A Multi-Layer Framework for Semantic Perception and Infrastructure Support in Autonomous Driving

A Multi-Layer Framework for Semantic Perception and Infrastructure Support in Autonomous Driving

Diego Gasco*, Franck Guillemard†, Saleh Bensator† Claudio Casetti*

* Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy

† Stellantis, Poissy, France

E-mail: {diego.gasco, claudio.casetti}@polito.it {franck.guillemard, saleh.bensator1}@stellantis.com

***Abstract*—Intelligent Transportation Systems (ITS) and Connected and Automated Vehicles (CAVs) represent the next frontier of future mobility. In this context, the perception of the surrounding environment becomes a key enabler for autonomous vehicles, with on-board sensors and Vehicle-to-Everything (V2X) communications playing complementary roles. However, Collective Perception itself must be integrated with intelligent processes to prioritize perceived actors that hold higher relevance for a specific vehicle; this approach is commonly known as semantic perception or semantic communication, which allows the vehicle to assess the potential impact of each object on its immediate driving decisions. This paper presents a four-layer semantic perception framework, supported by an ad-hoc infrastructure to enhance an ego vehicle's motion planning. Experimental results in complex urban scenarios show that the framework provides a median anticipation time of 1.3 seconds for occluded objects, reaching 4.6 seconds at the 95th percentile. Furthermore, the environmental awareness of the assisted vehicle is enhanced by enlarging its Operational Design Domain (ODD). Moreover, the semantic logic reduces communication load by transmitting only 40-50% of perceived objects, with only 20-30% classified as safety-critical, effectively preventing channel congestion.**

***Index Terms*—Semantic Communications, Task-Oriented Communications, V2X, Perception, Relevance, Content Selection.**

I. Introduction

Collective Perception is a key enabler for future application services in ITS. CAVs and Road Side Units (RSUs) are equipped with sensors such as cameras and LiDAR, whose fused outputs provide knowledge of surrounding vehicles and road users, including Vulnerable Road Users (VRUs).

While V2X-based perception sharing can significantly extend the ODD of individual vehicles, it also raises challenges related to communication load, scalability, and data relevance. Broadcasting all detected actors may saturate the wireless channel and provide receivers with information that is not useful for their current driving context. For this reason, recent research and standardization efforts have been focused on context-aware collective perception. This paradigm aims to select and share only a subset of perceived items based on their expected usefulness for the receiver. This selective, relevance-driven information exchange is referred to as Semantic Communication.

In this work, we present a multi-layer framework for semantic perception and infrastructure-assisted object selection, specifically targeting scenarios where RSUs are deployed in strategic locations to support autonomous driving vehicles. The proposed framework leverages semantic reasoning to prioritize perceived participants based on their contextual relevance and to provide vehicles with necessary and useful information, rather than exhaustive environment descriptions.

The considered use case assumes direct assistance from infrastructure nodes positioned at critical points of the road network, such as intersections or complex urban areas, where visibility constraints and traffic density limit the effectiveness of on-board perception alone. In this setting, RSUs act as semantic perception hubs, selecting and transmitting only those actors that are relevant for the assisted vehicle, according to its position, motion, and driving context.

The key contributions of this paper are:

- Multi-Layer Semantic Pipeline: A modular architecture that processes kinematic data into high-level semantic labels (obstruction, situational relevance, collision risk);
- Predictive Awareness Evaluation: Demonstration of how the early identification of occluded hazards and potentially dangerous situations leads to an increased safety margin and a broader ODD for the ego vehicle;
- Communication Efficiency: Demonstration of significant data reduction (up to 60%) compared to exhaustive broadcasting.

The remainder of this paper is organized as follows: Section II reviews related work and existing approaches. Section III describes the proposed framework architecture, composed of a multi-layer pipeline. Section IV details the environmental setup for both the infrastructure, the assisted autonomous vehicle, and the Vehicle-to-

Infrastructure (V2I) communication. Section V presents the experimental validation results. Finally, Section VI concludes the paper and Section VII outlines directions for future work.

II. Related Works

A systematic overview of the state of the art on collective perception is provided in [1], which highlights the main challenges related to scalability, communication efficiency, and relevance of shared information.

In this context, the latest release of Collective Perception Messages (CPMs) [2] emphasizes sharing only relevant participants, using Value of Information (VoI) to limit data and avoid channel overload. Similarly, [3] shows that filtering based on object dynamics, such as position and speed, helps reduce unnecessary communication and processing.

Authors in [4] developed an analytical model to estimate the number of participants contained in an environment model. The results show that the field of view (FOV) of on-board sensors is often limited, and how V2X can positively increase the detection and awareness ratio performance.

On a similar note, the work done in [5] defines the VoI as a weighted combination of proximity, timeliness, and quality of the information, with the specific weights adjusted according to the V2X use case. An example of VoI based on detection accuracy has been developed in the work done in [6], and in [7].

More recently, semantic information has been exploited to further enhance relevance-aware perception sharing. In [8] and [9], authors proposed the employment of a knowledge graph of the driving scene to generate semantic-based messages that contain situationally prioritized insights. This approach has shown to improve decision-making and communication efficiency in high-risk scenarios, even in the absence of direct visibility.

Along the same line, the work presented in [10] evaluates the potential of semantic communication in V2X with a framework that transmits only semantic spatiotemporal embeddings from a camera, instead of raw frames. The approach significantly reduces the payload of messages while maintaining high predictive accuracy.

The concept of value-anticipating communication has been explored in [11], and it consists in anticipating the value of each potentially shared data from the perspective of potential receiver(s). This approach leverages VoI based on relative entropy and has shown a significant improvement in cooperative tracking of items.

Finally, authors in [12] demonstrate that no single strategy for content-selection can absolutely maximize the performance of semantic and task-oriented V2X communications. In this context, the need for content-selection strategies to provide the receiver with all the necessary information from sender(s) is a key, not-negligible aspect.

III. Semantic Perception Framework

This section outlines the proposed framework, supporting assisted autonomous driving in complex urban environments. Throughout this work, the assisted autonomous vehicle is referred to as the ego vehicle (or simply, the ego).

The system exploits an RSU to perceive, process, and label surrounding traffic participants (e.g., vehicles and vulnerable road users), using the multi-layer semantic perception framework. The RSU processes the perceived kinematic data, including position, speed, acceleration, heading, and bounding-box dimensions of each detected actor, and transmits to the ego vehicle a structured message containing only the derived semantic annotations: *(i)* a Non-Line-of-Sight (NLoS) flag, indicating whether the object is occluded from the ego's perspective; *(ii)* a semantic relevance flag, capturing behavioral context such as a pedestrian approaching a crossing or a vehicle entering the roundabout; *(iii)* speed violation indicators for non-compliant road users; and *(iv)* a predicted time-to-collision, derived from trajectory intersection analysis.

Beyond receiving environmental insights, the ego vehicle contributes by continuously sharing its intended motion plan with the infrastructure. This bidirectional exchange forms a closed-loop architecture that maintains collaborative interaction between the ego vehicle and the infrastructure.

The framework is organized into a sequence of processing layers, as illustrated in Fig. 1, each addressing a specific task of the pipeline. In the following paragraphs, we will delve into the details of every single layer.

A. Data Acquisition And Proximity Filtering

The infrastructure continuously acquires real-time traffic data from the environment. For each detected object, the system retrieves its position, velocity, acceleration, orientation, lane assignment, and dimensions. To emulate realistic sensing conditions, we define some hypotheses, such as a circular perception region around the infrastructure, and a distance threshold (set to 100 meters). Moreover, Gaussian distance-dependent noise is injected into position, velocity, and shape measurements, following commonly adopted sensor noise models in the literature [13].

Let $\mathbf{g}_i(t)$ denote the ground-truth data (e.g., position, speed, dimension, etc..) retrieved by the simulation environment for object i at time t. The noisy observations are modeled as:

$$\tilde{\mathbf{g}}_i(t) = \mathbf{g}_i(t) + \mathcal{N}(0, \sigma_g(d_i)), \tag{1}$$

where d_i is the Euclidean distance between the infrastructure sensor and the perceived object, and $\sigma_g(d_i)$ is the distance-dependent standard deviation.

All observations are stored in the Local Dynamic Map (LDM) of the infrastructure, which maintains short-term temporal state history for each detected object.

Building upon this representation, the first processing stage of the framework performs an ego-centered distance-based filtering. At each time step, the current state of the ego vehicle is retrieved from the LDM.

For all detected actors, the Euclidean distance to the ego vehicle is computed. Only those items whose distance is below a predefined maximum proximity threshold (set to 50 meters) are retained. If no surrounding participants satisfy this proximity condition, the layer returns no candidates for further processing.

This filtering step ensures that subsequent layers process only spatially relevant participants, reducing computational effort for the receiver, while preserving nearby useful information.

B. Interaction Relevance

The concept of interaction relevance aims to identify traffic participants that are not only spatially close but also behaviorally and contextually significant for the ego vehicle's future motion. To this end, interaction areas are defined as spatial regions around conflict zones (e.g., junctions or crossings). This semantic reasoning enables the identification of vehicles or pedestrians that may either interfere with the ego vehicle's right-of-way or to whom the ego vehicle must yield priority.

The relevance of surrounding road users is assessed based on three distinct situations, determined by the current path of the ego vehicle:

- Ego is within the interaction area: approaching vehicles are considered relevant for a potential failure to yield;
- Ego is approaching the interaction area: other vehicles already within the interaction area are selected if they are holding priority;
- Pedestrian interactions: pedestrians are considered relevant if they are within the interaction area or if their motion state indicates an intention to enter it, such as approaching a crossing or moving toward the roadway.

This process assigns an interaction-relevance label to those agents that may lead to potential conflict scenarios, while leaving all other detected participants unaltered.

C. Obstruction Detection

To detect hidden hazards, this layer performs an obstruction detection to identify agents that are in line-of-sight (LoS) or in NLoS, using a geometric approach. For each ego–target pair, visibility is evaluated by testing intersection rays from the ego's position to the target's position against all static (e.g., buildings, walls) and dynamic (e.g., other vehicles) obstacles in the environment. Based on this geometric evaluation, each target object is classified as LoS (direct line of sight), NLoS-static (occluded by static entities), or NLoS-dynamic (occluded by moving entities).

In this context, the system needs to have access to both static information, such as the pre-loaded maps of the environment, and to dynamic data, for which the LDM storage plays a fundamental role. This classification allows the system to explicitly reason about hidden threats, which is particularly critical in dense urban scenarios, such as intersections, roundabouts, and crossings.

D. Predictive Risk Assessment

The final layer predicts future collision risks through multi-hypothesis trajectory estimation, conditioned on the observed behavior of surrounding agents. For each relevant vehicle, the system generates multiple candidate speed profiles representing both compliant and potentially aggressive driving behaviors.

To identify potentially dangerous drivers, the system leverages the short-term history stored in the LDM. At each time step, the instantaneous speed of each vehicle is compared against the context-dependent speed limit: $v_{\max}$. If the measured speed exceeds the applicable limit, the violation magnitude $\delta_t = v_t - v_{\max}$ is recorded with its timestamp. A vehicle is classified as dangerous driver if the mean magnitude of the violation in a sliding window of duration T_w exceeds a threshold $\delta_{\min}$:

$$\bar{\delta} = \frac{1}{|\mathcal{V}|} \sum_{(t_i, \delta_i) \in \mathcal{V}} \delta_i > \delta_{\min} \tag{2}$$

where $\mathcal{V} = \{(t_i, \delta_i) \mid t_i \geq t_{\text{now}} - T_w\}$ is the set of violations recorded within the window. In this work, $T_w = 5$ s and $\delta_{\min} = 2$ m/s.

For vehicles not flagged as non-compliant, three speed hypotheses are created to capture uncertainty in future driver intent: a conservative profile $(0.7 \cdot v)$, a nominal profile (v), and a mildly aggressive profile $(1.1 \cdot v)$, where v is the current measured speed of the vehicle.

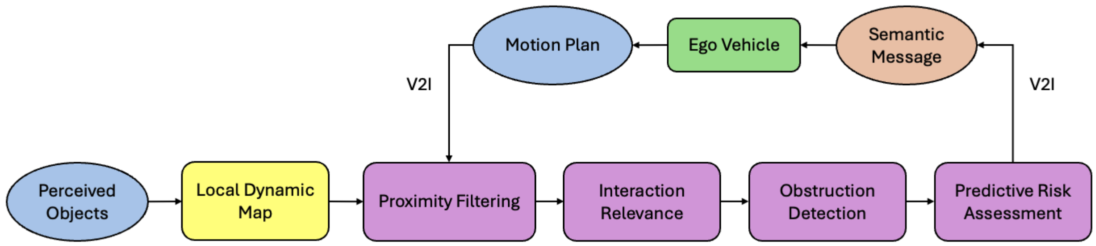

Fig. 1. Overview of the proposed framework architecture. The green rectangle represents the ego vehicle. Blue elements denote input data originating from both sensor-based perception and the ego vehicle itself (i.e., the planned motion transmitted via V2I communication). In yellow, the infrastructure Local Dynamic Map. Purple components illustrate the proposed multi-layer processing pipeline. The orange element indicates the framework output, encoded as a Semantic Message and transmitted back to the ego vehicle, always through V2I communication.

For vehicles flagged as non-compliant, candidate speed profiles are adjusted to account for higher-than-nominal behavior. Specifically, both a nominal profile and an aggressive profile are generated, with the latter reflecting the average speed exceeded within the window of time T_w. This modeling captures the likelihood of continued unsafe driving and allows the predictive layer to anticipate possible future emergencies.

Each speed hypothesis is propagated along precomputed road waypoints for a fixed $T_p = 3$ s horizon, assuming zero acceleration, a reasonable approximation given the short prediction window. When multiple feasible paths exist within the horizon, all alternatives are considered.

Predicted object trajectories are then evaluated against the ego's planned motion using a spatio-temporal conflict detection procedure. Let $\mathcal{T}_e(t)$ and $\mathcal{T}_o(t)$ denote the time-parameterized trajectories of the ego vehicle and a surrounding object, respectively. The conflict criterion is evaluated on the Euclidean distance between the predicted center positions of the two agents. A potential conflict is detected if:

$$\exists\, t \in [0, T_p] \quad \text{s.t.} \quad \|\mathcal{T}_e(t) - \mathcal{T}_o(t)\| \leq d_{\min}, \tag{3}$$

where $d_{\min} = 2$ m is the minimum safety distance threshold.

For pedestrians, future motion is predicted using a simplified constant-velocity model conditioned on crossing intent and inferred motion direction, estimated from recent position history. Only pedestrians actively crossing the roadway are considered for collision evaluation.

Detected conflicts are aggregated across all trajectory hypotheses and ranked according to time-to-conflict and minimum predicted separation distance.

IV. Environment Setup

The simulation environment is built using SUMO (Simulation of Urban Mobility) [14], a microscopic traffic simulator capable of modeling detailed vehicle and pedestrian behavior in urban road networks. The backend to control the ego vehicle motion and to implement the framework logic has been developed using the Python API of the Traffic Control Interface (TraCI), which offers a direct way to retrieve data and manage the SUMO environment. An urban scenario has been chosen as the case study, since it offers complex interactions among different items, including merging, yielding, and gap-acceptance, making it particularly suitable for evaluating the performance of our framework. To generate NLoS conditions from static items, the polygon elements available in SUMO have been used to mimic the presence of real buildings. Subsequently, a geometry-based approach is employed to systematically determine and classify LoS and NLoS (including both static and dynamic obstruction) conditions between the ego vehicle and surrounding targets. Traffic scenario includes heterogeneous traffic participants, such as trucks, buses, motorbikes, and passenger cars at varying densities, as well as pedestrians crossing the street in areas traversed by the ego vehicle. Furthermore, 40% of vehicles are modeled as non-compliant with traffic rules, displaying behaviors such as higher speeds, increased impatience, and greater driving variability. Hereafter, all traffic participants are referred to as “objects”.

Infrastructure placement is defined at key points along the ego vehicle's trajectory, where each unit operates independently within a 100 m sensing radius, reflecting the coverage of typical urban monitoring systems. The infrastructure evaluates the environment through the framework pipeline at a frequency of 10 Hz, matching the update rate of the simulation. In these experiments, we assume a V2X penetration rate of zero for the other participants, so only the ego vehicle and infrastructure are connected and can communicate, while the remaining entities are considered as non-connected and noncooperative road users.

To clarify the conceptual categorization of perceived objects in our scenario, Figure 2 presents a schematic Venn diagram highlighting:

- Objects perceived by the ego vehicle through its FOV;
- Objects detected by the infrastructure in its sensing range;
- Objects transmitted by the infrastructure;
- Objects labeled as potentially critical by the infrastructure, within the transmitted ones.

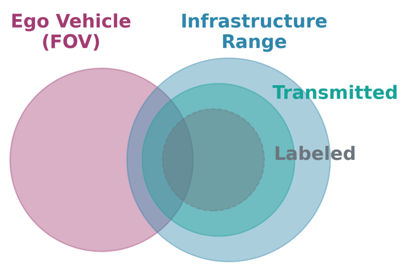

Fig. 2. Logical Venn diagram illustrating the relationships among objects categorization in the simulation environment: ego vehicle's FOV, infrastructure range, transmitted and labeled objects.

The ego vehicle operates along precomputed routes within the SUMO environment, and its main capabilities include a limited FOV, motion planning generation, and infrastructure integration. It perceives a restricted area spanning a maximum sensing distance of 80 meters and a horizontal angle of 70 degrees, with objects outside this range or in NLoS conditions ignored through dedicated API calls. Simultaneously, the vehicle generates a short-term motion plan based on its current state at each simulation step. In case the infrastructure signals an emergency, the ego vehicle adjusts its parameters to adopt a more cautious driving behavior, such as increasing following distances and reaction times.

V2I communication supports the bidirectional exchange of information, with the ego vehicle providing its motion plan and the infrastructure returning the corresponding framework outputs. A carrier frequency of 5.9 GHz is adopted, corresponding to the reference band used by IEEE 802.11p and C-V2X systems; all propagation and RSSI computations are performed accordingly. To emulate V2I communication, the Received Signal Strength Indicator (RSSI) and propagation delay are computed for each exchanged packet. The RSSI follows the stochastic propagation model specified in 3GPP TR 36.885 [15], suitable for urban environments, and accounts for distance-dependent path loss, stochastic shadowing, and fixed device losses:

$$\text{RSSI} = P_{\text{TX}} - L_{\text{propagation}}(d) - L_{\text{shadow}} - L_{\text{device}}, \quad (4)$$

where $P_{\text{TX}} = 23$ dBm is the transmit power, $L_{\text{propagation}}(d)$ is the path loss at distance d, L_{shadow} is a zero-mean Gaussian random variable with standard deviation $\sigma = 3$ dB, and $L_{\text{device}} = 3$ dB represents implementation losses. Successful reception occurs when RSSI $\geq RSSI_{\text{TH}}$, with $RSSI_{\text{TH}} = -90$ dBm (i.e., the receiver sensitivity); otherwise, the packet is discarded. Although propagation delay is explicitly modeled, it is negligible at the distances considered, enabling effective instantaneous data exchange.

V. Experiments

A. Anticipation Time

Anticipation time measures how much earlier the infrastructure detects an occluded object compared to the time it first becomes potentially observable by the ego vehicle's on-board sensors, i.e., when the object exits the occluded state and enters a LoS (and/or FOV) condition. This metric is critical in urban scenarios, where advanced awareness of occluded road users directly improves driving safety.

More technically, the anticipation time can be expressed as:

$$T_{\text{anticipation}} = t_{\text{ego\_visible}} - t_{\text{infrastructure\_detect}} \quad (5)$$

where $t_{\text{ego\_visible}}$ is the time at which the object leaves the occluded region and becomes potentially detectable to the ego vehicle, and $t_{\text{infrastructure\_detect}}$ is the time at which the infrastructure-based perception system detects the critical event.

To quantitatively assess the impact of anticipation time on an ego vehicle assisted with the infrastructure framework, we recorded the anticipation time values for all objects that were initially detected in NLoS conditions, both static and dynamic, and subsequently passed to a visible state for the ego vehicle, across several simulations with traffic densities ranging from 10 to 45 obj/km.

Figure 3 depicts the Gaussian Kernel Density Estimation (KDE) of the resulting anticipation time distribution. The distribution is right-skewed with a median value of 1.3 seconds, indicating that in at least half of the observed cases, the infrastructure provides more than this amount of additional awareness.

This temporal margin is particularly relevant for hazard detection in urban driving scenarios. Moreover, the extended right tail of the distribution, reaching the 95th

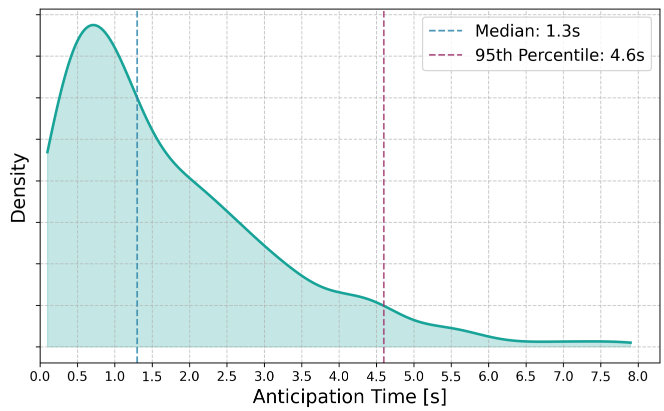

Fig. 3. Distribution of anticipation times estimated via Gaussian Kernel Density Estimation, showing the median and 95th percentile.

percentile at 4.6 seconds, highlights scenarios with substantially longer anticipation times, further underscoring the safety benefits of infrastructure-assisted perception in occlusion-prone environments.

B. Collective Perception and Semantic Evaluation

One of the primary goals of collective and collaborative perception is to expand the ODD of the ego vehicle, enhancing its understanding of the surrounding environment beyond the constraints of on-board sensors. Figure 4 illustrates this effect, showing how the expanded perception contributes to improved situational awareness.

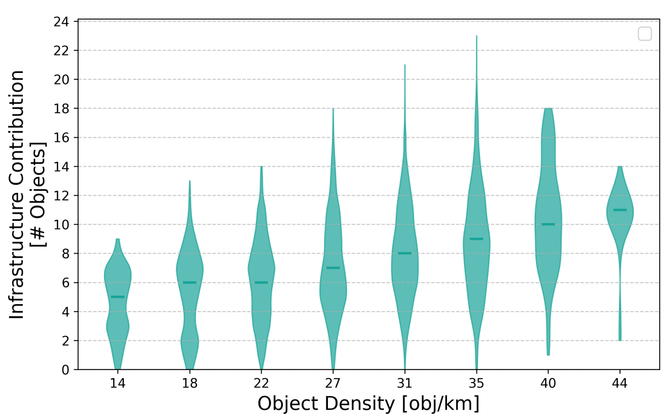

Fig. 4. Violin plots illustrating the infrastructure contribution, quantified as the number of additional objects perceived by the ego vehicle beyond its FOV, across increasing object densities.

In particular, the image shows the distribution of the infrastructure contribution to the ego vehicle's perception across different object density bins. The y-axis represents the number of additional objects detected and shared by the infrastructure that are not perceived by the ego vehicle, while the x-axis corresponds to the binned object densities in the environment, both in terms of vehicles and pedestrians. Each violin illustrates the full distribution of infrastructure contributions within a density bin, with the horizontal bar indicating the median value. The plot highlights a clear growth trend, showing that the infrastructure contribution generally increases with object density. At a density of 14 obj/km, the infrastructure contributes a median of $\sim$5 additional objects unseen by the ego; at 44 obj/km, this rises to $\sim$12, indicating a substantial ODD expansion in denser environments. The lower extremes likely reflect scenario-dependent configurations where limited occlusions or favorable ego-object positioning naturally reduce the infrastructure's contribution. Conversely, more complex spatial arrangements enable the infrastructure to provide a significantly higher number of additional detections. This plot emphasizes that infrastructure-assisted perception can significantly augment the ego vehicle's ODD, particularly under higher traffic density conditions.

Semantic communication aims to identify only the objects truly relevant to the ego vehicle, increasing information usefulness and preventing channel congestion. In our framework, perceived objects are first proximity-filtered and then semantically labeled, as described in Section III. Figure 5 presents the percentage of objects selected by the framework and subsequently labeled, relative to the total number perceived by the infrastructure sensors. It is important to specify that an object is considered labeled if at least one of the conditions presented in the framework is verified. The trends are shown as a function of object density on the x-axis. As shown by the purple line, approximately 40-50% of the perceived objects are selected and transmitted to the ego vehicle. This demonstrates that, even using only a proximity-based filter, a significant portion of non-relevant data can be excluded, reducing unnecessary communication load and processing effort by the receiver. The blue line represents the fraction of labeled objects over the total perceived. Only a limited subset, approximately 20% to 30%, is classified as potentially critical according to the rules defined in the framework pipeline. This observation indicates that the system can preserve safety-critical information, thereby avoiding high channel load and huge receiver computational effort.

VI. Conclusions

Collective perception and semantic communication are two of the main enablers for the evolution and success of Intelligent Transportation Systems. This work proposed

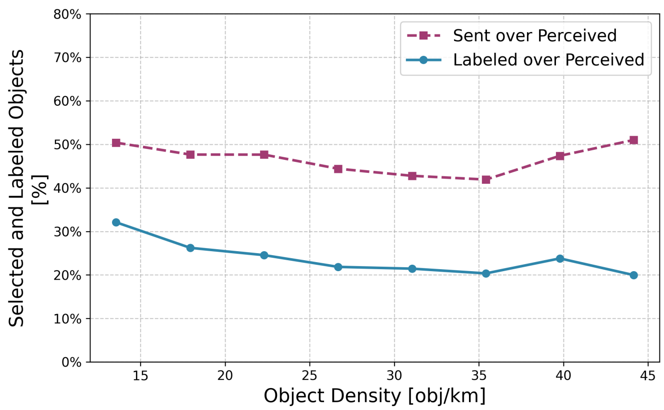

Fig. 5. Percentage of transmitted and labeled objects over the total perceived by the infrastructure, as a function of object density.

a multi-layer semantic perception framework to support infrastructure-assisted driving for an ego vehicle.

Results across varying object densities demonstrate that the framework can effectively enlarge the Operational Design Domain while semantically labeling the most relevant objects according to the ego vehicle's current state. The analysis highlights that the anticipation time gained for hidden objects through infrastructure sharing represents a significant safety enhancement, providing valuable additional reaction time in critical situations.

Importantly, the filtering function and semantic prioritization mechanisms show how environmental knowledge can be leveraged to avoid transmitting unnecessary information, particularly under high traffic load conditions and to reduce receiver workload.

VII. Future Works

Future work will extend the framework by considering varying V2X penetration rates among vehicles and pedestrians. Ablation studies, together with the adoption of different labels and parameters, will be performed to fine-tune each layer for diverse scenarios. Furthermore, the impact on the ego vehicle's driving performance will be analyzed in terms of total travel time, number of emergency events, and acceleration profiles.

Acknowledgment

This publication is part of the PNRR-NGEU project, funded by the Italian Ministry of University and Research (MUR) under Decree DM 117/2023.

References

[1] L. Wan, J. Zhao, A. Wiedholz, M. Bied, M. Martinez de Lucena, A. Dinkar Jagtap, A. Festag, A. Augusto Fröhlich, H. Ejaz Keen, and A. Vinel, "A systematic literature review on vehicular collaborative perception—a computer vision perspective," *IEEE Transactions on Intelligent Transportation Systems*, vol. 27, no. 1, p. 81–118, Jan. 2026. [Online]. Available: http://dx.doi.org/10.1109/TITS.2025.3631141

[2] ETSI, "Intelligent transport systems (its); vehicular communications; basic set of applications; collective perception servic; release 2," European Telecommunications Standards Institute (ETSI), Tech. Rep. TS 103 324 V2.1.1, 2023.

[3] Q. Delooz, A. Willecke, K. Garlichs, A.-C. Hagau, L. Wolf, A. Vinel, and A. Festag, "Analysis and evaluation of information redundancy mitigation for v2x collective perception," *IEEE Access*, vol. 10, pp. 47 076–47 093, 2022.

[4] F. Schiegg, I. Llatser, and T. Michalke, "Object detection probability for highly automated vehicles: An analytical sensor model," in *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems - Volume 1: VEHITS,*, INSTICC. SciTePress, 2019, pp. 223–231.

[5] M. Giordani, T. Higuchi, A. Zanella, O. Altintas, and M. Zorzi, "A framework to assess value of information in future vehicular networks," in *Proceedings of the 1st ACM MobiHoc Workshop on Technologies, MOdels, and Protocols for Cooperative Connected Cars*, ser. TOP-Cars '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 31–36. [Online]. Available: https://doi.org/10.1145/3331054.3331551

[6] V. A. Wolff, E. Xhoxhi, F. A. Schiegg, and J.-E. Neumann, "Uncertainty and prioritization: Empirical evaluation of a voi-based cpm generation pipeline using real-world data," in *2025 IEEE Vehicular Networking Conference (VNC)*, 2025, pp. 1–8.

[7] T. Lyu, S. Li, F. A. Schiegg, M. Noor-A-Rahim, D. Pesch, and A. O'Driscoll, "Accuracy and relevance: A value of information based prioritisation of perceived objects for the etsi collective perception service," in *2025 IEEE Vehicular Networking Conference (VNC)*, 2025, pp. 1–8.

[8] R. Tessier, O. Shagdar, and B. Monsuez, "Relevance-Aware Semantic Communication for Intelligent Collective Perception System," *International Journal On Advances in Software*, vol. 18, no. 3&4, Dec. 2025. [Online]. Available: https://hal.science/hal-05471632

[9] ——, "Context-aware collaborative perception: Estimating relevance through knowledge representation," in *Proceedings of the Fourteenth International Conference on Advances in Vehicular Systems, Technologies and Applications (VEHICULAR 2025).* Lisbon, Portugal: IARIA Press, 2025, pp. 1–5. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=vehicular_2025_1_10_30005

[10] M. A. Onsu, P. Lohan, B. Kantarci, A. Syed, M. Andrews, and S. Kennedy, "Spatiotemporal semantic v2x framework for cooperative collision prediction," 2026. [Online]. Available: https://arxiv.org/abs/2601.17216

[11] T. Higuchi, M. Giordani, A. Zanella, M. Zorzi, and O. Altintas, "Value-anticipating v2v communications for cooperative perception," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 1947–1952.

[12] L. Lusvarghi and J. Gozalvez, "The selection of relevance: Content-selection strategies in semantic and task-oriented v2x communications," in *2025 IEEE Future Networks World Forum (FNWF)*, 2025, pp. 1–6.

[13] F. A. Schiegg, I. Llatser, D. Bischoff, and G. Volk, "Collective perception: A safety perspective," *Sensors*, vol. 21, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/1/159

[14] 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 - Traffic Conference 2018," in *IEEE Intelligent Transportation Systems Conference (ITSC), 2018.* IEEE, 2018.

[15] 3GPP, "3GPP TR 36.885: Study on LTE-based V2X Services," 3rd Generation Partnership Project (3GPP), Technical Report 36.885, 2016, version 14.0.0. [Online]. Available: https://www.3gpp.org