ROADVIEW

Robust Automated Driving in Extreme Weather

Project No. 101069576

Deliverable D4.6

Final readiness assessment of specific datasets

WP4 - Data logging, filtering, annotation, and quality assurance

Authors	Ian Marsh (RISE)
Lead participant	RISE, Research Institutes of Sweden, AB
Delivery date	31 October 2024
Dissemination level	PU – Public
Type	R – Document, report

Version 01

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
Ian Marsh (RISE)	Draft V0.1	01/10/2024
Ian Marsh (RISE)	Version V0.9	16/10/2024
Tuomas Herranen (LUA)	First review	22/10/2024
Topi Miekkala (VTT)	Second review	28/10/2024
Ian Marsh (RISE)	Final version	29/10/2024
Johannes Ripperger (accelCH)	Final checks and formatting	31/10/2024

Partner short names

HH	Hogskölan Halmstad
LUA	Lapin Ammattikorkeakoulu OY
THI	Technische Hochschule Ingolstadt
CE	Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement
RISE	RISE Research Institutes of Sweden AB
FGI	Maanmittauslaitos – Finnish Geospatial Research Institute
accelCH	accelopment Schweiz AG
VTT	Teknologian Tutkimuskeskus VTT OY

Contents

Abbreviations

AD	Automated Driving
AI	Artificial Intelligence
AV	Automated Vehicle
CAV	Connected Automated Vehicle
DDI	Direct Data Injection
DRL	Data Readiness Level
EU	European Union
GA	Grant Agreement
GNSS	Global Navigation Satellite System
HD	High Definition
HD MAP	Detailed and highly accurate digital representation of the real-world environment
HiL	Hardware-in-the-Loop
IMU	Inertial Measurement Unit
ML	Machine Learning
MRM	Minimum Risk Manoeuvre
NER	Named Entity Recognition
NR-PCQA	No-Reference Point Cloud Quality Assessment
ODD	Operational Design Domain
OEM	Original Equipment Manufacturer
OTA	Over-the-Air
PCQA	Point Cloud Quality Assessment
PSNR	Peak Signal-to-Noise Ratio
RCS	Radio Cross Section
SSIM	Structural Similarity Index
TRL	Technology Readiness Level
ViL	Vehicle-in-the-Loop
VMAF	Video Multi-Method Assessment Fusion
WP	Work Package

Executive summary

Let's start with the title 'Final Readiness of Specific Datasets'. 'Final' means the second report in a 2-stage evaluation. 'Readiness' means we add a quality stamp to a dataset. 'Specific Datasets' means data gathered, curated, processed and released from, and by, the ROADVIEW project.

The datasets are provided by Finnish Geospatial Institute (FGI) in Finland and the Technical Institute of Ingolstadt (THI) in Germany. Using the techniques in this deliverable, we assign a score or Readiness, 1 (low) to 9 (high) to the data under study. We call the result of the evaluation a Data Readiness Level, or DRL. DRLs should be the seen as the 'data equivalent' of Technical Readiness Levels. In theory, one should be able to use TRL and DRL together to evaluate a system that makes use of data, such as Artificial Intelligence. Specifically, the ROADVIEW project uses Machine Learning, or ML, in the context of Autonomous Driving.

Autonomous Driving = Technology Readiness Level + Data Readiness Level

We have rated the THI dataset as **8.0** and the FGI dataset as **7.0**.

The rest of this report contains:

- How singular DRL values such as 8 and 7 were assigned.
- Improvements to the DRL assignments
  - Additional techniques for point-cloud quality
  - Weight assignment for modalities
  - Smoothing data
- Cross-verifying quality across modalities, simultaneously
- 1 additional dataset under review

Objectives

The objectives are an evaluation of the data quality supplied by the project. Data quality is a relatively new concept in computer science and IT, therefore a second objective how the quality is derived, especially with respect to autonomous driving. Given the novelty of data quality, it is important we explain the philosophy, methods and tools for experimentation and modification should improvements be derived from this work. We expect this and know of at least 1 institute (see Conclusions) using the DRL concept. A further objective will be a public release of the DRL software (D4.7).

Methodology and implementation

The main methodology is to define a simple, working assessment of the datasets used in ROADVIEW. Irrespective of the algorithms' performance (F-Scores, Confusion matrices and so on) the success of autonomous vehicles will largely depend on the quality of data used. Therefore, the data should be of sufficient quality for the scenarios and settings defined in the project. Since the goal of the project is to enable autonomous vehicles to operate in rain and snow, the coverage, quality, and settings will be a determining factor in the projects' success.

Quantitively, where the operation of the vehicle perception is too low, we can pinpoint at least where the data is/was insufficient using the main concept. Backtracing from an object misclassification (or missing at all) to the HD MAP, data source (Visible camera, Infrared, LiDAR, RADAR) to the pixels and potentially a sensor issue would be a truly valuable tool in this field. Comparing to a software development environment, we envisage the process akin to a debugger in code development. Figure 1 shows a simplified illustration of data under review. In the project data

accumulation (REHEARSE and FGI) and the DRL development was done in parallel. Therefore, these datasets were gathered before we could apply our DRL concept. That means the DRL level is static, improvements cannot be made to these datasets (nor should be as it would influence existing results). Therefore, an ongoing effort is to evaluate a dataset under collection. This is provided by VTT in the project. Note, this is not stated in the initial work plan, indeed is an extra. A second extra piece of work is an evaluation of synthetic data, more on this below.

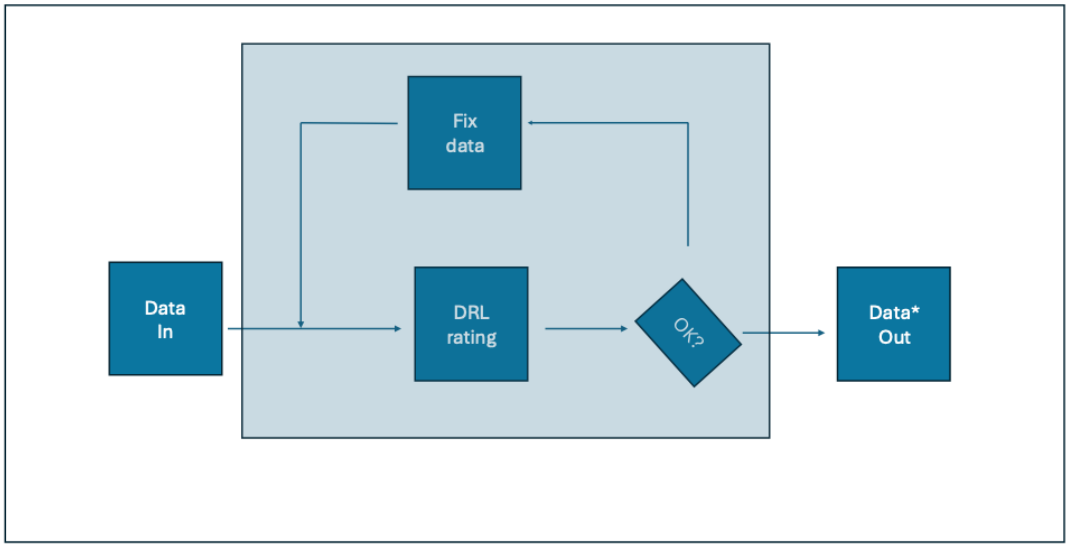


Figure 1. Using the DRL concept to improve data quality.

Outcome

The outcome of this second deliverable is an evaluation of the quality of two in-house produced datasets. Better quality evaluation of point clouds from LiDAR and 4G RADAR is included, as well as how to combine the quality of each modality. These methods are independent of the data we have access to here, so should be seen as a generic piece of original work. Furthermore, we deal with the time-varying aspects of image-lidar-radar quality during a drive. More significantly, a description of the set of tools used to do this evaluation and finally a description of a tool to find differences between the quality of modalities. Compared to the first deliverable (D4.5) regarding ROADVIEW data quality, we have refined the usage of these tools, added to them, as well added logging information to publicly available tools (a requirement).

Next steps

The next steps will be a release of our software to Github, an accompanying technical paper as well as the input evaluate improve cycle for 2 additional datasets.

Notation

Links in the text are underlined. Text in the `Consolas` font refers to a command that can be run, usually UNIX.

Part I: Preliminaries

What's in this deliverable?

This deliverable contains an evaluation of two datasets produced by the ROADVIEW project. A clear separation of the conditions inherent in the data. That is the controlled environment at the proving ground in Ingolstadt, Germany whilst the public roads of wintery conditions in Finland. Two methods for assessing the point cloud data which was not in the first incarnation of data quality work. An assessment of the contributions from the modalities, their relative magnitudes and time-varying nature is considered. Finally, a tool for comparing the quality of the modalities is explained. In a nutshell, we have evaluated two datasets and made improvements along our data quality journey. We have added some explanatory illustrations in this document that explains how we derive the data qualities of 7 and 8.

What's not in this deliverable?

Issues that are not included in this deliverable, are sensor calibration, cameras and LiDARs (though LiDAR calibration can be found in this Link). This is mostly due to the competence area, as well as access to specialised equipment. RADAR data is not discussed in detail this deliverable. We do not cover the preliminaries in the initial data readiness report (D4.5), nor a detailed coverage of the Finnish Dataset (FGI), that is in the first deliverable. Other project parts: hardware-in-the-loop, slipperiness estimation, ODDs and the architecture are not really covered. Explainable AI also is not really part of the Data Readiness but is included in a forthcoming technical paper. Basically, it is borderline case in terms of inclusion, for this deliverable.

A review of autonomous datasets (as of October 2024).

Recently companies, research institutions and universities have made their autonomous driving datasets open to the research community. A Medium post found and summarised 15, as of July 2021, by Alex Nguyen. **Audi A2D2** dataset 41K labelled images with 38 features, 2.3 TB split by annotation type, semantic segmentation, 3D bounding box. **ApolloScape**, 100K street view frames, 80k LiDAR point cloud and 1000km trajectories for urban traffic. 3D tracking annotations for 113 scenes and over 324,000 unique vehicle trajectories for motion forecasting. **Berkeley DeepDrive** 100k annotated videos and 10 tasks, 1000 hours driving, 100M frames plus geographic, environmental, and weather diversity. **Cityscapes** urban street scenes in 50 German cities. Semantic, instance-wise, and dense pixel annotations for 30 classes grouped into 8 categories, 5K images with fine annotations & 20K with coarse annotations. **Comma2k19**, 33 hours of commute time recorded on highway 280 in California. 1-minute scenes captured on 20km of highway between San Jose-San Francisco. Collected using comma EONs, i.e., a road-facing camera, phone GPS, thermometers, and a 9-axis IMU. **Google Landmarks** (2018) divided into two sets of images to evaluate recognition and retrieval of human-made and natural landmarks. 2M images of 30K unique world, 2019 saw Landmarks-v2, 5M images & 200K landmarks. **KITTI Vision data**, 2012**. LeddarTech** Dataset, 2021, cameras, LiDARs, radar, IMU + full waveform a 3D solid-state flash LiDAR sensor. Contains 29k frames in 97 sequences & 1.3M 3D boxes annotated. **Level 5 Open Data** (Lyft) 55K human-labelled 3D annotated frames, surface map, and an underlying HD spatial semantic map captured by 7 cameras and 3 LiDAR sensors. **nuScenes** dataset from Boston + Singapore using a full sensor suite, 32-beam LiDAR, 6 360° cameras and radars, the dataset with 1.44M camera images capturing a diverse range of traffic situations, driving manoeuvres, and unexpected behaviours. Examples are from clear weather night-time, rain & construction zones. **Oxford Radar RobotCar** Dataset contains 100+ recordings of a route through Oxford, UK, captured over 1 year. Captures different conditions, including weather, traffic & pedestrians + construction / roadworks. **PandaSet** the 1st open-source AV dataset for academic & commercial use. Contains 48K

camera images, 16K LiDAR sweeps, 28 annotation classes, and 37 semantic segmentation labels. **Udacity** Self Driving Car Dataset has open-sourced access to a variety of projects for autonomous driving, including neural networks trained to predict steering angles of the car, camera mounts, and dozens of hours of real driving data. **Waymo Open** Dataset is an open-source multimodal sensor dataset, covers a wide variety of driving scenarios and environments. It contains 1K types of different segments where each segment captures 20 seconds of continuous driving, corresponding to 200K frames at 10 Hz per sensor. A summary of datasets considered in a nuScenes paper is below. **WADS** (Winter Adverse Driving dataSet, 2021, by Kurup et al. in DSOR: A Scalable Statistical Filter for Removing Falling Snow from LiDAR Point Clouds in Severe Winter Weather. Collected in the snow belt region of Michigan's Upper Peninsula, WADS is the first multi-modal dataset featuring dense point-wise labelled sequential LiDAR scans collected in severe winter weather. Over 26 TB of multimodal data has been collected of which over 7 GB of LiDAR point clouds (3.6 billion points) have been labelled (semantic KITTI format). This outdoor dataset introduces falling snow and accumulated snow along with all the semantic KITTI classes to further AV tasks like semantic and panoptic segmentation, object detection and tracking, and localization and mapping in conditions of moderate to severe snow. The **Boreas** dataset from the university of Toronto, 2023, was collected by driving a repeated route over the course of 1 year resulting in stark seasonal variations. In total, Boreas contains over 350km of driving data including several sequences with adverse weather conditions such as rain and heavy snow. The Boreas data-taking platform features a unique high-quality sensor suite with a 128-channel Velodyne Alpha Prime lidar, a 360-degree Navtech radar, and accurate ground truth poses obtained from an Applanix POSLV GPS/IMU. We currently support live and open benchmarks for odometry, metric localization, and 3D object detection.

With respect to our contribution to the dataset ecosystem. The **ROADVIEW** project will release its dataset in 2025. Importantly from this work an internal quality stamp.

The ROADVIEW reference architecture (a summary of D2.3)

The work in this task is in essence adjunct to the architecture work done in the Architecture Reference Task, part of WP2. Figure 2 below shows the generic architecture ROADVIEW uses. It represents the core functional generic architecture of the ROADVIEW project. Novel ROADVIEW modules are shown in dark grey, and available OEM modules in light grey colours, respectively. There are 3 functionalities: Perception, Planning, and Control in red, purple, & blue. The **perception** block functions are for robust environment perception. Basically, filtering of noisy sensor data, low-level sensor fusion, object detection, free-space detection, weather-type detection, slipperiness detection, and visibility detection. The perception block receives input from various sensing modalities such as RGB cameras, LiDARs, RADARs, Thermal cameras, Inertial Measurement Units (IMU), and Global Navigation Satellite System (GNSS) modules. The **planning** block is for localisation, trajectory prediction & planning functionalities. The **controller** block focuses on weather-aware decision-making and control of the velocity and acceleration-related parameters. Each sensor reading is processed individually up to the low-level sensor fusion module, which feeds the downstream perception modules such as object detection. Dashed arrows in the figure represented sensor readings cleaned from noise (i.e., outliers) introduced by, for instance, falling snow particles or rain drops.

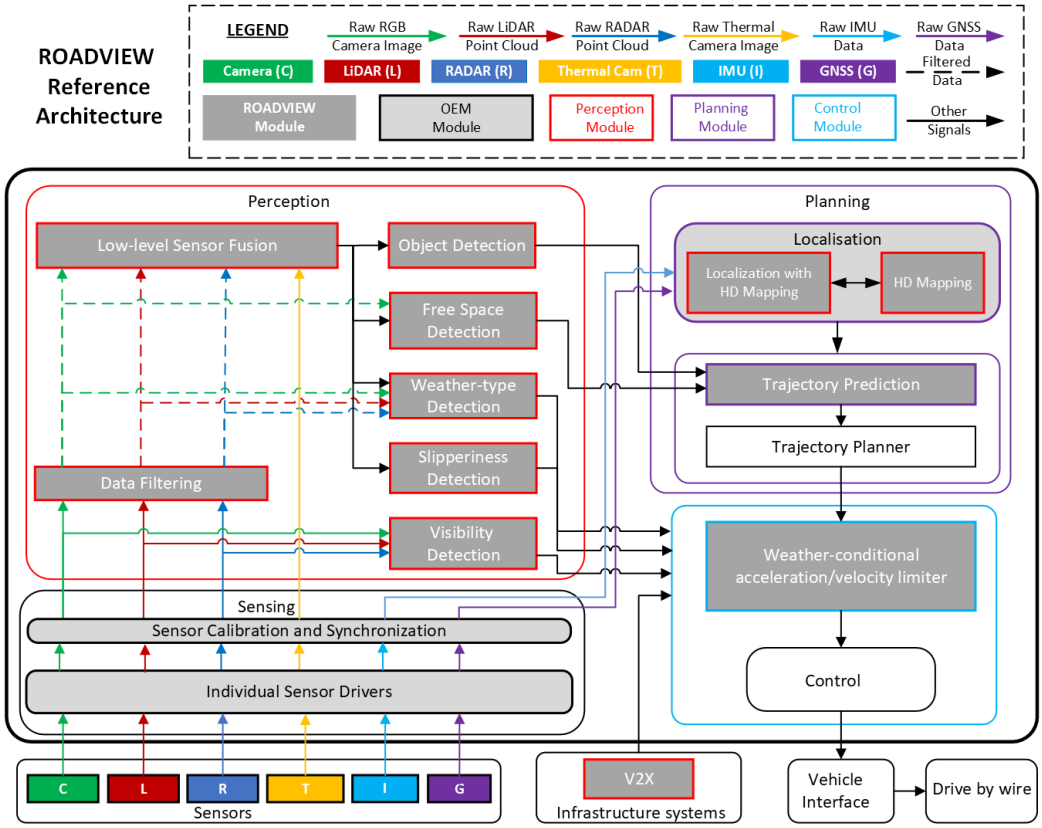


Figure 2 - The ROADVIEW reference architecture (source Deliverable D2.3)

The **reference architecture** in Figure 2 drives our Readiness Level development. Note, Autonomous driving can be thought of Sensing-Perception-Planning. The Readiness Levels impacts most on the Sensing and Perception stages. However, should the data be incorrect, misleading, missing, etc. When driving, i.e. the model cannot cope with the environment, it is conceivable that the planning stage be affected, a minimum risk manoeuvre should be invoked. To reiterate, the DRL work mostly affects the Sensing (lower DRLs) and Perception modules (higher DRLs) in training, and the Planning stage whilst driving.

Figure 3 shows the DRLs, and the sensor calibration and synchronisation above correspond to level 0 in our DRLs. That is before data is gathered, and per modality, the calibration needs to be done. We mean extrinsic calibration with respect to calibration in this case. Data filtering maps to data quality assessment in our DRL schema, that is levels 2-7. These are described in more detail below with an example. We do not look at the quality of the fused modalities in DRL in this first report, rather per modality.

The **project methodology** with respect to data is to find the best from each dataset and as a final dataset release a ROADVIEW dataset. Indeed, one of the objectives of ROADVIEW is 'stamp' the dataset with a quality value from 1 (lowest) to 9 (highest). We need to state here, the value attributed to a dataset, or indeed amalgamation should not be used to compare datasets at this stage. That is because the setup, calibrations, vehicles, conditions will not be the same from dataset to dataset. However, we will look at being able to provide DRL scores per configuration, in the latter parts of this task.

There are **no specific external datasets** required to assess the quality, nor specific modalities. Should there be more than one device type, the values will be averaged (see DRL calculation below) and should a modality is missing, their contributions or weights, are reassigned to the existing modalities.

Data sources in autonomous driving

From a dataset point of view the simplest view of the project is:

Data in ROADVIEW project = VISIBLE Camera data +
Thermal Cameras +
LiDAR data +
RADAR data +
IMU sensor positioning data +
GNSS data

To be clear, the project is large, complex and has many facets, weather modelling, testing, system integration, testing and finally working demonstrations using real trucks. That said, a vehicle that uses sensors to navigate, avoid and find its destination with respect to the road and weather conditions will be make heavy use of data, especially in the training phase. That is where the test vehicle performs journeys (see below) to learn of all possible situations it will encounter when licensed and driving passengers around. The goal from autonomous driving is to produce a “picture” of the environment, often called an **HD map**. In ROADVIEW this is discussed in additional detail in Tasks T5.4 and T8.3.

Part II: Sensors and data

This table persists from the first (D4.5) deliverable for reference. The reason is to show the modalities, the parameters and their range, for the reader to understand their role in general, and in the ROADVIEW project. The Data Readiness Level will be dependent on the setup, this is one example of what a setup will be.

Table 1. Minimum requirements on ROADVIEW sensor set (see Task 4.1).

Sensor	Used in / Required for ROADVIEW Innovation	Parameter	ROADVIEW minimum sensor requirement
RGB Camera	Data Filtering Low-Level Sensor Fusion Object Detection Free Space Detection Weather-Type Estimation Slipperiness Estimation Visibility Estimation	Number of Sensors	>= 3
		Field of View	TBD
		Resolution	>= 2MP
		Framerate (Hz)	>= 10
		Sensor Model	RGB Camera
		Placement	Front, (Rear optional), both Sides
LiDAR	Data Filtering Low-Level Sensor Fusion Object Detection Free Space Detection Weather-Type Estimation Slipperiness Estimation Visibility Estimation HD Mapping Localization with HD Mapping	Number of Sensors	>= 1
		Field of View	360° horizontal >= 40° vertical
		Resolution	128 bins
		Framerate (Hz)	>= 10
		Measurement Range (m)	>= 80
		Sensor Model	No Requirement
		Placement	Top (to minimize obstacles)
RADAR	Data Filtering Low-Level Sensor Fusion Object Detection Free Space Detection Weather-Type Estimation Visibility Estimation	Number of Sensors	>= 1
		Field of View	TBD
		Resolution	TBD
		Framerate (Hz)	>= 10
		Measurement Range (m)	>= 100
		Sensor Model	No Requirement
		Placement	Front
Thermal Camera	Low-Level Sensor Fusion Object Detection Free Space Detection Weather-Type Estimation Slipperiness Estimation	Number of Sensors	>= 1
		Field of View	TBD
		Resolution	TBD
		Framerate (Hz)	>= 10
		Sensor Model	No Requirement
		Placement	Front
GNSS	HD Mapping Localization with HD Mapping	Number of Sensors	1
		Framerate (Hz)	>= 10
		Sensor Model	No Requirement
		Placement	Antenna on Top
IMU	HD Mapping Localization with HD Mapping	Number of Sensors	1
		Framerate (Hz)	>= 100
		Sensor Model	No Requirement
		Placement	No Requirement

HD Mapping

An HD (High Definition) map is a detailed and highly accurate digital representation of the real-world environment, primarily developed for autonomous driving systems. These maps go beyond the traditional navigation maps that we use in our smartphones or cars. Here's what distinguishes HD maps from standard maps:

- Detail and Precision: HD maps provide centimetre-level precision, enabling self-driving cars to understand their surroundings better and make informed decisions.
- Layers: Unlike standard maps that might only provide roads and points of interest, HD maps come with multiple layers of information, including:
- Road Profile: This includes details like road curvature, gradient, and width.
- Lane Information: It offers specifics about each lane, its boundaries, type (e.g., turning lane, straight lane), and associated rules.
- Traffic Signs & Signals: HD maps will have exact positions of all traffic signs, signals, and other regulatory markers.
- Infrastructure Details: This may include crosswalks, barriers, guardrails, pedestrian areas, and more.

Dynamic Updates: Given the rapidly changing nature of roads due to construction, accidents, or other events, it's crucial for HD maps to be frequently and dynamically updated. Some systems aim to update in near real-time. 3D Representation: While many standard maps offer 3D views for a better user experience, HD maps can provide a true 3D representation, accounting for elevation changes and including structures like bridges, tunnels, and buildings. Sensors & Integration: HD maps are typically developed considering the suite of sensors (like LIDAR, radar, cameras) on autonomous vehicles. The data from these sensors can be cross-referenced with HD maps for tasks such as precise localisation. HD maps play a crucial role in making autonomous driving safer and more reliable. By giving vehicles, a comprehensive understanding of their environment, they help ensure that the vehicle can handle complex driving scenarios even if sensors face temporary obstructions or difficulties.

A ROADVIEW data pipeline

Requirements (WP2) ⇒ Sensor capture (WP4) ⇒ Format wrangling (WP4) ⇒ Validation checks (WP4) ⇒ Fused data (WP4) ⇒ Model training (WP5) ⇒ Object detection (WP5)

Data readiness, or quality can be defined for the online or offline case. In the online case could be a situation that was not anticipated or not seen before causing the vehicle to make a poor decision (the classic Uber vehicle crash in the US, unfortunately killing a pedestrian). The offline case is pertinent to highlighting of internal practices, where the software system can be ‘debugged’ from a data point of view. Should object detection fail we can look at the causes due to data issues. Note, the algorithms themselves could be issue. These should be tested on a ‘perfect dataset’ possibly machine generated (see below).

An alternative view is to use the DRL in assessing datasets gather in progress and feedback to the gathering.

Data quality & autonomous vehicles: a short (& selected) state of the art

Since autonomous vehicles are essentially ‘computers on wheels’ data gathering, processing & navigating there is a substantial amount of material. Often citating each sensor type, and increasingly more often, fused data. The photometric society of America produced a paper at the link about Geometric Inter-Swath Accuracy and Quality of LiDAR Data, which is not related to driving but an interesting take on LiDAR data quality. Data quality issues can be found in [1-3]. Within the consortium [4-6] cover vehicle segmentation in LiDAR point clouds. Adverse weather autonomous driving is in [7-13]. Each public datasets have its own set of publications, where nuScenes and mmDetection3D attempt to summarise the others, NuScences ‘corruptions’ a method is close to our work [20]. Around ML pipelines references [15-18] are relevant, where we point to [15] as a lightweight approach. [14] is video quality estimation done by video quality experts, rather than vehicle developers or ML specialists. Point cloud quality is an active area, not only for LiDAR data, but also for 4G RADAR as these sensors produce point clouds. Corruptions

feature in other works, too, notably the Robo* data from the National University of Singapore and the European MultiCorrupt work [29]. Note, we restrict our referenced work to papers with open-source code available. We weight slightly to code that is solid, robust and maintained. Unfortunately code co-published with papers is often not maintained, or in some cases removed from public repositories for commercialisation.

Data quality for autonomous vehicles: a software overview

A key aspect in this project is the use of multiple sensors in adverse or poor weather conditions. It is anticipated that driving in conditions with non-differentiating backgrounds, water particles in the air (attenuating sensors) make it more difficult for the vehicle to assess surrounding obstacles and obstructions. Visible image quality issues come from lack of sharpness due to focus or dirty sensors, blurriness from lack of focus, confusing ranges and even the speed of the measurement vehicle itself. Obviously over exposed images from bright or low sun / lights as well as reflectance from piles of snow (see citations). If video processing is done on a sequence of images, then single frame 'images' can be factor especially over several ones, also due to compression of both images and video. Some form of downscaling is needed, either in frame rates or redundant coding. Table 2 below shows image quality assessment tools.

Table 2. Image & point cloud quality assessment with open-source repos.

Image Attribute	Source Lang.	Comment	GitHub Stars	Forks	Links / Notes.
Blurriness	Python	Provides a quick & accurate method for scoring blurriness. See Figure 8.	277	1	https://github.com/WillBrennan/BlurDetection2
Multiple	Python	PyTorch Image Quality (PIQ) is a collection of measures and metrics for image quality assessment. The library contains a set of measures and metrics that is continually getting extended	1100	107	https://github.com/photosynthesis-team/piq
Blurriness	Python	Judge if an image is blurred or not using a mathematical method.	-	-	Own development.
Exposure	Python	No reference image sharpness assessment based on local phase coherence.	40	7	https://github.com/elejke/awesome-defocus-detection
Sharpnes[1]	Python	Sharpness methods.	31	7	https://github.com/topics/image-sharpness
Sharpnes[2]	Python	Sharpness detection	-	-	Library calculates the variance of the Laplacian for each greyscale image.
Brightness	Python	Mean of the pixel values for the greyscale images over and underexposure	-	-	Own development.
Noise	Python	Library to calculate Shannon entropy of each image using `skimage`.	Many and varied.		https://github.com/topics/skimage

LiDAR Noise	Source Lang.	Comment	GitHub Stars	Forks	Links / Notes.
Point-Cloud Noise	Python	A complete toolbox from Open mmDetection3D.	27.9K	9.2K	https://github.com/open-mmlab/mmdetection Corruption / noise in point clouds. See Figure 18.
Point-Cloud Noise	Python	OpenPCDet is a clear, simple, self-contained open-source project for LiDAR-based 3D object detection.	4.3K	1.2K	https://github.com/open-mmlab/OpenPCDet
Point-Cloud Noise	Python	MultiCorrupt: Weather effects in the point clouds [29].	44	5	https://github.com/ika-rwth-aachen/MultiCorrupt
LiDAR single-sided metrics	**Source Lang.**	**Comment**	**GitHub**	**Forks**	**Links / Notes.**
COPP-NET	Python	A no-reference point cloud quality assessment (NR-PCQA) method with local area correlation analysis capability, denoted as COPP-Net [26]	2	0	https://github.com/philox12358/COPP-Net

For still versus motion quality assessment, a quality assessment tool, `ffmpeg`, is an open-source audio and video codec, probably is the best-known for video, notably in cross platform players such as VLC for decoding (usually) and HandBrake (coding). `Ffmpeg` and its associated tools, `ffprobe` include three quality metrics:

- Peak Signal-to-Noise Ratio (PSNR), which measures the difference between the original and compressed videos. A higher PSNR generally indicates that the reconstruction is of higher quality.
- Video Multi-Method Assessment Fusion (VMAF) developed by Netflix, VMAF is a perceptual quality metric that considers both human vision system models and machine learning models to provide a more accurate quality score.
- Video Structural Similarity Index (SSIM) is another metric that evaluates the perceptual difference between two videos. A value closer to 1 means the videos are more similar.

We can use `ffmpeg` as the external tool videos, as it has `ffprobe` and `ffplay` as sister tools to play, examine and examine the coding, these tools help when dealing with many files. In the VMAF case, a separate codebase and tool is available. Note, however a reference or “best quality” is needed, to make the comparison. Note others exist [14], as well as separate ones for LiDAR [20, 21]. A no reference video quality paper is in [26]. If videos are recoded, or resized or even format changed (mkv, mp4, mov, wmv) one should use the quality comparison tools above. An interesting, and yet unexplored option is to rate images from good weather conditions. In other words, ideal conditions against poor/harsh/adverse conditions, again in other words again, the reference image is the good weather, and the degraded image is in poor conditions.

A simple method commonly used in AD showing frame sequences / point cloud data is to use gif, a format that shows a series of still frames as a sequence. We do not look specifically at this format, as it often used to demonstrate a sequence, rather than use it in the perception and detection processes within AD.

Data Readiness Levels: A concept and history

Our overall data quality builds on Lawrence's data readiness levels (DRL) concept [1] and is a significant part of project planning and development. It is not by chance that up to 80% of the total project time is spent on pre-processing data, basically following the Pareto principle. The principle states "it takes 20% of the time to do 80% of the work, and 80% of the time to do the remaining 20%", even simpler stated, "one can do the large parts, but the fiddling consumes a lot of time (and effort)".

The main challenges of constructing meaningful data readiness levels are: i) assigning a single DRL (1–9) to large, complex datasets, often with imperfections [2], such as missing values, inaccuracies, and incomplete readings; ii) different readiness suggests different implications to different users, since data is often context sensitive; iii) some data consumers may have methods to handle imperfections, for example, in the missing data case, methods may, or may not, have been coded to handle missing values. Depending on the upstream capabilities, the DRL may be inaccurate [3,4]; and iv) production of quality sensor datasets (real car readings) are currently available for use.

One way to see data readiness is the values 1-9 indicate a quantitative measure of the time / effort / cost to repair / replace produce new values. Lower values are in principle easier to fix than those higher up the scale. Note, if errors in the lower end will propagate through the data pipeline causing issues, and probably harder to detect. This is because transformations (scaling, fusing, ML algorithms etc.) are applied and debugging the ML pipeline is more time consuming. Indeed, WP5 deals with making the ML pipeline as transparent as possible.

Lawrence's initial concept defines data readiness in three different bands—A (utility), B (validity), and C (accessibility)—depending on the knowledge and understanding of the available data and their usefulness for a given objective. We go back to the 9 levels present in TRLs and use them as shown in Figure 3.

Data quality in ROADVIEW: a perspective

In this first deliverable on data quality, we split the initial data readiness into 3 sections:

1. Introduce data readiness concepts
2. Details about the data sets under assessment
3. Assessment quality

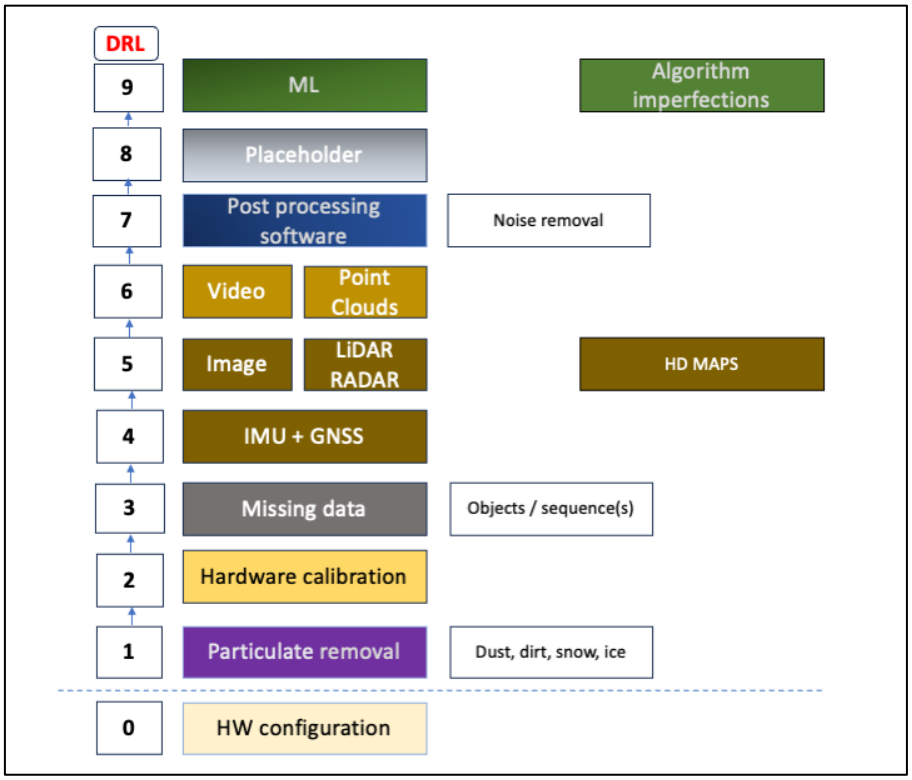


Figure 3. A generic view of data of the data processing stages in autonomous driving. Note, the lower numbers indicate the stage is processed before or previously in a ML pipeline. To attribute a numeric DRL level the box to the right, e.g. "HW configuration" must be performed successfully. This is repeated from DRL 0 to DRL 9 means all previous steps have performed as expected.

In the initial readiness assessment D4.5, RISE introduced the concept of DRLs. Data plays a fundamental role in the ROADVIEW project, making the completeness, correctness and interoperability of data are utmost important aspects of the data processing pipeline. RISE investigated the multimodal sensors that are used in ROADVIEW, summarizing the physical nature of each sensor, but focusing on camera image quality for the initial DRL definition in D4.5. As shown in Figure 4, from the lower DRL levels, **DRL 1** is about the sensors and perturbations therefrom. We know snow, ice, dust, and dirt inhibit signals from the LIDAR and RADAR. This is very much the focus of other tasks, with a slight emphasis on RADAR. Calibration at **DRL 2** is an important aspect, calibration, which we separate into 2 types:

Intrinsic calibration
Intrinsic calibration is basically sensor calibration, lens alignment, which can result in distorted / rotated images and point clouds, the resulting matrices contain wrong values with respect to reality, often abbreviated to NAs. The abbreviation can stand for Not Applicable, Not Available or Not Assessed. In the case of LiDAR, the laser sweeps should be within a range and intrinsic calibration is done to a known point cloud, physically adjusting the LiDARs. Often RGB cameras are seen as the trusted modality. Intrinsic calibration is only done once, rechecks and calibration are redone if something breaks during trials.

Extrinsic calibration
It is important to position the sensors with respect to each other. In vehicles with sensors, often the RGB camera is seen as the 'base' sensor or modality. One reason is that one can infer 'where' the camera is on the vehicle using internal and external imaging solutions. Speciality software takes pictures outside of the vehicle and inside the vehicle to calculate the position of the RGB cameras. The Radio Cross Section (RCS) of a target is the equivalent area as seen by a RADAR. It is the fictitious area intercepting that amount of power which, when scattered equally in all directions, produces an echo at the RADAR equal to that from the target.

At **DRL 3** we deal with missing data, as when streamed from sensors in real time (even when buffered) gaps in the sequence can occur. Data rates exceed capacity, sensors do not always pick up information, noise in the data overrides the signal and so on. The ratio of signal to noise is a key factor in any sensing situation. Theoretical results and extensive practice and calibration can to some degree predict the performance of a sensor, or modality in autonomous driving. However, in practice, and real driving situations it is more difficult to ascertain the signal and noise ratios.

At **DRL 4** we have the positioning and inertia measurement units. For now, we have them at the same level. This could be for discussion later but knowing where we are (GNSS) and how the vehicle accelerates, brakes, or vibrates is important in post-processing. Repeated runs along the same route in different conditions produces repeatability but also changing, a limited, but known, number of parameters. Indeed, in ROADVIEW runs are done when the weather has changed. An extension is to create the weather artificially, as done in the THI and Cerema test site in Germany and France.

At **DRL 5** we have images and the LIDAR / RADAR. They are different modalities, but in essence are similar in terms of sensing. We look at several different open-source tools to assess them.

At **DRL 6** we have the collated 'data'. That is video and LiDAR scans, processed (frame rates, resolutions) single images combined into videos and LiDAR point clouds, spatially and temporally collated. We look at the video quality as well in this Task, using `vmaf` which the rest of the project does not do. LiDAR point clouds are heavily discussed in this report. With respect to data quality, we have access to the FGI & THI datasets and are evaluating those, and we can introduce weather effects into public datasets and evaluate them, this is done with tools such as MultiCorrupt [29] and Robo3D [33].

At **DRL 7** we have the processing of the point clouds, removing noise from the point clouds to make object recognition effective, this is being done in other parts of the project.

We have left one **DRL 8** free as some processing in the ML pipeline, that can affect the final quality. Research is being done to rasterise point clouds, make the point clouds more sense or using multiple sources to enhance quality. Further, research on explainable AI might require a level. Therefore, we leave **DRL 8** for each advancement.

DRL 9 is strictly about the object detection, in free space or in scenes that is coupled to the algorithms themselves. We have implicitly assumed that the algorithms are perfect and only data is a detractor, which is strictly not true. The algorithm and often the algorithm and the data it is trained, validated, and tested upon can have significant impacts on the results in AD. This is why different techniques are evaluated, often on several datasets, and with quite different outcomes. The mmDetection suite tests their algorithms on different datasets showing a fair difference. Therefore, the algorithms themselves need to be considered, and this is being done in WP5.

Three data sources

Table 3. Datasets procured in the ROADVIEW project (details).

#	Data provider	Dataset Name	Size (RISE s3)	Modalities	Type	Purpose
1	Finnish Geospatial Institute, (FI)	FGI	110 GB	RGB Camera, Thermal Camera, LIDAR, IMU, GNSS	Open roads	Driving on real roads under snowy conditions. One rural and one urban journey. Assess data as gathered.
2	Tech. Institute Ingolstadt, (DE)	REHEARSE	320 GB	RGB Camera, Thermal Camera, LIDAR, IMU, GNSS	Controlled environment, testing ground	Testing ground in Southern Germany. Assess data as gathered.
3	Lapland University of Applied Sciences (FI)	-	Configurable	Images	Synthetically generated	Test case for DRL and for calibration. Ongoing work. Tool link. (WinterSim)

Data sources in terms of complexity

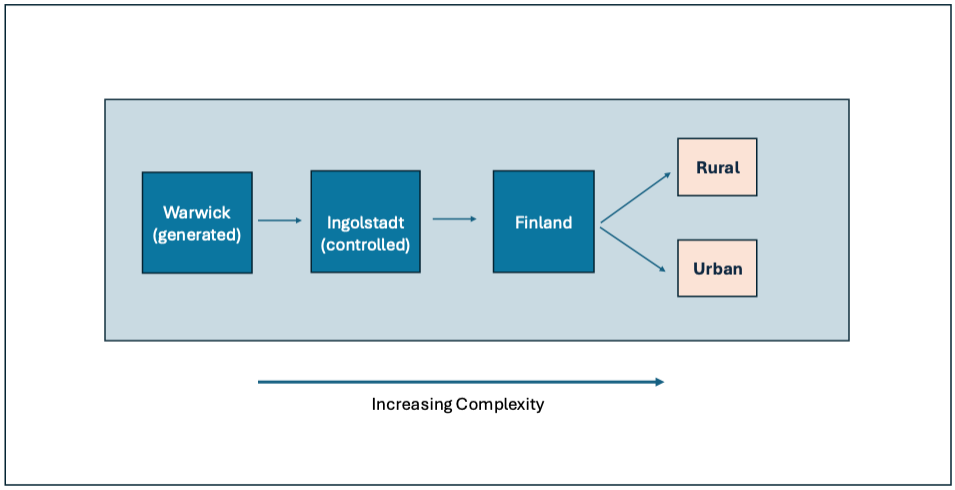


Figure 4. Datasets used in terms of their scene complexity.

Data source I: The Finnish Geospatial Institute (FGI) training data

We take a little deeper dive into a dataset available in the ROADVIEW project. Kindly made available by the Finnish Geospatial Institute, we list some of the attributes of the data made available. The scope of the description is quite high-level formats, visualisation to the lower-level formatting and how to unpack and use the four sensor types. Table 4 shows the sensor, the recorded attribute and format and notes on the data.

FGI sensors, formats, and files

Table 4. Data formats from the FGI dataset.

Sensor	Format	Notes
Camera	PNG	Anonymised (Blurred faces and number plates). Unrestricted license compressed format.
Thermal Camera	TIFF	Raw, unrestricted license format (16 bit)
LiDAR	RG	Binary packed range and reflection data
Road weather	RW	Binary packed road conditions data (see below)
RADAR	-	Not available from FGI

Road conditions

In the FGI dataset, the road conditions were categorized as in Table 5. Bool represents Boolean, if available and measured or not.

Table 5. Road weather, conditions & data fields in the FGI Dataset.

Attribute	Format
Surface Temperature	Bool
State	Bool
Water	Bool
Grip	Bool
Ice	Bool
Snow	Bool
EN15518 State	Bool
Air Temperature	Bool
RH	Bool
Dew Point	Bool
Frost Point	Bool
Data Warning	Bool
Data Error	Bool
Unit Status	Bool
Error Bits	Bool

A master's thesis on the topic has recently been produced, which includes slipperiness and grip [27].

Urban and rural driving data in Finland

The Finnish Geospatial Institute (FGI) has produced a data set, shown in Table 4. It was used to generate the video shown in Figure 5. The whole data set size was about 34 Gigabytes for a 10-minute drive. Table 8 below shows the data rates and message sizes computed in kB / s and in MB / minute. Content-wise the data is divided into a Rural and Urban drives, 1810 and 3027 in PNG, TIFF images and RW 'format' (see above) respectively.

Table 6. Two autonomous driving routes from FGI (https://www.maanmittauslaitos.fi/en/research).

Rural	Sizes	No. of files	Urban	Sizes	No. of files
Camera anonymised	2.4G	1810	Camera anonymised	4.7G	3027
LiDAR	913M	1810	LiDAR	1.5G	3027
Thermal camera	6.7G	1810	Thermal camera	39M	3027
Road Weather	21M	1810	Road Weather	39M	3027

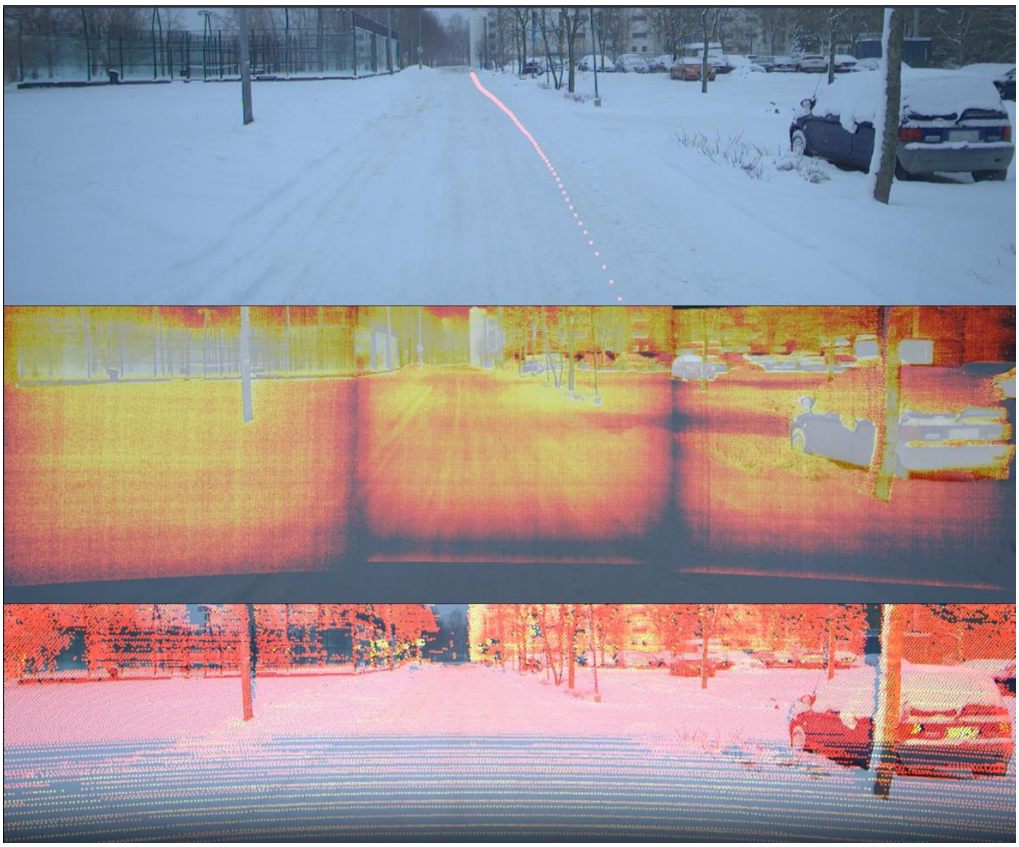


Figure 5. Example of FGIs Urban 3 data sources fused, Video link (5fps).

The output of a processed urban dataset is a 250 MB video around 5 minutes long at 10 frames per second. The rural one is 60 MB around 3 minutes also at 10 frames per second. More on rates and sizes in Table 6 above. Note this is without RADAR data and does not say anything about training, validating an AI model. The start frames can also be specified and in the rural case this is offset by 300 'frames' as the capturing started early. Table 7 below

shows the sizes and times to fuse data (offline) the two datasets. Processing was done on a 2021 Apple MacBook Pro M1 Max (Apple Silicon), 32 GB RAM, 16 GPUs and 16 CPUs.

Table 7. Output characteristics using the FGI dataset above.

Environment	Output video (@ 10 frames / sec)	Fusion Time
Rural	60 MB	30 mins
Urban	250 MB	50 mins

In terms of performance issues, Python will produce binary files *.pyc rather than text to make the interpreted code into a binary. For large programs (rather than data) this can speed up execution. For optimising the code, one can use `timeit` to find functions or sequences that consume time. Alternatives are `cProfile` and very recently, `Scalene` from Umass, Amherst, USA (Aug 28th, 2023). Table 8 below gives the breakdown per sensor and how the data streams are reduced by down sampling.

Table 8. Sensor sizes and rates from the FGI dataset.

Sensor	Data per message (kB)	Freq. (Hz)	Data rate (kB/s)	MB/min	Data rate at 5Hz low sampled data (kB/s)	MB/min
Novatel GNSS + IMU positioning	0,09	205,0	19	1,10	19	1,10
Front colour camera	1399	10,0	13984	819	6993	410
Thermal camera centre	241	60,0	14459	847	1204	71
Thermal camera left	240	60,0	14407	844	1200	70
Thermal camera right	235	60,1	14126	828	1176	69
Vaisala MD30	0,04	40,0	2	0,09	2	0,09
Velodyne VLS128 LiDAR	617	9,1	5630	330	3087	181
SUM			**62626**	**3669**	**13680**	**802**

As can be noted, thermal cameras have a lot higher data rate than other sensors (60Hz vs 10Hz). Now there is 3.6 GB of raw data per minute. If only the frames which are used to generate the sensor-fused frames at 5Hz are listed, the data rates decrease significantly. Computed is a low sampled example in the two last columns, where only 5 samples per second have been taken from all cameras and the LiDAR. In this low-sampled version one minute of raw data is 802 MB. The project will take different data sets under various weather conditions. Using a low sampled would take a few 100's GBs, but if full resolution is needed, an estimate is > 1TB.

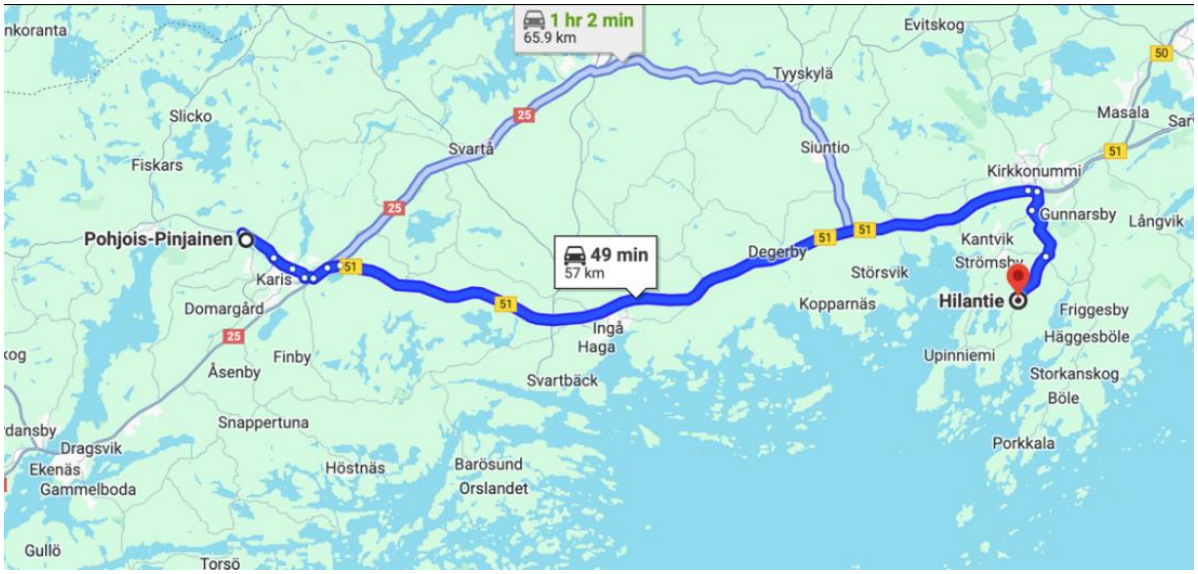


Figure 6. Otaniemi. A rural route between Hilantie-Pohjoiseen in southern Finland.

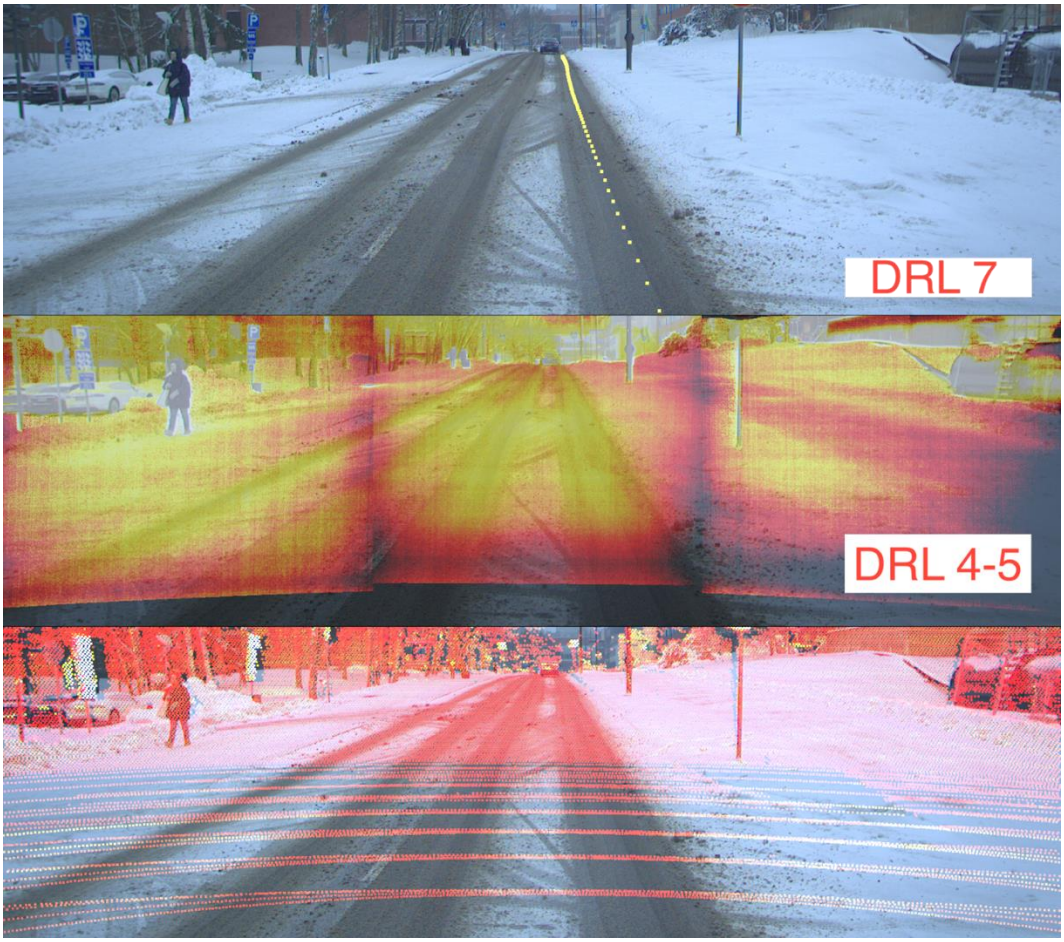


Figure 7. DRLs applied to the images from FGI urban journey (credit: The Finish Geospatial Institute, fgi.fi). Top: Simple RGB image. Mid: thermal camera image projected into the RGB camera image. Bottom: LiDAR point cloud projected into RGB camera image.

For this dataset, Figure 7 shows one representative frame of the dataset, showing the RGB camera as well as thermal counterpart and LiDAR point cloud data, all projected into the RGB camera image. The RGB camera image was rated with DRL 7 while the thermal camera frame only got up to 4-5. This is mainly due to the edges of the thermal camera image, simply caused by the working principle of the sensor itself. These edges do cause issues in the image quality evaluation though, hence less DRL rating. Lastly the LiDAR point cloud is yet to be evaluated since the methodology for point cloud rating is not yet finalized. The projected point cloud into the image plane in the bottom image of Figure 7 leads to the question of rating extrinsic calibration, which needs to be further discussed also. Ongoing work will look at the dataset from THI and VTT which has been received by RISE. Together with HH, the LiDAR point cloud data will be projected into spherical coordinate system to generate image representations for assessment. RISE also plans to apply the DRL concept to at least 1 public dataset.

Data source II: The Technical University of Ingolstadt, Germany

As a controlled environment, ROADVIEW used the testing facilities in Ingolstadt, Germany. Figure 8 shows a test track of the CARISSMA department at Ingolstadt outdoor test facilities. There is an acceleration zone of 210m and a dynamic area with a 60m x 70m area. The maximum speed is 100km/h. Different parts of the track are watered with different intensities. The image shows the watering and measurement instrumentation points. Initial data sets with images and point clouds have been produced and published, in this work we refer to the REHEARSE dataset. The full work will be available in Feb. 2024.

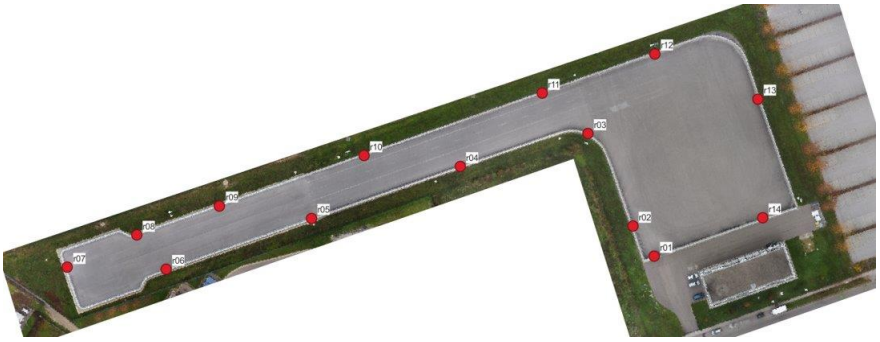


Figure 8. Outdoor test track of the Technical University of Ingolstadt, Germany.

Figure 9 shows artificial weather conditions (rain) and a LiDAR point cloud in these conditions.

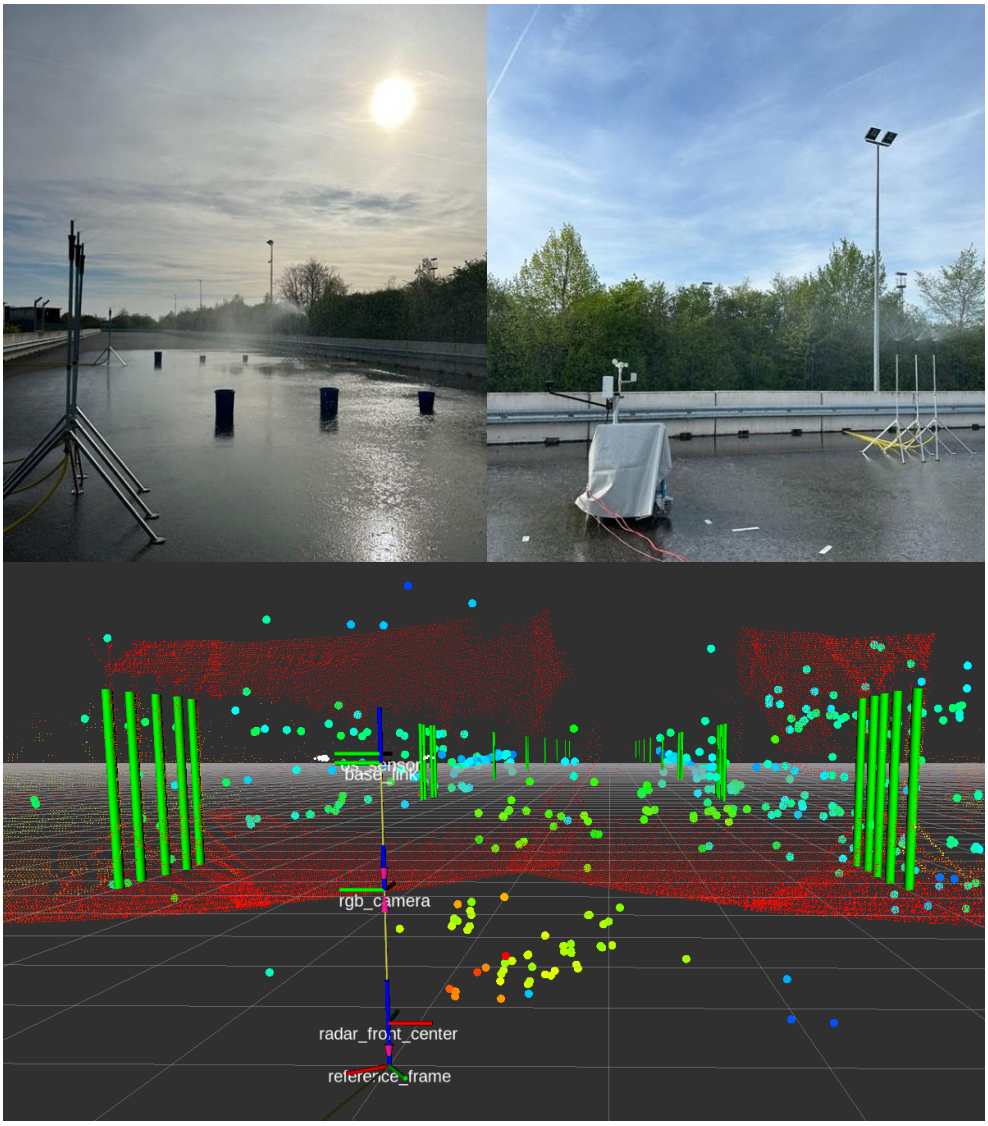


Figure 9. Reference dataset of measured weather characteristics, THI D3.2, WP3

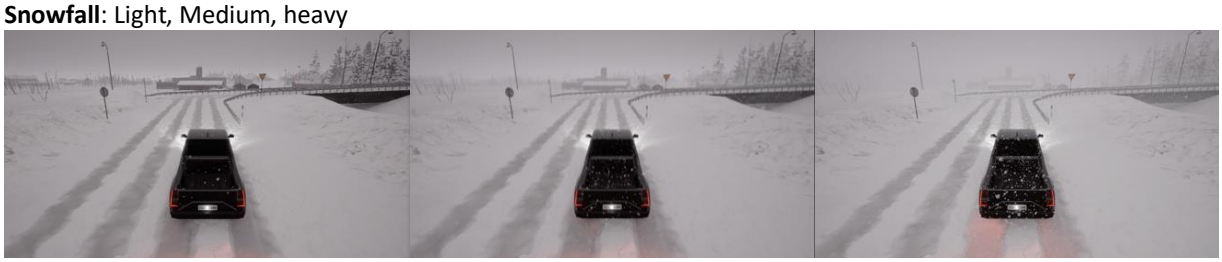
The datasets were created to observe the 3 weather conditions rain, clear, and fog different conditions on different sensors. The utilised sensors in this case were VISIBLE Camera (LUCID), FLIR (Thermal) Camera, RADAR (ZF ProWave), and LiDAR (Innoviz One and Ouster OS1). This dataset has a total of 68.4 minutes of recording. The recording took place in the CARISSMA outdoor proving ground in Ingolstadt Germany, and in the CE proving ground in Clermont Ferrand France. Each with different intensities. The amount of rain was also measured and calibrated using three different methods, litres per square meter, drop shape and amount, and direct weather measurements.

Data Source III: Simulating weather through image generation

An alternative to using lots of data, plus expense of buying and instrumenting sensors as well as measuring, storing, processing, and driving around areas, one can generate the scenes as a sensor would see. As examples of visible images that can be generated, see Figure 10 below. Credit for the images are to Tuomas Herranen at Lapland University of Applied Sciences using the WinterSim tool.

Snow: Light, heavy

Snowfall: Light, Medium, heavy



Time of day: Winter daytime, morning, night

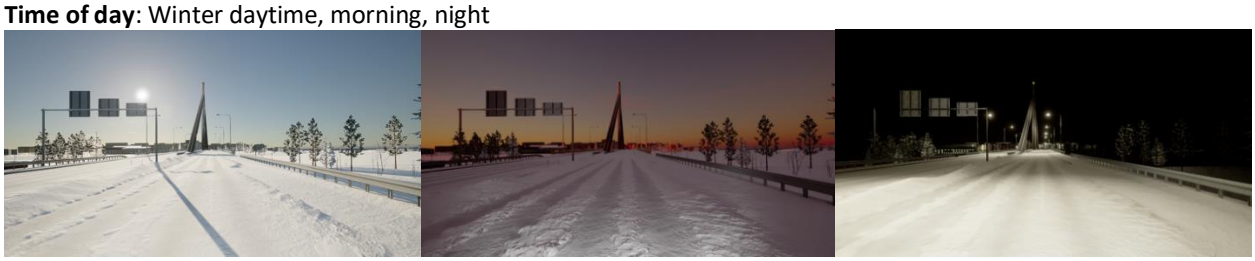


Figure 10. Machine-generated scenes from WP2 (source: Lapland University of Applied Sciences, credit: Tuomas Herranen).

Three ways forward using these images, is to

1. Calibration of data quality
    a. Generate a perfect image (DRL score 9)
    b. Generate a very noisy, blurred image (DRL score 1)
2. Backtrace from object misclassification to pixels
3. Generate scenes that we do not have access to, i.e. conditions, weather, etc.

The DRs of the images is high, up to 9, we have not tested them in WP4 and object detection, but work is in progress. Since no deliverables are promised that are planned, the results will appear in a technical paper.

Constructing the DRL 'score': a walkthrough

To make the concept clear to the reader, we step through the construction of the score. Step-by-step it will be clearer how the score is derived. Furthermore, it should be clear how to amend or improve the score. We acknowledge this is a research topic and improvements can be proposed. In particular using the disciplines using time series analysis.

The main fulcrum of this work is how the DRL is calculated. For the moment, it is a linear combination of the inputs from the sensors. Should one be missing, RADAR for example, its contribution is distributed amongst the others. Similarly, so with multiple numbers of the same, for example 4 cameras. For forward and backward sensors, we will (for now) exclusively use the forward ones. Side ones will be given 25% max. However, this is a heuristic value and will be explored further.

It is important to reiterate that the Data Readiness or Quality is really for off-line use, to ensure the data is of sufficient quality. It is assumed that the algorithms are 'perfect' which is unreasonable, but there are many algorithms and measures to determine how good the algorithm is. Estimating the data quality known the performance (in %) of algorithms is possible, but over complicates the issue and might place emphasis on the wrong place.

An alternative would be start from the other end and say when the object detection is correct in the training data, then the data is perfect and for each 'imperfection' in the image correlate this with the classification success. Or a misclassification is back traced to an image that is not clear enough. There is also the issue of the training and validation sets, how they are selected. In the standard image sets this is done, however in our internal datasets this is being assessed.

A linear combination of modalities: Constructing the DRL 'score'

DRL = RGB + Thermal + LiDAR + RADAR

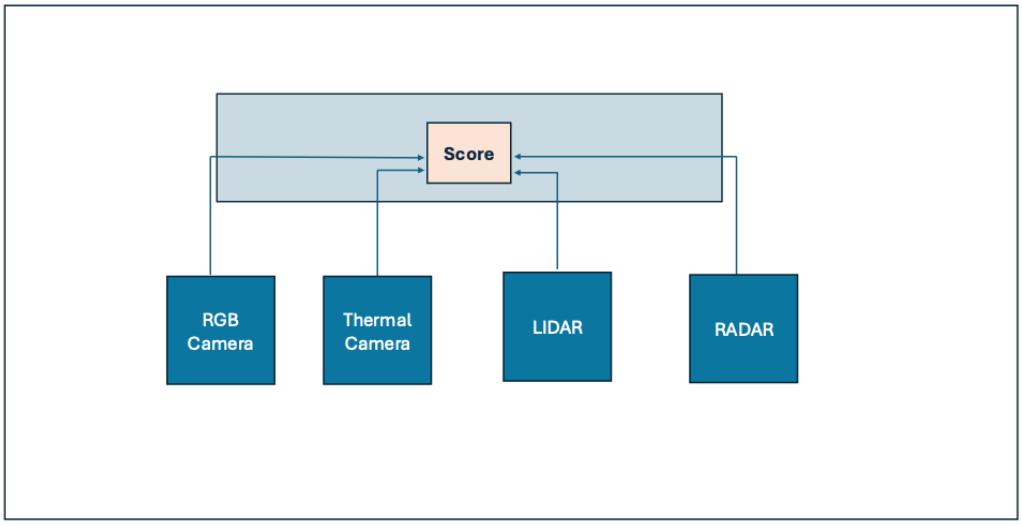


Figure 11. Modalities combine additively to produce an amalgamated dataset score.

A time varying combination of data quality: Constructing the DRL 'score'

$$DRL(t) = RGB(t) + Thermal(t) + LiDAR(t) + RADAR(t)$$

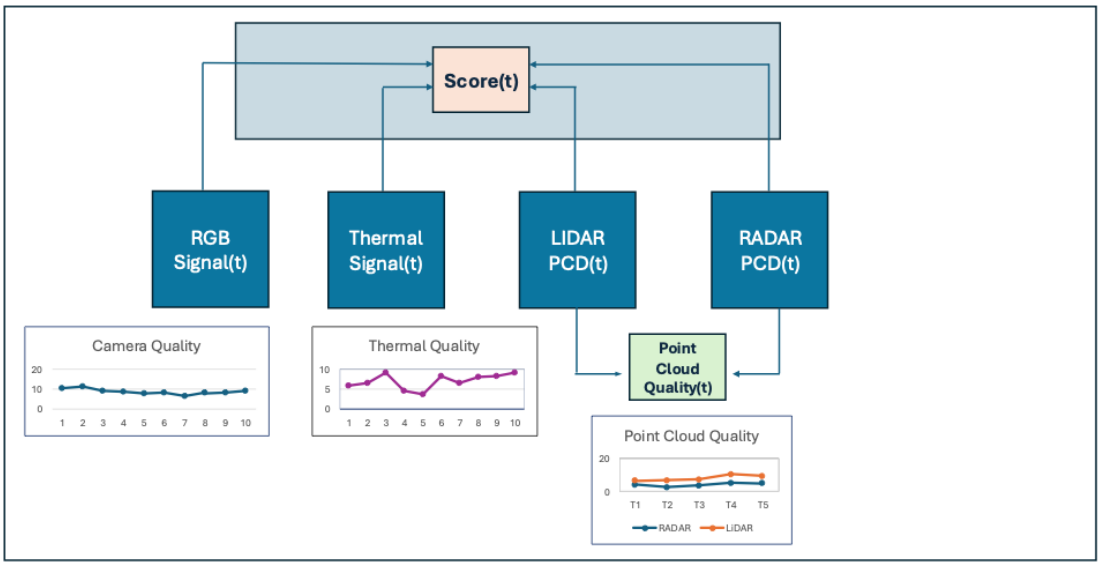


Figure 12. Modalities depend on time. We combine the modalities per time unit (at the lowest frequency) to produce a score per time unit.

As mentioned before, once normalised we combine the individual scores into a singular value. This is show in Figure 12 above. It is important the time perspective and to work at the lowest common sampling frequency. Whilst recently using AI, upscaling has made an entrance into image and video processing we keep to the traditional method and downsample the samples to those of the lowest frequency source, that is typically the LiDAR (see Table 1).

Smoothing the DRL scores

Our final objective is to assign a score to several modalities, derived from several methods over a period. In other words, the final DRL score is "an average of an average of an average". Furthermore, the tools to assess the images and point clouds may change or be added to, or removed, or even bug fixed. One must consider the batching interval for the operations below. At a LiDAR frequency of 10Hz that would be 1/10th of a second per sample. 5 Hz is also quoted as a sampling period for some LiDARs too. Batching processing per second would be a trade-off between reacting to every value and not looking over too long a period. One second in AD could be critical. One needs to normalise the values to the same range, e.g. 1-100. One can also remove the standard deviation and divide by the mean to move (and anonymise) the values to the 1-100 range. There are several methods to process the values from the tools (see Figure 13).

Some methods include z-score normalization, min-max normalization, range normalization, decimal scaling, and max_scale normalization. Other are:

- Choose the median values from the tools and combine those
- One is to carry both the mean and deviation and combine those

- One is too smooth the values with an Exponential weighting,
- One chooses a value in the range [0-1] to weight the historical or more recent values
- Differencing the values, removing the previous value from the current one can be a good method to detect outliers. Some time series methods. In AR(I)MA methods the I indicate incremental, which essentially means differencing. The MA is the moving average which is the same as the exponential weighting mentioned.

Figure 13 shows the same time series with one of the methods above, applied.

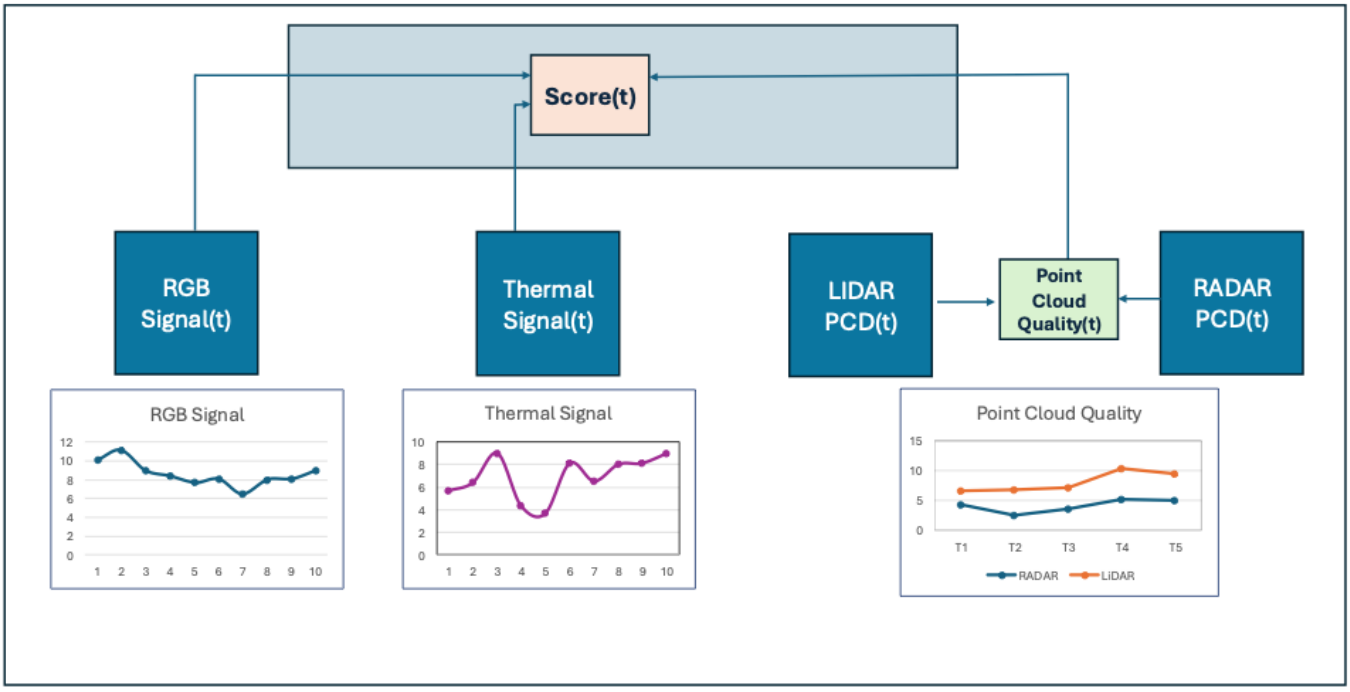


Figure 13. Modalities 'smoothed' over time.

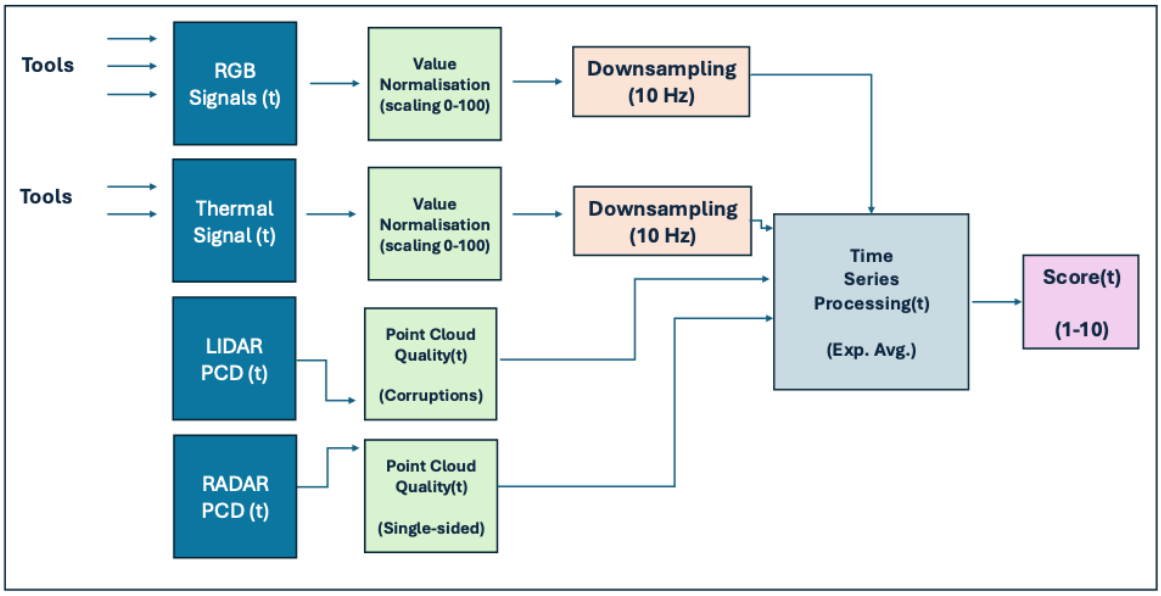


Figure 14. Calculating DRL scores over time. A minimum rate of 10Hz for the LiDAR is assumed.

Image Tools

Figure 14 show shows the processing steps. To link this with Table 2, the list of open-source tools, in the top right of the figure Tools → we rely on `BlurDetection` with some adaptions by us. PIQ the PyTorch Image quality package as well. We have our own tool too, which was described and illustrated in D4.5 based on the variance of the Laplacian for each greyscale image. Noise in the images, useful for cleaning images with respect to perception we measure via the Shannon entropy of each image using the popular `skimage` toolkit. Note, we didn't differentiate in the processing in the thermal and RGB images. That would be improved in the future, particularly in the colour spectrum parts, which we ignore to be fair (i.e. weigh equally between n RGB + Thermal).

Data Readiness Level processing (Levels 3-5 only)

Our final objective is to assign a score to several modalities, derived from several methods over a period. In other words, the final DRL score is "an average of an average of an average". Furthermore, the tools to assess the images and point clouds may change or be added to, or removed, or even bug fixed. One must consider the batching interval for the operations below. At a LiDAR frequency of 10Hz that would be 1/10th of a second per sample. Batching processing per second would be a trade-off between reacting to every value and not looking over too long a period. One second in AD could be critical. One needs to normalise the values to the same range, e.g. 1-100. One can also remove the standard deviation and divide by the mean to anonymise the values to the 1-100 range.

We will discuss these in more detail in the upcoming technical paper, with examples of course. In terms of data quality, they do not impact the ratings severely, but some decisions need to be made. If one would like to choose 'the best' modality at a given time, "LiDAR is performing best" these methods become more important. One way to see them is to make contributions fair. We have not discussed modality selection in this report but do describe an introspection tool next.

		RGB		Normalised			Downsampled			Time weighted		Score		Score (1-100)
vmaf														
	4.7			0.40										
	12.0			1.00			0.40							
	1.9			0.15			0.15			0.32				
	0.0			0.00			0.13							
	1.5			0.13										
piq	56			0.51										
	12			0.00			0.00							
	99			1.00			1.00			0.5				
	NA			NA			0.00		Note NA removal					
	12			0.00										
RISE Tool														
	1.0			0.00										
	1.2			0.18			0.00							
	1.7			0.63			0.63			0.3		0.34		34%
	2.1			1.00			0.00							
	1.0			0.00										
		Point Cloud												
Corruptions														
	-1			0.00										
	0			0.25			0.00							
	1			0.50			0.50			0.5				
	2			0.75			1.00							
	3			1.00										
Single-sided														
	1.10			0.00										
	1.21			0.35			0.00							
	1.21			0.35			0.35			0.1				
	1.41			1.00			0.00							
	1.10			0.00										

Figure 15. Example Excel sheet illustrating DRL scores for level 3-5 (images and point clouds). Missing data (NAs) are handled. The downsampling, averaging and overall weighting have a number of options and are not discussed further in this report.

Cross validation over modalities

A central question that arises is how we can adjudge readiness levels where there are several modalities? Not only the modalities DRLs over frames and driving scenes. We have chosen to use an open-source tool called `rerun`. There are, of course, several options and the criteria chosen, more or less in order, were: open-source, and reasonable licensing, regular updates, (an active community), fast (which means compiled and forward and reverse for a large dataset), extensible (to add our DRLs per frame), configurable (different users have different needs). In terms of the content, point matching per modality, time format conversions, and fine-grained media control, we may want to identify where the DRL drops across modalities. Advantageous is a modern language, such as Rust, fast-in browser visualisation, e.g. WebAssembly and a programmable tile-able layout. Platform portability is desired, which means that at least Windows, Mac, and Linux should be supported. Note `rerun` is not necessarily specific to autonomous driving but supports forwarding through large amounts of serial data. Open3D++, rviz and FoxGlove are other tools for viewing point cloud data.

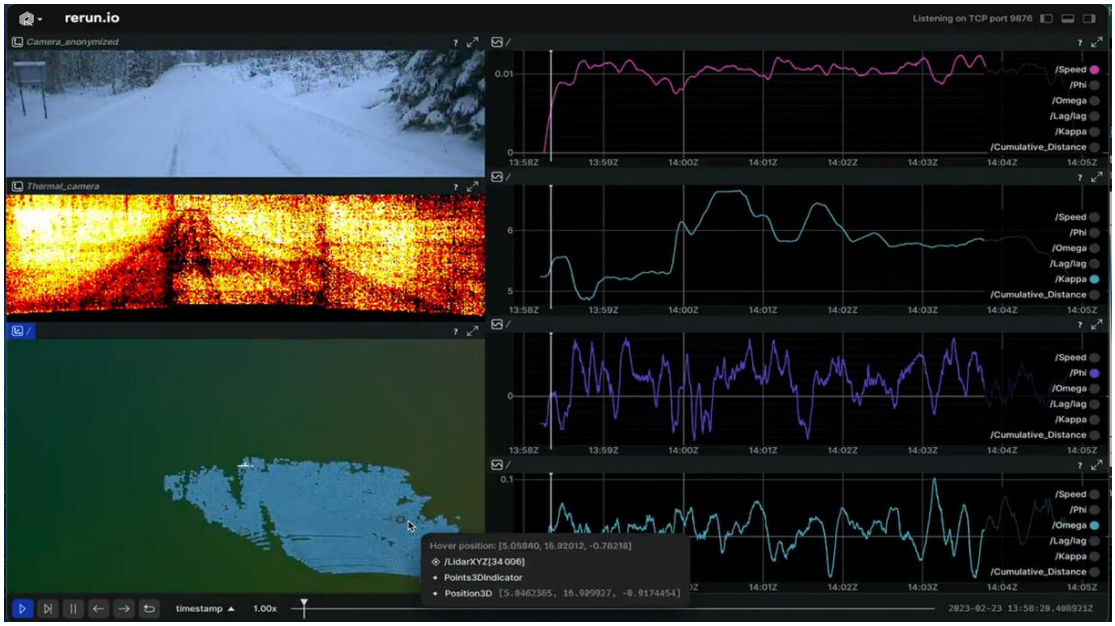


Figure 16. DRL introspection and visualisation of the Finnish Geospatial Institute (FGI) data using the rerun tool. Image impairments, sharpness, exposure and noise are shown as time series to the right.

Data annotations

Whilst most of this report is about data quality, annotating data is a significant issue. In machine learning, data annotation is the process of labelling data. This labelled data is then used to train supervised learning models. Data annotation is a crucial step in many machine learning projects, especially in projects such as ROADVIEW: vision, scene detection. Obviously, the quality and accuracy of the annotated data affects the performance of the resulting machine learning models. Within ROADVIEW we use:

- Image Annotation: This involves marking various objects within images. Types of image annotations include:
- Bounding boxes: Drawing rectangles around objects of interest.
- Semantic segmentation: Labelling each pixel of an image with a class label.
- Polygonal segmentation: Drawing polygons around objects, especially those with irregular shapes.
- Key point annotation: Marking specific points of interest on objects, often used for pose estimation.

- Named Entity Recognition (NER): Labelling words or sequences of words as specific entities like names, locations, or dates.
- Video Annotation: Labelling objects or actions within video sequences. This can involve bounding boxes over time or tagging entire video clips with action labels.
- Audio Annotation: Marking sections of audio data to label different sounds, words, or other audible events.

The process of data annotation is time-consuming, requires domain expertise and ROADVIEW has a Task for this purpose. An interesting angle is active learning, where a machine learning paradigm where the model itself decides which data points should be annotated furthermore, based on where it predicts the annotation would be most valuable. This can help reduce the amount of manual annotation.

Point cloud data quality

Data Corruptions

Figure 17 shows some functions of corruptions in 3D object detection. The 3D corruptions project is built upon MMDetection3D and OpenPCDet with code modifications. The authors identify 32 LiDAR corruptions and 14 camera corruptions. They test the impact of the corruptions on the Kitty dataset [25].

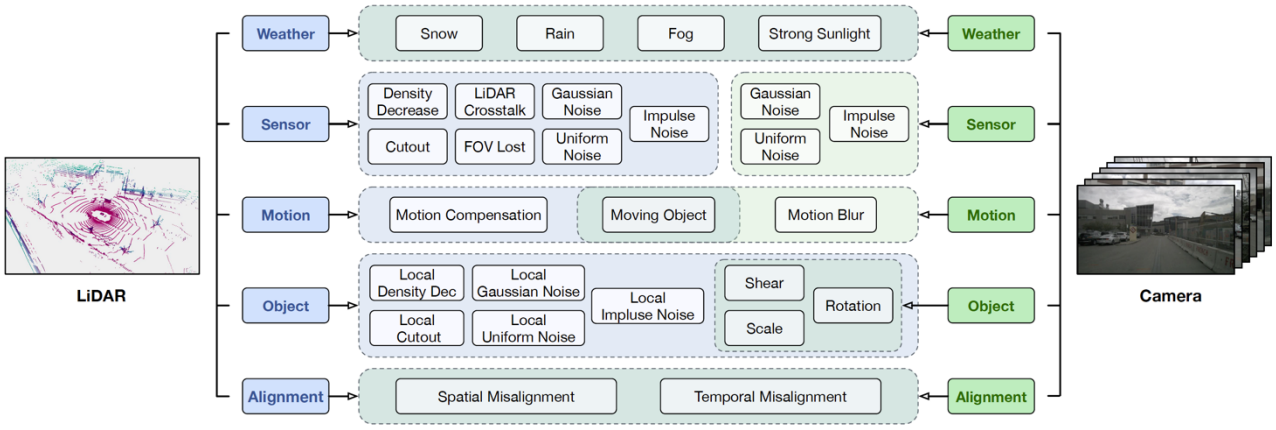


Figure 17. Benchmarking Robustness of 3D Object Detection to Common Corruptions in Autonomous Driving.

In this work we use 2 sources of data corruptions to validate the effect of noise in the point cloud.

- Robo* tools from National University of Singapore.
- Multi-Corrupt from Aachen in Germany.

Single Sided Metric(s) approach

And 1 single-sided metric;

- COPP-NET

Prior point cloud quality assessment (PCQA) methods largely ignored the effect of local quality variance across different areas of the point cloud. A no-reference point cloud quality assessment (NR-PCQA) method with local area correlation analysis capability is the approach here. They split a point cloud into patches, generate texture and structure features for each patch, and fuse them into patch features to predict patch quality. The features of all the patches of a point cloud for correlation analysis, to obtain the correlation weights. Finally, the predicted qualities and correlation weights for all the patches are used to derive the final quality score.

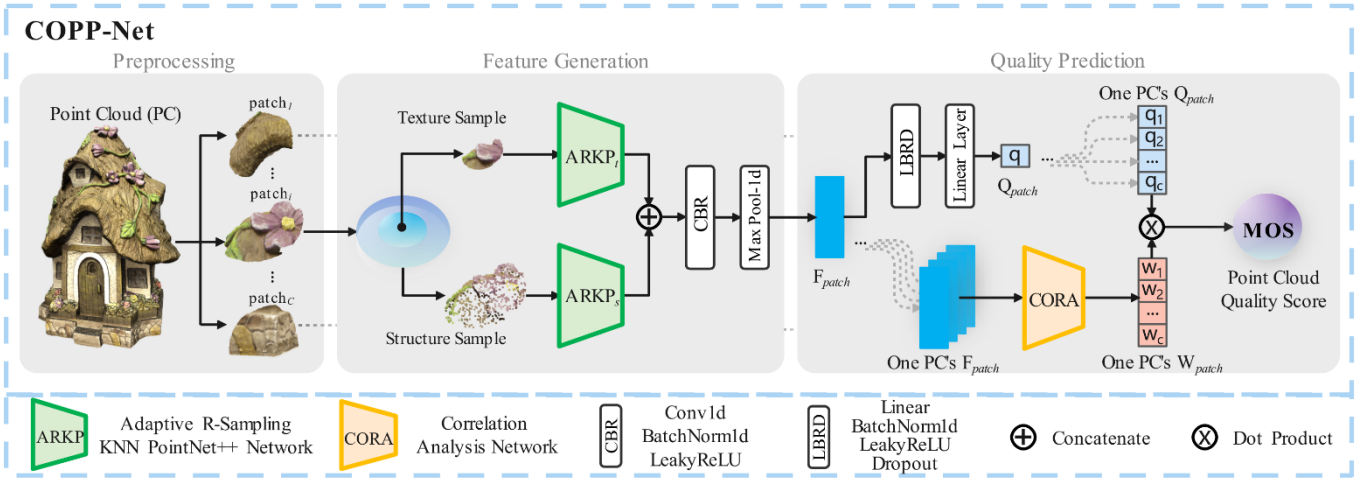


Figure 18. A single-sided (i.e. no reference) approach to point cloud data with respect to quality.

MMDetection3d

An interesting library is mmdetection3d. Not as cited as the others, but with over 1500 forks it is a feature rich dataset for experimentation, in paperswithcode, there are 63 such papers. MMDetection3D is an open-source object detection toolbox based on PyTorch, towards the next-generation platform for general 3D detection. It is a part of the OpenMMLab project. It supports multi-modality/single-modality detectors out of box using detectors including MVXNet, VoteNet, PointPillars, etc. It directly supports popular indoor and outdoor 3D detection datasets, including ScanNet, SUNRGB-D, Waymo, nuScenes, Lyft, and KITTI. For the nuScenes dataset, they also support the nuImages dataset. All the 300+ models and methods of 40+ papers as well as modules supported in MMDetection3D can be trained or used. The many authors claim it trains faster than other codebases. Like MMDetection3D and MMCV, MMDetection3D can also be used as a library to support different projects. Note it can detect images using pixel-based classic bounding boxes, semantic segmentation as well as panoptic image recognition. Note, The Robo* tools used in our point cloud quality assessment are based on the MMDectection3D.

Part III: Data Management Plan

This part of D4.6 is concerned with the computing aspects of the datasets and their quality. This includes the storage format, de-facto standards and alternatives, a little about the storage. In terms of respecting EU rules, we expect each partner to anonymise the(ir) data with respect to privacy. That means blurring number plates, faces or other objects that can lead to personal information being exposed and potentially misused.

ROADVIEW Storage: Holding the ROADVIEW data: RISE ICE CLOUD, The Finnish Geospatial Institute and Collaboration between the Technical Institute of Ingolstadt and Cerema research institute (REHEARSE).

One of the key features of the ICE data centre is its use of innovative cooling solutions. The facility takes advantage of the cold climate in Luleå, Sweden to cool the servers with outside air, significantly reducing the energy required for cooling. This makes the data centre more environmentally friendly and cost-effective, as it minimises energy consumption and operational costs.

The ICE data centre at RISE in Luleå serves as a testbed and research platform for various projects related to sustainable computing, including:

- Energy-efficient data storage and processing
- Renewable energy integration and smart grid solutions
- Edge computing and distributed data centres
- Cloud computing and resource optimisation
- Data centre security and resilience

By providing a state-of-the-art facility for the research and development of green data storage technologies, RISE in Luleå contributes to the global efforts towards more sustainable and energy-efficient IT infrastructure. In terms of data upload an initial speed test upload to the ICE data centre showed that 650 megs (about a data CDs worth) was uploaded in about 9 secs. A DVD (4.3 GB worth of data) took about 140 seconds. The raw dataset from FGI is about 42 Gig, after processing, 60 Gig, and a merged 30 second video 2 Gig. No RADAR data is included in the FGI dataset. The S3 data format, developed by Amazon, is used in the centre, albeit an open-source implementation thereof called Ceph. After some testing, Stockholm to ice.ri.se (S3 buckets) 650 MB in 9 secs. Adding NVMRe memory sped up access.

Part IV: Concluding remarks

Discussion

There can be a philosophical discussion about data quality. One is what is data quality? What is the quality for? Can it be generalised, which to some degree TRLs have been. People know what each TRL level is, and which applies to their systems. Indeed, European Union projects stipulate the starting and often the finishing level.

One caveat of this work is we assess a dataset but are reluctant to use the DRLs to compare datasets. Since they vary over time, and equipment changes, as do conditions, therefore it is not simple or fair to compare datasets using the DRL concept.

There is no (4G) RADAR data in the FGI dataset, but in the REHEARSE (Ingolstadt) + Cerema there is. We have made some progress in point cloud quality assessment compared to D4.5. One is to introduce corruptions in the data (Robo3D and MultiCorrupt) the other is a single sided measure.

In terms of data and quality, it should be of no surprise that the data from the controlled environment of THI is of slightly higher quality. Not only are the light levels, stationary equipment and controlled distances, which are convivial to higher quality, but are also controllable to some degree. This is to ROADVIEW's advantage, and to some degree, the same can be said of the Warwick (Carla-generated) data. For data quality control, the concept, the evaluation tools, as well as ascertaining the DRL level (0-9), controlled to 'real' datasets in terms of quality, is necessary. With respects to RADAR, we are talking about the output, or data generation stage, but where RADAR data can / could be, improved we are in progress. RISE has RADAR experts who are looking at heating, ice buildups in lab environments. One point is how the RADAR unit is mounted and calibrated.

Public datasets are curated, making the datasets cleaner. This is important when releasing a dataset for experimentation. Remember 80% of the time one performs less useful work, hence the work in a public dataset, however, these may not include lots of extra work needed in a real scenario. Curated (nuScenes) versus non-curated (FGI) artificial data (Warwick simulator), often clip long range sensing (LiDAR and RADAR). Now we fuse data as separate streams, there is a philosophy where data can be collected from all sensors and fused in a single pass. This will be investigated later in the project and is one branch in the MMdetection3D dataset. Datasets for cars versus trucks. One concern from the data gathered is that we are using data gathered from vehicles, thus far. However, other WPs will produce data from a truck setup. This will enable ROADVIEW to assess the view from the correct point of view. It will also be an opportunity to create a unique dataset.

Future Work

The next steps are to collate the licenses for the software used. If changes are made to the code bases, we will add them in-situ to the code bases. To really test the idea of Data Readiness Level we really need to have a live dataset, feedback the quality (see Figure 1) and then reevaluate when changes have been made to the dataset. This makes more sense as a feedback mechanism rather than to 'quality stamp' an existing dataset.

Conclusions

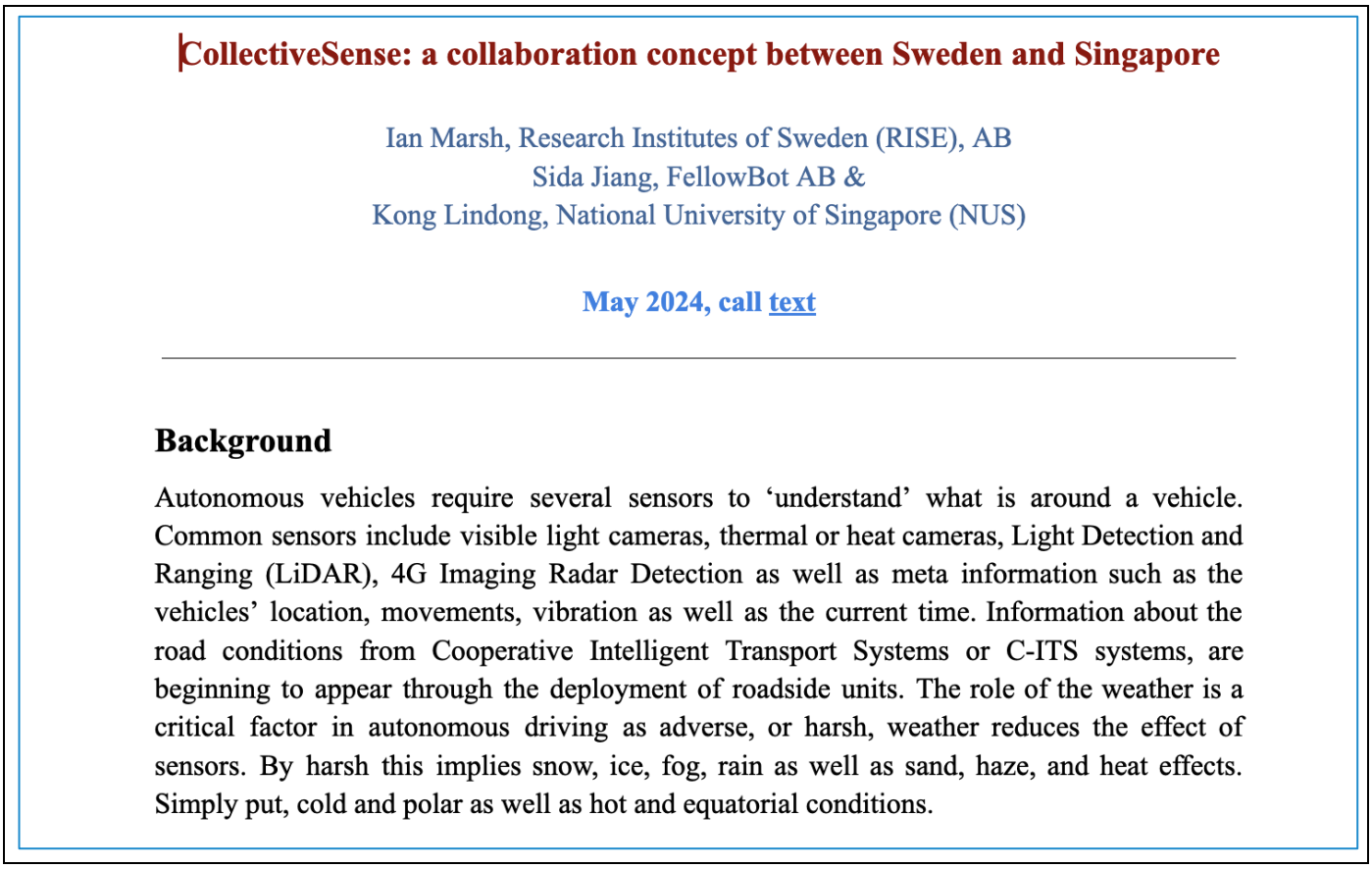
This document fulfils two purposes the readiness level of the data used in ROADVIEW. It is primarily around the sensors autonomous vehicles will carry and use, their purpose, rates, format access and quality are discussed, mostly for the first less experienced user. Subsequent deliverables will go into more detail and start to annotate data sources and ultimately assign a DRL level. This is not the final result, an indication of where to improve the data where needed. Basically, more introspection will be needed but will provide a convenient 'score'.

ROADVIEW goes beyond the State of the Art by not only applying the DRL concept (moving from Lawrence's three bands to DRL) to the project datasets, but also by using the largest set of 'tests' for each dataset. Data input and

output rates will be managed by machine learning pipelines and complex issues, such as batching the streamed data. ROADVIEW follows a holistic approach from the sensor hardware to the perception models presented to the decision-making system. The context of each dataset as well as their usage in the ROADVIEW system integration will define all data quality assessments.

Curated datasets such as nuScenes provide high quality data, however, are to some degree curated, for example the distance of the LiDARs is reduced to produce clear, less noisy point clouds. The real world and data are more complex with longer distances being detected.

As a result of this work, RISE AB and the National University of Singapore are collaborating, in that RISE is introducing NUS's Robo* software in the DRL and RISE is helping NUS in providing datasets with inclement weather for their challenges [31].



CollectiveSense: a collaboration concept between Sweden and Singapore

Ian Marsh, Research Institutes of Sweden (RISE), AB
Sida Jiang, FellowBot AB &
Kong Lindong, National University of Singapore (NUS)

May 2024, call text

Background

Autonomous vehicles require several sensors to 'understand' what is around a vehicle. Common sensors include visible light cameras, thermal or heat cameras, Light Detection and Ranging (LiDAR), 4G Imaging Radar Detection as well as meta information such as the vehicles' location, movements, vibration as well as the current time. Information about the road conditions from Cooperative Intelligent Transport Systems or C-ITS systems, are beginning to appear through the deployment of roadside units. The role of the weather is a critical factor in autonomous driving as adverse, or harsh, weather reduces the effect of sensors. By harsh this implies snow, ice, fog, rain as well as sand, haze, and heat effects. Simply put, cold and polar as well as hot and equatorial conditions.

Figure 19. A small spinoff project between RISE AB, FellowBot AB (an SME) and the National University of Singapore (NUS) on harsh weather and data corruptions in point cloud data.

We have rated 2 datasets, showing similar ratings, from which ML models can be trained. Also, we are refining the techniques with these and 2 other datasets. These will be reported in scientific articles, a blog post and in the last deliverable in this series, software.

References

[1] Lawrence, N. D. (2017). Data Readiness Levels. Available from link.

[2] Selvans, Z. (2021, June 17). Automated Data Wrangling. Catalyst Cooperative. https://catalyst.coop/2021/05/23/automated-data-wrangling.

[3] Rekatsinas, T., et al. (2017). HoloClean: Holistic Data Repairs with Probabilistic Inference. arxiv.org/pdf/1702.00820.pdf, https:/github.com/HoloClean/holoclean.

[4] Aksoy, E. E., Baci, S., & Cavdar, S. (2020, October). Salsanet: Fast Road and vehicle segmentation in LiDAR point clouds for autonomous driving. In 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 926-932). IEEE.

[5] Cortinhal, T., Tzelepis, G., & Aksoy, E. E. (2020). SalsaNext: fast, uncertainty-aware semantic segmentation of LiDAR point clouds for autonomous driving. arXiv preprint arXiv:2003.03653.

[6] Cortinhal, T., Kurnaz, F., & Aksoy, E. E. (2021). Semantics-aware multi-modal domain translation: From LiDAR point clouds to panoramic color images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3032-3048).

[7] Maanpää, J., Taher, J., Manninen, P., Pakola, L., Melekhov, I., & Hyyppä, J. (2021, January). Multimodal End-to-End Learning for Autonomous Steering in Adverse Road and Weather Conditions. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 699-706). IEEE.

[8] Bijelic, M., Gruber, T., & Ritter, W. (2018, June). A benchmark for LiDAR sensors in fog: Is detection breaking down? In 2018 IEEE Intelligent Vehicles Symposium (IV) (pp. 760-767). IEEE.

[9] Pfeuffer, A., & Dietmayer, K. (2019, July). Robust semantic segmentation in adverse weather conditions by means of sensor data fusion. In 2019 22nd International Conference on Information Fusion (FUSION) (pp. 1-8). IEEE.

[10] Pfeuffer, A., & Dietmayer, K. (2018, July). Optimal sensor data fusion architecture for object detection in adverse weather conditions. In 2018 21st International Conference on Information Fusion (FUSION) (pp. 1-8). IEEE.

[11] Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., & Heide, F. (2020). Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11682-11692).

[12] Jianqing Wu; Hao Xu; Jianying Zheng; Junxuan Zhao. Automatic Vehicle Detection with Roadside LiDAR Data Under Rainy and Snowy Conditions, IEEE Intelligent Transportation Systems Magazine, Volume: 13, Issue: 1, Spring 2021, link.

[13] Teja Vattem, George Sebastian; Luka Lukic, Rethinking LiDAR Object Detection in adverse weather conditions, 2022 International Conference on Robotics and Automation (ICRA), link.

[14] Alban Marie, Karol Desnos, Luce Morin and Lu Zhang, Metrics for Semantic Segmentation in a Machine-to-Machine Communication Scenario, 15th International Conference on Quality of Multimedia Experience (QoMEX) Evaluation of Image Quality Assessment, QoMEX, 2023.

[15] Richard Marcus, Niklas Knoop, Bernhard Egger, Marc Stamminger, A Lightweight Machine Learning Pipeline for LiDAR-simulation, 2022, arXiv.

[16] Jiyang Chen, Simon Yu, Rohan Tabish, Ayoosh Bansal, Shengzhong Liu, Tarek Abdelzaher, and Lui Sha, LiDAR Cluster First and Camera Inference Later: A New Perspective Towards Autonomous Driving, arXiv.

[17] Richard Marcus, Niklas Knoop, Bernhard Egger, Marc Stamminger, A Lightweight Machine Learning Pipeline for LiDAR-simulation, arXiv.

[18] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S., Ecker Matthias Bethge, Wieland Brendel, Benchmarking Robustness in Object Detection: Autonomous Driving when Winter is Coming, arXiv.

[19] LiDAR Principles, Processing and Applications in Forest Ecology, 2023.

[20] Benchmarking Robustness of 3D Object Detection to Common Corruptions in Autonomous Driving https://github.com/thu-ml/3D_Corruptions_AD.

[21] Simple, self-contained open-source project for LiDAR-based 3D object detection. https://github.com/open-mmlab/OpenPCDet.

[22] Caesar, Holger, et al. "nuscenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[23] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," in IEEE Access, vol. 8, pp. 58443-58469, 2020, doi: 10.1109/ACCESS.2020.2983149.

[24] Torres Vega, Maria & Sguazzo, Vittorio & Mocanu, Decebal & Liotta, Antonio. (2016). An experimental Survey of No-Reference Video Quality Assessment Methods. International Journal of Pervasive Computing and Communications. 12. 10.1108/IJPCC-01-2016-0008.

[25] Yinpeng Dong, Caixin Kang, Jinlai Zhang, Zijian Zhu, Yikai Wang, Xiao Yang, Hang Su, Xingxing Wei, Jun Zhu, Benchmarking Robustness of 3D Object Detection to Common Corruptions, CVPR2023. Link.

[26] Deepti Ghadiyaram, Chao Chen, Sasi Inguva, and Anil Kokaram, A No Reference Video Quality Predictor for compression and scaling artifacts. Link.

[27] Julius Pesonen**,** Pixelwise Road Surface Slipperiness Estimation for Autonomous Driving with Weakly Supervised Learning, Aalto University School of Science, 2023.

[28] Vargas J, Alsweiss S, Toker O, Razdan R, Santos J. An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions. Sensors. 2021; 21(16):5397. https://doi.org/10.3390/s21165397.

[29] T. Beemelmanns, Q. Zhang, and L. Eckstein. Multicorrupt: A multi- modal robustness dataset and benchmark of lidar-camera fusion for 3D object detection, 2024. Link.

[30] Andreas Geiger, Philip Lenz and Raquel Urtasun, Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite, Conference on Computer Vision and Pattern Recognition (CVPR), 2012. Link.

[31] RoboDrive Challenges Robodrive2024, 2025, Report, Slides

[32] The MapBench project for robust multi-modal HD map construction: http://mapbench.github.io

[33] Lingdong Kong, et. Al Robo3D: Towards Robust and Reliable 3D Perception against Corruption, link.