Ref. Ares(2025)7017869 - 29/08/2025

ROADVIEW

Robust Automated Driving in Extreme Weather

Project No. 101069576

Deliverable 4.9

Report on data annotation

WP 4 – Data Logging, Filtering, Annotation, and Quality Assurance

Authors	Tuomas Herranen, Kalle Niemi, Sofie Väisänen, Miroslav Ivanov
Lead participant	LUA
Delivery date	29 August 2025
Dissemination level	Public
Type	Report

Version 1.0

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
Sofie Väisänen (LUA)	Draft deliverable	15/07/2025
Tuomas Herranen ((LUA)	Draft deliverable	18/08/2025
Harry Chan (QMUL)	Revision 1	19/08/2025
Torbjörn Norlander (Repli5)	Revision 2	22/08/2025
Tuomas Herranen (LUA)	Revision	26/08/2025
Eren Erdal Aksoy (HH)	Final version	29/08/2025

Contents

List of Figures

List of Tables

Partner short names

HH	Hogskolan Halmstad
LUA	Lapin Ammattikorkeakoulu OY
RISE	RISE Research Institutes of Sweden AB
FGI	Maanmittauslaitos – Finnish Geospatial Research Institute
AVL	AVL Software and Functions GmbH

Abbreviations

AI	Artificial Intelligence
D	Deliverable
IoU	Intersection Over Union
mIoU	mean Intersection Over Union
LiDAR	Light Detection and Ranging
WP	Work Package

Executive summary

Objectives

In the development of autonomous vehicles, access to high-quality, diverse, and accurately annotated data is essential for the effective training and validation of machine learning models. This deliverable supports the creation of datasets that serve as a foundation for advancing AI systems in autonomous driving. By providing annotated multimodal sensor data collected in complex real-world environments, it contributes to the development of perception algorithms, decision-making systems, and overall vehicle intelligence, in line with the objectives of the ROADVIEW project.

The primary objective for this delivery was to annotate the raw sensor data collected in Task 4.2, "Data Logging." Particular emphasis was placed on capturing a wide range of adverse weather conditions and developing a method to label weather-induced noise in the data. The resulting annotated dataset was subsequently meant to be used in Task 4.4, "Data Filtering," to support the development of advanced data filtering algorithms.

Methodology and implementation

The datasets used in Task 4.5 "*Data annotation*" have been collected collaboratively by ROADVIEW project partners across a range of environmental and traffic conditions. The data acquisition process was conducted using various sensor modalities under different weather scenarios (e.g., rain, fog, snow, and cloudy), as well as at varying times of day and in diverse traffic situations. This variability ensures that the datasets reflect realistic and complex driving environments, which are essential for robust AI model development.

Data annotation was approached using both manual and automatic workflows. A smaller portion of the data was labelled and fine-tuned manually with high accuracy, while the rest of the data was provided with coarser, semi-automatically generated labels. Labelled data includes synchronised LiDAR and Camera data that are provided with both semantic segmentation and bounding-box labels. For dynamic objects, instance IDs are also included. The 35 classes used follow the popular Semantic KITTI categorisation but were expanded to include ROADVIEW-specific classes. These new classes include weather-related categories such as snow, rain, and fog, which are instrumental in advancing data filtering algorithms.

LUA utilized the Segments.AI platform for multi-sensor data labelling and fine-tuning, and machine learning algorithms for automatic pre-annotation. The Segments.AI platform enabled the accurate labelling of noise caused by precipitation, such as snowfall, in the point clouds. LUA also collaborated with the company Advian Oy in producing the quality annotations.

Outcomes

This deliverable has resulted in 10k frames of accurately annotated LiDAR point clouds and Camera images, encompassing 35 distinct classes. Additionally, 40k additional frames are included with weaker automatic annotations. The annotated data captures a wide range of traffic scenarios under diverse weather conditions, including rain, snow, fog, and low light. As one of the few labelled datasets to move beyond clear-weather data, it provides a valuable resource for developing and benchmarking robust perception systems in autonomous vehicles.

The provided data includes comprehensive annotations, such as semantic segmentation and 3D bounding boxes for LiDAR data, as well as panoptic segmentation and 2D bounding boxes for RGB camera images. By being openly available, it fosters transparency, collaboration, and reproducibility, thereby supporting both the ROADVIEW project and the broader research community in advancing safe and reliable autonomous mobility.

Additionally, during the task, the ROADVIEW consortium also released a joint paper called the Snowy Scenes, which featured part of the annotated winter data.

Next steps

The results of this deliverable have been utilized in Task 4.4 "*Data Filtering*" for the development of data filtering algorithms. LUA continues to support partners that use the data and help fix any issues related to it.

1 Introduction

This report presents the annotation process and results for the ROADVIEW data produced in WP4. The aim was to annotate data representing a wide range of adverse weather conditions and traffic scenarios, and to establish a method for categorising noise in the data caused by these conditions. The data annotation task was performed primarily by LUA, with support provided by HH. Additionally, LUA collaborated with Advian Oy in producing high-quality annotations.

The initial data collection was conducted by ROADVIEW partners AVL and FGI in Task 4.2 across diverse environmental and traffic contexts. FGI provided datasets from winter environments, while AVL contributed data from a wider range of conditions, like snowy, rainy, foggy, and clear weather. The datasets included multiple sensor modalities and were prepared for annotation.

Chapter 2 provides a brief description of the datasets used for annotation. Chapter 3 continues with the tools employed. It then outlines how the data was annotated and the categories applied. Chapter 4 presents the annotation results, illustrated with example images and comparisons. The final chapter summarises the conclusions. The labels are made publicly available through the ROADVIEW project website (https://roadview-project.eu/)

2 Datasets

The datasets provided for annotation were collected by ROADVIEW partners FGI and AVL as part of WP4 task 4.2 "Data Logging." The datasets covered a wide variety of adverse weather conditions and traffic scenarios. In addition to the varying weather conditions, the sensor configurations adopted by the two partners differ substantially, with the principal distinctions relating to both the quantity and the type of cameras and LiDAR sensors utilized. This also influenced the annotation process, as, for example, the LiDAR resolution determines the level of detail with which objects are represented in the point cloud. The datasets presented in the following subsections were recorded at a sampling rate of 10 frames per second (10 Hz). Each frame contains synchronized measurements from all sensors. The size of each dataset is expressed in terms of the total number of frames it includes. All of the data was also already anonymised, motion compensated and prepared for use with the chosen annotation tools described in Chapter 3.

2.1 FGI dataset

The FGI data was collected in southern Finland at Espoo and Helsinki. The annotated dataset encompasses both urban and highway environments during daytime in winter, under varying levels of snowfall and accumulated snow. The data selected for annotation includes sequences from three separate sections representing data collection trips.

The first section of the annotated data was recorded under clear winter weather conditions in the Otaniemi district of Espoo, Finland (Figure 1), with snow present on the ground but no active snowfall. The sky was clear, providing consistent lighting throughout the recordings. The first section was divided into two separate continuous sequences with a total of 1011 frames.

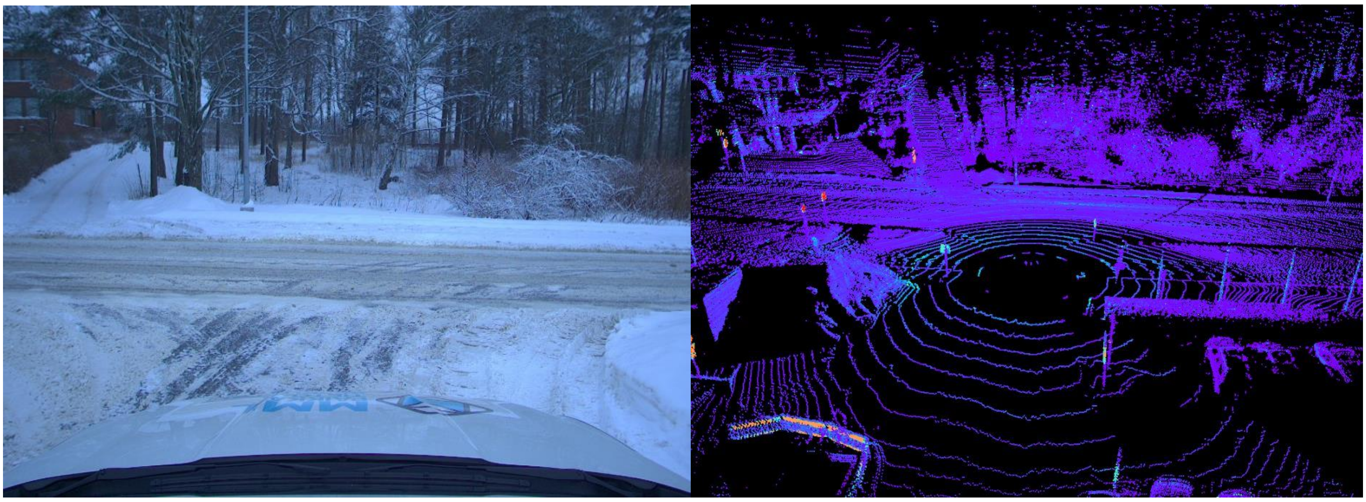

Figure 1 Example of synchronised RGB front camera image (Left) and LiDAR point cloud (Right) from the same frame in clear winter weather.

The second section was also collected in snowy weather conditions in the same district as the first section. During the recording sessions, snow was both present on the ground and actively falling, providing a challenging environment for perception systems (Figure 2). The section includes only one continuous sequence, which is 3016 frames long. The section also includes a variety of traffic elements, such as buses, trucks, cars, pedestrians, and cyclists, captured in challenging winter conditions. The lidar data also included a lot of noise caused by the falling snowflakes, which needed to be included in the annotations.

Figure 2 Example of synchronised RGB front camera image (Left) and LiDAR point cloud (Right) in snowfall.

Unlike the previous two sections, the third section was divided into 7 shorter, non-continuous segments that contain a total of 1,000 frames. The smaller sequences were chosen based on the presence of particularly interesting and diverse scenes. The primary objective was to select portions of the original recording that contain a high density of pedestrians, cyclists, and both moving and stationary vehicles, to ensure a wide range of annotated object classes. An example of the Highway data is shown in Figure 3.

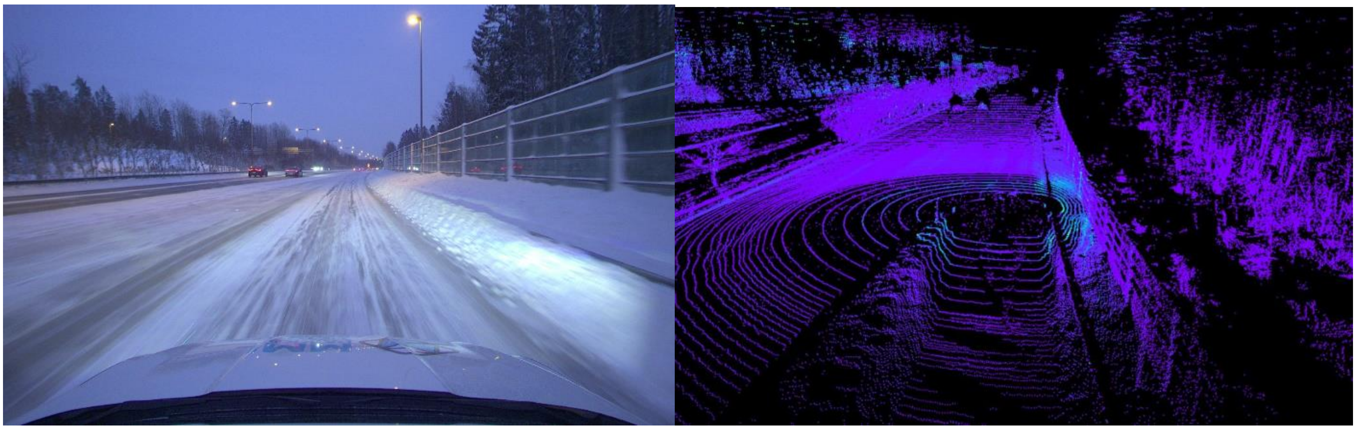

Figure 3 Example of synchronised RGB front camera image (Left) and LiDAR point cloud (Right) in a winter Highway scene.

2.2 AVL dataset

The AVL datasets were collected using a vehicle operating in Regensburg, Germany. It encompasses a broad range of traffic scenarios and environmental conditions. The data was divided into six sections according to the prevailing weather conditions at the time of collection. The six different sections are:

1. Cloudy
2. Rainy #1
3. Rainy #2
4. Snowy Day
5. Snowy Night
6. Foggy

Each of the sections included several minutes and thousands of frames of data for annotation. This meant that for each section, only the most important parts were chosen for fine-tuning. This included mainly the sections with noticeable noise in the data caused by adverse weather conditions.

The AVL data collection vehicle had a very different sensor setup compared to FGI. AVL included left, right and middle LiDAR sensors as well as front, left, right and rear RGB cameras (Figure 4). LiDARs used were of lower resolution compared to FGI, with two 16-channel and one 32-channel. However, the four cameras provide stronger support for multisensory annotation, as they capture views from all sides of the vehicle.

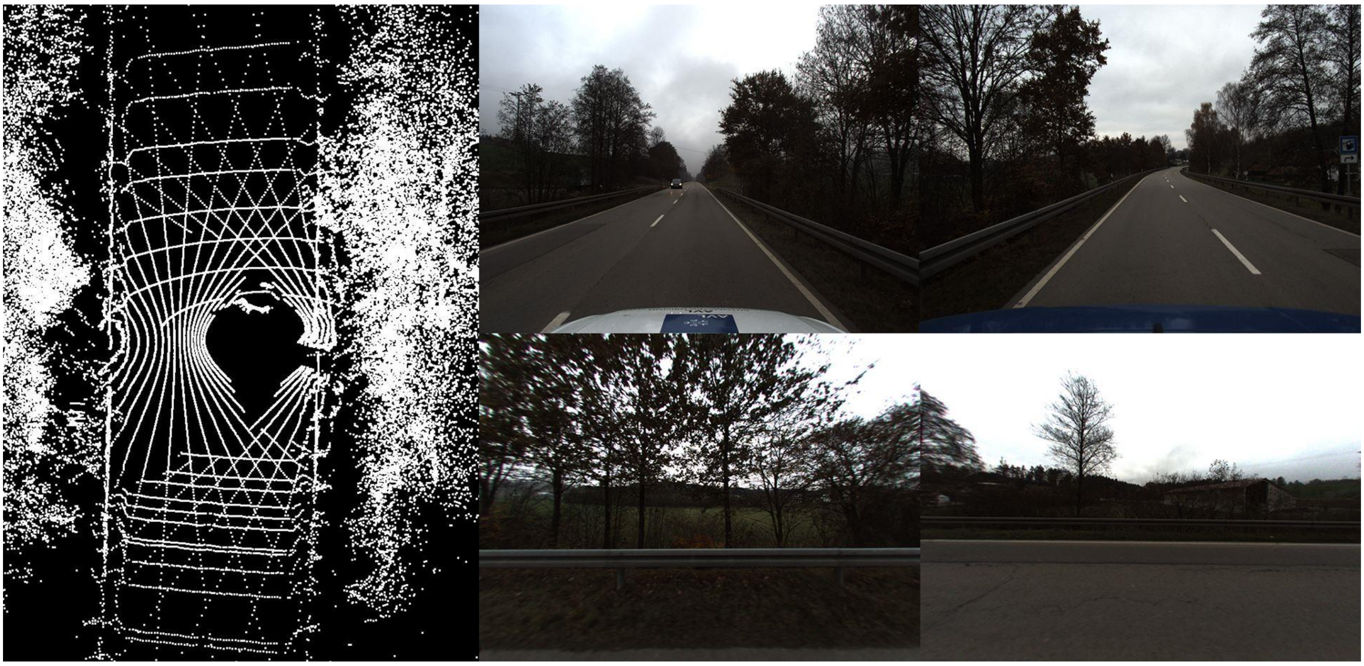

Figure 4 Example of unlabelled AVL LiDAR data (Left) and all four cameras (Right)

3 Annotation process

This chapter outlines the methods and tools employed for data annotation. The primary objective was to maximize the amount of annotated data from different weather conditions. Given that producing fine-tuned annotations is highly time-consuming, only the most relevant segments of the dataset were selected for manual refinement, while the remaining data were annotated using faster, automated methods. The following sections detail the procedures adopted for both automated labelling and manual fine-tuning.

3.1 Tools

The annotation fine-tuning work was conducted using the Segments.AI platform, which was introduced in 2020 by Otto Debals and Bert De Brabandere [1]. Segments.AI was specifically selected for the ROADVIEW project due to its user-friendly interface and efficient support for multisensory annotation.

The platform enables streamlined workflows for labelling complex multimodal data, including LiDAR sequences, camera images, and segmentation masks, making it well-suited for large-scale autonomous driving datasets. As an online platform, Segments.AI offers the benefit of being cross-platform and portable across various workstations. It was also essential to have technical support for the annotation tools.

Python was used to interface with Segments.AI and to upload and manipulate datasets in bulk. With the Segments.AI Python API, it is possible, for example, to normalise the sequence to start from the origin point in 3D space.

3.2 Labelling workflow

The manual labelling and fine-tuning of both LiDAR and Camera data were carried out using the Segments.AI platform. Datasets were uploaded to the Segments.AI environment using the Python SDK, which contained images alongside corresponding point cloud data in binary (.bin) format. The data provided from Task 4.2 was already prepared for Segment.AI, with no further alignment required. This setup enabled synchronized multi-sensor annotation. The labelled data was subsequently reviewed manually and returned for fine-tuning when necessary. Camera images and the majority of the AVL data were annotated automatically using algorithms for object detection and semantic segmentation. These automatically generated annotations were also subject to review, with selected portions uploaded for manual refinement. The steps of the process are illustrated in Figure 5.

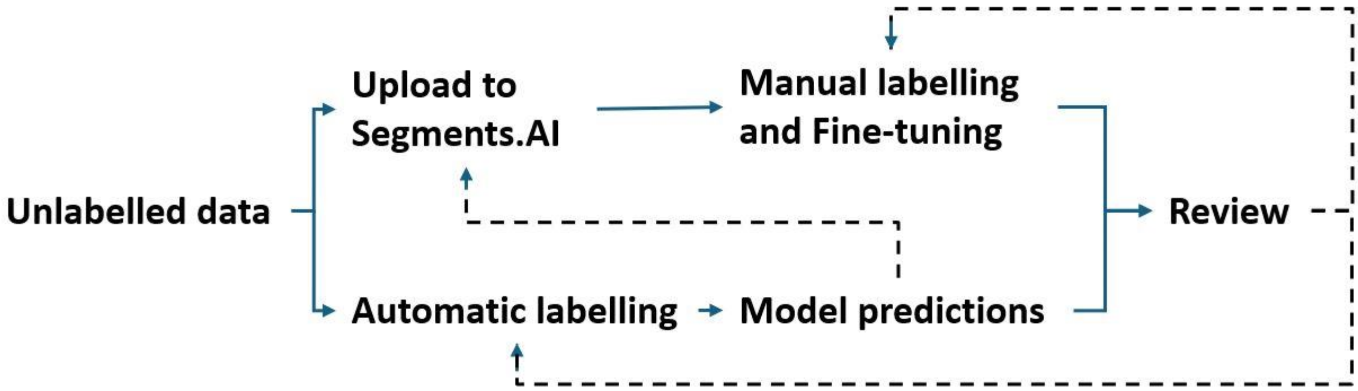

Figure 5 General overview of the different stages in the annotation process.

The following subsections describe the different labelling methods used in the annotation process.

3.2.1 Point cloud semantic segmentation

For 3D point cloud semantic segmentation, target sequences were divided into sample groups, and frames were labelled in a merged view to facilitate efficient labelling of static objects such as buildings, traffic signs, and vegetation. Identification of these objects was supported by the front camera view in combination with the dataset's geospatial reference in Google Maps, ensuring higher accuracy in visually complex or, e.g. snowy environments. The goal was to label each point in the point cloud with a distinct class, which are described in section 3.3. Labelling a 3D point cloud manually is very time-consuming, but it enables precise identification of each point using the accompanying RGB camera images as a reference. Segments.Ai's sequence-based annotation tools were further leveraged to maintain temporal consistency, particularly in scenes affected by occlusion or dynamic weather conditions.

A key task involved detecting and labelling noise introduced by adverse weather. Snowfall, rain, and fog were often captured as point clusters where particles were sufficiently dense. Using Segment.AI's annotation tools, these weather-induced artefacts were accurately labelled, as illustrated in Figure 6.

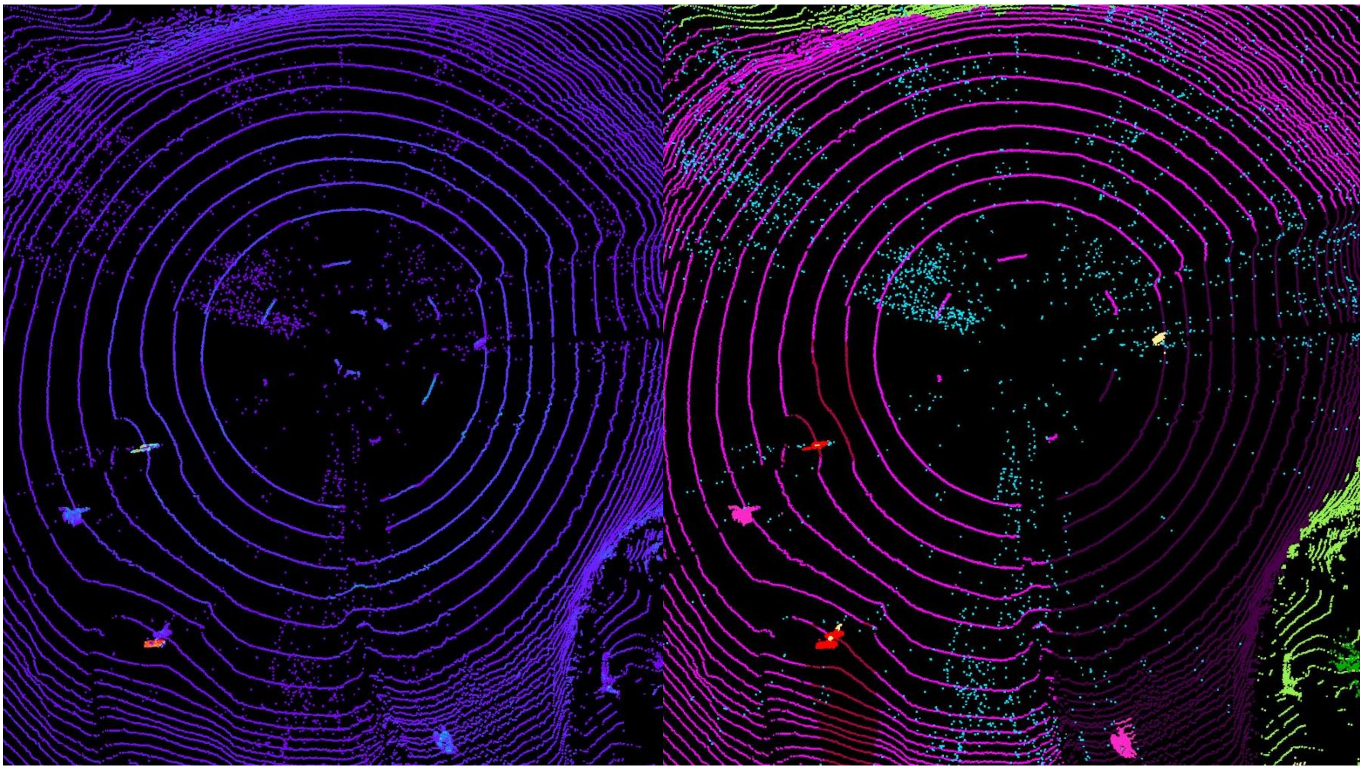

Figure 6 Original lidar frame from FGI with noise caused by snowfall (left) and labelled counterpart (right). Snowfall is visible as clusters of points near the center of the image (cyan color).

Furthermore, with the manual workflow, it was decided that points that could not be identified or were below the ground were marked as outliers. This included points that were caused by reflections and power lines over the driveway. Unlabelled included the points of the ego vehicle.

3.2.2 Point cloud 3D Bounding box annotation

3D Bounding box annotations were performed on the same samples that were used for semantic segmentation. A separate project was created specifically for bounding box annotation. This workflow allows for efficient reuse of the original data structure while applying a different annotation strategy.

In the bounding box dataset, annotations focused on three main object categories: stationary vehicles, moving vehicles, and pedestrians. All relevant vehicle classes, such as trucks, cars, and buses, were included along with their dynamic counterparts. An example of 3D bounding boxes is shown in Figure 7.

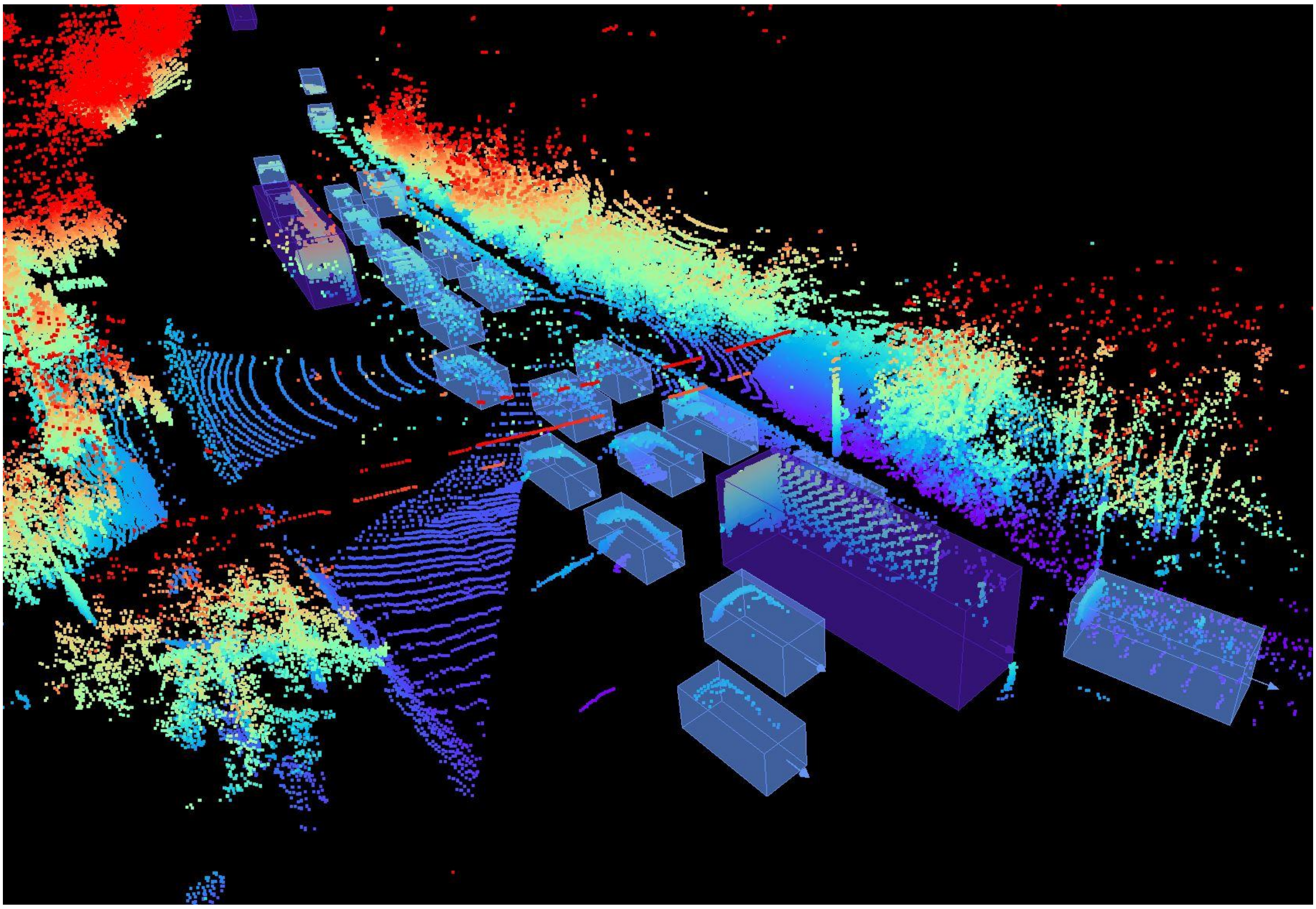

Figure 7 3D Bounding box development in Segments.AI

The annotation process was carried out using tools, like batch mode for dynamic objects. Initially, each visible vehicle and pedestrian was manually labelled in the frame. Afterwards, only minor adjustments were needed for the automatic tools to adjust the bounding box.

3.2.3 Automatic point cloud annotation

Automatic point cloud annotation was conducted in collaboration with Advian Oy. The objective was to annotate the data collected by AVL using the same categories applied in manual annotation. The process began with an evaluation of several deep learning frameworks for point cloud semantic segmentation pretrained on SemanticKITTI [2] data. Based on this assessment, SalsaNext [3] was selected and fine-tuned with manually annotated samples. After generating point-wise labels for all data points, automated post-processing was applied, including outlier removal, range filtering, and snow detection in the model outputs. The final step involved computing 3D bounding boxes and validating the generated results.

Given the low resolution of the left and right LiDAR sensors (16-channel) on the AVL vehicle, only the middle 32-channel LiDAR was used for automatic annotation. This decision was motivated by the fact that the pretrained SemanticKITTI [2] models were originally developed for higher-resolution (64-channel) LiDAR data

Segments.AI provides a built-in propagation feature that automatically interpolates bounding box positions across frames, allowing for more efficient annotation of moving objects while maintaining temporal consistency.

Also, automatically annotated data only included labels for snowflakes, and other weather-related noise was marked as outliers. For this reason, a portion of the data was sent to manual fine-tuning so the weather noise could be annotated as well.

3.2.4 2D Image annotation

Image annotations were performed on the same set of samples used for point cloud sequence segmentation. The annotation workflow is designed to balance automation and manual refinement, ensuring both efficiency and high-quality results across large volumes of data.

The workflow consists of the following four main steps:

1. Pre-annotation using object detection models

   Initial bounding box annotations were generated using the GroundingDINO model [4], focusing on two primary categories: vehicles and pedestrians. At this stage, all vehicles were treated as static objects, with no distinction made between moving and stationary instances.

2. Manual correction of bounding boxes

   Pre-annotations were refined by correcting class labels and adding missing objects. To support this, a custom bounding box annotation application was developed using:

   - Gradio[5] with a box annotation plugin
   - Hugging Face Transformers [6] for local model inference
   - GroundingDINO[4] and YOLO [7] models to assist with object detection, class search, and temporal interpolation between frames

Figure 8 below illustrates the custom bounding box annotation tool. This tool enables annotators to manually correct errors made by automated annotation models, ensuring higher accuracy and consistency in the final dataset.

Figure 8 Custom bounding box annotation tool

3. Semantic and instance segmentation

   Mask2Former[8] was used to generate semantic segmentation masks. Subsequently, SAM 2.1 (Segment Anything Model) [9] was applied for instance-level segmentation, and the outputs were combined to produce panoptic segmentation results. Figure 9 below illustrates how semantic and instance segmentation results are visualized within the annotation interface.

Figure 9 Instance and semantic segmentation combined.

4. Final upload and correction in Segments.AI

   The resulting annotations, including segmentation bitmaps, were uploaded to the Segments.AI platform. Model prediction errors were manually corrected using the segmentation interface to ensure label accuracy. Due to the variety of detection algorithms used, some objects were prompted in separate parts and mapped to the closest SemanticKITTI [2] class. Non-instanced annotations had to be fixed by deleting false detections and combining certain objects, like the lamp and pole parts of a street light.

   Using tools reminiscent of image editing programs, the already annotated objects were locked, and fine boundaries were drawn on the edges of trees and terrain. This way, the rest of the image was left with the machine annotations, which could also be fine-tuned.

3.3 Categories

The datasets utilize the semantic class definitions introduced in the SemanticKITTI[2], which includes a total of 28 standard categories. However, in the context of the ROADVIEW project, an extended label taxonomy consisting of 35 classes has been adopted to better represent the complexity of real-world driving scenarios. These classes are used for both point cloud and image semantic segmentation.

This extended set introduces dedicated classes for dynamic objects, such as moving vehicles, cyclists, and pedestrians, enabling finer-grained semantic distinction between static and dynamic instances of the same object type. These additional categories are particularly important for training and evaluating perception systems in environments with varying traffic density and motion.

An important addition was the inclusion of dedicated classes for noise caused by adverse weather conditions. Thus, three new categories were created to represent noise caused by snow, rain and fog in the data. These categories were only added for the LiDAR 3D point clouds since the snow or rain particles are barely visible in the camera images (apart from raindrops or ice on the lens itself). This was particularly important for task 4.4 “Data Filtering.”

The full list of categories used in ROADVIEW for 3D point cloud annotations is presented in Figure 10 below.

Figure 10 List of all the classes used in the ROADVIEW datasets for 3D point clouds.

Since bounding box annotations are provided only for dynamic objects such as vehicles and pedestrians, only the corresponding semantic categories are used for both 2D and 3D bounding boxes. For automatically generated semantic segmentation masks of camera images, labels for sky, lane markings, and curbs were additionally included. While these categories are not annotated in the 3D point cloud data, they were deliberately incorporated into the image annotations.

4 Results

The task produced 10k frames of fine-tuned point cloud data, comprising 5k FGI and 5k AVL frames. Fine-tuned RGB camera images include 800 semantic segmentation and 3k 2D bounding box labels for AVL data. Additionally, 40k frames of AVL data were automatically annotated with lower precision.

For LiDAR point cloud frames, semantic label files were generated, assigning a class to each point in the cloud. These label files were stored in binary (.bin) uint8 format. Figure 11 shows the final class distribution for all points in the fine-tuned data.

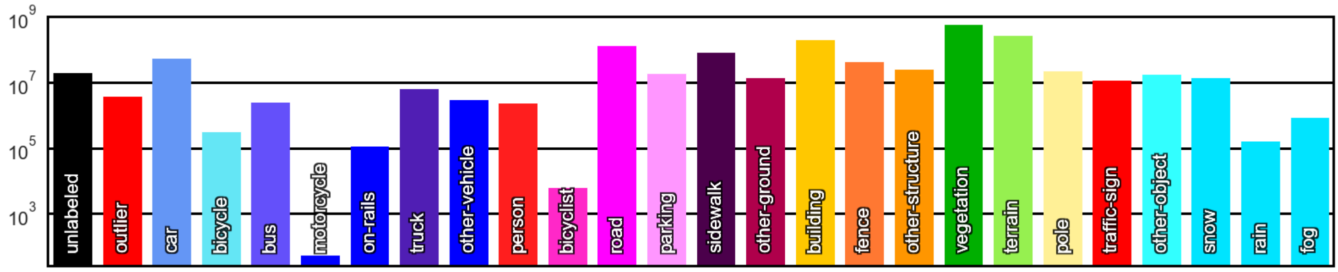

Figure 11 Distribution of annotated points per class in the fine-tuned point clouds.

In addition, all 3D bounding boxes were exported in (.txt) format using a 7-degree-of-freedom (7-DoF) representation. In this format, the coordinates (x, y, z) define the center position of the bounding box, (dx, dy, dz) represent its length, width, and height, and yaw specifies its orientation. Additionally, a class ID was assigned to each bounding box. Figure 12 demonstrates the class distribution for all 3D bounding boxes in the fine-tuned data.

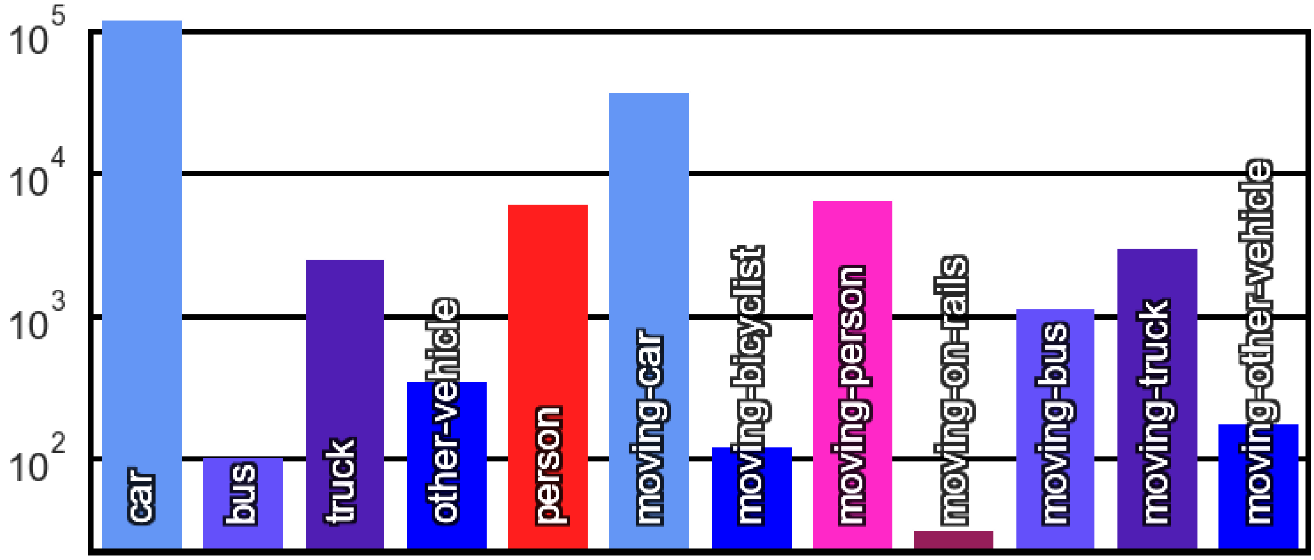

Figure 12 Distribution of 3D bounding boxes per class in the fine-tuned data

For camera images, panoptic labels were saved in (.png) format, where pixel values encode the instance ID. Each set of panoptic labels was accompanied by a JSON file specifying the mapping between instance IDs and category IDs. In addition, 2D bounding box annotations were provided in the form (x1, y1, x2, y2), where (x1, y1) correspond to the top-left corner and (x2, y2) to the bottom-right corner of the bounding box.

4.1 Benchmarking

Benchmarks were done using an autonomous driving dataset called Snowy Scenes [10], which was released by the ROADVIEW consortium during Task 4.5. The dataset included the winter data captured by FGI in Espoo, Finland (Figure 13). Three different perception tasks were benchmarked: 3D object detection, 3D semantic segmentation and 3D point cloud denoising.

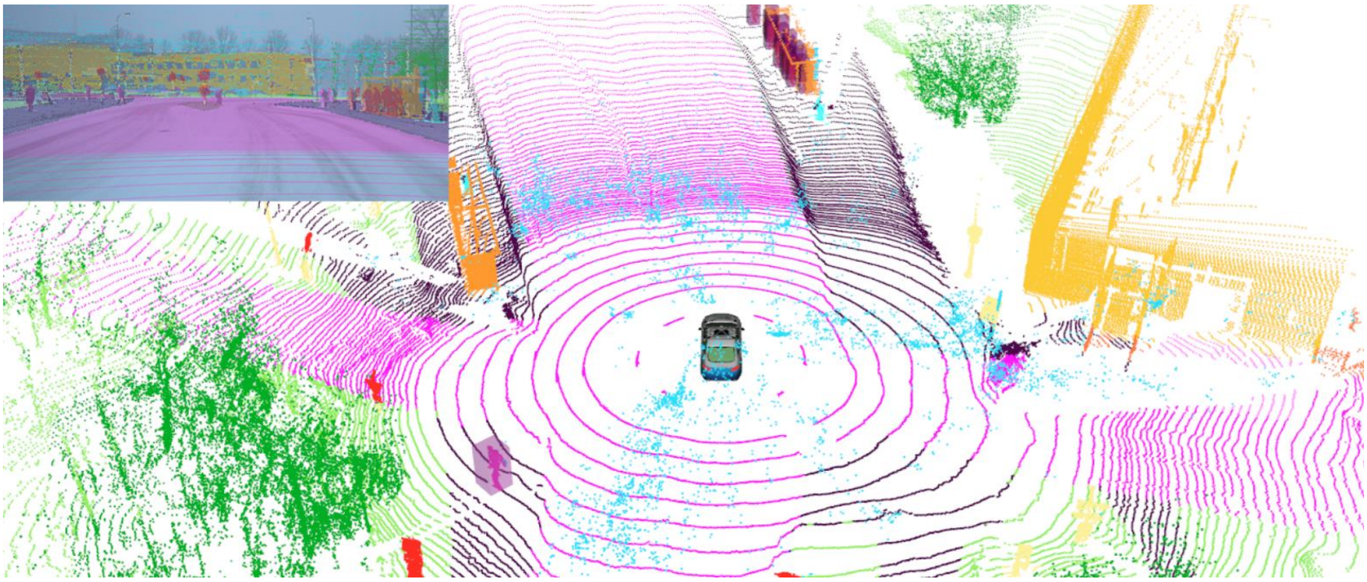

Figure 13 Example image from the Snowy scenes dataset showing the labelled LiDAR point cloud and front camera image with the point cloud cast on top of it. Snow noise is labelled with cyan color.

The first benchmark studied 3D semantic segmentation, where the objective was to assign a unique class for each point in the 3D point clouds. The benchmark included 3 different methods: point-based PTv3 [11], projection-based SalsaNext [3] and voxel-based Cylinder3D [12]. Benchmark evaluated the Intersection-over-Union (IoU) and mean IoU (mIoU). Results show that voxel-based Cylinder3D achieved the highest mIoU of 78.6, followed by PTv3 (73.6) and SalsaNext (68.1) [3]. Results indicate that voxel-based methods are more effective for semantic segmentation than point-based methods or range-view presentation in SalsaNext [3].

The labelled snow noise in the FGI data was used for benchmarking denoising in point clouds. The benchmark focused on evaluating both supervised and unsupervised models. The results showed that the unsupervised models performed poorly compared to the supervised counterparts. This unexpected result is likely caused by the fact that the architectural design of some of the networks isn’t suited for high-resolution LiDAR used in FGI data. However, semantic segmentation models provided good results, as shown in Table 1.

Model	Type	Precision↑	Recall↑	F_1↑	IOU↑
SalsaNext	SL	72.7	84.9	78.3	64.3
Cylinder3D	SL	88.8	75.0	81.3	68.5
3D-OutDet	SL	93.2	65.0	76.6	62.1
DROR	UN	7.2	64.5	12.9	6.9
SSR	UN	37.3	64.8	47.4	31.0
LiSnowNet-L1	UN	26.8	23.0	24.7	14.1
LiSnowNet-L2	UN	30.7	22.2	25.7	14.7
3D-UnOutDet	UN	80.2	64.1	71.2	55.3

Table 1 Results of the snow noise removal on the Snowy Scenes test split. SL stands for supervised and UN for unsupervised learning. The symbol ↑ indicates that higher is better.

3D object detection benchmark evaluated three different LiDAR-based object detection methods for their performance with the Snowy scenes data. This included BEV (Birds-eye-view)-based, voxel-based and point-based methods. The benchmark also assessed multimodal detection using PoinPainting [13] with front-view RGB camera pixels. The results indicated that due to the noise caused by snowfall in the data, when applying the PointPainting [13] method, a significant performance boost is gained over the other methods. This highlights the importance of sensor fusion. With LiDAR-only methods, voxel-based methods seem to work the best, as shown in Table 2.

Method	Modality	AP↑	ATE↓	ASE↓	AOE↓
PointPillars [36]	L	75.4	0.25	0.071	0.57
CenterPoint [37]	L	69.4	0.26	0.082	0.41
TransFusion-L [38]	L	79.7	0.26	0.14	0.57
VoxelNeXt [39]	L	82.1	0.20	0.073	0.58
IA-SSD [40]	L	78.9	0.23	0.062	0.46
PointPainting (Centerpoint) [41]	LC	86.4	0.18	0.064	0.27

Table 2 Results of the Object detection performance on the Snowy Scenes test set.

The results of the extensive benchmarking show that with state-of-the-art models, voxel-based methods are superior in object detection, semantic segmentation and noise removal. Results also show that unsupervised point cloud denoising methods are not resolution-agnostic, unlike the voxel-based methods.

4.2 Weather noise

The annotated AVL data included various adverse weather conditions. To ensure a high-accuracy annotation for the noise in the 3D point clouds, selected parts were manually fine-tuned. Figure 14 shows the results of the fine-tuned annotations in foggy, rainy and snowy conditions.

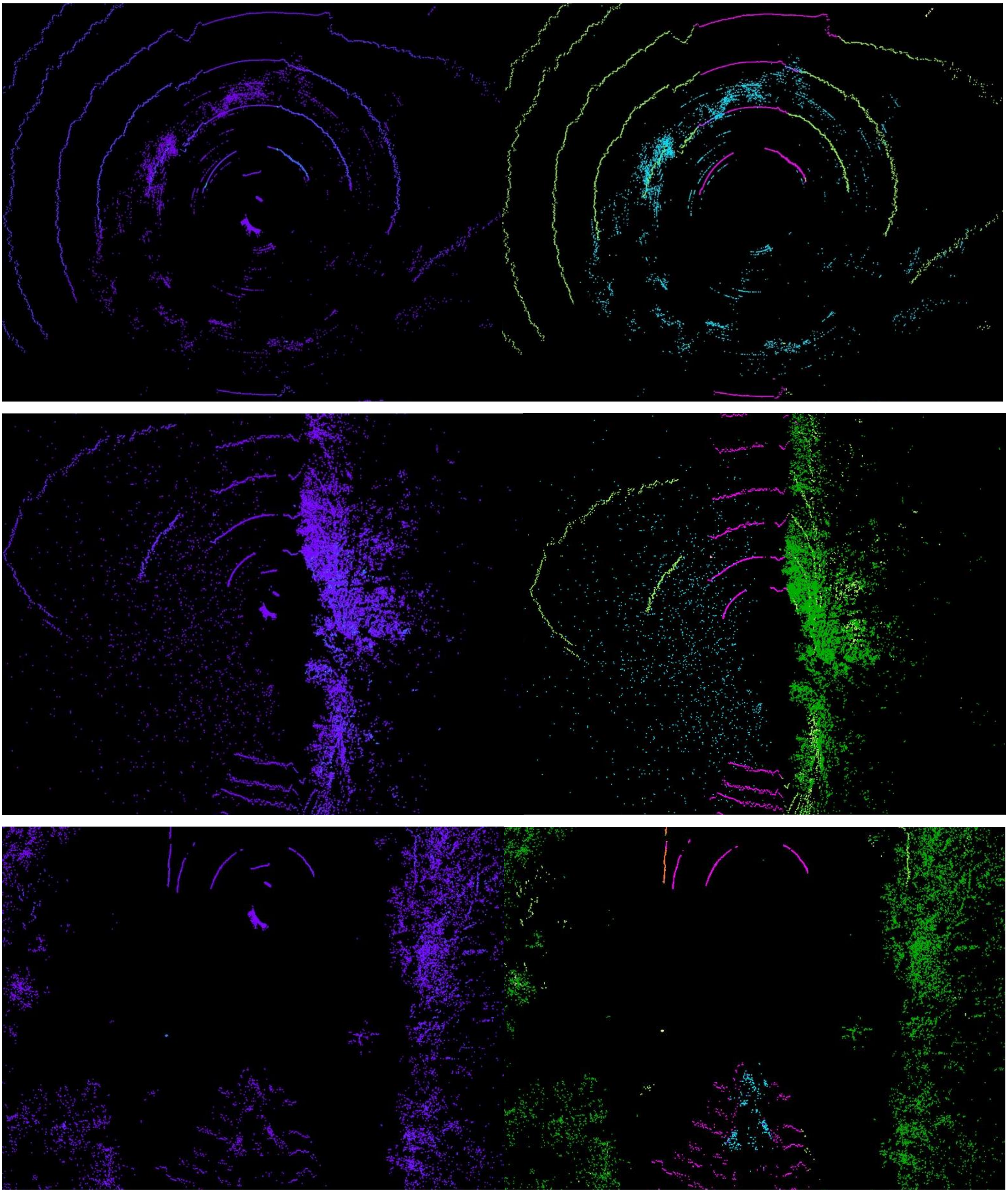

Figure 14 An example collection of images demonstrating what the labelled noise caused by adverse weather looks like in the point clouds. Images to the left show the original point cloud, and right side the labelled counterpart. The top images show noise in heavy fog, the middle shows snowfall, and the bottom shows road spray in rainy weather.

Although the annotated point cloud data covers a wide range of challenging weather conditions, most data was annotated for snowy conditions. Table 3 shows the number of frames included in each weather category.

Table 3 The table shows how the annotated frames were distributed according to weather conditions.

Weather condition	FGI (number of frames)	AVL (number of frames)
Rain	-	1400
Fog	-	900
Snow	5000	1100
Cloudy	4000	1000
Low-light	-	700
Clear	1000	-

The AVL RGB camera data contains both 2D bounding box annotations and semantic segmentation for all scenes, with the exception of the snowy night sequence. Due to severe underexposure, the images from this sequence were unsuitable for semantic segmentation; however, 2D bounding boxes for vehicles were generated where feasible. The fine-tuned semantic segmentation dataset consists of 200 sequential frames captured under cloudy conditions across all four cameras, resulting in a total of 800 refined images. The remaining images were annotated automatically, achieving a high level of precision. Some limitations were observed in cases where raindrops or ice on the camera lens affected image quality, as illustrated in Figure 15.

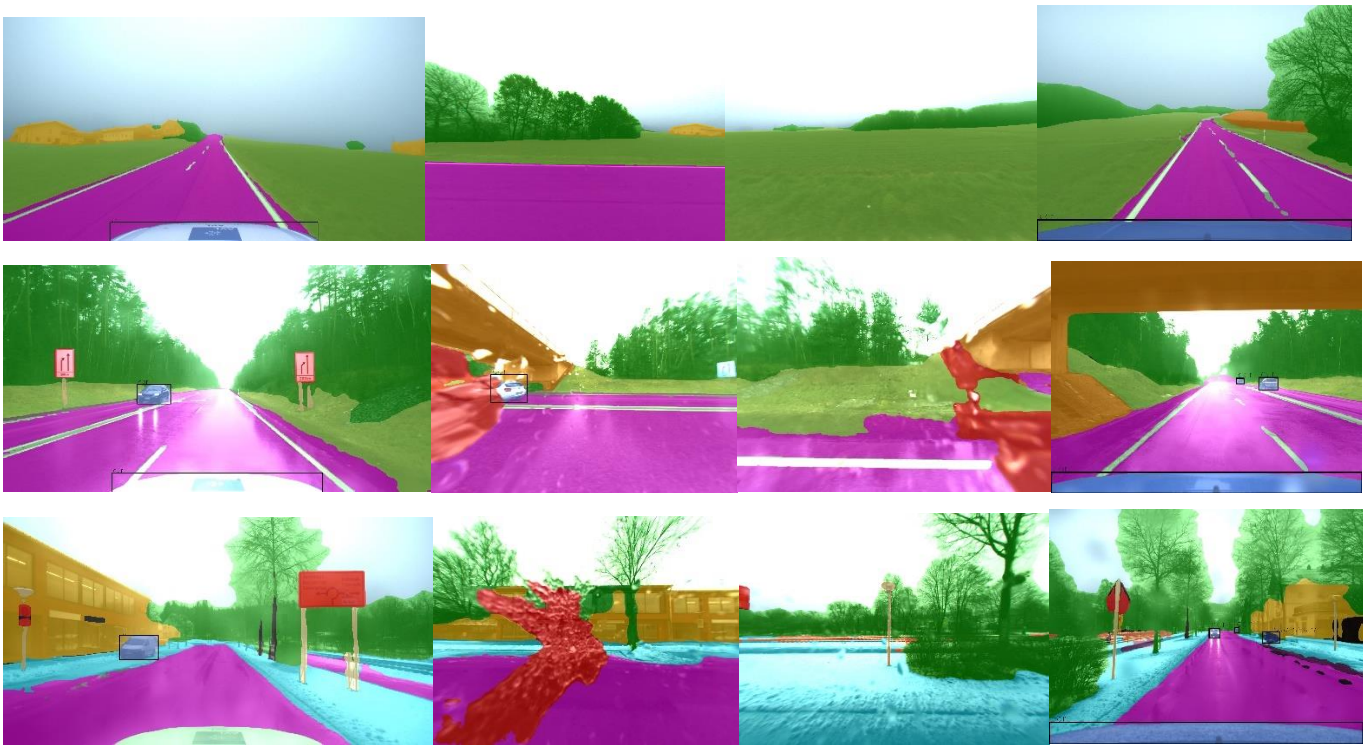

Figure 15 An example of automatically annotated images under three different weather conditions is presented. The top row illustrates results in foggy conditions, where no issues were observed. The middle row shows raindrops on the left and right RGB camera lenses, leading to false detections (highlighted in red). A similar effect is evident in the bottom row, where ice accumulated on the lens of the left camera during snowy conditions.

Although ice and raindrops obstructing the lens could, in principle, be labelled, this was excluded from the annotations due to the time-intensive nature of image semantic segmentation. Moreover, snowy scenes posed the greatest challenge for automatic annotation and would require substantial manual fine-tuning to achieve reliable results.

Conclusions

This deliverable resulted in a substantial amount of annotated high-quality multimodal sensor data that captures a wide range of traffic scenarios under diverse weather conditions, including rain, snow, fog, and low light. The data can be accessed through the ROADVIEW project website (https://roadview-project.eu/).

The labelled data was divided into two parts, of which the first one includes highly fine-tuned data and the latter weaker automatically generated annotations. The fine-tuned part includes a total of 10k frames of 3D point cloud data with high-quality point-wise labels and 3D bounding boxes. 800 frames of this data also include fine-tuned panoptic segmentation and 3000 frames with 2D bounding boxes for RGB camera images. The second part adds 40k frames of automatically generated point cloud and camera labels with lower precision.

The data was successfully annotated using the Segments.AI platform as well as fine-tuned AI models. The annotated point cloud data includes 35 label categories following the Semantic KITTI format but expanded with new classes, including three label categories for noise caused by adverse weather. This provides valuable ground-truth data for the development and testing of machine learning algorithms used, for example, for noise filtering.

Both fine-tuned and automatically annotated data were manually reviewed by visually inspecting the quality. The ROADVIEW consortium also released a joint paper that included benchmarks for the fine-tuned data, testing different 3D perception tasks with state-of-the-art models.

5 References

[1] Segments.ai. 2024. Our History. Referenced 15.07.2025. https://segments.ai/company/

[2] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, J. Gall, C. Stachniss, “Towards 3D LiDAR-based semantic scene understanding of 3D point cloud sequences: The SemanticKITTI Dataset,” The International Journal on Robotics Research. Volume 40, 8–9, 959–967. 2021.10.1177/02783649211006735

[3] T. Cortinhal, G. Tzelepis and E. E. Aksoy, “Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds,” in International Symposium on Visual Computing, 2020.

[4] S. Liu, Z. Zeng, T. Re, F. Li, H. Zhang, J. Yang, Q. Jiang, C. Li, J. Yang, H. Su, J. Zhu, L. Zhang, ”Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection,” arXiv:2303.05499, 2024.

[5] A. Abid, A. Abdalla, A. Abid, D. Khan, A. Alfozan, J. Zou, “Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild,” arXiv:1906.02569, 2019.

[6] T. Wolf, et al. "Huggingface's transformers: State-of-the-art natural language processing." *arXiv preprint* arXiv:1910.03771, 2019.

[7] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", arXiv:1506.02640, 2016.

[8] C. Bowen, et al. "Masked-attention mask transformer for universal image segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022.

[9] A. Kirillov et al, “Segment Anything”, arXiv:2304.02643, 2023.

[10] I. Gouigah, A. M. Raisuddin, T.-T. Ngo, P. Litkey, P. Manninen, H. Hyyti, J. Maanpaa, H. Tuomas, V. Sofie, N. Kalle, L. Pernickel and E. E. Aksoy, “Snowy Scenes: A Multimodal Dataset Toward Snow-tonomous Vehicles,” IEEE RA-L (Under Review), 2025.

[11] X. Wu, L. Jiang, P.-S. Wang, Z. Liu, X. Liu, Y. Qiao, W. Ouyang, T. He, and H. Zhao, “Point transformer v3: Simpler faster stronger,” in Proceedings of the IEEE/CVF CVPR, 2024, pp. 4840–4851.

[12] H. Zhou, X. Zhu, X. Song, Y. Ma, Z. Wang, H. Li, and D. Lin, “Cylinder3d: An effective 3d framework for driving-scene lidar semantic segmentation,” arXiv preprint arXiv:2008.01550, 2020.

[13] S. Vora, A. H. Lang, B. Helou, and O. Beijbom, “Pointpainting: Sequential fusion for 3d object detection,” in IEEE/CVF CVPR, 2020, pp. 4604–4612.