Computing and Informatics, Vol. 45, 2026, 1075–1099, doi: 10.31577/cai_2026_5_1075

EARLY REDUNDANCY ASSESSMENT IN COOPERATIVE PERCEPTION USING A MODIFIED YOLO ARCHITECTURE

Nina Masarykova*, Matej Halinkovic, Marek Galinski, Peter Truchly

Institute of Computer Engineering and Applied Informatics
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovicova 2, 842 16 Bratislava, Slovakia
e-mail: nina.masarykova@stuba.sk

Abstract. Cooperative perception enables connected and autonomous vehicles to overcome the inherent limitations of individual sensors, such as occlusion and a restricted field of view, by sharing perceived objects among vehicles. While this enhances situational awareness, it introduces significant redundancy in dense traffic, leading to unnecessary computation and communication overhead. This paper proposes a novel redundancy-aware perception framework that integrates explicit object-level redundancy assessment directly into the detection stage. We extend the YOLOv11x architecture to ingest lightweight occupancy and inverse-distance masks derived from shared data, enabling the model to perform simultaneously 2D object detection and binary redundancy classification in a single inference step. On V2X-Sim, the proposed model achieves a redundancy F1-score of 0.972 with a measured forward-pass latency increase of 7.70 % (0.687 ms) relative to the RGB detector. By identifying redundant objects early, the framework enables resource-saving strategies, such as context-aware downsampling and selective offloading.

Keywords: Cooperative perception, object detection, redundancy mitigation, connected vehicles, autonomous driving

Mathematics Subject Classification 2010: 68T45, 68T07, 93C85

* Corresponding author

1 INTRODUCTION

Advanced Driver Assistance Systems (ADASs) have the potential to significantly improve road safety by reducing human error, a leading cause of traffic accidents. Using advanced artificial intelligence (AI) models and computer vision, ADASs can perceive and navigate their surroundings. The fusion of data from multiple sensors, such as RGB cameras, LiDAR, and radar, improves the overall situational awareness of a single vehicle [1]. However, processing such large-scale sensor data in real time may exceed the computational capabilities of individual vehicles or impose undesirable energy demands, particularly in electric vehicles. These constraints have motivated research into edge and cloud offloading as a potential solution for maintaining high-performance perception while alleviating on-board computational and energy constraints [2].

Although integrating data from multiple sensors improves depth perception and sensing under adverse conditions [3, 4], it cannot mitigate limitations such as occlusion, a limited field of view, or individual sensor failures. To further improve safety and efficiency, cooperative perception has emerged as a key solution, enabling connected vehicles to share their sensor data and collectively enhance the accuracy of their perception systems [5, 6]. Sharing perceived objects among connected participants significantly reduces the number of missed detections and allows vehicles to become aware of nearby objects or hazards sooner. However, in high-density traffic scenarios, many objects detected by one vehicle are redundantly perceived and processed by multiple others.

Redundancy in cooperative perception arises in multiple forms, including spatial redundancy, in which the same object is perceived from several viewpoints, and temporal redundancy, in which consecutive frames capture nearly identical information. While redundancy can provide additional context for object tracking and validation, excessive repetition burdens both the perception and communication subsystems. Some existing approaches address redundancy through heuristics, such as distance-based filtering or non-maximum suppression across shared detections, without explicitly learning contextual relationships among observations [7]. In our previous work, we explored lightweight frame filtration for perception offloading by identifying frames containing vehicles using a computationally inexpensive convolutional model [8]. However, that approach did not account for redundancy in multi-vehicle environments, in which the same objects may be perceived repeatedly by multiple agents.

This paper focuses on assessing redundancy rather than removing it directly. Our approach evaluates redundancy during the initial RGB detection stage, before subsequent multimodal processing or image offloading. By identifying redundant observations early, the ego vehicle can make informed decisions about which sensor data are worth processing or sharing, potentially reducing downstream computational and communication overhead.

To enable this, we introduce a modified YOLOv11x model that incorporates lightweight contextual information from connected vehicles in the form of two mask

channels that encode the spatial locations and approximate distances of shared objects. The proposed model simultaneously predicts 2D bounding boxes and performs binary redundancy classification within a single forward pass. This design allows the system to determine, for each detected object, whether it has already been observed by another connected vehicle. By shifting redundancy assessment to the perception stage, the method provides a practical means of optimizing computational and bandwidth use. It also offers a novel alternative to existing frameworks that do not explicitly reason about redundancy at detection time.

The rest of the paper is organized as follows. Section 2 reviews related work on cooperative perception and redundancy mitigation strategies. Section 3 describes the proposed method and the modifications made to the YOLOv11x model. Section 4 presents the evaluation; Section 5 discusses the results; and Section 6 concludes the paper and outlines directions for future research.

2 BACKGROUND

Several studies address redundancy mitigation in cooperative perception. The solution proposed by Lobato et al., called VILE, aims to mitigate data redundancy in collective perception systems. The mechanism divides the map into a honeycomb-like grid and uses a data-selection algorithm to designate one vehicle in each hexagonal region to share its perception with nearby vehicles. Simulation-based evaluation shows that VILE can reduce object redundancy by up to 85 % and the number of generated packets by up to 55 % while preserving the original number of uniquely detected objects. Importantly, the other agents in the hexagonal area are still required to perform full object detection independently [9].

Luo et al. focus on minimizing redundancy in shared intermediate features through contractive learning and mutual-information upper-bound estimation. Although their method improves object detection accuracy, feature filtration occurs after the features have been transmitted to connected participants and therefore does not reduce the required transmission bandwidth [10].

Delooz et al. compare several redundancy-mitigation rules, including distance-based, dynamics-based, frequency-based, and entropy-based rules, that are also subject to European standardization. Distance- and dynamics-based rules achieve promising results in terms of channel busy ratio, environmental awareness, and redundancy level. However, these rules are applied after the full detection pipeline when determining whether a perceived object should be included in a Collective Perception Message (CPM), and therefore cannot avoid the cost of local detection [7].

The Priority-Aware Collaborative Perception (PACP) framework proposed by Fang et al. uses a bird's-eye view (BEV) of the scene to optimize communication. It assigns higher sharing priority to data from connected vehicles that increase coverage of the perceived area and reduce blind spots, thereby reducing redundant data transmission by up to 13.6 % [11].

Recent intermediate-feature methods address communication redundancy before collaborative fusion. Where2comm constructs spatial confidence maps from locally extracted features and communicates only spatially sparse regions considered important for perception. The selected regions can be adjusted according to the available communication bandwidth, and the received features are subsequently fused for collaborative object detection [12]. How2comm also targets the trade-off between detection performance and communication volume but uses a mutual-information-aware mechanism with spatial-channel filtering to retain informative intermediate features. It additionally employs flow-guided delay compensation and a collaboration transformer to address temporal misalignment and integrate information from multiple agents [13].

Where2comm and How2comm therefore determine which intermediate features should be transmitted and fused to improve collaborative detection under bandwidth constraints. They do not formulate redundancy as an explicit per-object prediction indicating whether the same physical object has already been observed by a neighbouring vehicle. In contrast, our method uses compact masks derived from shared object information and jointly predicts object detections and binary redundancy labels during the ego vehicle's RGB detection pass. These labels are intended to guide subsequent decisions, such as whether an image should be offloaded or whether more computationally intensive downstream processing is required.

The reviewed methods consequently address complementary stages and representations: standardized rule-based methods filter object messages after local detection, intermediate-feature methods reduce the feature payload exchanged for collaborative detection, and the proposed approach assesses object-level redundancy as part of detection. To the best of our knowledge, no prior cooperative perception method jointly predicts detections and explicit per-object redundancy labels within a single detector forward pass. The proposed model addresses this gap through a dedicated redundancy-classification head conditioned on shared contextual information.

3 PROPOSED SOLUTION

The key idea behind the proposed cooperative-perception-based frame-filtration strategy is that, in environments with a sufficient density of connected vehicles, many objects observed by the ego vehicle have already been detected and reported by neighbouring participants. After appropriate coordinate transformation and post-processing, the ego vehicle can use these shared detections to reduce the need to repeatedly execute computationally expensive multimodal perception models. This requires the ego vehicle to identify whether another connected vehicle has already observed each perceived object.

To achieve this, we extend the YOLOv11x detection model by incorporating additional channels that represent masks of objects perceived and shared by nearby vehicles. This modification allows the model to simultaneously perform 2D object

detection and binary redundancy classification using RGB data and lightweight contextual information, thereby avoiding the complexity and high computational cost of depth-aware or multimodal detectors. Based on the model's redundancy predictions, the ego vehicle can determine whether a given frame contains any previously unseen objects.

If all detected objects in a frame are classified as redundant, the frame becomes a candidate for omission from further processing by more advanced models at the edge or in the cloud, potentially saving computational resources. Conversely, if any non-redundant objects are detected, the frame is transmitted for further processing. The image can nevertheless be selectively downsampled using the context-aware strategy proposed in [14] to reduce bandwidth consumption while preserving relevant regions. This redundancy-aware approach can guide the allocation of both on-board computational and communication resources.

This section describes the main components of the redundancy-aware filtration framework. First, we introduce the V2X-Sim dataset and the filtering steps and assumptions used to generate ground-truth labels for the multi-agent task. We then quantify the prevalence of redundant observations to assess the opportunity for resource savings in dense cooperative scenarios. Next, we describe the coordinate transformations required to project shared 3D bounding boxes from neighbouring vehicles onto the ego vehicle's 2D image plane. This information is converted into two compact representations of shared objects: an inverse-distance mask and an occupancy mask. Finally, we explain how the standard YOLOv11x architecture is adapted with a 5-channel input and a dedicated redundancy-prediction head to perform simultaneous object detection and redundancy classification.

3.1 Dataset

The proposed method is developed and evaluated using the V2X-Sim dataset [15], which contains 100 cooperative driving scenes simulated in CARLA. Each scene captures 20 seconds of multi-agent interaction at 5 frames per second, resulting in a temporal resolution of 200 ms between frames. The dataset follows the NuScenes [16] data structure and includes a comprehensive sensor suite consisting of six RGB cameras, one LiDAR sensor, an IMU, and a GNSS receiver. At each timestamp, synchronized sensor data and annotations are provided for five participating vehicles (agents) operating within the shared environment.

To emulate realistic communication between connected vehicles, including the effects of latency, we treat objects perceived by other agents at timestamp $t-1$ as the shared information available to the ego vehicle at timestamp t. This corresponds to a temporal delay of 200 ms. For simplicity and to ensure consistent supervision during training, we use ground-truth annotations as a proxy for detections shared by neighbouring agents.

To construct redundancy supervision and generate the corresponding redundancy masks, object identity across agents is established using the `instance_token` provided by the V2X-Sim dataset. This token uniquely identifies a physical ob-

ject over time and across all agents within a scene. An object detected by the ego vehicle is labelled as *redundant* if another agent reports an object with the same `instance_token` at timestamp $t-1$; otherwise, it is labelled as *non-redundant*. This mechanism defines redundancy precisely at the object-instance level, independently of spatial-overlap heuristics, and enables consistent generation of the redundancy masks used as input to the proposed model.

Because the annotations describe all objects present in the global CARLA scene at each time step, they must be filtered to identify those that are visible to a given agent. Following standard practice in the literature, object visibility is determined by counting the LiDAR points associated with each annotated 3D bounding box. Only objects with at least five LiDAR points are considered as visible to an agent. This filtering ensures that the shared detections reflect realistic perceptual capabilities rather than omniscient scene-level labels. Although each vehicle is equipped with six cameras, only images from the three forward-facing cameras are used because they contain most of the objects relevant to the detection task.

Finally, because the dataset provides object annotations as 3D bounding boxes, whereas the YOLO-based model operates on 2D images, all labels are projected onto the image plane. Each 3D bounding box is converted to a 2D rectangular bounding box by projecting its corner points using camera calibration parameters and then taking the minimum and maximum image-plane coordinates along both axes.

Despite its richness and multi-agent structure, the V2X-Sim dataset has inherent limitations that must be considered when interpreting the results. First, all scenes are generated in the CARLA simulator, which, although highly configurable and photorealistic, cannot perfectly replicate the full complexity of real-world driving. Second, the visibility-filtering strategy based on LiDAR point counts, although widely adopted, is imperfect. Objects may be excluded if they receive too few LiDAR returns, especially small, distant, or highly reflective objects, despite being clearly visible in the camera images. Conversely, LiDAR-based visibility does not fully reflect camera visibility because the two sensors have different occlusion patterns. Consequently, the process of determining which objects are visible to each agent may introduce noise into the training labels.

3.2 Redundancy Characteristics of the Dataset

To quantify perceptual redundancy, we evaluate new and redundant objects in the V2X-Sim dataset for one to four neighbouring connected vehicles. The analysis includes 29 700 camera images and 145 677 annotated object appearances. An image is considered fully redundant when it contains no non-redundant objects. The three image views acquired at the same instant are grouped into a synchronized timestamp, yielding 9 900 timestamps. A timestamp is considered fully redundant only when all three images are fully redundant.

As the number of neighbouring vehicles increases, the redundant-object ratio rises from 43 % to 68 %, while the proportion of fully redundant images increases from 20.55 % to 44.64 %. The stricter timestamp-level ratio increases from 0.86 %

to 10.44 %. Table 1 summarizes these results, while Figure 1 shows the object- and image-level statistics.

This high level of redundancy presents an opportunity for computational and bandwidth savings. However, it also creates a highly imbalanced learning signal if used directly for supervision. Therefore, during model training, we consider only a single connected vehicle, which yields a more balanced ratio of new to redundant objects and facilitates stable supervised learning.

#CVs	New	Redundant	Redundant Ratio	Fully Redund. Images	Fully Redund. Timestamps
1	82 889	62 788	0.43	6 104 (20.55 %)	85 (0.86 %)
2	57 920	87 757	0.60	10 559 (35.55 %)	591 (5.97 %)
3	49 560	96 117	0.66	12 611 (42.46 %)	924 (9.33 %)
4	46 925	98 752	0.68	13 258 (44.64 %)	1 034 (10.44 %)

Table 1. Redundancy statistics for one to four neighbouring connected vehicles

These observations support the motivation for the proposed redundancy-aware perception framework. As the number of connected vehicles increases, a large fraction of perceptual information becomes redundant. Identifying this redundancy early can help vehicles avoid unnecessary downstream computation, reduce upload volume, and prioritize the processing of novel and safety-critical objects.

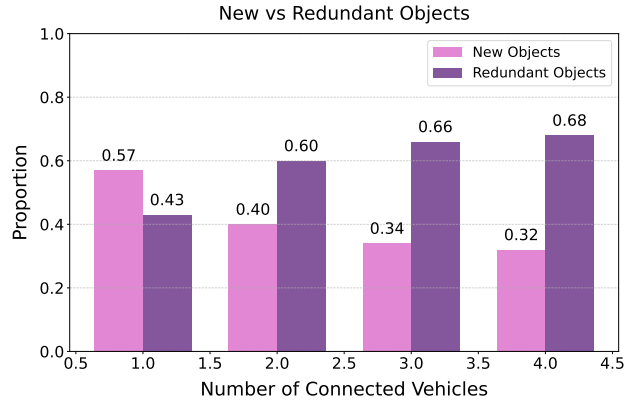
3.3 Coordinate System and Projections

A crucial component of the proposed redundancy assessment framework is the ability to transform object detections received from connected vehicles into the coordinate system of the ego vehicle's camera. Accurate coordinate alignment ensures that shared objects projected onto the ego vehicle's image plane correspond to their correct visual location, enabling the model to correctly interpret the contextual mask information.

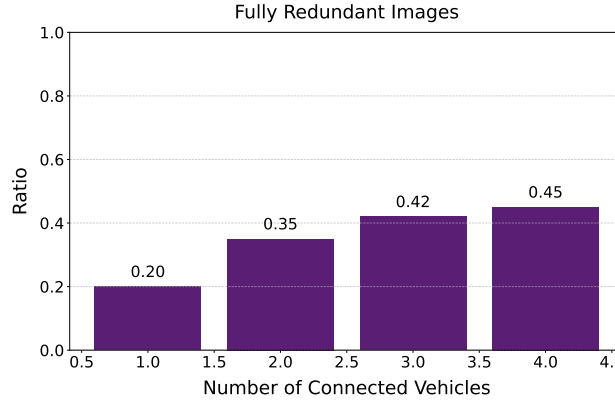
For dataset preparation, these transformations are handled using the nuScenes devkit, which provides utilities for converting 3D bounding boxes between global, ego, and camera coordinate frames. In real-world cooperative perception systems, detections transmitted by connected vehicles are typically expressed in a global reference frame, including the object centre coordinates $\mathbf{p}_{\text{global}} \in \mathbb{R}^3$, dimensions, and orientation. The ego vehicle must therefore convert these global-frame detections into its camera coordinate frame using its own pose and the camera's intrinsic and extrinsic parameters.

Assuming $\mathbf{R}_{\text{ego}} \in \mathbb{R}^{3\times3}$ and $\mathbf{t}_{\text{ego}} \in \mathbb{R}^3$ denote the rotation and translation of the ego frame expressed in the global frame, and $\mathbf{R}_{\text{cs}} \in \mathbb{R}^{3\times3}$ and $\mathbf{t}_{\text{cs}} \in \mathbb{R}^3$ denote the rotation and translation of the camera frame expressed in the ego frame, the transformation of a point from global coordinates to the camera frame is

$$\mathbf{p}_{\text{cam}} = \mathbf{R}_{\text{cs}}^{\top}\left(\mathbf{R}_{\text{ego}}^{\top}\left(\mathbf{p}_{\text{global}} - \mathbf{t}_{\text{ego}}\right) - \mathbf{t}_{\text{cs}}\right), \tag{1}$$



a) Normalized proportion of new vs. redundant objects



b) Fraction of fully redundant images

Figure 1. Statistical analysis of redundancy in the dataset for different numbers of connected vehicles. Figure 1 a) shows the normalized proportion of newly observed and redundant objects as the number of connected vehicles increases, while Figure 1 b) presents the fraction of images that contain only redundant objects.

where $\mathbf{p}_{\text{cam}} \in \mathbb{R}^3$ is the point expressed in the camera frame. This follows the inversion of the global $\rightarrow$ ego and ego $\rightarrow$ camera transforms used in nuScenes, where $\mathbf{R}^\top$ corresponds to the inverse rotation.

Orientation must be transformed consistently across frames. Using unit quaternions to represent rotations, the object's orientation expressed in the camera frame is obtained as

$$\mathbf{q}_{\text{cam}} = \mathbf{q}_{\text{cs}}^{-1} \otimes \left(\mathbf{q}_{\text{ego}}^{-1} \otimes \mathbf{q}_{\text{global}} \right), \tag{2}$$

where $\mathbf{q}_{\text{global}}$ encodes the object's orientation in the global frame, $\mathbf{q}_{\text{ego}}$ represents the ego vehicle's orientation in the global frame, and $\mathbf{q}_{\text{cs}}$ represents the camera orientation relative to the ego frame. Here, $\otimes$ denotes quaternion multiplication.

3.4 Masks of Shared Objects

After projecting the 3D bounding boxes into the 2D image plane, the next step is to convert the shared detections from connected vehicles into mask-based representations that can be ingested by the modified YOLOv11x model. Instead of allocating a fixed number of per-object mask channels, which scales poorly and introduces sparsity, the final design employs two compact aggregated masks that summarize the spatial presence and proximity of all shared objects.

First, a binary mask is generated for each projected object by rasterizing the 2D polygon obtained from its transformed 3D bounding box. These per-object masks are then combined to form two complementary representations. The *inverse-distance mask* is constructed by assigning each pixel the inverse-normalized distance of the closest object covering that pixel. Distances are logarithmically scaled into the range $[0, 1]$ using predefined minimum and maximum bounds ($d_{\min} = 2\,\text{m}$, $d_{\max} = 120\,\text{m}$), ensuring that nearby objects are emphasized.

The second representation, the *occupancy mask*, encodes object overlap by computing the number of projected objects covering each pixel $p_{i,j}$ and mapping this count through a smooth saturation function:

$$p_{i,j} = 1 - \exp(-k_{\text{occ}} c), \tag{3}$$

where c is the pixel-wise object count and k_{occ} is a scaling factor that controls how rapidly the occupancy value saturates as overlap increases. In our implementation, k_{occ} is set to 1.

To create spatially coherent masks and account for small projection errors, the occupancy masks are refined using constant-radius morphological dilation (radius $r = 4\,\text{pixels}$), which expands object regions in a controlled manner. This is followed by Gaussian blurring with a standard deviation of $\sigma = r/2$ to soften mask edges and improve continuity. These dilation and blurring steps also help compensate for the one-frame communication latency inherent in the cooperative perception setup. Because shared detections originate from timestamp $t - 1$, their projected masks may be slightly misaligned with the ego vehicle's view at timestamp t. The softened mask boundaries reduce the model's sensitivity to spatial discrepancies caused by object motion between frames, thereby improving robustness during training.

The resulting two-channel mask representation provides the model with compact yet expressive contextual information that captures both the spatial extent and relative proximity of objects detected by connected vehicles. An example of the resulting RGB image and corresponding masks is shown in Figure 2.



a) RGB image b) Inverse-distance mask c) Occupancy mask

Figure 2. RGB camera image and corresponding masks. Figure 2 a) shows the original RGB image, Figure 2 b) shows the inverse-distance mask, and Figure 2 c) shows the occupancy mask used as an additional input to the model.

3.5 Model Modification

The proposed framework, shown in Figure 3, extends the original YOLOv11x architecture from Ultralytics [17] to jointly perform 2D object detection and redundancy assessment using cooperative perception priors. While the baseline pretrained model operates on 3-channel RGB images and predicts 80 object classes, the modified architecture incorporates additional geometric context from connected vehicles and introduces a dedicated redundancy-classification head, enabling multitask learning within a single inference pass. The detection head is configured for the four classes present in the dataset: vehicles, bicycles, emergency vehicles, and motorcycles.

3.5.1 Extended Multi-Channel Input Representation

To integrate information shared by connected vehicles, the standard RGB input is augmented with the two previously described mask representations: an occupancy mask and a normalized inverse-distance mask. These form a 5-channel input tensor that combines appearance cues with spatial priors derived from cooperative perception. All data-augmentation operations used in YOLOv11x, such as Mosaic, LetterBox, RandomPerspective, horizontal flips, and colour transformations, are extended to operate consistently on all five channels. This includes applying identical geometric transformations across channels, enforcing appropriate padding for sparse mask inputs, and ensuring that the interpolation strategies preserve the semantic meaning of the mask values.

3.5.2 Dual-Branch Feature Extraction with Mid-Level Fusion

Because the mask channels encode geometric and distance-related structure rather than visual texture, directly concatenating them with RGB at the network stem would be suboptimal. To address this, the modified architecture employs a dual-branch feature extraction design. The RGB channels are fed through the original YOLOv11x backbone stem, which is optimized for appearance-based feature learning, while the two mask channels pass through a lightweight parallel convolutional

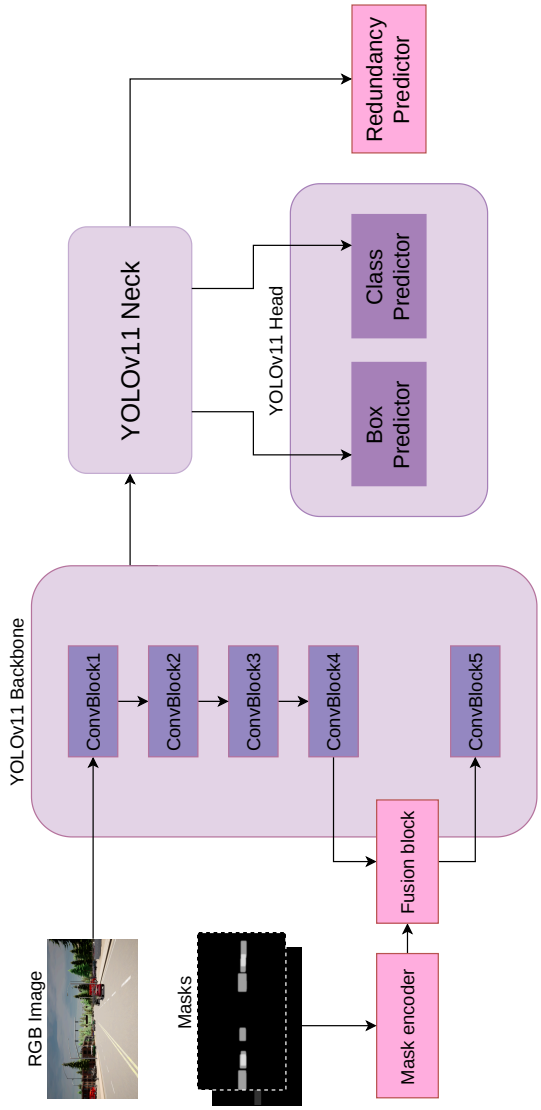


Figure 3. Architecture of the proposed redundancy-aware detection model. The backbone and detection components are based on the original YOLOv11x architecture and are shown in purple. The proposed modifications, shown in pink, introduce additional processing branches for the inverse-distance and occupancy masks and a redundancy-prediction head. These additions enable the model to jointly perform object detection and redundancy estimation from multimodal inputs.

branch. The resulting feature maps are fused midway through the backbone via channel-wise concatenation. This mid-level fusion strategy ensures that early layers, which specialize in fine-grained texture extraction, are not disrupted by the mask inputs, while allowing deeper layers to integrate high-level geometric context from cooperative perception.

3.5.3 Preservation of the Original Detection Head

The original YOLOv11x detection head is preserved without modification, maintaining its bounding box regression and 4-class classification outputs (*vehicle*, *bicycle*, *emergency vehicle*, *motorcycle*). Retaining the unaltered detection head ensures compatibility with standard downstream perception modules and preserves the model's ability to perform accurate object detection while introducing additional redundancy reasoning.

3.5.4 Redundancy Prediction Head

To assess whether an object perceived by the ego vehicle is redundant, i.e., already observed and shared by a connected vehicle, an additional prediction head is introduced. This head outputs a binary redundancy logit for each anchor and reuses the same anchor assignments as the bounding box regression and classification branches. By sharing early and mid-level features with the detection task, the model benefits from multi-task learning, enabling joint optimization of detection and redundancy assessment. Ground-truth annotations are extended with a sixth label field that specifies the redundancy status of each object and directly supervises this new head.

3.5.5 Training Procedure Modifications

The training pipeline is extended to jointly optimize detection and redundancy estimation. Redundancy labels are incorporated into the same ground-truth–to–anchor assignment process used for bounding box regression and classification. This aligns the object-level redundancy supervision with the detections, preserving consistency across prediction branches. After anchor assignment, the redundancy ground-truth values are aligned with the matched anchor indices on a per-image basis, enabling precise supervision of the redundancy logit for each positive anchor.

The redundancy head is trained using a weighted binary cross-entropy loss. Positive anchors receive their matched ground-truth redundancy labels and are weighted by their detection target scores. Background anchors receive a non-redundant target with a small weight of 0.05 to prevent the model from collapsing to a solution that predicts all objects as non-redundant. The weighting scheme accounts for background anchors and keeps the redundancy task balanced when the number of redundant objects varies significantly across frames. All other YOLOv11x loss components remain unchanged. The final training objective is the sum of the standard

YOLOv11x losses and the redundancy loss, enabling unified optimization of detection accuracy and redundancy classification within a single framework.

4 EVALUATION

This section evaluates the effectiveness of the proposed redundancy-aware detection framework on the V2X-Sim dataset. The dataset is partitioned into training, validation, and test subsets using a scene-wise 80 %/10 %/10 % split, such that no scene appears in more than one subset. We report both standard object-detection performance and the model's ability to classify objects as redundant or non-redundant based on cooperative perception masks. Here, mAP denotes mean average precision over IoU thresholds from 0.50 to 0.95 in steps of 0.05, whereas mAP_{50} uses an IoU threshold of 0.50.

4.1 Object Detection Performance

Table 2 summarizes the detection performance of the modified YOLOv11x model. Despite the architectural changes introduced to incorporate cooperative perception priors, the model maintains strong detection performance, achieving a precision of 0.781 and a recall of 0.734. The mAP_{50} and mAP values are 0.803 and 0.687, respectively. The comparison with the RGB-only baseline below shows comparable overall mAP with a different precision–recall trade-off.

Precision	Recall	mAP_{50}	mAP
0.781	0.734	0.803	0.687

Table 2. Object detection performance

4.2 Redundancy Classification Performance

The redundancy-classification head achieves strong aggregate performance on the test set. As shown in Table 3, the model achieves an accuracy of 0.969, with a redundancy precision of 0.952 and a notably high recall of 0.994. The resulting F1-score of 0.972 indicates that the model can reliably differentiate between new and redundant observations, effectively leveraging the geometric and distance-based cues encoded in the mask representations. The false-positive rate (FPR), the fraction of non-redundant objects incorrectly classified as redundant, is 0.060.

Accuracy	Precision	Recall	F1-score	FPR
0.969	0.952	0.994	0.972	0.060

Table 3. Redundancy evaluation metrics with the default 0.5 classification threshold

From a safety perspective, however, the relative importance of prediction errors is asymmetric. A false-positive redundancy prediction (i.e., predicting a non-redundant object as redundant) is substantially more harmful than a false negative because it may cause the system to suppress a previously unseen object and prevent it from being forwarded to downstream perception modules. This could create a critical hazard. In contrast, false negatives only reduce the achievable computational or communication savings under the proposed filtering policy.

To reduce false-positive redundancy predictions, the redundancy classifier's decision threshold was adjusted to favour the correct detection of *non-redundant* objects. The threshold of 0.72 was selected by analysing the receiver operating characteristic (ROC) curve and identifying an operating point with a lower false-positive rate (FPR) while retaining a high F1-score. This reflects a safety-oriented trade-off that prioritizes low FPR over maximal recall of redundant objects.

Table 4 reports the resulting performance with the threshold set to 0.72. The false-positive rate decreases from 0.060 to 0.017, representing an approximately 3.5-fold reduction, while the overall accuracy and F1-score remain high. This conservative bias classifies more detections as non-redundant, reducing the risk of suppressing novel objects.

Accuracy	Precision	Recall	F1-score	FPR
0.969	0.985	0.948	0.966	0.017

Table 4. Redundancy evaluation metrics with the safety-oriented threshold of 0.72

4.3 Comparison with an Occupancy-Based Rule

To compare the learned redundancy head with a simple alternative, we combine the RGB-only detector with an occupancy-based rule. For each predicted box, clipped to the image boundaries, the rule computes the mean value of the occupancy mask within that box and classifies it as redundant when the mean is at least a threshold τ. The rule uses the same occupancy representation available to the proposed model but requires no additional training. The threshold $\tau = 0.279$ is selected by maximizing redundancy F1 on the validation set over a grid in increments of 0.001 and is then fixed for test evaluation. The detection-confidence and non-maximum suppression IoU thresholds are 0.25 and 0.45, respectively, matching the RGB baseline evaluation.

Table 5 compares the resulting test metrics with the proposed model at its default threshold of 0.5. The rule-based comparator uses class-aware, one-to-one matches at IoU ≥ 0.5; unmatched predictions and missed ground-truth objects are excluded from redundancy classification. Each method is evaluated on its own matched detections. The proposed model has a higher redundancy F1-score (0.972 versus 0.942) and a lower FPR (0.060 versus 0.111), supporting learned redundancy assessment over this occupancy-only rule under the evaluated conditions.

Method	Accuracy	Precision	Recall	F1-score	FPR
RGB + occupancy rule	0.937	0.906	0.981	0.942	0.111
Proposed	0.969	0.952	0.994	0.972	0.060

Table 5. Test-set redundancy classification: occupancy-based rule and proposed model

4.4 Ablation Studies

4.4.1 Effect of Cooperative Masks on Detection

To quantify the impact of the proposed mask-based cooperative priors on detection accuracy, we trained a baseline YOLOv11x model on the same dataset using only RGB input and the standard detection loss. The same RGB-detector settings are used by the occupancy-based comparator in Section 4.3. Table 6 compares test-set detection performance between the baseline and the proposed redundancy-aware model.

Model	Precision	Recall	mAP_{50}	mAP
Baseline (RGB only)	**0.825**	0.691	0.789	**0.692**
Proposed (RGB + masks)	0.781	**0.734**	**0.803**	0.687

Table 6. Detection performance on the test set: baseline vs. redundancy-aware model

The proposed model exhibits a slightly different precision–recall trade-off: recall increases from 0.691 to 0.734, while precision decreases from 0.825 to 0.781. The mAP_{50} improves marginally, whereas the overall mAP remains comparable. These results indicate comparable overall detection performance while enabling redundancy estimation within a single inference pass. The differences are descriptive results from the reported runs; statistical significance across training runs has not been established.

4.4.2 Feature Fusion Strategy

To evaluate the importance of the proposed dual-branch architecture, we conduct an ablation study in which the redundancy-mask branch is removed. Instead, the inverse-distance and occupancy masks are concatenated directly with the RGB image at the input, resulting in a 5-channel tensor that is processed by the standard YOLOv11 backbone.

Table 7 reports the results. Compared to the proposed mid-level fusion design, direct mask concatenation leads to a decrease in detection performance, particularly in recall and overall mAP. Redundancy classification performance remains comparable across both configurations, with direct mask concatenation achieving an accuracy of 0.974, precision of 0.959, recall of 0.995, and an F1-score of 0.977.

Redundancy classification remains comparable between the two configurations. However, the lower detection performance with direct concatenation supports the dual-branch input architecture in this experiment.

Method	Params	Precision	Recall	mAP_{50}	mAP
Proposed	58.8 M	0.781	**0.734**	**0.803**	**0.687**
Concatenation	58.1 M	**0.789**	0.654	0.763	0.652

Table 7. Ablation study on feature fusion strategy

4.5 Computational Complexity and Runtime

Computational profiling and runtime measurements are performed in FP32 using PyTorch 2.9.1+cu128 (CUDA 12.8) on an NVIDIA GeForce RTX 4090 GPU. The stored 450×800 images are resized with preserved aspect ratio and padded to 384×672 under the input-size setting of 640. All three architectures are profiled at this resolution with a batch size of one.

Because learned weight values do not change the inference structure, latency is measured using randomly initialized models and fixed synthetic inputs already resident on the GPU. Each of three rounds comprises 30 warm-up passes followed by 100 timed forward passes, with CUDA synchronization around each measurement. Table 8 reports the median over the resulting 300 measurements per model. The measurements cover the model forward pass and exclude image preprocessing, mask generation, data transfer, and non-maximum suppression.

Model	Parameters (M)	Est. GFLOPs	Median (ms)
RGB baseline	56.878	123.145	8.922
Direct concatenation	58.099	127.016	9.360
Proposed fusion	58.767	128.279	9.609

Table 8. Model complexity and forward-pass latency at 384×672, batch size one

Relative to the RGB baseline, the proposed architecture increases the parameter count by 3.32 % and estimated FLOPs by 4.17 %. Its additional median latency is 0.687 ms per image, corresponding to a 7.70 % increase on the reference GPU. Direct concatenation increases estimated FLOPs by 3.14 % and median latency by 4.92 %. These results quantify the cost of the architectural additions; net resource savings depend on the downstream processing and communication avoided.

5 DISCUSSION

The results demonstrate that the proposed model can accurately detect objects while simultaneously assessing redundancy based on cooperative perception cues. However, several important considerations arise when interpreting its performance, particularly with respect to safety, dataset characteristics, and the inherent limitations of the V2X-Sim perception setup.

5.1 Safety-Oriented Redundancy Assessment

As highlighted in Section 4, the redundancy classifier achieves high accuracy and an F1-score above 0.97. Nevertheless, the relative importance of error types differs considerably in safety-critical domains. A false-positive redundancy prediction, incorrectly labelling a non-redundant object as redundant, may prevent the object from being forwarded to downstream perception modules, potentially causing the system to overlook a previously unseen obstacle. In contrast, false negatives primarily result in reduced computational savings under the proposed filtering policy.

To mitigate this risk, adjusting the redundancy decision threshold offers a practical and effective solution. A higher threshold biases the classifier toward treating uncertain cases as non-redundant, thereby reducing the likelihood of suppressing novel objects. This trade-off is evident in the evaluation with a threshold of 0.72, which reduces the false-positive rate from 0.06 to 0.017 while maintaining strong overall performance. Such threshold tuning provides a configurable safety mechanism that enables deployment-dependent calibration.

5.2 Dataset Limitations

While V2X-Sim provides rich cooperative multi-agent annotations, its visibility-filtering mechanism introduces noise into the ground-truth labels. Because the filtering relies on LiDAR point counts to determine object visibility, some objects that are clearly visible in the RGB images may be excluded from the dataset labels if they receive insufficient LiDAR returns. This leads to situations in which the model correctly detects an object, but the annotation treats it as background, thereby artificially lowering the measured precision.

The opposite scenario also occurs. Certain objects, particularly vehicles for which only the roof is visible because of occlusion, may pass the LiDAR visibility threshold and thus appear in the ground-truth annotations, even though they are not meaningfully visible in the camera image. In such cases, the detector may fail to predict these objects.

The evaluation is limited to V2X-Sim and uses ground-truth annotations from neighbouring vehicles with a fixed 200 ms delay. This setup does not capture the full effects of missed detections or false positives from neighbouring vehicles, localization errors, variable latency, or packet loss. Because V2X-Sim is a simulated dataset, the present experiments also do not evaluate robustness to the full range of road, lighting, weather, and connectivity conditions encountered in real-world driving [18]. Generalization to real-world cooperative perception therefore remains to be evaluated.

5.3 Impact of Architectural Design Choices

The ablation studies provide insight into how cooperative perception priors should be integrated into real-time object detectors. Directly concatenating the redun-

dancy masks with RGB input degrades detection performance, particularly recall, suggesting that sparse geometric cues interfere with early appearance-driven feature extraction. In contrast, the proposed dual-branch architecture enables modality-specific feature learning and preserves detection accuracy through mid-level fusion.

Although redundancy classification performance remains high even with early fusion, object detection reliability is a primary requirement for safety-critical perception systems. Since missed detections can lead to hazardous behaviour, maintaining detection performance while enabling redundancy reasoning is essential, and the proposed fusion strategy achieves this balance.

Redundancy classification is evaluated only on detections matched to annotated objects at IoU ≥ 0.5. As a result, configurations with lower detection recall may implicitly bias redundancy metrics by excluding hard or borderline objects from evaluation. In particular, the direct concatenation variant detects fewer objects overall, which may partially inflate redundancy classification scores by limiting evaluation to easier detections.

These findings support architectural separation and mid-level fusion for incorporating cooperative perception cues while retaining comparable overall detection performance.

5.4 Practical Implications and Use Cases

The ability to determine whether an object has already been perceived by connected vehicles enables several practical applications in cooperative perception systems, particularly those operating under communication or computation constraints. The following use cases (UC) illustrate potential integrations into V2X pipelines; their end-to-end resource savings are not evaluated in this work.

UC1: Camera-only ego vehicle: In scenarios where the ego vehicle is equipped only with an RGB camera, redundancy predictions can serve as an efficient and lightweight mechanism to identify objects that are novel from the perspective of other agents. Even without depth sensors, the inverse-distance mask provides a coarse spatial prior that can be further refined through monocular depth estimation. The ego vehicle may then transmit a minimal alert message indicating the presence and approximate location of newly detected objects, improving situational awareness for connected vehicles whose line of sight may be occluded.

UC2: Selective offloading with camera–LiDAR perception: When the ego vehicle has both camera and LiDAR sensors, redundancy estimation can guide selective offloading strategies. New objects, those not detected by neighbouring vehicles, can be forwarded for full-resolution processing at an edge server or remote computing unit, while redundant observations can be compressed or downsampled. Techniques such as context-aware spatial downsampling [14] can

reduce the resolution of image regions containing only redundant objects, potentially lowering bandwidth and computational requirements while prioritizing critical regions.

UC3: Temporal LiDAR refinement instead of repeated full inference: If the current frame contains only redundant objects, the redundancy decision can be used as a gate for a lightweight temporal-refinement branch instead of immediately invoking another full 3D detection or offloading cycle. Our previous work [19] demonstrates a closely related mechanism in which previous-frame 3D bounding boxes are propagated using object velocity, an enlarged region of interest is extracted from the current LiDAR scan, and the box centre and yaw are refined geometrically. On nuScenes, this procedure achieved a mean BEV IoU of 0.665; 75.74 % of the refined boxes exceeded IoU ≥ 0.5, and 41.77 % exceeded IoU ≥ 0.75, with an average CPU runtime of 0.016 s per object. In the proposed framework, the redundancy classifier could therefore provide the trigger for identifying candidate frames for refinement-only processing, while any detected non-redundant object would force the full downstream perception or offloading path. These values should be viewed as evidence of feasibility rather than expected end-to-end savings because the refinement study used annotations as a proxy for detector outputs and did not quantify end-to-end communication savings.

UC4: Prioritization in collaborative fusion: In cooperative multi-agent fusion pipelines, redundancy-aware labels can be used to prioritize processing on non-redundant observations. For example, edge servers can allocate more computational resources to novel objects, or fuse only non-redundant detections first to reduce overall latency. This is particularly beneficial in time-critical scenarios where processing budgets are limited.

UC5: Integration with predictive scheduling and energy-aware perception: In electric or hybrid vehicles, computation incurs a measurable energy cost. By identifying frames containing only redundant objects, the vehicle could reduce the activity of downstream perception hardware for short intervals, subject to validation of the resulting perception performance. Such predictive-scheduling strategies could improve energy efficiency during long-duration autonomous driving.

Overall, the proposed redundancy-evaluation framework provides a foundation for a broad range of cooperative perception optimizations, supporting potential communication and computation savings through the prioritization of newly observed objects.

5.5 Potential Savings in Selective Offloading

To quantify the opportunity for selective offloading, we use the fully redundant image and timestamp percentages obtained from the V2X-Sim redundancy analysis in

Section 3.2. We consider a reference system that uploads all three forward-facing camera images and one LiDAR scan at every timestamp. A redundancy-aware system can omit an observation when it contains no non-redundant objects. The dataset analysis supports two selection granularities: individual camera images and synchronized timestamps. The resulting percentages characterize the prevalence of redundant observations in the dataset and are used below to estimate the corresponding potential reduction in upload traffic.

At the image level, each camera view can be selected independently. The idealized reduction in the number of uploaded images for k neighbouring vehicles is

$$p_k^{\text{img}} = \frac{N_{\text{fully redundant images}}(k)}{N_{\text{images}}}. \tag{4}$$

For an image size B_{img} in bytes, three camera views, and a sampling rate of $f = 5\,\text{Hz}$, the corresponding reference image-upload rate is

$$R_{\text{img}} = \frac{3f\,8B_{\text{img}}}{10^6}, \quad R_{\text{img,saved}}(k) = p_k^{\text{img}} R_{\text{img}}, \tag{5}$$

where the rates are expressed in Mbit/s. An uncompressed 800×450 8-bit RGB image contains 1.080 MB, resulting in a reference rate of 129.60 Mbit/s. To provide a compressed reference, the RGB image was encoded as JPEG at quality 90. The encoded image has a size of approximately 0.104 MB, corresponding to 12.45 Mbit/s for the three-view stream. Table 9 reports the resulting idealized rate reductions. The uncompressed values follow directly from the fixed image dimensions, whereas the JPEG values are illustrative because the encoded size varies with image content and encoder settings.

#CVs	Images omitted	Uncompressed rate saved (Mbit/s)	JPEG Q90 rate saved (Mbit/s)
1	20.55 %	26.63	2.56
2	35.55 %	46.07	4.43
3	42.46 %	55.03	5.29
4	44.64 %	57.85	5.56

Table 9. Idealized image-upload savings for uncompressed RGB and a representative JPEG encoding

LiDAR offloading requires a stricter timestamp-level decision because one scan is associated with the complete three-view observation. We therefore use a fully redundant timestamp, for which all three evaluated camera images contain no non-redundant objects, as a proxy for a LiDAR omission opportunity. Across the V2X-Sim sweeps, the mean LiDAR size was 0.518 MB, corresponding to 20.71 Mbit/s at 5 Hz. Let p_k^{ts} denote the fraction of fully redundant timestamps. The idealized LiDAR rate reduction is

$$R_{\text{LiDAR,saved}}(k) = p_k^{\text{ts}} R_{\text{LiDAR}}. \tag{6}$$

Table 10 summarizes the timestamp-level estimates.

#CVs	Fully redundant timestamps	LiDAR rate saved (Mbit/s)
1	0.86 %	0.18
2	5.97 %	1.24
3	9.33 %	1.93
4	10.44 %	2.16

Table 10. Idealized LiDAR-upload savings based on fully redundant timestamps

These estimates quantify the available offloading opportunity rather than the savings achieved by the evaluated classifier. They assume correct identification of fully redundant observations and exclude protocol overhead and additional V2X messages.

6 CONCLUSION

This paper presented a redundancy-aware perception framework that integrates redundancy assessment directly into the object-detection stage. By augmenting the YOLOv11x architecture with lightweight spatial masks derived from cooperative perception data, the proposed model can jointly detect objects and identify whether they have already been perceived by connected vehicles. Experiments on the V2X-Sim dataset show that the method maintains strong detection performance while achieving an F1-score of 0.97 in redundancy classification, demonstrating that the mask-based representations provide meaningful context for reasoning about shared observations.

The redundancy analysis shows substantial perceptual overlap, with up to 68 % redundant objects, 44.64 % fully redundant images, and 10.44 % fully redundant timestamps when four neighbours are available. Under the stated assumptions, this corresponds to an idealized image-rate reduction of 57.85 Mbit/s for uncompressed RGB or 5.56 Mbit/s for the representative JPEG encoding, together with approximately 2.16 Mbit/s of avoided LiDAR traffic. The proposed architecture adds 4.17 % in estimated FLOPs and 0.687 ms in median forward-pass latency relative to the RGB baseline, indicating that redundancy assessment can be introduced with limited inference overhead while enabling selective offloading.

As a direction for future work, the same redundancy-assessment mechanism could be extended to single-vehicle perception. Instead of constructing masks from detections shared by connected vehicles, the masks could be derived from the ego vehicle's own detections in the previous frame. This would allow the model to reason about temporal redundancy and determine whether an object is newly observed or has already appeared in recent frames. Such a capability could enable adaptive perception pipelines in which computationally expensive modules are activated only when the scene contains novel information, thereby further reducing processing demands in resource-constrained settings.

Future work will also evaluate the method on real-world cooperative perception datasets under imperfect neighbouring detections and communication conditions. Deployment on representative automotive and edge hardware will enable measurement of end-to-end latency, communication volume, and energy consumption.

Acknowledgements

This project was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under project No. 09I05-03-V02-00014 and by the APVV-23-0519 project, "Legal and Technical Challenges of Smart Mobility to Increase Road Traffic Safety".

The authors acknowledge the use of generative AI tools for language editing and grammar refinement during the preparation of this manuscript. These tools were used to improve clarity and readability, but all content, ideas, and interpretations are solely those of the authors.

REFERENCES

[1] Halinkovic, M.—Vinel, A.—Benesova, W.: Li-Vip3d: Enhancing End-to-End Perception and Prediction with Camera-LiDAR Fusion. 2025 IEEE International Automated Vehicle Validation Conference (IAVVC), 2025, pp. 1–6, doi: 10.1109/IAVVC61942.2025.11219625.

[2] Xie, J.—Zhou, X.—Cheng, L.: Edge Computing for Real-Time Decision Making in Autonomous Driving: Review of Challenges, Solutions, and Future Trends. International Journal of Advanced Computer Science & Applications (IJACSA), Vol. 15, 2024, No. 7, Art. No. 59, doi: 10.14569/ijacsa.2024.0150759.

[3] Zang, S.—Ding, M.—Smith, D.—Tyler, P.—Rakotoarivelo, T.—Kaafar, M. A.: 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, 2019, No. 2, pp. 103–111, doi: 10.1109/MVT.2019.2892497.

[4] Zhang, Y.—Carballo, A.—Yang, H.—Takeda, K.: Perception and Sensing for Autonomous Vehicles Under Adverse Weather Conditions: A Survey. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 196, 2023, pp. 146–177, doi: 10.1016/j.isprsjprs.2022.12.021.

[5] Xu, R.—Xiang, H.—Tu, Z.—Xia, X.—Yang, M. H.—Ma, J.: V2X-Vit: Vehicle-to-Everything Cooperative Perception with Vision Transformer. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G. M., Hassner, T. (Eds.): European Conference on Computer Vision – ECCV 2022. Springer, Cham, Lecture Notes in Computer Science, Vol. 13699, 2022, pp. 107–124, doi: 10.1007/978-3-031-19842-7_7.

[6] Qiao, D.—Zulkernine, F.—Anand, A.: CoBEVFusion Cooperative Perception with LiDAR-Camera Bird's Eye View Fusion. 2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2024, pp. 389–396, doi: 10.1109/DICTA63115.2024.00064.

[7] Delooz, Q.—Willecke, A.—Garlichs, K.—Hagau, A. C.—Wolf, L.—Vinel, A.—Festag, A.: Analysis and Evaluation of Information Redundancy Mitigation for V2X Collective Perception. IEEE Access, Vol. 10, 2022, pp. 47076–47093, doi: 10.1109/ACCESS.2022.3170029.

[8] Masarykova, N.—Galinski, M.—Truchly, P.: Single-Filter CNN for Vehicle Recognition. 2024 International Symposium ELMAR, 2024, pp. 5–8, doi: 10.1109/ELMAR62909.2024.10694374.

[9] Lobato, W.—Mendes, P.—Rosário, D.—Cerqueira, E.—Villas, L. A.: Redundancy Mitigation Mechanism for Collective Perception in Connected and Autonomous Vehicles. Future Internet, Vol. 15, 2023, No. 2, Art. No. 41, doi: 10.3390/fi15020041.

[10] Luo, G.—Zhang, H.—Yuan, Q.—Li, J.: Complementarity-Enhanced and Redundancy-Minimized Collaboration Network for Multi-Agent Perception. Proceedings of the 30th ACM International Conference on Multimedia (MM '22), 2022, pp. 3578–3586, doi: 10.1145/3503161.3548197.

[11] Fang, Z.—Hu, S.—An, H.—Zhang, Y.—Wang, J.—Cao, H.—Chen, X.—Fang, Y.: PACP: Priority-Aware Collaborative Perception for Connected and Autonomous Vehicles. IEEE Transactions on Mobile Computing, Vol. 23, 2024, No. 12, pp. 15003–15018, doi: 10.1109/TMC.2024.3449371.

[12] Hu, Y.—Fang, S.—Lei, Z.—Zhong, Y.—Chen, S.: Where2comm: Communication-Efficient Collaborative Perception via Spatial Confidence Maps. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (Eds.): Advances in Neural Information Processing Systems 35 (NeurIPS 2022). Curran Associates, Inc., 2022, pp. 4874–4886, doi: 10.52202/068431-0352.

[13] Yang, D.—Yang, K.—Wang, Y.—Liu, J.—Xu, Z.—Yin, R.—Zhai, P.—Zhang, L.: How2comm: Communication-Efficient and Collaboration-Pragmatic Multi-Agent Perception. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (Eds.): Advances in Neural Information Processing Systems 36 (NeurIPS 2023). Curran Associates, Inc., 2023, pp. 25151–25164, doi: 10.52202/075280-1093.

[14] Prabhune, O.—Chen, T.—Kim, Y.: Content-Aware Input Scaling and Deep Learning Computation Offloading for Low-Latency Embedded Vision. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024, pp. 2218–2226, doi: 10.1109/CVPRW63382.2024.00227.

[15] Li, Y.—Ma, D.—An, Z.—Wang, Z.—Zhong, Y.—Chen, S.—Feng, C.: V2X-Sim: Multi-Agent Collaborative Perception Dataset and Benchmark for Autonomous Driving. IEEE Robotics and Automation Letters, Vol. 7, 2022, No. 4, pp. 10914–10921, doi: 10.1109/LRA.2022.3192802.

[16] Caesar, H.—Bankiti, V.—Lang, A. H.—Vora, S.—Liong, V. E.—Xu, Q.—Krishnan, A.—Pan, Y.—Baldan, G.—Beijbom, O.: nuScenes: A Multimodal Dataset for Autonomous Driving. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11618–11628, doi: 10.1109/CVPR42600.2020.01164.

[17] Jocher, G.—Qiu, J.: Ultralytics YOLO11. 2024, `https://github.com/ultralytics/ultralytics`.

[18] Galinski, M.—Milesich, T.—Janeba, M.—Danko, J.—Lehoczký, P.—Magdolen, Ľ.—Šoltés, L.—Tomčala, A.: Camera and Telemetry Data from Slovak Roads in Various Light and Weather Conditions. Scientific Data, Vol. 11, 2024, No. 1, Art. No. 1450, doi: 10.1038/s41597-024-04276-y.

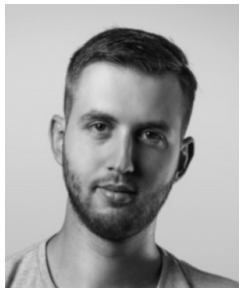
[19] Fech, B.—Masarykova, N.—Galinski, M.—Truchly, P.: Towards Bandwidth-Efficient Perception Offloading in Vehicular Networks via Temporal LiDAR-Based Bounding Box Refinement. 2026 IEEE Vehicular Networking Conference (VNC), 2026, pp. 1–4, doi: 10.1109/VNC69225.2026.11629146.



Nina Masarykova is Ph.D. student at the Slovak University of Technology in Bratislava, Slovakia, where she also received her B.Sc. degree in informatics and her M.Sc. degree in applied informatics. Her research focuses on computer vision and artificial intelligence, with an emphasis on automated driving, cooperative perception, and the optimization of detection pipelines in computational offloading scenarios.



Matej Halinkovic is Ph.D. student at the Slovak University of Technology in Bratislava, Slovakia, where he also received his B.Sc. degree in informatics and his M.Sc. degree in applied informatics. His research focuses on computer vision and artificial intelligence in the domain of automated driving and systems, with an emphasis on multimodal deep learning.



Marek Galinski is Associate Professor and Head of the 5G and Automotive Innovation Lab at the Slovak University of Technology in Bratislava. His research interests include next-generation wireless communication networks, with a focus on connected vehicles, V2X communication, and low-latency, high-reliability solutions for intelligent automated mobility. He is the co-founder of Encoderium, a software development and technology advisory company, and a co-organizer of Bratislava OpenCamp, a community conference on open technologies.



Peter Truchly is Associate Professor at the Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava. He received his Master's degree in informatics in 1995 and his Ph.D. in applied informatics in 2004 from the same university. He is actively involved in research in the field of vehicular ad-hoc networks, software-defined networks, transport layer protocols, wireless and satellite networks as well as heterogeneous networks. He is also interested in the development of educational materials and e-learning platforms as well as their pilot testing. He has participated as a researcher in many international projects (H2020, Erasmus+, FP7, LdV, COST) and national projects.