Ref. Ares(2025)8929528 - 20/10/2025

ROADVIEW

Robust Automated Driving in Extreme Weather

Project No. 101069576

Deliverable 4.7

GitHub release of ROADVIEW Readiness levels

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

Authors	Ian Marsh (RISE)
Lead participant	RISE
Delivery date	31 August 2025
Dissemination level	PU – Public
Type	DATA – data sets, microdata, etc.

Version 1.0

Co-funded by the European Union

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)	Initial document created	01/02/2025
Ian Marsh (RISE)	Revision 1	01/08/2025
Eren Aksoy (HH)	Review	30/08/2025
Ian Marsh (RISE)	Final version	31/08/2025
Johannes Ripperger (accelCH)	Final checks and formatting	31/08/2025

Table of Contents

Partner short names

HH	Hogskolan Halmstad
THI	Technische Hochschule Ingolstadt
RISE	RISE Research Institutes of Sweden AB
FGI	Maanmittauslaitos – Finnish Geospatial Research Institute
accelCH	accelopment Schweiz AG

Abbreviations

AD	Automated driving
D	Deliverable
GNSS	Global Navigation Satellite System
IMU	Inertial measurement unit
KRCC	Kendall Rank Correlation Coefficient
LLM	Large multimodal models
NR-IQA	No-reference image quality assessment
NR-PCQA	No-reference point cloud quality assessment
NSS	Natural scene statistics
PLCC	Pearson Linear Correlation Coefficient
RMSE	Root Mean Square Error
SRCC	Spearman's Rank Correlation Coefficient
WP	Work Package

Executive summary

This deliverable is a short report on the image and point quality assessment within work package 4 (WP4). The premise is that automated driving requires a considerable amount of data in order to train machine learning models on to detect a number of road conditions in order to be able to navigate around them, stop or avoid. Irrespective of the sensors, inclement or adverse weather make sensing all the more difficult, and to cover most cases is difficult. In this deliverable we describe the conditions, how we augmented images and point clouds to cover most weather conditions. The road conditions are derived from existing data from 2 sets, an open drive in Finland and from a test track in Southern Germany. A necessary prerequisite therefore for autonomous driving is quantitative data quality assessment.

This deliverable investigates methods for evaluating the quality of images and point clouds without relying on reference data, known as no-reference image quality assessment and no-reference point cloud quality assessment. We evaluate five methods for image quality and two for point cloud quality. Our image data was sourced from two drives in winter conditions, and the point clouds were obtained from a test track setting. This was sensor data captured under controlled weather conditions. The images and point clouds used in this deliverable were synthetically distorted by artificial fog and rain.

Our findings suggest that large multimodal models (LLM) and natural scene statistics (NSS) are promising approaches for assessing the quality of weather-distorted images. The results also suggest that reference point cloud methods are not yet mature enough to reliably evaluate point cloud quality in the automated driving domain.

Finally, as part of this deliverable, all the code for this, and the Data Readiness Level work is available.

1 Introduction

This deliverable systematizes degradations in multimodal data sources for Automated Driving (AD). The focus of this paper is on the image and point cloud quality using open-source metrics. Our motivation in this work is to ensure the data gathered during the data collection phase is complete and of sufficient quality. Simply, iterating the data collection phase is costly, and when incomplete potentially dangerous when the AD vehicle is in production and use.

A prominent research area within general-purpose no-reference image quality assessment (NR-IQA) is the field of natural scene statistics (NSS). NR-IQA based on NSS relies on the insight that high-quality natural images exhibit consistent statistical properties, and distorted images possess measurable deviations from such statistical properties [1].

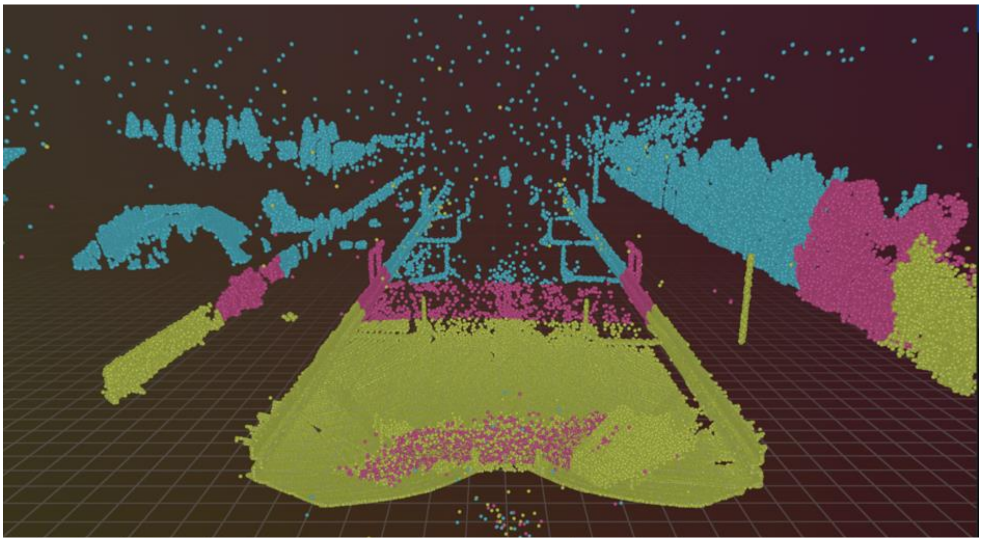

Figure 1. A LiDAR point cloud from our test track in THI Ingolstadt, Germany. The point cloud is augmented by fog-based noise to evaluate quality tools. The colour indicates the amount of fog noise added.

From an originally large set of 42 image quality methods and five-point cloud methods, we narrowed the choices down to just five image quality methods, and two point cloud methods in this work. One criterion for reduction is certain tools output very similar results. Alternatively, they use similar approaches. Indeed, Table 1 not only mentions the tools, but the underlying technology behind each. To do a thorough quality assessment evaluation across large datasets, see Table 1, by selecting a subset of quality assessment tools, we could save considerable time and expense plus keep the results simple and meaningful. The selected tools are listed in Table 1. A paper outlining the reduction from 42 to five tools is available from link.

Table 1. List of image quality and point cloud tools. Open source was a necessity, and preferably code with an accompanying publication. For the 5 image quality methods we selected those using different underlying assessment methods.

Modality	Method	Type
Image Quality	IL-NIQE	NSS-based
Image Quality	DBCNN	CNN-based
Image Quality	TOPIQ	Attention-based
Image Quality	QualiCLIP	CLIP-based
Image Quality	Q-Align	LMM-based
Point Cloud	MM-PCQA	Projection- and model-based
Point Cloud	MS-PCQE	Projection-based

2 Related Work

Liu et al. postulate that AD dataset surveys are not sufficiently extensive [2]. Therefore, they present an exhaustive study of 200+ datasets encompassing sensor modalities, varying sizes, tasks, and contextual conditions. Vargas et al. look at automated vehicle sensors and their vulnerability to weather [3]. Sheikh et al. developed the LIVE dataset, one of the first and most widely used synthetic benchmarks for NR-IQA [4]. It contains 779 distorted images that were generated by distorting 29 pristine reference images. Five distortion types were used in the process: JPEG compression, JPEG2000 compression, white noise, Gaussian blur, and simulated fast fading. Despite having a limited range of distortion types, LIVE remains a standard benchmark in the field of NR-IQA. Another prominent synthetic benchmark by Ponomarenko is TID2013, which comprises 3000 distorted images generated from 25 reference images using 24 different distortion types [5]. In addition to standard degradations, Gaussian noise and JPEG compression, TID2013 includes specialized distortions, such as chromatic aberrations, mean shift, and non-eccentricity pattern noise.

Ghadiyaram and Bovik introduced CLIVE, a widely used NR-IQA benchmark dataset containing authentic distortions of images [6]. It consists of 1162 images captured under natural conditions using mobile phones. The images in CLIVE have complex and varied real-world distortions such as low-light conditions, unfocused images, low resolutions, or motion blur. The authors conducted subjective quality scores through a large-scale crowdsourced study involving 8000 participants, which provided some ground-truth values. This work distinguishes itself from prior work by focusing on the evaluation of NR-IQA and NR-PCQA methods in the realm of automated driving. While previous studies assess models on general-purpose image and point cloud benchmarks, we limit ourselves to images and point clouds from real environments and augmented to capture a whole range of weather conditions. A continually updated list of adverse weather datasets is at link.

3 Test tracks

Table 2. Open road (FGI in Finland) and testing track (REHEARSE in Germany) adverse-weather conditions drives and datasets.

Provider	Name	Size	Details
Finnish Geospatial Institute, FI	FGI	14 GB	One urban & one rural journey ~49 km Recorded 03-12-2023.
Technische Hochschule Ingolstadt, DE	REHEARSE	320 GB	H_2O-generated (not snow and ice) weather conditions. Data collection ongoing.

3.1 A Finnish open road setting

The FGI dataset includes data from two drives: one urban drive around Otaniemi, Espoo, and another covering both urban and rural scenery, from Espoo to Munkkivuori in Helsinki, Finland. These two drives are referred to as Otaniemi and Munkkivuori, respectively. The Otaniemi drive was captured during the day, resulting in brighter images, whereas the Munkkivuori drive was captured in the evening, resulting in darker images. Both drives were recorded on the 3rd of December 2023, naturally during winter conditions in Finland. The modalities were LiDAR, RGB and thermal cameras, as well as IMU movement and vibrations and GNSS location services. The Munkkivuori drive has of 6915 PNG images, whilst the Otaniemi drive 7851 PNG images.

The FGI dataset was selected for the NR-IQA evaluation due to its authentic conditions, providing realistic winter road scenes encountered during real-world driving. Although the REHEARSE dataset contains image data, it is only images captured on a test track or in a tunnel, not in authentic driving conditions. From each drive of the FGI dataset, 10 images were selected as reference images for the synthetic distortions using Python's random package using a known seed for reproducibility. This results in a total of 20 reference images. The number of images was chosen to balance scene diversity with computational efficiency. The selected reference images are shown in Figure 2.

Figure 2. *Left*: Ten reference images used from the Finnish (Munkkivuori) dataset. Note the higher contrast caused by the other vehicles' brake lights in the image sequence. *Right*: Ten reference images used from the Finnish (Otaniemi) dataset. Note, the relatively similar scenes, with respect to intensity and contrast, and the low light conditions.

3.2 A German test track setting

The REHEARSE dataset comprises LiDAR, image, and RADAR data collected from an outdoor test track and an indoor tunnel using static sensors, see [8]. The data is collected under clear, rainy, and foggy conditions, with the rain and fog artificially generated using sprinklers. The point clouds are stored as binary files, each containing a sequence of points of type (x,y, z, i) where (x,y,z) are the Cartesian coordinates and i is the intensity value.

For the point cloud data, the REHEARSE dataset was selected over the FGI dataset. The REHEARSE dataset was recorded under controlled conditions, outdoors and known distances, therefore was preferred. A total of 40-point clouds were randomly selected from the outdoor test track, comprising 20 captured in clear weather and 20 in rainy conditions. One other dimension we considered is normal and low light conditions in this case of the 40-point clouds, 27 were recorded during the day, and 13 were recorded during the nighttime. Only one distortion type was used in the point cloud quality evaluation, see Section 4.

4 Data augmentation

4.1 Background

We can generate weather impairments to cover additional cases compared with those *solely* gathered from driving alone. Also, purely synthetically produce images, although getting better, still produce 'cartoon-ish' scenes.

The Albumentations Python library [albumentations] was used to synthetically apply weather distortions to images. It is an image augmentation library featuring over 100 transformations for augmenting images. The selection of Albumentations is motivated by its support for weather-related distortions, particularly artificial rain and fog, as well as its ability to generate image sets with monotonically increasing noise severities. Also, some image augmentation tools use depth information to create more realistic noise, but that requires that every pixel in the image has a depth value. Albumentations doesn't require depth information, an advantage in this work. Albumentations does not require depth information, which is not available for the images and is a prerequisite for synthetic distortion techniques. It has also been successfully used as a tool for applying synthetic weather distortions in prior studies [10]-[12].

This deliverable uses two types of synthetic weather distortions for images: rain and fog. Noise from falling snow was excluded because Albumentations does not implement it, and no suitable library capable of generating image sets with monotonically increasing snow noise was found.

4.2 Rain distortion

Within Albumentations, the functional transform `add_rain` is used to create synthetic images with rain effects. The `add_rain` transform adds semi-transparent streaks to an image, simulating the appearance of raindrops. This is achieved by first defining a set of coordinates as the starting points for the raindrops, along with their slant angle, width, length, and colour. Using this information, `add_rain` utilizes OpenCV to draw a line for each specified raindrop. `add_rain` also blurs the image using a box blur with a kernel size determined by the `blur_value` parameter. Finally, it changes the brightness of the image by converting it to Hue, Saturation, Value (HSV) format and multiplying the value component by the `brightness_coefficient` parameter.

To simulate a gradual and uniform increase in rain-induced distortion, droplet_share is progressively increased to simulate denser rainfall, while brightness_coefficient is decreased to reflect reduced visibility due to the intensified rain, i.e.

- Slant
- Drop length
- Drop width
- Blur value

are kept constant to maintain consistent raindrop geometry and avoid abrupt visual changes. Example of two versions of the same image with varying amounts of rain distortion are shown in Figure 3 below.

To simulate a gradual and uniform increase in rain-induced distortion, droplet_share is progressively increased to simulate denser rainfall, while brightness_coefficient is decreased to reflect reduced visibility due to the intensified rain. slant, drop_length, drop_width, and blur_value are kept constant to maintain consistent raindrop geometry and avoid abrupt visual changes. One image with rain distortion is shown in Figure 3.

Figure 3. Rain distortion, Left droplet share = 0.007 & brightness 0.7. Right: droplet share = 0.003, brightness 0.9

4.3 Fog distortion

Fog effects are synthesized using the `add_fog` transform from Albumentations. `add_fog` simulates fog by taking an input image, a fog intensity coefficient, `fog_intensity`, a transparency value, `alpha_coef` for the fog particles, a list of coordinates specifying their positions, and a list of their respective radii. For each specified fog particle, add_fog uses OpenCV to draw a circle on a copy of the image. The image with the circle is then blended in the original image.

For each reference image across both locations, 100 versions with synthetic fog were synthesized. The progressive increase was linearly spaced, following the same approach as for the rain-distorted images. The choice of 100 distortion levels follows the reasoning of Ahar et al. to capture subtle image quality differences. An example is below.

Figure 4. Fog distortion: A fog visual strength = 0.53.

5 Evaluation

5.1 No-reference image quality evaluation

No-reference quality assessment methods encompass both no-reference image quality and no-reference point cloud quality methods. They are typically assessed by comparing their predicted scores with human evaluations of the same images or point clouds. Evaluations, asking humans to adjudge datasets is done, in this domain called subjective assessment, on a five-point scale. We need statistical measures to state if the five considered methods are relevant, unambiguously. Using the subjective scores, NR-IQA methods are typically assessed using the following:

- Pearson Linear Correlation Coefficient (PLCC)
- Spearman's Rank Correlation Coefficient (SRCC)
- Kendall Rank Correlation coefficient (KRCC)
- Root Mean Square Error (RMSE)

PLCC and RMSE assess the accuracy of predicted scores by measuring their deviation from corresponding subjective scores. In contrast, SRCC and KRCC evaluate how well the predicted scores preserve the rank order of the subjective evaluations. SRCC measures the monotonic relationship between the predicted scores and the ground-truth quality scores by comparing their ranks, KRCC and SRCC, use ranks as opposed to absolute scores for its evaluation. It is computed by comparing all possible pairs of observations (i, j) such that i < j, and classifying them as concordant, discordant, or tied. For two variables X = {X_1, ..., X_n} and Y = {Y_1, ..., Y_n}, we consider all pairs (i, j): i < j <=n. A pair is concordant, discordant, or tied.

5.2 No-reference point cloud quality evaluation

Figure 1 showed a point cloud augmented with noise. As with image quality assessment, point cloud quality assessment methods are categorized as full reference, reduced reference, or no reference. Since LiDAR point clouds collected by automated vehicles typically lack reference data, only no-reference PCQA models were considered in this work. Categories within no-reference modelling exist; namely model-based, projection-based, and hybrid approaches. Model-based NR-PCQA techniques extract colour and geometry information directly from 3D point clouds without projecting them onto a 2D plane [15]. Projection-based methods project points in the point cloud onto 1+ 2D planes. The quality evaluation is then done on the projections, resulting in (relative) computationally efficient assessments. Hybrid NR-PCQA combines projection and model-based methods to capture distortions a single modality cannot. Zhang et al. state structural degradation and geometry downsampling are better detected in point clouds using a hybrid approach [16].

6 Results

6.1 Image quality

Figure 5 shows the Spearman correlations. For Q-Align and IL-NIQE the results are good, with the caveat of the execution time of Q-Align.

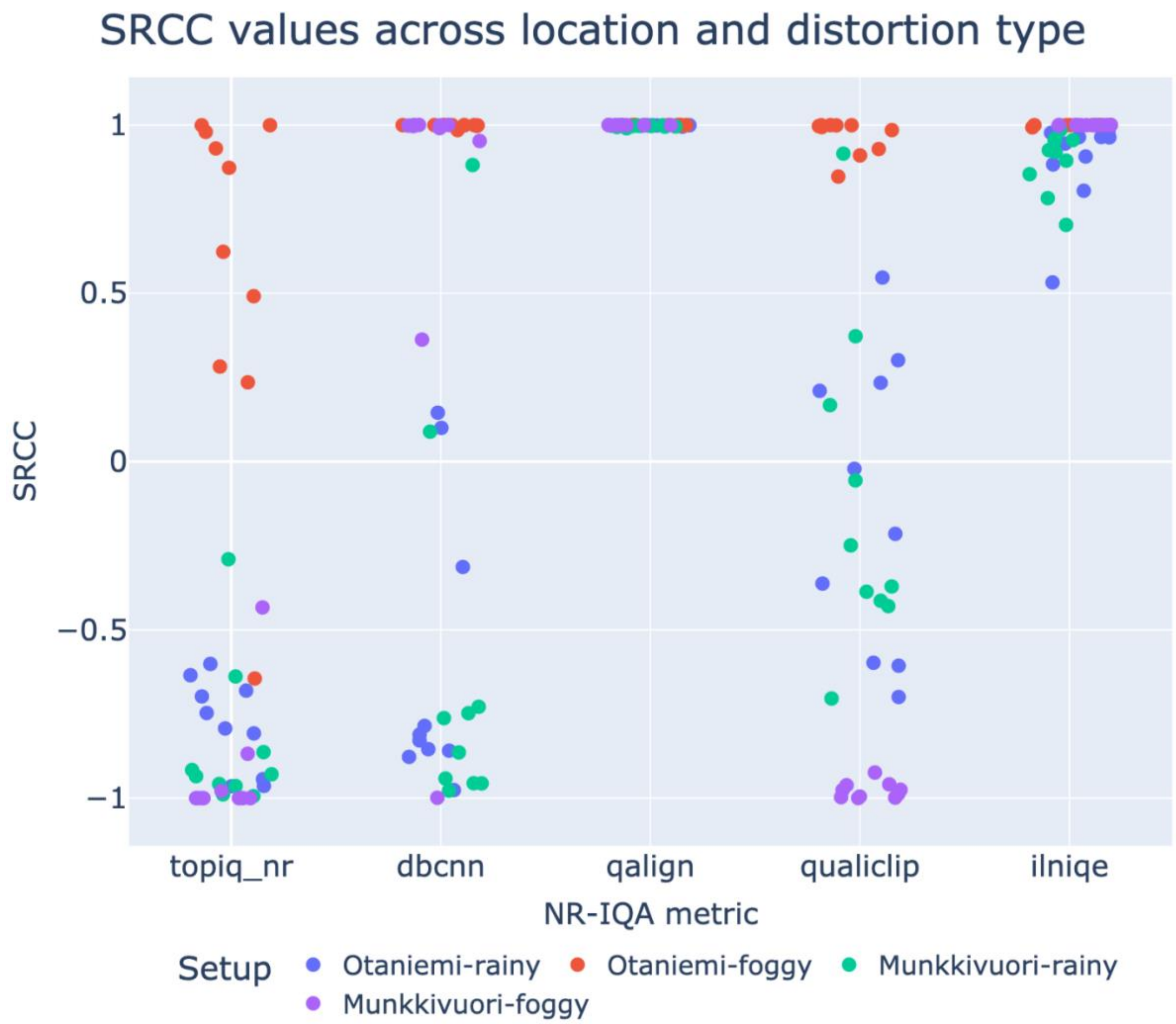

Figure 5. Image quality metrics using Spearman's rank per image quality type. Colours indicate the weather conditions. High average correlation (a mean close to 1) with minimal deviation, a standard deviation shows Q-align performing best.

6.2 Point cloud quality

Figure 6 shows the Spearman correlations (SRCC) in good agreement, i.e. their means within 0.1 of each other. However, the quality for a no-reference point cloud method requires additional research from our findings.

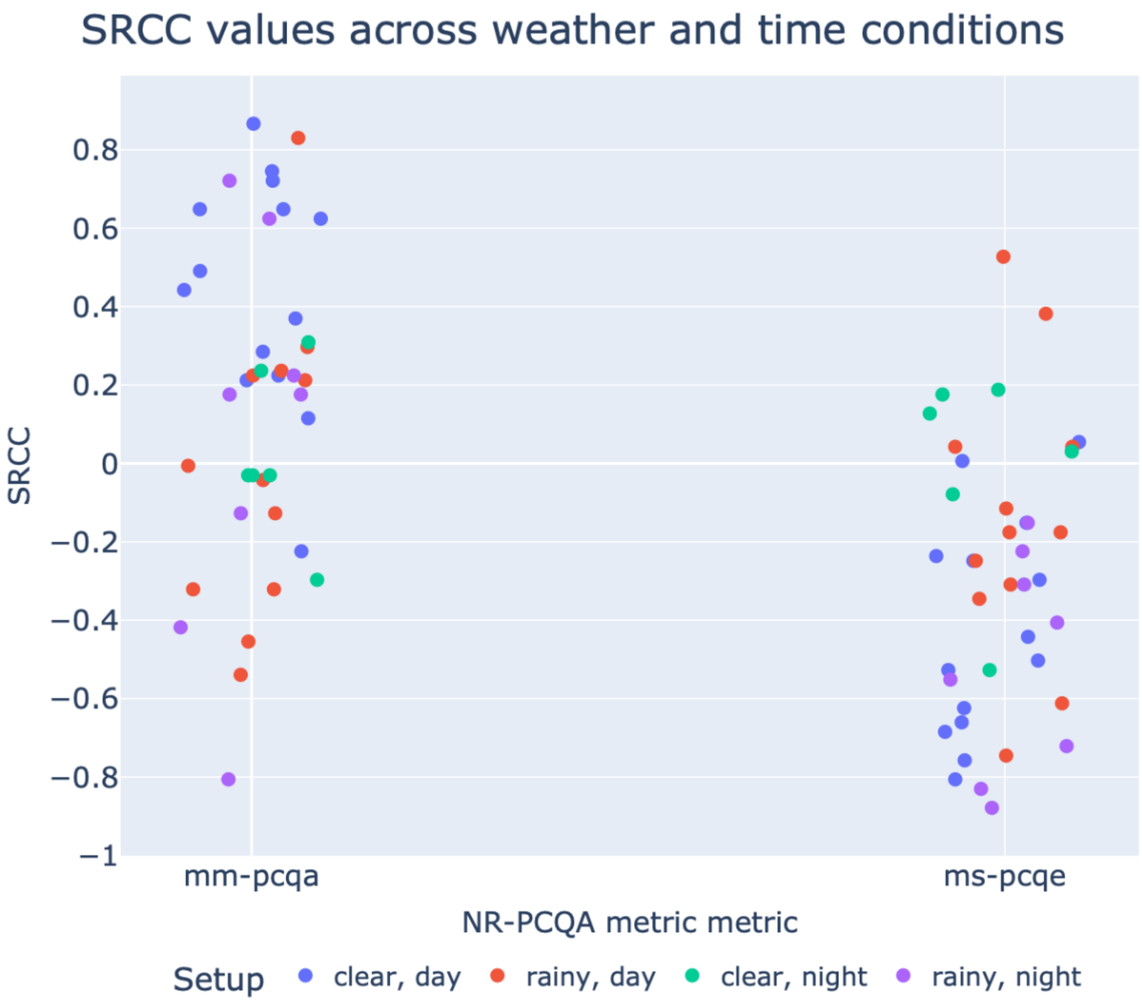

Figure 6. Point cloud results for two methods across differing weather (colour coded) and light (day-night) conditions. Note the low mean values and relatively high standard deviation. Both Spearman and Kendall report similarly low (poor in this case) correlation.

7 Discussion

General. A first point for this work is, with reference images and point clouds, would the results be better or usable? Clearly yes, however in an urban real-world drive, a reference is hard to obtain. That said, in the test track case, conditions, weather (generated) and distances are known. Therefore, we expect a reference in these cases would be better than those reported here.

Image quality. For the NR-IQA results Q-Align outperforms the four other evaluated methods. As a state-of-the-art, large multimodality model, Q-Align has demonstrated very strong performance on conventional NR-IQA benchmarks [17]. Q-Align was also trained on 12 NR-IQA datasets that contained both authentic and synthetic distortions. Training on both authentic and synthetic datasets reduces the accuracy of an NR-IQA model [18], the authors of Q-Align report that their model avoids this issue. This exposure to such a wide range of image content likely contributed to the superior performance of Q-Align.

Point cloud quality. The poor results of the NR-PCQA models indicate that the field of NR-PCQA is not yet mature enough to accurately assess AV point clouds. MS-PCQE exhibited a negative correlation with the ground-truth rankings, indicating a failure to accurately rank point clouds wrt. distortion levels.

Resources. Q-Align was the slowest model, inference times being approx. 70% higher than IL-NIQE. Q-Align with peak GPU memory being x10 greater than that of IL-NIQE. Given these differences in computational cost, and that IL-NIQE achieves good results with much lower inference times and GPU requirements, it may be preferable to use IL-NIQE instead of Q-Align in certain situations. We also evaluated quality over day-night conditions; however, space-time limitations necessitate a whole new article.

8 Conclusions

This paper evaluated a set of image and point cloud quality methods on weather-distorted data used in the field of automated vehicles. The evaluation centred on synthetic weather distortions of images and point clouds. In the case of images, synthetic rain and fog were 'added', whereas in point clouds, only synthetic fog was used.

Among the five evaluated no reference image quality, Q-Align and IL-NIQE showed solid results across fog and snow distortions, hence large multimodality and NSS models are suitable for automated driving. While both are acceptable, Q-Align outperforms IL-NIQE, however has (much) higher computational needs. Our no reference point cloud evaluations indicates both selected methods (MM-PCQA, MS-PCQE) showed low correlations, high variability, meaning unacceptable quality estimates.

To conclude, our statistically backed findings state that no reference image quality assessment is mature and reliable, whilst no reference point cloud quality assessment requires further research for applications in the driving domain.

9 Code

The code for this work and Data Readiness Levels can be found in the ROADVIEW GitHub repository:

https://github.com/roadview-project/No_reference_image_and_point_cloud_quality

10 Acknowledgements

Thanks to the Finnish Geospatial Institute (FGI) for the recording, processing, and sharing the data sets from Otaniemi and Munkkivuori. The work was co-funded by the European Union under grant number 101069576. 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. UK participants in this project are co-funded by Innovate UK under grant number 10045139.

11 References

[1] Q. Mao, S. Liu, Q. Li, G. Jeon, H. Kim, and D. Camacho, “No-reference image quality assessment: Past, present, and future,” Expert Systems, vol. 42, no. 3, p. e13842, 2025.

[2] M. Liu, E. Yurtsever, X. Zhou, J. Fossaert, Y. Cui, B. L. Zagar, and A. C. Knoll, “A survey on autonomous driving datasets: Data statistic, annotation, and outlook,” 2024.

[3] J. Vargas, S. Alsweiss, O. Toker, R. Razdan, and J. Santos, “An overview of autonomous vehicles sensors and their vulnerability to weather conditions,” Sensors, vol. 21, no. 16, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/16/5397

[4] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” IEEE Transactions on image processing, vol. 15, no. 11, 2006.

[5] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti et al., “Image database tid2013: Peculiarities, results and perspectives,” Signal processing: Image communication, vol. 30, pp. 57–77, 2015.

[6] D. Ghadiyaram and A. C. Bovik, “Massive online crowdsourced study of subjective and objective picture quality,” IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 372–387, 2015.

[7] I. Marsh, “Initial readiness assessment of specific datasets,” https://roadview-project.eu/wp-content/uploads/sites/59/2024/05/ROADVIEW Deliverable-4.5 v04.pdf, May 2024.

[8] Y. Poledna, M. F. Drechsler, V. Donzella, P. H. Chan, P. Duthon, and W. Huber, “Rehearse: adverse weather dataset for sensory noise models,” in 2024 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2024, pp. 2451–2457.

[9] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, “Albumentations: Fast and flexible image augmentations,” Information, vol. 11, no. 2, 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/2/125

[10] H. Mokayed, A. Nayebiastaneh, K. De, S. Sozos, O. Hagner, and B. Backe, “Nordic vehicle dataset (nvd): Performance of vehicle detectors using newly captured nvd from uav in different snowy weather conditions.” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 5314–5322.

[11] Y. Kim, B. Park, and I.-Y. Moon, “A study on object detection performance through data augmentation under adverse weather conditions,” in 2024 International Conference on Sustainable Technology and Engineering (i-COSTE). IEEE, 2024, pp. 1–3.

[12] A. Shiran, J. Li, Y. Liu, M. Hummel, O. Jenewein, and K. Bezboruah, “Water level detection in adverse weather conditions using security cameras,” in SoutheastCon 2024. IEEE, 2024, pp. 187–193.

[13] A. Ahar, S. Mahmoudpour, G. Van Wallendael, T. Paridaens, P. Lambert, and P. Schelkens, “A just noticeable difference subjective test for high dynamic range images,” in 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), 2018, pp. 1–6.

[14] D. C. Lepcha, B. Goyal, A. Dogra, and V. Goyal, “Image superresolution: A comprehensive review, recent trends, challenges and applications,” Information Fusion, vol. 91, pp. 230–260, 2023.

[15] S. Porcu, C. Marche, and A. Floris, “No-reference objective quality metrics for 3d point clouds: A review,” Sensors (Basel, Switzerland), vol. 24, no. 22, p. 7383, 2024.

[16] G. Zhai and X. Min, “Perceptual image quality assessment: a survey,” Science China Information Sciences, vol. 63, pp. 1–52, 2024.

[17] Z. Zhang, W. Sun, X. Min, Q. Zhou, J. He, Q. Wang, and G. Zhai, “Mm-pcqa: Multi-modal learning for no-reference point cloud quality assessment,” arXiv preprint arXiv:2209.00244, 2022.

[18] H. Wu, Z. Zhang, W. Zhang, C. Chen, L. Liao, C. Li, Y. Gao, A. Wang, E. Zhang, W. Sun et al., “Q-align: Teaching lmms for visual scoring via discrete text-defined levels,” arXiv:2312.17090, 2023.