ROADVIEW

Robust Automated Driving in Extreme Weather

Project No. 101069576

Deliverable D5.1

SW on Adaptive Sensor Fusion and Perception Solutions - First report

WP5 – Robust perception system

Authors	Eren Erdal Aksoy (HH), Idriss Gouigah (HH), Tiago Fernandes Cortinhal (HH), Abu Mohammed Raisuddin (HH), Siddhant Som (HH)
Lead participant	HH
Delivery date	21 May 2024
Dissemination level	Public
Type	Document, report

Version 03

Co-funded by the European Union

Innovate UK

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. Project grant no. 101069576.

UK participants in this project are co-funded by Innovate UK under contract no.10045139.
Swiss participants in this project are co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract no. 22.00123.

Revision history

Author(s)	Description	Date
Eren Erdal Aksoy (HH)	First draft, report structure	15/02/2024
Maytheewat Aramrattana (VTI)	Revision 1 (for internal review)	19/02/2024
Eren Erdal Aksoy (HH)	Revised based on Revision 1	22/02/2024
Heikki Hyyti (FGI)	Revision 2 (update after review)	26/02/2024
Eren Erdal Aksoy (HH)	Revised based on Revision 2	28/02/2024
Andreia Cruz (accelCH)	Formatting and formal check (v01)	27/02/2024
Eren Erdal Aksoy (HH)	Revised and Section 2 added based on the reviewer's feedback. (v02)	08/05/2024
Mario Ceccarelli (accelCH)	Logo & acknowledgement update (v03)	21/05/2024

Contents

List of Figures

List of Tables

Partner short names

HH	Halmstad University
LUA	Lapin Ammattikorkeakoulu
THI	Technische Hochschule Ingolstadt
VTI	Statens Vag- och Transportforskningsinstitut
CE	CEREMA
RISE	RISE Research Institutes OF Sweden
FGI	Maanmittauslaitos – Finnish Geospatial Research Institute
S4	Sensible 4 OY
R5	Repli5
KO	Konrad GMBH
FORD	Ford Otomotiv Sanayi A.S
CRF	Canon Research Centre France S.A.S.
ZF	ZF Friedrichshafen AG
accelCH	accelopment Schweiz AG
WMG	Warwick Manufacturing Group, University of Warwick
VTT	Technical Research Centre of Finland
AVL	AVL Software and Functions GmbH

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
AD	Automated Driving
BEV	Bird-Eye-View
D	Deliverable
GA	Grant Agreement
GNSS	Global Navigation Satellite System
HD	High Definition
IMU	Inertial Measurement Unit
LiDAR	Light Detection and Ranging
M	Month
MS	Milestone

NA	Not Available
RADAR	Radio Detection And Ranging
SW	Software
T	Task
WP	Work Package
ZOD	Zenseact Open Dataset

Executive summary

The present document is a public deliverable of the ROADVIEW project funded under the Horizon Europe Framework Programme within the call HORIZON-CL5-2021-D6-01-01 “More powerful and reliable on-board perception and decision-making technologies addressing complex environmental conditions (CCAM Partnership)”. This deliverable provides the results of Task 5.1 “Adaptive Sensor Fusion and Perception Solutions”, developed within WP5 dedicated to “Robust perception system”.

This deliverable outlines the first report on algorithms and software implementation of the following four modules:

- **Sensor Fusion:** To fuse Camera, LiDAR, RADAR, and Thermal Camera data streams.
- **Object Detection:** To detect pedestrians, vehicles, and other agents (e.g., cyclists) in the environment.
- **Free Space Detection:** To detect free drivable areas in the scene.
- **Weather type Detection:** To distinguish between different types of weather (e.g., fog, rain, snow, etc.).

The deliverable is a compilation of four different scientific reports covering key topics and algorithms. Some of these algorithms described here were already reported as scientific manuscripts in international peer-reviewed conferences/journals. For some of the approaches, we only provide the obtained early results. All four of these manuscripts are attached to this deliverable to provide a more in-depth explanation.

The first paper introduces a new multimodal domain translation framework and contributes to the concept of adaptive sensor fusion by allowing the recovery of missing camera data using already available LiDAR data. This proposed multimodal domain translation framework is the first of its kind that generates photo-realistic RGB-D images from raw 3D point clouds by solely relying on the scene semantics. This is of utmost importance for the selection of an optimal and cost-effective sensor setup for information fusion. Based on this domain translation work, the second paper shows how to augment the LiDAR data to boost 3D object detection. The final output of this work is semantically segmented dense full-scan point clouds with drivable free space and 3D bounding boxes for each detected object (e.g., cars) in the scene. This work is unique as it can return augmented and semantically segmented LiDAR scans without camera sensors during inference. The third paper proposes a novel model that jointly learns various tasks (e.g., object detection and free space detection) by fusing information coming from different sensor modalities, including camera, LiDAR, and RADAR. This work goes beyond the state-of-the-art by solving multiple tasks simultaneously without the need to run different networks, each dedicated to a specific task. Finally, the fourth paper rather focuses on the prediction of different weather conditions by relying on camera and RADAR data. This paper employs a state-of-the-art model as a baseline to distinguish between five different weather types (clear, cloudy, rain, fog, and snow) by fusing camera and LiDAR data.

The present deliverable D5.1 contributes to Milestone (MS) 11 (due in Month (M) 18) “Intermediate check of the perception modules”. We note that the dissemination level of this deliverable D5.1 is public.

Objectives

The main objective of this deliverable is to provide an intermediate update on T5.1, which is being developed under the umbrella of WP5. The provided algorithms and perception solutions are planned to be tested in ROADVIEW demonstration vehicles in WP5 to give an overview of the expected outcome of T5.1.

Methodology and implementation

The development phase starts with the revisit of deliverable D2.4 “Requirements of the physical system setup”, which provides a comprehensive description of the perception- and control-related functional components, their requirements, and priority levels in the implementation and integration phases in the ROADVIEW reference architecture. These perception-related functional requirements form the actual content of this deliverable.

Outcomes

The outcome of this deliverable is a set of perception solutions to be used in different ROADVIEW demonstrations to solve, for instance, sensor fusion, object detection, and free-space detection, among others.

Next steps

Some of the perception solutions presented in this deliverable are still in the implementation phase. Therefore, further improvements in the model design, accuracy, and runtime are expected in the upcoming project lifecycle. The algorithms developed in this deliverable will be integrated into different demo vehicles for the upcoming demonstrations in WP8. Note that the final version of the ROADVIEW perception solutions will be reported in D5.2 “SW on Adaptive Sensor Fusion and Perception Solutions - Final report” in M36.

1 Introduction

In deliverable D2.4 “Requirements of the physical system setup”, we defined several fundamentally important functional requirements (such as sensor fusion, object detection, and visibility estimation functionalities), along with their sensor modalities and hardware needs, guiding the design of the ROADVIEW reference system architecture introduced in D2.3 “ROADVIEW system reference architecture”. These requirements form the fundamental perception-related functionalities in ROADVIEW.

For the sake of clarification, Figure 1 ROADVIEW Functional Modules. See D2.4 for more details illustrates all the functional components introduced in D2.4. Red blocks indicate perception modules with required specifications such as the expected sensor types (e.g., LiDAR and/or Camera), processing time, spatial resolution, and region of interest information that each functional module needs to cover. For instance, the functional module named “Free Space Detection” can receive input from a LiDAR sensor only, however, the full scan of the environment should be provided in high spatial resolution, such as 128-channel. On the other hand, the “Object Detection” functionality relies on the synchronized LiDAR and Camera data streams. We here note that only red blocks filled with a dark gray background color are considered new ROADVIEW perception modules, while those in light gray blocks represent other relevant standard modules already present in OEM vehicle platforms. Consequently, this deliverable is responsible for the implementation of the following four modules:

- **Sensor Fusion:** To fuse Camera, LiDAR, RADAR, and Thermal Camera data streams.
- **Object Detection:** To detect pedestrians, vehicles, and other agents (e.g., cyclists) in the environment.
- **Free Space Detection:** To detect free drivable areas in the scene.
- **Weather type Detection:** To distinguish between different types of weather (e.g., fog, rain, snow, etc.)

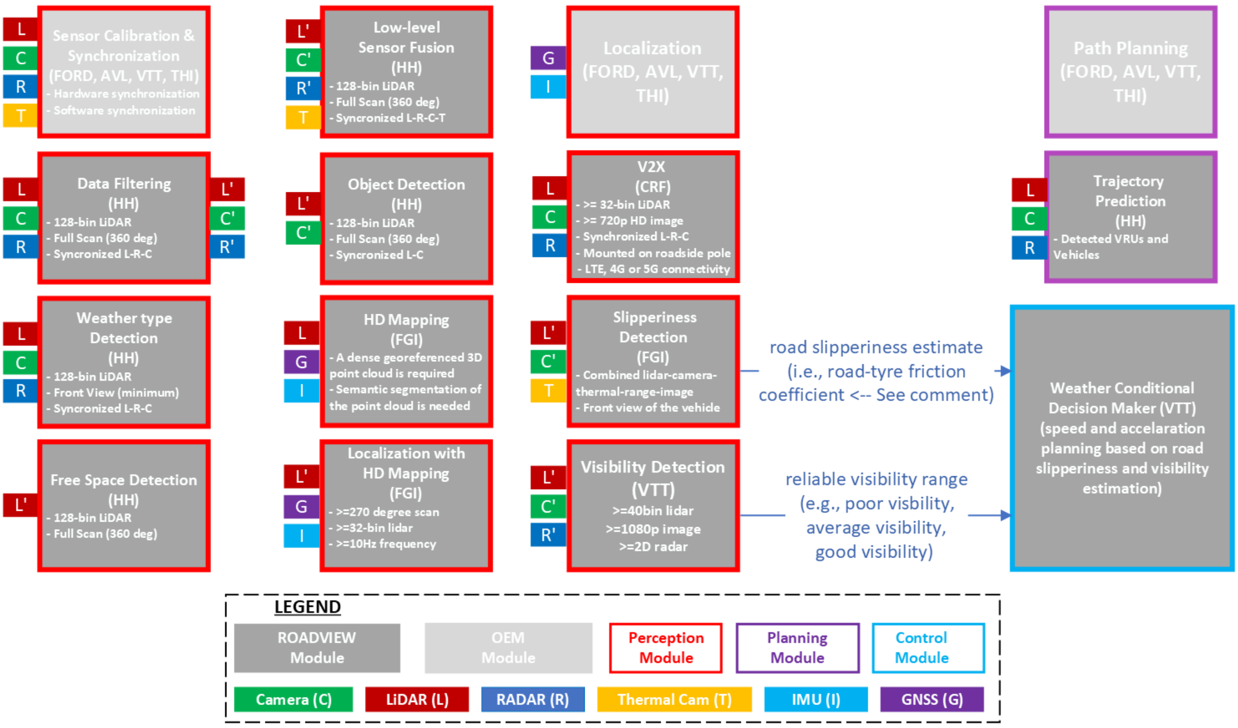

Figure 1 ROADVIEW Functional Modules. See D2.4 for more details. Note that $C', L', and\ R'$ refer to the denoised versions of Camera (C), LiDAR (L), and RADAR (R) readings.

The remaining perception solutions are to be described in the other relevant deliverables. For instance, the Data Filtering module will be detailed in D4.8, whereas D5.6 and D5.7 will provide the technical implementations of HD Mapping and Localization with HD Mapping, respectively. In addition, D5.3 and D5.4 are dedicated to the implementation of V2X-related collaborative perception solutions. Finally, D5.5 will introduce the Slipperiness and Visibility Detection modules in detail. Note that the Trajectory Prediction functionality in the Planning module requires object tracking, which will also be detailed in the next version of this deliverable in D5.2 in M36.

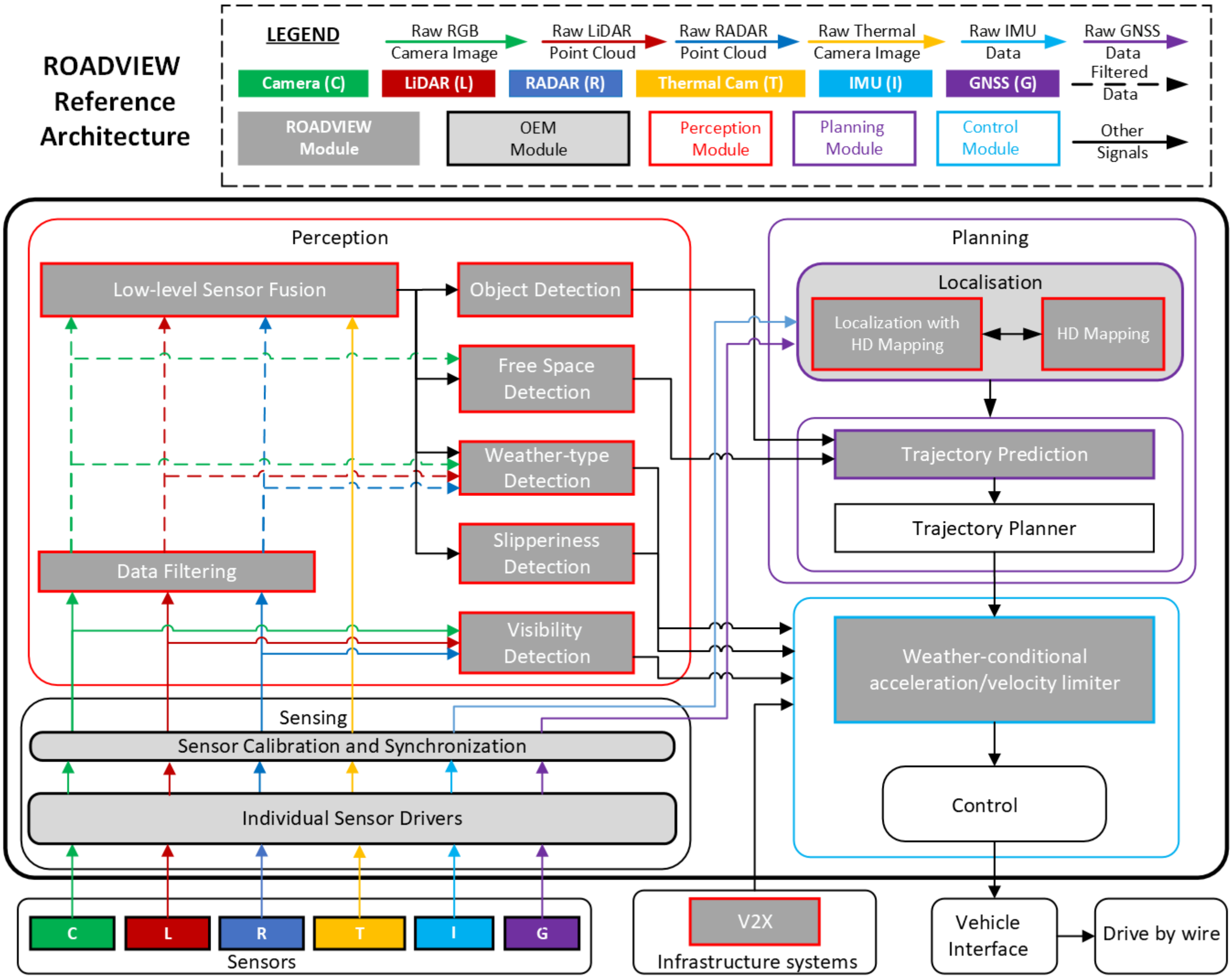

Figure 2 ROADVIEW Reference Architecture: A Functional Generic Architecture of the Intended Functionalities. Red blocks represent perception solutions. See D2.3 for more details.

Figure 2 depicts the ROADVIEW reference architecture introduced in D2.3, which particularly details the interdependencies and information flow between all these perception modules (see the red box in Figure 2). For instance, the Object Detection, Free Space Detection, and Weather type Detection modules receive the output of the Sensor Fusion module, which should process the output of the Data Filtering module (See Figure 2). Since we are still in the early stages of implementation, this data flow between the modules is yet to be established in full scale.

In this deliverable, we go a step further and present the early implementation results of the four modules mentioned above: **Sensor Fusion, Object Detection, Free Space Detection,** and **Weather type Detection**. In this regard, this deliverable comprises four scientific manuscripts that provide in-detail descriptions of these four perception modules. Each manuscript is attached to this deliverable (See the Appendix in Section 6), and a brief overview of each is provided below:

- **Paper 01:** The first article [1] is entitled "Depth- and Semantics-aware Multi-Modal Domain Translation: Generating 3D Panoramic Color Images from LiDAR Point Cloud". This study introduces a novel model called TITAN-Next, which is a depth- and semantics-aware conditional generative model designed for domain translation between LiDAR and camera sensors in a multimodal context. TITAN-Next employs scene semantics to transform 3D LiDAR point clouds into RGB-D images, a first in its field. Although this model relies on camera and LiDAR sensors during training, it receives LiDAR-only data for synthesizing camera images. This work is an important contribution to enhancing the functionality of autonomous vehicles, as it provides fail-safe features and enriches the image domain with additional data. The proposed model is evaluated on the large-scale and challenging Semantic-KITTI dataset [2], and experimental findings show that it considerably outperforms the original TITAN-Net [3] and other strong baseline models by 23.7% margin in terms of accuracy. Note that this work was already published in the journal Robotics and Autonomous Systems (RAS). For more details about the implementation, we refer to the publicly available source code https://github.com/TiagoCortinhal/TITAN-Next.

  In the context of ROADVIEW, this work contributes to adaptive **Sensor Fusion** by allowing the recovery of missing camera data using already available LiDAR scans. This is of utmost importance for the selection of an optimal and cost-effective sensor setup for information fusion. For instance, in the case of a camera failure, the vehicle can adapt to the situation by fusing the synthesized camera image (instead of the missing real counterpart) with the corresponding LiDAR stream. The subsequent work (**Paper 02** presented below) also shows how such a cross-domain generative model can boost downstream tasks such as object detection.

- **Paper 02:** The second article [4] is named “Semantics-aware LiDAR-only Pseudo Point Cloud Generation for 3D Object Detection”. This study is based on our previous generative model in **Paper 01** [1] and introduces a unique LiDAR-based framework for augmenting raw scans with synthetic denser pseudo-point clouds by using only LiDAR sensor readings and scene semantics. The proposed framework starts with segmenting raw point clouds to extract semantic maps of the scene. Next, a domain translator (TITAN-Next [1]) is employed to generate synthetic image segments and depth cues without requiring real cameras during inference, resulting in semantically enriched dense pseudo-point clouds. In particular, a novel semantically guided projection method is introduced to improve object detection by focusing only on relevant pseudo points. The framework, applied to various 3D object detection methods, shows up to a 2.9% performance increase and yields comparable results to state-of-the-art LiDAR-only detectors on the KITTI 3D object detection dataset [5]. As in the case of **Paper 01** [1], this framework also relies only on LiDAR data during inference.

  In ROADVIEW, this work plays an important role in **Object Detection** and **Free Space Detection.** The final output of this proposed method is semantically segmented dense full-scan point clouds with 3D bounding boxes for each detected object (e.g., cars) in the scene. These semantic segments not only indicate where each object (e.g., cars, pedestrians, cyclists, etc.) is located but also highlight the drivable free space in the scene. This approach is also unique in that it jointly solves the problem of semantically labeling point clouds and generating dense pseudo counterparts. We note here that this work does not propose a new object detection method, but rather introduces a new point cloud enrichment method to enhance subsequent object detection methods with little to no modification.

- **Paper 03:** The third article titled “Multimodal Multitask Learning for Autonomous Driving” proposes a novel framework that jointly learns various tasks by fusing information coming from different sensor modalities, including camera, LiDAR, and RADAR. More specifically, this work integrates recent advancements in 3D object detection, 3D semantic segmentation, and 3D panoptic segmentation into a unified model to address all these three tasks. Since each sensor has a unique data format (2D images vs. 3D point cloud data), the proposed framework processes each data stream in different branches to extract relevant descriptive features accordingly. An extensive experimental study on the large-scale multimodal nuScenes dataset [6] is conducted to compare obtained results with the state-of-the-art methods.

Although the ultimate goal is to learn 3D object detection together with 3D semantic and panoptic segmentation tasks, in this paper, we only report experimental findings for 3D object detection and 2D free space detection in top-view, i.e., Bird-Eye-View (BEV). It is noted that the implementation of the overall framework is still in progress, therefore, it is planned to include 3D semantic segmentation as well as panoptic segmentation tasks together with the 3D object detection in the final version. This work reports the initial results obtained on a limited set of tasks, namely 3D object detection and 2D free space detection. The framework formulation includes 3D semantic and panoptic segmentation for the sake of generality, highlighting the direction of the ongoing implementation phase. The final version of the framework is also planned to be benchmarked on the ROADVIEW dataset.

For the ROADVIEW project, this paper plays an important role in **Sensor Fusion, Object Detection,** and **Free Space Detection.** It is important to note that the first two papers (**Paper 01** [1] and **Paper 02** [4]) heavily rely on the LiDAR modality during inference, although both Camera and LiDAR modalities are required in the training phase. In contrast, this third paper proposes a multimodal learning framework that can receive input from both Camera, LiDAR, and RADAR. We here note that this study is still in progress and has not yet undergone peer review in any scientific publication.

- **Paper 04:** This last article is entitled "Camera- and LiDAR-based Weather-type Classification for Autonomous Vehicles". In this paper, we provide the preliminary findings obtained using a state-of-the-art artificial neural network model. More specifically, our model is a pre-trained vision transformer [7] that receives input from both camera and LiDAR sensors and predicts five different weather classes, including cloudy, clear, snow, rain, and fog. We apply early fusion of both modalities and study the representation of LiDAR readings. For this purpose, the raw 3D LiDAR point cloud is transformed into a 2D panoramic view image representation. Furthermore, the front view of the point cloud is cropped to cover only the central 120° to examine the effects of the visual coverage area. We conduct an extensive experimental study on the large-scale Zenseact Open Dataset (ZOD) [8]. The ZOD data is collected from 14 European countries over multiple years and includes various traffic scenarios, adverse weather conditions (such as rain, fog, and snow), road types, and lighting conditions.

  In the context of ROADVIEW, this work contributes to **Sensor Fusion** and **Weather type detection.** In this preliminary study, we particularly focus on the early fusion of both modalities. We plan to combine this work with the more sophisticated sensor fusion framework introduced in **Paper 03,** which also includes the RADAR modality. It is important to note that this study is still in progress and has not yet undergone peer review in any scientific publication.

2 Functional Implementation

In the previous section, we provided a brief overview of the core processing architecture introduced in D2.3. This architecture plays a crucial role in the functional implementation of the perception modules, as it details the sensor types to be employed, their spatial and temporal resolutions, and the information flow between each perception module. For instance, this core architecture conveys the message that object and free space detection modules should rely on fused versions of the thermal camera and filtered RGB, LiDAR, and RADAR data streams. Based on this high-level architectural design, this section provides additional information about the functional implementation of sensor fusion and other perception algorithms.

Figure 3 shows the fusion strategy to be employed in Paper III. As inspired by the work in LiRaFusion [9], BEVFusion [10], 2D image and 3D point cloud data streams are processed individually in two different streams. As highlighted in red background color, LiDAR and RADAR point clouds are initially voxelized and combined in an **early fusion** manner. After extracting individual BEV features in both point cloud and RGB image spaces, a **mid-level fusion** is employed (green background) to integrate camera information with that of LiDAR and RADAR sensors. The reason for having a mix of early and mid-level fusions is that at each level there exists a unique set of features, which are crucial for the task to be learned. A similar fusion strategy has already been investigated in relevant works such as LiRaFusion [9] and BEVFusion [10]. Our earlier ablation studies (provided in Table II in Paper III) show that combining both fusion strategies boosts the performance of the learned task such as segmentation. We propose to use the fused features to jointly learn both tasks: 2D semantic segmentation (including drivable free space detection) and object detection. This idea emerges from the works [1], [11] where the authors show that joint learning of multiple relevant complementary tasks (such as depth prediction and semantic segmentation) boosts performance by a large margin.

All perception functionalities are implemented using the well-known machine learning library PyTorch[1]. We train our fusion framework for multi-task learning using a large-scale multimodal nuScenes dataset [6] and compare our results to the state-of-the-art methods BEVFusion [10], Lidar Multinet [11]. We also plan to benchmark the performance of our model on individual tasks (e.g. semantic segmentation using only LiDAR) using different datasets such as SemanticKITTI [2]. Once the implementation is done, we plan to retrain our model using the ROADVIEW dataset collected and annotated in WP4. Around M30, we plan to start the integration of our fusion model into the demonstration vehicles in WP8. Therefore, during the functional implementation, we pay particular attention to the real-time processing capacity of the pipeline.

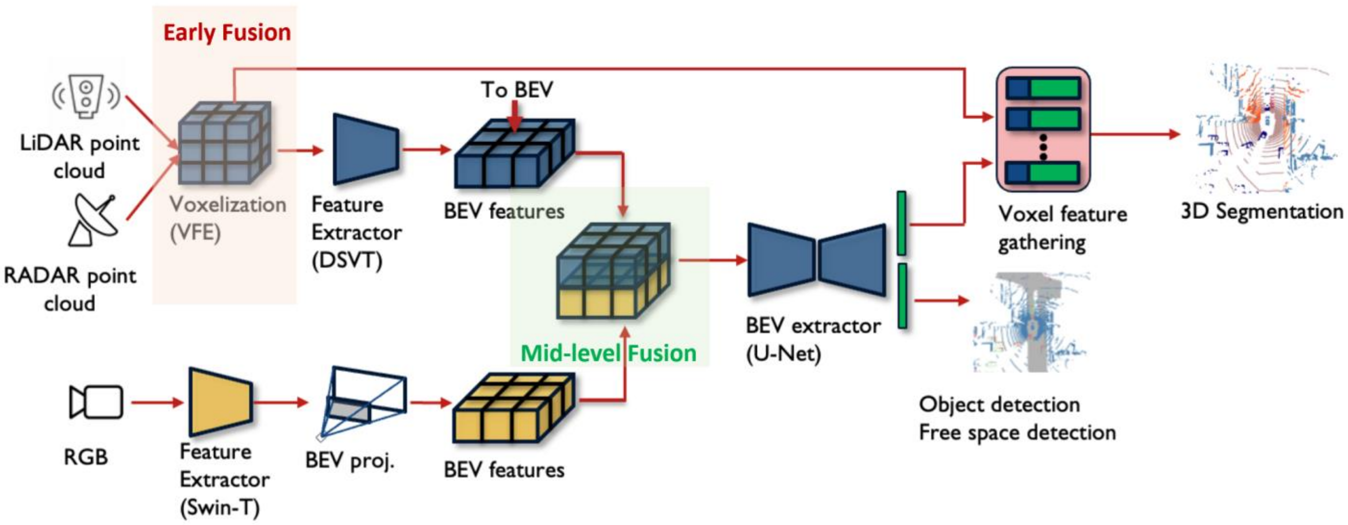

Figure 3 Early and mid-level fusion of different sensor modalities.

[1] https://pytorch.org/

In the following, to facilitate functional implementation and integration, we provide **ROS modules and signals** of each perception solution developed in this deliverable. Note that the modules listed below are the methods presented in the attached 4 papers (See the Appendix in Section 6). These module interfaces and signals play a crucial role in the ROADVIEW system integration in WP8. It is important to note that some interfaces and signals defined below may have different forms since the development of these modules is still ongoing.

ROS Modules

Sensor Fusion Module

Role of the module	This module will fuse the filtered camera, LiDAR and RADAR data into a single data representation. The format of the fused data is TBD.
Responsible parties	HH
Work Package and Tasks	WP5: T5.1
Computational resource usage	Moderate RAM and GPU usage

Signals produced:

- fused/roadview_fusion_data

Signals consumed:

- filtered/camera_image
- filtered/lidar_point_cloud
- filtered/radar_point_cloud

Object Detection Module

Role of the module	Object Detection Module will only use fused data to detect objects around the ego vehicle
Responsible parties	HH
Work Package and Tasks	WP5: T5.1
Computational resource usage	Heavy RAM and GPU usage

Signals produced:

- combined/objects

Signals consumed:

- fused/roadview_fusion_data
- filtered/camera_image

Free Space Detection Module

Role of the module	Free Space Detection module will semantically segment the given data and publish the free space / drivable area.
Responsible parties	HH
Work Package and Tasks	WP5: T5.1
Computational resource usage	Heavy RAM and GPU usage

Signals produced:

- combined/free_space

Signals consumed:

- fused/roadview_fusion_data
- filtered/camera_image

Weather-Type Detection Module

Role of the module	This module will utilize already existing weather type data from the Data Filtering module and use the fused data to detect the weather type again, based on all four predictions, this module will make the final verdict on the weather type.
Responsible parties	HH
Work Package and Tasks	WP5: T5.1
Computational resource usage	Low RAM and GPU usage

Signals produced:

- combined/weather_type

Signals consumed:

- filtered/weather_type/camera
- filtered/weather_type/lidar
- filtered/weather_type/radar
- fused/roadview_fusion_data

Signals

Fused Sensor Data (fused/roadview_fusion_data)

Role of the signal	Gives fused data from camera, LiDAR, RADAR and Thermal Camera.
Producer module	Data Filtering Module
Real-time requirements	10Hz.
ROS topic name	roadview_msgs/fused/roadview_fusion_data (proposed)

Signal schema:

Field name	**Data type**	**Required / optional**	**Comment**
header	std_msgs/Header	Required	Timestamp
fused_point_cloud	sensor_msgs/PointCloud	Required	x, y, z coordinates and reflectance value
fused_rgbd	sensor_msgs/Image Message	Optional	Equivalent rgbd of the fused_data_pc point cloud

fused_range_view	sensor_msgs/Image Message	Optional	Equivalent range view of the fused_data_pc point cloud

Detected Object List (combined/objects)

Role of the signal	Gives the detected objects
Producer module	Object detection (proposed)
Real-time requirements	10Hz.
ROS topic name	roadview_msgs/combined/objects(proposed)

Signal schema:

Field name	Data type	Required / optional	Comment
header	std_msgs/Header	Required	Timestamp
roadview_objects	vision_msgs/Detection3DArray Message	Required	object detection result, bounding boxes can be found inside the message.

Free Space Estimate (combined/free_space)

Role of the signal	Gives the boundaries of free space wrapped inside the Detection message
Producer module	Free Space Detection Module (proposed)
Real-time requirements	10Hz.
ROS topic name	roadview_msgs/ combined/free_space (proposed)

Signal schema:

Field name	Data type	Required / optional	Comment
header	std_msgs/Header	Required	Timestamp
n_bbox	std_msgs/Int32	Required	number of bounding boxes
free_space_bbox	std_msgs/Float32MultiArray	Required	free space detection result, bounding boxes can be found inside the message.
free_space_pc	sensor_msgs/PointCloud Message	optional	x, y, z coordinates and reflectance value for the free space

Weather Type from Camera (filtered/weather_type/camera)

Role of the signal	Detects the type of weather e.g., 0=normal, 1=snowy, 2=foggy, 3=rainy.
Producer module	Data Filtering Module (Proposed)
Real-time requirements	10Hz.
ROS topic name	roadview_msgs/filtered/weather_type_camera (proposed)

Signal schema:

Field name	Data type	Required / optional	Comment
header	std_msgs/Header	Required	Timestamp
camera_weather_type	std_msgs/Int8	Required	will return a value between 0 to 3(inclusive) where 0 means normal weather, 1 means snowy, 2 means foggy and 3 means rainy weather

Weather Type from LiDAR (filtered/weather_type/LiDAR)

Role of the signal	Detects the type of weather e.g., 0=normal, 1=snowy, 2=foggy, 3=rainy.
Producer module	Data Filtering Module (Proposed)
Real-time requirements	10Hz.
ROS topic name	roadview_msgs/filtered/weather_type_lidar (proposed)

Signal schema:

Field name	Data type	Required / optional	Comment
header	std_msgs/Header	Required	Timestamp
lidar_weather_type	std_msgs/Int8	Required	will return a value between 0 to 3(inclusive) where 0 means normal weather, 1 means snowy, 2 means foggy and 3 means rainy weather

Weather Type from RADAR (filtered/weather_type/radar)

Role of the signal	Detects the type of weather e.g., 0=normal, 1=snowy, 2=foggy, 3=rainy.
Producer module	Data Filtering Module (Proposed)
Real-time requirements	10Hz.
ROS topic name	roadview_msgs/filtered/weather_type_radar (proposed)

Signal schema:

Field name	Data type	Required / optional	Comment
header	std_msgs/Header	Required	Timestamp
radar_weather_type	std_msgs/Int8	Required	will return a value between 0 to 3(inclusive) where 0 means normal weather, 1 means snowy, 2 means foggy and 3 means rainy weather

Combined Weather Type (combined/weather_type)

Role of the signal	Detects the type of weather e.g., 0=normal, 1=snowy, 2=foggy, 3=rainy.
Producer module	Weather-type Detection Module (Proposed)
Real-time requirements	10Hz.
ROS topic name	roadview_msgs/combined/weather_type(proposed)

Signal schema:

Field name	Data type	Required / optional	Comment
header	std_msgs/Header	Required	Timestamp
combined_weather_type	std_msgs/Int8	Required	will return a value between 0 to 3(inclusive) where 0 means normal weather, 1 means snowy, 2 means foggy and 3 means rainy weather

In Figure 1Figure 4, we show the expected input and output of the perception modules introduced in this deliverable. Some of these modules still require improvement to receive additional modalities. For instance, Sensor Fusion needs to receive the Thermal camera information as well, as highlighted by the black frame and darker background in Figure 4. The same is true for the RADAR input in the module Weather type detection.

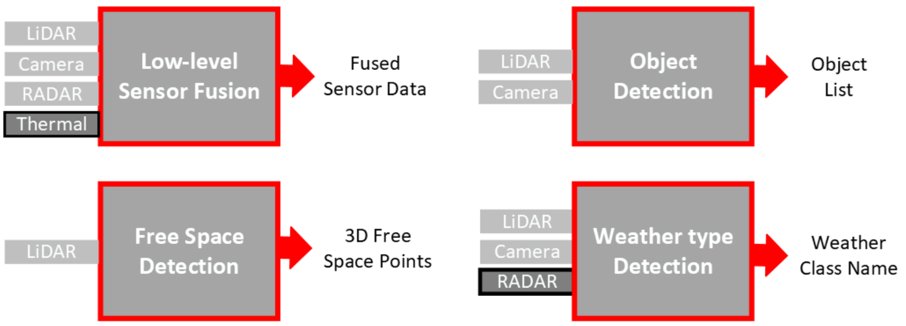

Figure 4 Perception solutions introduced in this deliverable

3 Relation to other deliverables

The basic skeleton of this deliverable was formed in D2.4 "Requirements of the physical system setup", where the ROADVIEW functional requirements were detailed. Each perception-related requirement forms a unique functional component to be implemented in this deliverable. The concurrently submitted deliverable D5.3 "SW on Collaborative Perception Solutions - First report" in WP5 also provides various perception solutions, which rather focus on collaborative perception by sharing information from roadside infrastructure sensors. Another highly relevant deliverable submitted in M18 is D5.6 "SW on Improved Localization Using High-density Map Updating - First report" in WP5, which also relies on LiDAR-based semantic segmentation methods introduced in this deliverable. We also note that this deliverable will play an important role in the system integration and demonstration activities in WP8.

4 Conclusions and Continuous Improvement

This deliverable presents the initial implementation results of the ROADVIEW perception solutions. It is anticipated that there will be further enhancements in the model design, accuracy, and runtime during the upcoming project lifecycle. All these introduced perception solutions are also planned to be evaluated using the ROADVIEW dataset.

This deliverable includes four scientific manuscripts that provide comprehensive explanations of the ROADVIEW perception solutions. Table 1 shows the correlation between different perception functions and the papers introduced. It is worth noting that the functionality **Object Detection** is a generic term and refers to the detection of all types of obstacles covering vulnerable road users, cars, etc. We also note that the final versions of these perception modules, together with an additional **Object Tracking** functionality required for the Trajectory Prediction module in path planning (See Figure 1) will be elaborated on in the forthcoming version of this deliverable in D5.2 in M36.

One of the challenges faced during the implementation of these perception functionalities is the lack of a dataset that covers multiple sensors under harsh weather conditions. To the best of our knowledge, there exists no public multimodal dataset (Camera, LiDAR, RADAR, and Thermal Camera) providing annotations for multiple tasks such as object detection, semantic segmentation, and weather-type classification. Therefore, each functionality proposed here is assessed on a range of datasets such as nuScenes [6] and ZOD [8]. We plan to train and test all our perception solutions on the annotated multimodal ROADVIEW dataset to be released in D4.4 and D4.9 in M36.

Table 1 Relationship between the ROADVIEW perception functions and the introduced papers.

	Paper 01	Paper 02	Paper 03	Paper 04
Sensor Fusion	✓		✓	✓
Object Detection		✓	✓	
Free Space Detection		✓	✓	
Weather-type Detection				✓

5 References

[1] T. Cortinhal and . E. E. Aksoy, "Depth- and semantics-aware multi-modal domain translation: Generating 3D panoramic color images from LiDAR point clouds," Robotics and Autonomous Systems, 2024.

[2] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss and J. Gall, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," in IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[3] T. Cortinhal, F. Kurnaz and E. E. Aksoy, "Semantics-Aware Multi-Modal Domain Translation: From LiDAR Point Clouds to Panoramic Color Images," in IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2021.

[4] T. Cortinhal, I. Gouigah and E. E. Aksoy, "Semantics-aware LiDAR-Only Pseudo Point Cloud Generation for 3D Object Detection," in IEEE Intelligent Vehicles Symposium (In press), 2024.

[5] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[6] H. Caesar, V. Bankiti, A. . H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan and O. Beijbom, "nuScenes: A Multimodal Dataset for Autonomous Driving," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations, 2021.

[8] M. Alibeigi, W. Ljungbergh, A. Tonderski, G. Hess, A. Lilja, C. Lindstrom, D. Motorniuk, J. Fu, J. Widahl and C. Petersson, "Zenseact Open Dataset: A large-scale and diverse multimodal dataset for autonomous driving," in IEEE/CVF International Conference on Computer Vision, 2023.

[9] J. Song, L. Zhao and a. S. K.A., "LiRaFusion: Deep Adaptive LiDAR-Radar Fusion for 3D Object Detection," in IEEE ICRA, 2024.

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

[11] D. Ye, Z. Zhou, W. Chen, Y. Xie, Y. Wang, P. Wang and a. F. H., "Lidarmultinet: Towards a unified multi-task network for lidar perception," in AAAI Conference on Artificial Intelligence, 2023.

6 Appendix