relevant regions of the feature map are identified, they are saved to a new feature map called object feature map (OFM), which maintains the same dimensions as the original. This new feature map essentially acts as a mask, highlighting regions corresponding to objects with non-zero values, while non-objects remained zeroed out. The idea is to focus on retaining high-value feature data of objects that correspond to detected objects, and effectively generating a condensed feature map that emphasizes regions of interest.

The OFM containing these extracted regions is then packaged and transmitted to another vehicle, shown in Figure 3:(2). Upon receiving, the vehicle must align the received feature map with its existing feature maps. The fusion of these maps can be implemented using different methodologies, based on the intended result and the capabilities of the system, shown in Figure 3:(3). In our evaluation, we followed F-Cooper [5] by using the maxout function during the fusion process. After fusion, the feature map that has been enriched is reintroduced into the detection model to undergo further evaluation aiming to enhance object detection.

TABLE I
AVERAGE FILE SIZE COMPARISON (KB) FOR POINT CLOUD AND SELECTIVE POINT CLOUD

File Format	Average PC	Average SPC
Text	8,788.08 (KB)	153.72 (KB)
Gzip	3,035.98 (KB)	55.12 (KB)
Text Precision 0.2	1,979.38 (KB)	33.77 (KB)
Gzip Precision 0.2	518.73 (KB)	8.15 (KB)

TABLE II
AVERAGE FILE SIZE COMPARISON (KB) FOR FULL FEATURE MAP AND OBJECT FEATURE MAP

File Format	Average FM	Average OFM
Text	220,000.00 (KB)	220,000.00 (KB)
Gzip	9,075.21 (KB)	665.74 (KB)
Text Precision 0.2	44,000.00 (KB)	44,000.00 (KB)
Gzip Precision 0.2	759.43 (KB)	81.76 (KB)

D. Benefits of Hybrid Cooperative Perception Fusion

In cooperative perception, where bandwidth restrictions could impact object detection performance, effective data transfer remains a significant challenge. The adaptive hybrid fusion framework offers two key advantages:

- Reduction in Network Congestion: Transmits only critical data, ensuring efficient bandwidth usage.
- Reliable Perception Performance: Maintains high object detection accuracy across varying environments.

This is essential for enhancing safety and maintaining robust perception performance in diverse environments, from densely populated urban centers to sprawling suburban areas.

The hybrid LiDAR-bounding box fusion framework is similar to the Cooper [4] approach by enhancing detection capabilities, when it comes to occluded or partially visible objects in the receiving vehicle's environment. However, this framework employs selective point cloud extraction to reduce data transmission requirements while maintaining the object detection performance. Similarly, the hybrid feature map-bounding box fusion framework operates on the same principles of data reduction by focusing on transmitting critical regions containing objects rather than transmitting the full point clouds or feature maps. As a result, it greatly minimizes the amount of resources required to transmit within densely populated areas, while ensuring efficient and reliable perception.

We observed significant reductions in data transmission on the KITTI dataset [8] with results varying dependently on the environment's complexity based on the number of objects detected. By transmitting only SPCs, it will achieves a reduction of 55x in data size, as shown in Table I from 8,788 KB to 153 KB, a 98% reduction. While transmitting only OFMs will achieve a approximate 14x reduction in data size, as shown in Table II from 9,075 KB to 665 KB, a 92% reduction. This significantly lowers bandwidth requirements, thus data reduction enables vehicles to comprehensively expand their ability to detect additional objects or improving initial detection confidence without sacrificing performance.

From our evaluations, the accuracy of our maxout-CNN method demonstrates an improvement in object detection performance, detecting additional objects without negatively overwhelming the model. Further investigation into adopting alternative approaches, such as transformer-based fusion methods, could further enhance performance and detection capacity. By integrating hybrid fusion methods, autonomous vehicles can achieve optimized perception accuracy, lower network usage, while adapting to complex environments to advance towards safe and efficient cooperative perception systems.

IV. THE FUTURE OF COOPERATIVE PERCEPTION

Technologies and methods related to machine learning and data fusion techniques are anticipated to advance significantly as we explore the cooperative perception domain. Enhancing the situational awareness and predictive capacity of autonomous systems will be possible in several key areas:

- Generative AI for Predictive Capabilities
- Advancements in Spatial and Temporal Fusion
- Hybrid Deep Learning Architectures for Spatial-Temporal Data

A. Generative AI for Predictive Capabilites

Generative AI models present a promising opportunity to improve predictive abilities. Renowned for its capacity to generate new data instances that closely resemble the training data, generative models are utilized to forecast and simulate future environmental states. This function could enable vehicles to create accurate representations of their environment in real-time, allowing them to predict future conditions and adjust their strategies accordingly.

For instance, generative models could simulate a variety of pedestrian behaviors, weather conditions, or traffic patterns

utilizing both historical and real-time data. By incorporating these simulations into decision-making processes, allowing vehicles to effectively anticipate unexpected scenarios, thus improving road safety and efficiency. This approach enables vehicle's capability to navigate dynamic environments with greater confidence while supporting strategic planning and risk assessment.

B. Advancements in Spatial and Temporal Fusion

Spatial and temporal fusion developments are fundamental to enhance cooperative perception systems. Utilizing these methods would allow models to comprehend the future states of the environment by evaluating data over time and space.

Spatial fusion method combines input from multiple sensors to create a unified, multidimensional representation of the environment. Addressing challenges such as synchronization and data inconsistencies across sensors will improve the accuracy of object recognition and classification. For instance, advancements in 3D object reconstruction techniques will provide comprehensive detail which can improve the decision-making processes.

By including time-based data, Temporal fusion methods can enhances spatial analysis enabling the system to track objects in a dynamic environment. Developing real-time dynamic models that continuously update with new data will enables autonomous systems to anticipate and respond to environmental changes effectively. With addition to refining sequence prediction models can improve the accuracy of object path predictions by anticipating future states and behaviors.

C. Hybrid Deep Learning Architectures for Spatial-Temporal Data

An innovative hybrid architectures that combine spatial and temporal data processing offer potential to advance cooperative perception. These architectures could leverage the complementary strengths of CNNs (for spatial processing) and RNNs (for temporal data analysis), which are capable of processing detailed spatial hierarchies and temporal dependencies more effectively. Another potential approach is the utilization of graph-based models to depict the changing environment as a graph, with nodes symbolizing objects and edges indicating relationships and interactions over time. Applying GNNs to process these spatial-temporal graphs enables autonomous systems which could effectively comprehending complex relations, improving decision-making and prediction accuracy.

In the future, the development of generative AI and spatial-temporal fusion will be important factors in determining the development of autonomous vehicle technologies. These developments will improve the predictive and perceptive capabilities of autonomous systems, providing new opportunities for safer and more reliable vehicles. Expanding these technological boundaries for cooperative perception will pave the way for full autonomous driving systems capable of human-like reasoning and predictive abilities.

V. Conclusions

This study has shown different hybrid fusion techniques to solve bandwidth constraints for cooperative perception systems in autonomous vehicles. Our frameworks would dynamically adapt to varying network conditions by transmitting critical data in real-time without overwhelming bandwidth capacity. For practical applications, our research demonstrates the future of autonomous vehicles through scalable, bandwidth-efficient data fusion techniques. By selectively transmitting critical data, our frameworks represent key advancing techniques to ensure safe, and reliable autonomous systems. Through hybrid LiDAR-bounding box fusion and hybrid feature map-bounding box fusion, our experimental testing significantly reduced data transmission usage without compromising detection performance.

Acknowledgment

The authors thank the United States Department of Energy and PACCAR Inc. for the support for this project. This material is based upon work supported by the Department of Energy, under Grant Number DE-EE0009861.

Legal disclaimer: The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

[1] Fred Lambert (electrek). *New damning footage shows several Tesla vehicles on Autopilot crashing into police*. Aug. 2023. URL: https://electrek.co/2023/08/09/damning-footage-shows-tesla-vehicles-autopilot-crashing-into-police/. (accessed: 11.14.2024).

[2] Fred Lambert (electrek). *Tesla Autopilot is again under NHTSA investigation after doubts over recall remedy*. Apr. 2024. URL: https://electrek.co/2024/04/26/tesla-autopilot-under-nhtsa-investigation-doubts-recall-remedy/. (accessed: 11.14.2024).

[3] National Highway Transportation Safety Administration (NHTSA). *Standing General Order on Crash Reporting: For incidents involving ADS and Level 2 ADAS*. 2024. URL: https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting. (accessed: 11.14.2024).

[4] Qi Chen et al. *Cooper: Cooperative Perception for Connected Autonomous Vehicles based on 3D Point Clouds*. 2019. arXiv: 1905.05265 [cs.CV]. URL: https://arxiv.org/abs/1905.05265.

[5] Qi Chen et al. "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*. SEC '19. Arlington, Virginia: Association for Computing Machinery, 2019, pp. 88–100. ISBN: 9781450367332. DOI: 10.1145/3318216.3363300.

[6] Cruise. *Cruise Releases Third-Party Findings Regarding October 2*. Jan. 2024. URL: https://www.getcruise.com/news/blog/2024/cruise-releases-third-party-findings-regarding-october-2/. (accessed: 11.14.2024).

[7] Sudip Dhakal et al. *VirtualPainting: Addressing Sparsity with Virtual Points and Distance-Aware Data Augmentation for 3D Object Detection*. 2023. arXiv: 2312.16141 [cs.CV].

[8] Andreas Geiger, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 2012.

[9] CBS News. *Electric Ford SUV driver was using automated system before fatal crash, investigators say*. Apr. 2024. URL: https://www.cbsnews.com/detroit/news/electric-ford-suv-driver-was-using-automated-system-before-fatal-texas-crash-investigators-say/. (accessed: 11.14.2024).

[10] Su Pang, Daniel Morris, and Hayder Radha. "CLOCs: Camera-LiDAR object candidates fusion for 3D object detection". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 10386–10393.

[11] Deyuan Qu et al. *HEAD: A Bandwidth-Efficient Cooperative Perception Approach for Heterogeneous Connected and Autonomous Vehicles*. 2024. arXiv: 2408.15428 [cs.CV].

[12] Deyuan Qu et al. *SiCP: Simultaneous Individual and Cooperative Perception for 3D Object Detection in Connected and Automated Vehicles*. 2024. arXiv: 2312.04822 [cs.CV].

[13] Vishwanath A. Sindagi, Yin Zhou, and Oncel Tuzel. "MVX-Net: Multimodal VoxelNet for 3D Object Detection". In: *2019 International Conference on Robotics and Automation (ICRA)*. 2019, pp. 7276–7282. DOI: 10.1109/ICRA.2019.8794195.

[14] Honghui Yang et al. "Graph R-CNN: Towards Accurate 3D Object Detection with Semantic-Decorated Local Graph". In: *Computer Vision – ECCV 2022*. Ed. by Shai Avidan et al. Springer Nature Switzerland, 2022, pp. 662–679.