



Degree Project in Technology

Second cycle, 30 credits

Evaluating NR-IQA and NR-PCQA Methods on Weather-Distorted Data in Autonomous Driving

A subtitle in the language of the thesis

VICTOR STENMARK

Evaluating NR-IQA and NR-PCQA Methods on Weather-Distorted Data in Autonomous Driving

A subtitle in the language of the thesis

VICTOR STENMARK

Bachelor's Programme in Information and Communication Technology

Date: June 12, 2025

Supervisors: Domova Veronika, Ian Marsh

Examiner: Haibo Li

School of Electrical Engineering and Computer Science

Host company: Företaget AB

Swedish title: Detta är den svenska översättningen av titeln

Swedish subtitle: Detta är den svenska översättningen av undertiteln

Abstract

Ensuring the use of high-quality data is essential in autonomous driving to develop safe and reliable self-driving models. A necessary prerequisite for this is to accurately assess data quality. This thesis investigates methods for evaluating the quality of images and point clouds without relying on reference data, known as no-reference image quality assessment (NR-IQA) and no-reference point cloud quality assessment (NR-PCQA), respectively. Five NR-IQA methods (IL-NIQE, TOPIQ, DB-CNN, QualiCLIP, and Q-Align) and two NR-PCQA methods (MM-PCQA and MS-PCQE) were selected and evaluated based on their ability to accurately assess weather-distorted images and point clouds. The image data were sourced from the FGI dataset, which consists of two drives in winter conditions, and the point clouds were obtained from the REHEARSE dataset, comprising sensor data captured under controlled weather conditions. The images and point clouds used in the thesis were synthetically distorted with artificial fog and rain. The NR-IQA and NR-PCQA methods were evaluated based on their ability to accurately rank different versions of the same image or point cloud by distortion level. Among the surveyed NR-PCQA methods, none demonstrated reliable performance in ranking the distorted point clouds. While MM-PCQA outperformed MS-PCQE, it still struggled to produce accurate rankings. In contrast, Q-Align and IL-NIQE achieved the best performance among the NR-IQA methods, while the others performed significantly worse. These findings suggest that large multimodal models and natural scene statistics are particularly promising approaches for assessing the quality of weather-distorted images. Conversely, the results also suggest that NR-PCQA methods are not yet mature enough to reliably evaluate point cloud quality in the autonomous driving domain.

Keywords

Canvas Learning Management System, Docker containers, Performance tuning

Sammanfattning

Nyckelord

Canvas Lärplattform, Dockerbehållare, Prestandajustering

Acknowledgments

I would like to thank xxxx for having yyyy. Or in the case of two authors:
We would like to thank xxxx for having yyyy.

Stockholm, June 2025

Victor Stenmark

Contents

1	Introduction	1
1.1	Background	2
1.2	Problem	3
1.3	Purpose	3
1.4	Ethics and Sustainability	3
1.5	Structure of the thesis	3
2	Background	5
2.1	Data quality in autonomous driving	5
2.2	No-reference image quality assessment	7
2.2.1	Natural scene statistics	7
2.2.2	Deep learning approaches	8
2.3	No-reference point cloud quality assessment	9
2.3.1	Model-based NR-PCQA	10
2.3.2	Projection-based NR-PCQA	10
2.3.3	Hybrid methods	11
2.4	Related work	12
2.5	Summary	13
3	Method	15
3.1	Literature search	15
3.2	Considered methods	15
3.3	Chosen approach	17
3.4	Dataset	18
3.5	Preprocessing images	20
3.5.1	Synthetic rain distortion	20
3.5.2	Synthetic fog distortion	21
3.6	Preprocessing point clouds	24
3.6.1	Point cloud distortion implementation	28

3.7	Selection of NR-IQA techniques	28
3.8	Selection of NR-PCQA techniques	30
3.9	Evaluation methods	31
3.10	Scoring images and point clouds	32
3.11	Analysis	33
4	Results and Analysis	37
4.1	NR-IQA results	37
4.1.1	Statistical tests	46
4.2	NR-PCQA results	47
4.2.1	Statistical tests	49
5	Discussion	52
6	Conclusions and Future work	57
6.1	Conclusions	57
6.2	Limitations	57
6.3	Future work	58
	References	61

List of Figures

3.1	Example where OD confidence does not correlate with differences in image quality.	16
3.2	The ten reference images used from the Otaniemi drive.	19
3.3	The ten reference images used from the Munkkivuori drive.	20
3.4	The same image from Otaniemi with different amounts of rain-related noise.	22
3.5	The same image from Munkkivuori with different increasing amounts of foggy noise.	25
3.6	The algorithm used to apply fog distortion to a particle p with intensity i [43].	26
3.7	Visualization of three point clouds subjected to increasing distortion levels. The blue point cloud corresponds to $\alpha = 0.03$, the pink to $\alpha = 0.06$, and the yellow to $\alpha = 0.09$	29
4.1	Distribution of SRCC values for the surveyed NR-IQA methods.	40
4.2	Boxplot of DB-CNN scores for every fifth image index, showing score distributions across weather and location variations	41
4.3	Boxplot of TOPIQ scores for every fifth image index, showing score distributions across weather and location variations	42
4.4	Boxplot of Q-Align scores for every fifth image index, showing score distributions across weather and location variations	43
4.5	Boxplot of QualiCLIP scores for every fifth image index, showing score distributions across weather and location variations	44
4.6	Boxplot of IL-NIQE scores for every fifth image index, showing score distributions across weather and location variations	45

4.7	Distribution of SRCC values for the surveyed NR-PCQA methods.	49
4.8	MM-PCQA score distributions across distortion levels and conditions.	50
4.9	MS-PCQE score distributions across distortion levels and conditions,	51
5.1	Example where QualiCLIP and Topiq assign a higher score to a perceptually worse-quality image	54

List of Tables

3.1	Parameters for the synthetic rain distortion	23
3.2	Parameter values used for synthetic fog distortion.	27
3.3	Parameters for the synthetic fog distortion of point clouds.	29
3.4	NR-IQA methods included in the study	35
3.5	Table of NR-PCQA methods included in the study	36
4.1	Mean SRCC and KRCC values of the NR-IQA techniques across both distortion types and locations, with standard deviations shown in parentheses. The p-values are obtained from permutation tests evaluating whether the mean SRCC and KRCC differ significantly from 0	37
4.2	Results of the Wilcoxon signed-rank tests after Holm-Bonferroni corrections.	47
4.3	Mean SRCC and KRCC values of the NR-PCQA techniques on the set of fog-distorted point clouds. The p-values are obtained from permutation tests evaluating whether the mean SRCC and KRCC differ significantly from 0.	48

Listings

Chapter 1

Introduction

Autonomous vehicles (AVs) are becoming increasingly prevalent as more companies invest in self-driving technology. In 2024-2025, 45% of car manufacturers are either investing or planning to invest in autonomous vehicle technology [1]. Currently, the market for autonomous vehicles is valued at \$122 billion, with projected revenues of \$400 billion in 2035 [1].

The growing development of AVs is enabled by rapid advancements in deep learning and other technologies that make sophisticated autonomous systems possible [2]. The core of these systems lies in two key components: a sensor suite that collects data about the vehicle's surroundings and motion, and an autonomous driving system (ADS) that interprets this data to control the vehicle. Common sensors used in the sensor suites of AVs include cameras, LiDAR, and RADAR. Cameras capture visual information of the environment, which is typically used for tasks such as object detection and path planning. LiDAR and RADAR generate point clouds of the surrounding environment, enabling accurate estimation of distances to nearby objects.

A key challenge in autonomous driving is striking the right balance in the quality of the training data used to develop the ADS. Training exclusively on high-quality data captured in clear-weather conditions may limit the ADS's ability to generalize to environments with fog, rain, or snow [3, 4, 5]. However, relying too much on low-quality data, such as weather-distorted images and degraded point clouds, may degrade model performance in clear conditions [6]. To ensure robust ADS performance, it is therefore important to maintain a balance in the quality of data used during training, which in turn requires reliable methods for assessing data quality.

This thesis aims to improve the current understanding of evaluating the quality of both camera and LiDAR data. More specifically, it compares

methods for evaluating the quality of images and LiDAR point clouds. The thesis is carried out at RISE in collaboration with ROADVIEW [7], an EU-funded project focused on developing perception and decision-making systems for automated vehicles in harsh weather conditions. The thesis will primarily interest ROADVIEW, as they seek to evaluate the quality of the data collected by their vehicles. The project may also benefit autonomous researchers who want to estimate the quality of image and point cloud datasets.

1.1 Background

The evaluation of images and point clouds can be divided into three fields: full-reference quality assessment (FR-QA), reduced-reference quality assessment (RR-QA), and no-reference quality assessment (NR-QA) [8, 9]. FR-QA refers to methods where an image or point cloud's quality is determined by comparing it to a reference image or point cloud with assumed perfect quality. RR-QA works by comparing the image to a set of extracted features from the perfect reference image or point cloud. Finally, NR-QA refers to methods where the evaluated images or point clouds are assessed independently, without any reference information. Images and point clouds captured by cameras and LiDARs on AVs lack corresponding reference data. This means that only NR-QA methods will be useful in the quality evaluation.

The thesis is concerned with evaluating image and point cloud data affected by weather-related noise. Weather events like rain, snow, and fog all affect the performance of cameras and LiDARs. Cameras can be incapacitated by rain or snow droplets on the lens [10]. In foggy conditions, cameras also struggle to capture clear images, which complicates the process of object detection. LiDAR performs better than cameras in light rain but struggles in moderate to heavy rain [10]. In those conditions, the rain droplets can be mistaken for obstacles, generating false points in the point cloud.

As previously mentioned, this thesis is conducted in collaboration with ROADVIEW and is, therefore, aligned with its objective of improving the performance of ADS in Nordic weather conditions. Accordingly, only weather distortions common in the Nordics are considered within the scope of this work. These weather distortions include rain and fog.

1.2 Problem

- **RQ1:** What methods are suitable for no-reference quality assessment of images captured by vehicle-mounted cameras under Nordic weather distortions?
- **RQ2:** What methods are suitable for no-reference quality assessment of point clouds generated by vehicle-mounted LiDARs under Nordic weather distortions?

1.3 Purpose

The overarching purpose of this thesis is to enhance the robustness and reliability of autonomous vehicles in Nordic conditions. Specifically, the aim is to improve the dependability of ADS by identifying methods for accurately assessing the quality of weather-distorted data.

1.4 Ethics and Sustainability

1.5 Structure of the thesis

Chapter 2

Background

This chapter begins with an overview of data quality and the research fields of no-reference image quality assessment (NR-IQA) and no-reference point cloud quality assessment (NR-PCQA). Subsequently, it examines the evaluation strategies, methods explored throughout the thesis, and the approach ultimately adopted. It ends with a discussion of the related work in the field. Throughout this chapter and the remainder of the thesis, the term no-reference quality assessment (NR-QA) is used to refer to both no-reference point cloud quality assessment (NR-PCQA) and no-reference image quality assessment (NR-IQA) methods.

2.1 Data quality in autonomous driving

Autonomous vehicles use a suite of sensors to navigate safely on the road, with LiDAR and RGB cameras being among the most widely utilized. A LiDAR operates by scanning the environment using one or several laser beams and then detecting the reflected signal [11]. In most LiDAR systems used in autonomous vehicles, distances to points are estimated using time-of-flight (ToF) measurements. This means that the distances to surrounding objects are calculated based on the time it takes for the laser pulse to travel to the surface and then return to the LiDAR. In addition to the measured distance, the LiDAR system records the corresponding azimuth (horizontal) and elevation (vertical) angles, denoted as ϕ and θ , respectively. These angles, together with the distance measurement, are used to compute the 3D Cartesian coordinates of each point [11], creating a point cloud representation of the surrounding environment. Furthermore, the system records the intensity of each reflected laser pulse, which is the strength of the returned signal. This means that

each point in the LiDAR output is represented in the form (x, y, z, i) , where (x, y, z) denotes the Cartesian coordinate and i represents the return intensity. RGB cameras are also commonly used as sensors for autonomous vehicles, as they can capture images with rich visual details of the environment. This makes them particularly well-suited for tasks such as object detection and classification, including the identification of road signs, pedestrians, and surrounding vehicles [12].

Collecting autonomous driving data consists of several key steps. A set of sensors has to be selected and mounted on a vehicle. The sensors must also be calibrated to ensure that the collected data accurately represents the surrounding environment. The next step involves driving the car along a set of planned routes and collecting data via its sensors. This raw data is then usually processed in several ways. One way this is done is by annotating images with 2D bounding boxes and point clouds with 3D cuboids to identify objects of interest [12]. Alternatively, each pixel or each point cloud is assigned a label describing which object or region it belongs to [12].

Several weather-related factors influence the quality of the collected images and point clouds. Rain can significantly impair camera performance as individual raindrops may stick to the lens, resulting in occlusions or blockages that render images unusable. Additionally, fog can cause uniform visual distortion, making it difficult to discern details and objects in the scene. Foggy conditions also create problems for LiDARs as the LiDAR beam is attenuated and backscattered. Attenuation occurs when fog droplets absorb and scatter portions of the LiDAR's laser energy, leading to a reduction in signal strength. Backscattering occurs when part of the laser signal is reflected by fog particles, resulting in false returns and noise in the point cloud. LiDAR typically performs better in rainy conditions, as attenuation and scattering are negligible in low to moderate rainfall. However, heavy and irregular rain causes lumps of fog, which leads to similar problems with attenuation and backscattering.

Current approaches to evaluating the quality of AV data primarily focus on assessing annotation quality [12, 13]. One key metric is consistency, which refers to the degree of consistency in labeling objects throughout a dataset. For example, if one type of vehicle is labeled as a car in one instance, it should be labeled as a car in all other cases in which it appears within the dataset [12]. Another key metric is annotation precision, which refers to the extent to which the assigned labels in a scene accurately represent the actual state of that scene.

This thesis distinguishes itself from prior research on AV data quality by focusing on the intrinsic quality of the data itself, rather than on the quality of

annotations or labeling.

2.2 No-reference image quality assessment

Methods of NR-IQA can be classified into distortion-specific and general-purpose methods [14]. Distortion-specific methods are designed to evaluate the quality of images affected by known distortions. Examples of such distortions include JPEG and JPEG2000 compression, for which several distortion-specific NR-IQA methods have been developed [15]. In contrast, general-purpose methods are designed to evaluate the quality of images regardless of the type of distortion. As weather-related noise comes in multiple forms (snow, rain, and fog), only general-purpose methods will be considered in this thesis.

2.2.1 Natural scene statistics

A prominent research area within general-purpose NR-IQA is the field of natural scene statistics (NSS). IQA based on NSS relies on the insight that high-quality natural images exhibit consistent statistical properties [16]. Likewise, distorted images induce measurable deviations from these properties. Using these insights, Mittal et al. [17] developed BRISQUE, an NSS metric in the spatial domain that works by calculating mean subtracted contrast normalized coefficients. These coefficients are subsequently fitted to a generalized Gaussian distribution model and an asymmetric generalized Gaussian distribution model. The resulting features are then fed to a support vector machine that predicts the score.

While BRISQUE was shown to be superior to contemporary NR-IQA metrics, one problem is that BRISQUE is an opinion-aware model that only recognizes distortion types that appear in the model's training data. This makes it and other opinion-aware models incapable of accurately scoring images with unknown distortion types. To address this issue, Mittal et al. [18] introduced NIQE, an NSS metric similar to BRISQUE that doesn't rely on human-labeled training data. Whereas BRISQUE is trained on features extracted from both natural and distorted images along with human judgements, NIQE is exclusively trained on the NSS features of natural images. It works by first extracting NSS-features from the test images and then fitting those features to a multivariate Gaussian (MVG) model. The score is computed as the distance from the fitted model to an MVG model of NSS-features extracted from a set of natural images [18].

Extending this approach, Zhang et al. [19] introduced ILNIQE, which enhances NIQE by introducing three additional NSS features: quality-aware gradient features, log-Gabor filter responses, and color distortions. ILNIQE also computes quality scores for localized segments of the image and averages those to get the score of the entire image.

2.2.2 Deep learning approaches

One consistent limitation of handcrafted approaches, such as NSS, is finding representations that accurately model complex images and various distortion types [20]. The advantage of a deep learning approach is that the network automatically learns optimized feature representations without human intervention. This makes deep learning approaches more robust, and as a result, they typically outperform NSS models [21].

Kang et al. [22] introduced CNNIQA, a convolutional neural network (CNN) for the task of NR-IQA. The network consists of one convolutional layer, one max-min pooling layer, two fully connected layers, and finally, a linear regression layer used for predicting quality scores. CNNIQA outperforms several NSS IQA models, including BRISQUE, on the LIVE dataset [22].

While CNN approaches, such as CNNIQA, have demonstrated strong performance on NR-IQA, one problem is that the need for labeled data increases as the networks become more complex. This is problematic as acquiring large labeled NR-IQA datasets is both difficult and expensive. Liu et al. [23] were able to circumvent this problem by generating image datasets where the quality ranking between each pair of images is known. The authors then use this data to train a Siamese network to learn to rank two images. This knowledge is then transferred to a traditional CNN that does the actual quality prediction.

Bosse et al. [24] developed WaDIQaM, a deep CNN capable of performing both full-reference and no-reference IQA. The network, inspired by VGG16 [25], consists of ten convolutional layers and five max pooling layers for feature extraction. It also uses two fully connected layers for regression. The authors were able to overcome the problem of limited IQA data by training the network on randomly sampled image patches.

One problem with existing CNNs in IQA is that training images have to be of the same shape in batch training. This typically results in images being resized to fit the specific network, which causes degradation. Ke et al. [26] developed MUSIQ, a multi-scale image quality transformer that overcomes

this problem. The network can process full-size images with different aspect ratios and perform multi-scale feature extraction. This is done to capture image quality across different regions of an image.

Other deep learning methods utilize CLIP [27], a neural network that learns visual concepts from natural language descriptions. CLIP was trained using contrastive representation learning on 400 million pairs of images and text snippets. In representation learning, the goal is to learn an embedding space where similar image-text pairs are close while dissimilar ones are far apart. Wang et al. [28] created CLIP-IQA, an extension of CLIP for NR-IQA using antonym prompt pairing. CLIP-IQA scores an image using the following process. First, the cosine similarities of the image to the prompts "good photo" and "bad photo" are computed. The cosine similarities are then used to compute a softmax function, which becomes the final score. Using CLIP for NR-IQA, CLIP-IQA achieves competitive scores on image benchmarks despite not being trained on human-annotated data [28].

A recent research trend is applying large multimodal (LMM) models for NR-IQA. This is done as LMMs are capable of capturing complex relationships between images and text information. Wu et al. [29] developed a methodology for teaching LMMs to score images based on quality and used that methodology on an extension of the mPLUG-Owl-2 LMM. The methodology is based on the insight that LMMs, like humans, are better at rating images using qualitative adjectives instead of using absolute scores. This is because LMMs are trained to understand and generate human-like text. Thus, the researchers converted the human mean opinion scores into discrete levels: 'bad', 'poor', 'fair', 'good', and 'excellent'. The researchers employed this method and extended the mPLUG-Owl-2 LMM to create Q-Align, an LMM that can perform both IQA and image aesthetic assessment. Q-Align performs comparably to state-of-the-art models while using less training data.

2.3 No-reference point cloud quality assessment

NR-PCQA methods are typically classified into three types: model-based, projection-based, and hybrid approaches combining both.

2.3.1 Model-based NR-PCQA

Model-based NR-PCQA techniques extract color and geometry information directly from 3D point clouds without projecting them onto a 2D plane [30]. Liu et al. [31] introduced ResSCNN, a model-based convolutional neural network for NR-PCQA. ResSCNN is composed of three distinct modules, each serving a specific purpose: hierarchical feature extraction, pooling and concatenation, and score prediction. The hierarchical feature extraction module consists of four blocks with three sparse convolutional layers each. A sparse convolutional layer is a convolutional layer that only works on the non-zero elements of the input data [32]. The point of using sparse convolutional layers as opposed to standard convolutional layers is that point clouds generally contain many empty voxels that do not contain any data. After performing hierarchical feature extraction, the network pools the output into 64×1 feature vectors, which are then concatenated to form a 256×1 vector. Lastly, the scoring module consists of two fully connected layers, PC-1 and PC-2. PC-1 has 256 input nodes and 32 output nodes. PC-2 has 32 input nodes and one final output node.

In addition to using deep learning, some NR-PCQA techniques take inspiration from IQA and employ NSS to point clouds. Zhang et al. [33] introduced 3D-NSS, which uses NSS and entropy on geometry and color features of point clouds. 3D-NSS projects point clouds into several different geometry feature domains. For each point in a point cloud, the 10 nearest neighbors are calculated using the k-nearest neighbors (k-NN) algorithm. These neighbors are used to calculate a 3×3 covariance matrix. Using the covariance matrix, three eigenvectors are computed for each point in the point cloud. These eigenvectors are then used to compute the curvature, anisotropy (a measure of how geometrical properties vary depending on direction), linearity, planarity, and sphericity. Furthermore, the point cloud's color information is converted from the RGB to the LAB color space, as LAB better aligns with human color perception. The geometry and color features are then used to compute a set of statistical parameters, including mean, standard deviation, generalized Gaussian distribution (GGD), asymmetric GGD (AGGD), and the shape-rate Gamma distribution. Finally, these properties are used to train a support vector regression model.

2.3.2 Projection-based NR-PCQA

Projection-based methods for NR-PCQA differ from model-based ones in that they start off by projecting the points in the point cloud onto a 2D

plane. Most projection-based NR-PCQA methods employ V-PCC [34], an encoding technique that decomposes a point cloud into a set of patches that are orthogonally projected onto a 2D grid. The patches are then merged into two video sequences: one containing geometry information and another containing texture information.

Liu et al. [35] introduced PQA-Net, one of the first no-reference metrics for point cloud quality assessment. The proposed methodology consists of three modules: projection and feature extraction, distortion type classification, and quality prediction. The projection is made by projecting the point cloud onto six 2D image planes. Feature extraction is done via a CNN that consists of four convolutional blocks that convert the image to a 64×6 feature vector. The resulting feature vector is used to classify the distortion type, producing a distortion probability vector. It also serves as input to the quality prediction module, which consists of two fully connected layers. The quality prediction module's output is then multiplied with the distortion probability vector to compute the point cloud's score.

Chai and Shao [36] introduced MS-PCQE, a projection-based NR-PCQA technique that uses projections of two different focal lengths. For each focal length, the model projects the point cloud onto six planes (front, back, left, right, up, and down). These projected images are then processed by a residual neural network to generate feature maps, which are subsequently fed into a ConvGRU module. The output is then passed through a mask-aware transformer block with multi-head attention. This process is repeated four times before the final output is processed by a multi-layer perceptron to predict the point cloud's score.

2.3.3 Hybrid methods

Hybrid NR-PCQA methods combine both projection-based and model-based techniques. The purpose of this multi-modal approach is to capture distortions that a single modality cannot. For instance, Zhang et al. [37] note that structural damage and geometry downsampling are more easily detected in the point cloud modality, while color quantization and noise are better detected in the image modality.

Zhang et al. [37] introduced MM-PCQA, the first multi-modal approach to NR-PCQA. MM-PCQA works by splitting the point cloud into sub-models and then using a point cloud encoder to create quality-aware embeddings. The projected images are rendered directly and encoded using an image encoder. The point cloud and image embeddings are then fused using a symmetric

cross-modality attention block. The results of this block are then fed to a quality regression module. The authors emphasize that this multi-modal technique outperformed all state-of-the-art NR-PCQA techniques it was tested against. In a subsequent article, the authors introduced MM-PCQA+, an enhanced version of MM-PCQA that integrated improved encoding techniques and a sampling strategy that increases the proportion of visible points in projections.

2.4 Related work

Several studies have surveyed and compared the performance of NR-IQA methods. Manap and Shao [38] provides a systematic analysis of NR-IQA approaches based on natural scene statistics (NSS) and machine learning, evaluating their performance on the LIVE IQA database. Among the NSS-based methods surveyed, BRISQUE had the best performance on non-distortion-specific datasets and on three out of five distortion-specific datasets. Among the learning-based methods, CBIQ, GRNN, SFLNIA, and CORNIA each achieved the best performance on one or two distinct datasets.

Yang et al. [20] advanced this research by evaluating the performance of deep neural network-based methods for NR-IQA. The authors benchmarked deep neural network-based NR-IQA methods on three synthetic image quality databases — LIVE, TID2013, and CSIQ — as well as on the authentic LIVE Challenge database. The authors found that DNN-based approaches generally outperformed conventional methods for NR-IQA. They also found that BLNDER and HIQA had the best performance overall on the selected datasets.

Yang et al. [16] present a thorough survey of NR-IQA techniques and the benchmark datasets used in their assessment. The survey distinguishes itself from earlier surveys, in part, by covering recent advancements in the field, including the emergence of advanced deep neural networks (DNNs) and large multimodal models (LMMs). However, the survey does not provide performance results of the NR-IQA techniques it discusses on benchmark datasets.

Evaluations of NR-PCQA techniques have also been performed. Zhou et al. [9] surveyed the area of PCQA techniques, looking at model-based and projection-based methods and their performance on four PCQA datasets. While the review is not specific to NR-PCQA, it benchmarked several NR-PCQA techniques on two NR-PCQA datasets. The results show that MM-PCQA has superior performance, outperforming all other NR-PCQA metrics.

Porcu et al. [30] surveyed the field of NR-PCQA directly. The authors review the state-of-the-art in NR-PCQA concerning both techniques and datasets. The authors don't directly compare the performance of NR-PCQA methods but instead use the results of [36], a paper that compares a novel NR-PCQA metric to five existing metrics. The results of the comparison indicate that MM-PCQA and MS-PCQE achieved the highest performance across the five point cloud datasets used in the study.

2.5 Summary

Chapter 3

Method

3.1 Literature search

A comprehensive literature search is conducted both to gain an overview of the key topics of the thesis and to identify the most suitable methods within each area. The key topics include NR-IQA, NR-PCQA, image distortion, and point cloud distortion. The primary tool used for finding literature is Google Scholar due to its broad coverage of scientific papers and journals. In the literature search, both survey articles and original research papers are considered. Survey articles are primarily used to gain a broad understanding of the fields of NR-IQA and NR-PCQA, while original research papers provide detailed information on specific methods. Both types of articles are also used to compare the performance of NR-IQA and NR-PCQA techniques. During the literature search, factors such as citation count, publication venue, year, and language were considered to ensure the quality of the chosen articles.

3.2 Considered methods

Before selecting the method described in Chapter 3.3, two alternative methods were considered. The first involved directly evaluating NR-IQA and NR-PCQA methods on weather-distorted images and point clouds. The problem with this approach, however, is that no dataset exists with subjective scores for weather-distorted images or point clouds. Pursuing this option would therefore require conducting a large-scale survey with human participants to obtain subjective scores. Given the substantial effort required, this method was ultimately rejected.

Another considered approach was to use object detection performance as



(a) OD confidence on the two leftmost bounding boxes is 0.78 and 0.77



(b) OD confidence on the two leftmost bounding boxes is 0.85 and 0.87

Figure 3.1: Example where OD confidence does not correlate with differences in image quality.

a proxy for setting ground truth values for image quality. This idea relies on the intuition that since images in AVs are ultimately used in object detection algorithms, it is logical to assess image quality based on the confidence of the object detection algorithm. The problem with this approach is that object detection performance is not always correlated with image quality. For instance, in figure 3.1, the two images are captured within a very short interval and appear to have precisely the same quality. However, because the vehicles on the right side of the scene are positioned closer to the camera in the second image, the object detection algorithm assigns higher confidence scores compared to the first image. Specifically, it assigns 0.85 and 0.87 in the second image and 0.78 and 0.77 in the first. These confidence scores represent the model's estimated probabilities that the detected objects are cars. If the ground truth was set based on these scores, the second image would be assigned a higher quality score than the first, despite the two images exhibiting identical quality. Thus, object detection confidence was deemed unreliable in setting ground truth scores, and this approach was rejected.

3.3 Chosen approach

NR-QA metrics are typically evaluated by comparing their scores with subjective scores provided by human subjects. However, this approach is not feasible for this thesis due to the lack of AV image and point cloud datasets scored by human subjects. Notably, no major survey on NR-IQA or NR-PCQA mentions the existence of a dataset tailored explicitly to the AV domain [16, 20, 30, 9].

The lack of suitable data motivates a different approach to the problem of NR-QA evaluation. This approach, developed in collaboration with my supervisor, Ian Marsh, relies on the insight that synthetic noise distortions at various levels can generate datasets with known quality rankings. To accomplish this, a reference image or point cloud is used and then distorted several times with increasing levels of noise. This produces a set of images or point clouds where the absolute quality of the samples is unknown, but the relative quality rankings are known.

With this information, the method of evaluating an NR-QA method is as follows. First, an image or point cloud set with known quality rankings is generated using the method described above. The NR-QA method is then applied to each sample in the set, generating a list of scores, which are subsequently ranked. Finally, the performance of the NR-QA is evaluated by comparing its rankings with the ground-truth rankings using Spearman

rank correlation coefficient (SRCC) and Kendall rank correlation coefficient (KRCC). The evaluation of NR-IQA and NR-PCQA using these metrics is described in greater detail in section 3.9.

The approach of applying synthetic distortions to generate image sets with known quality rankings has previously been used in the field of NR-IQA. Liu et al. [23] utilized it to generate RankIQA, an NR-IQA method trained to rank image pairs distorted by four types of distortion. Additionally, Agnolucci et al. [39] synthetically distorted images of varying levels and trained CLIP to rank the images according to distortion severity.

The approach in this thesis differs from that of Liu et al. and Agnolucci et al. in a few distinct ways. First, it utilizes larger image sets with known relative quality rankings rather than relying on sets with two or five images. Secondly, whereas Liu et al. and Agnolucci et al. use conventional image distortions such as GB, GN, JPEG, and JP2K, the synthetic distortions in this work will include weather-related distortions. Lastly, while previous work has used the approach to train an NR-IQA method, this thesis employs it to evaluate NR-IQA and NR-PCQA methods.

3.4 Dataset

The thesis utilizes two datasets for the evaluation: the Finnish Geospatial Institute (FGI) dataset [40] and the REHEARSE dataset [41]. Each dataset was chosen for different parts of the evaluation. The FGI dataset [40] includes data from two drives: one urban drive around Otaniemi, Espoo, and another covering both urban and rural scenery, extending from Espoo to Munkkivuori in Helsinki. 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 December 3, 2023, during winter conditions and include LiDAR, RGB images, and thermal camera data. The dataset is stored in ICE S3 storage and retrieved using the boto3 API. The Munkkivuori drive consists of 6915 PNG images, while the Otaniemi drive consists of 7851 PNG images. The REHEARSE dataset [41] comprises LiDAR, image, and RADAR data collected from an outdoor test track and an indoor tunnel using static sensors. The data is collected under clear, rainy, and foggy conditions, with the rain and fog artificially generated using sprinklers. Like the FGI dataset, the point clouds in the REHEARSE dataset are stored in ICE S3 Storage and were retrieved using the boto3 API. The point clouds are stored as binary files, each containing a sequence of points

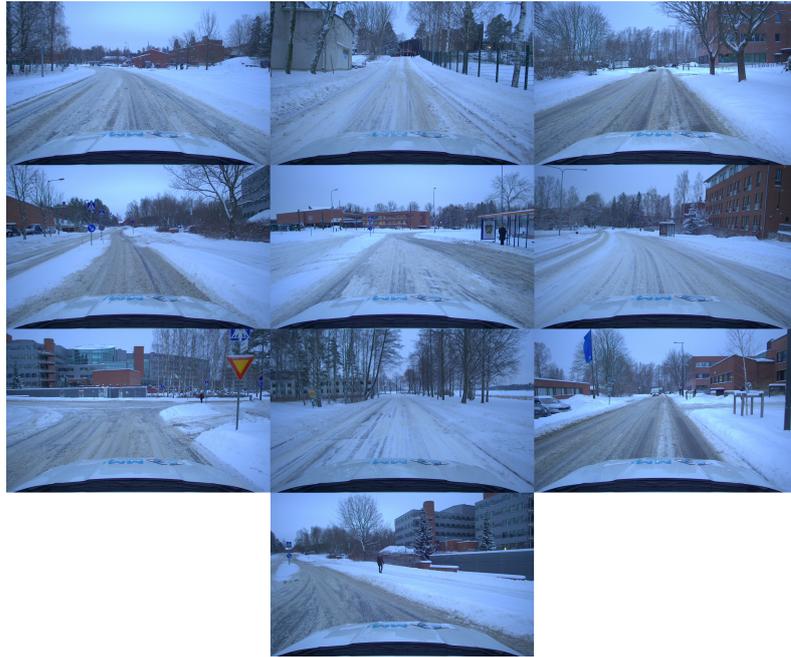


Figure 3.2: The ten reference images used from the Otaniemi drive.

of type (x, y, z, i) where (x, y, z) are the Cartesian coordinates and i is the reflectance value.

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 are randomly selected as reference images for the synthetic distortions. The selected images can be seen in Figures 3.2 and 3.3. For the point cloud data, the REHEARSE dataset was selected over the FGI dataset. The reason is that initial experiments applying NR-PCQA methods to the FGI point clouds yielded poor results. The REHEARSE dataset [41], recorded under more controlled conditions, was therefore used instead. Only point clouds from the outdoor test track were used. The indoor tunnel data were excluded because extensive preprocessing was required to make the data compatible with the NR-PCQA evaluation. 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. Of these, 27 were recorded during the day, and 13 were recorded during the night.



Figure 3.3: The ten reference images used from the Munkkivuori drive.

3.5 Preprocessing images

The Albumentations Python library [42] is used to synthetically apply weather distortions to images. Albumentations 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 handle a wide range of noise severities.

3.5.1 Synthetic 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 color. 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 HSV format and multiplying the

value component by the `brightness_coefficient` parameter.

For each image selected in section 3.4, 100 versions with synthetic rain are generated using the parameters outlined in Table 3.1. Parameters listed as constants remain the same for all versions of the image. For parameters listed as ranges $[a, b]$, the values are linearly spaced across the n versions. Thus, the first version has a value of a , the last version has a value of b , and the step size is given by $(|b - a| / (n - 1))$. 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. An example of three versions of the same image with varying amounts of rain distortion is shown in Figure 3.4.

3.5.2 Synthetic fog distortion

Fog effects are synthesized using the `add_fog` transform from Albumenations. `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 with the original image using the following formula:

$$S(i, j, c) = \alpha F(i, j, c) + (1 - \alpha)S(i, j, c)$$

where $S(i, j, c)$ refers to the pixel value at location (i, j) and color channel c in the original image, and $F(i, j, c)$ refers to the corresponding pixel in the fog circle. The blending factor α is given by the formula $\alpha = \text{alpha_coef} \times \text{fog_intensity}$.

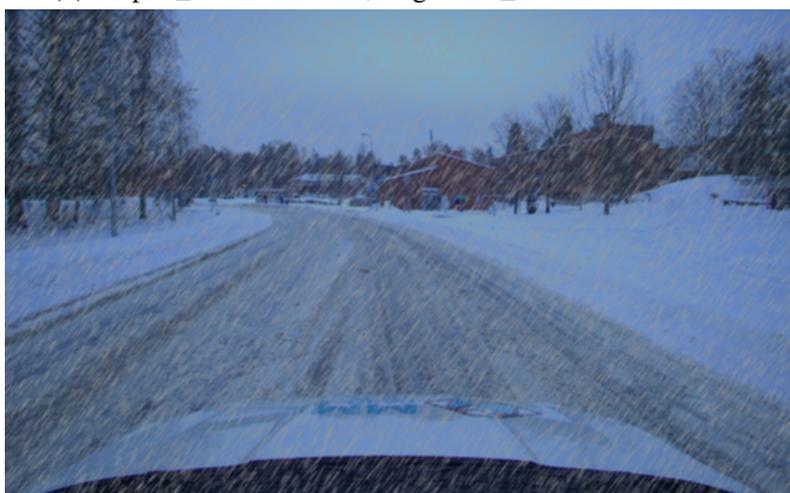
Table 3.2 shows the parameters used to add synthetic fog noise to the images. For each reference image across both locations, 100 versions with synthetic fog are synthesized using the parameters in Table 3.2. The progressive increase is linearly spaced, following the same approach as for the rain-distorted images described earlier. Examples of an image distorted three times with increasing fog severity can be seen in Figure 3.5



(a) droplet_share = 0.003, brightness_coefficient = 0.924



(b) droplet_share = 0.005, brightness_coefficient = 0.848



(c) droplet_share = 0.007, brightness_coefficient = 0.788

Figure 3.4: The same image from Otaniemi with different amounts of rain-related noise.

Parameter name	Explanation	Value
droplet_share	Proportion of pixels where a droplet is placed. Since droplets are larger than a single pixel, the total share of affected pixels is greater.	[0.0, 0.01]
rain_drops	Array of pixel coordinates indicating where droplets are placed. The droplet placement is randomized (seeded for reproducibility), and the number of raindrops is controlled by droplet_share.	
slant	The angle in degrees at which the synthetic raindrops are applied.	15
drop_length	Length of each raindrop in pixels.	20
drop_width	Width of each raindrop in pixels.	1
blur_value	The value N used in an NxN box filter applied to the image. This is done since rainy images are typically blurrier [42].	7
brightness_coefficient	Coefficient applied to each pixel value to reduce brightness, as rainy images are typically darker.	[1.0, 0.7]

Table 3.1: Parameters for the synthetic rain distortion

3.6 Preprocessing point clouds

The work of Hahner et al. [43], in this thesis referred to as `LiDAR_fog_sim`, is used as the tool for synthetically distorting point clouds. `LiDAR_fog_sim` is a tool for augmenting LiDAR point clouds with simulated fog. The choice to use `LiDAR_fog_sim` for point cloud distortion is motivated by two factors. Firstly, it is a state-of-the-art tool that simulates fog in a physically accurate manner [43]. Secondly, the authors provide the code necessary to distort any point cloud, meaning that the tool requires minimal configuration.

`Lidar_fog_sim` simulates fog by extending an optical LiDAR model proposed by Rasshofer et al. [44]. In the extended model, the authors introduce two significant concepts: hard target and soft target. The hard target refers to the surface that reflects the LiDAR pulse under clear weather, while the soft target refers to the fog particles. The algorithm Rasshofer et al. use to distort fog particles is outlined in Figure 3.6. For each point p , the algorithm inputs the point's intensity i , differential reflectivity β_0 , and the half-power pulse width τ_H . The differential reflectivity models how much of the LiDAR pulse is reflected back by the target object. The half-power pulse width is a parameter that defines the duration over which a LiDAR pulse maintains significant power around its peak.

The algorithm starts by calculating the distance R_0 based on p . Lines 3-5 compute the response intensity from the hard target i_{hard} , factoring in the attenuation coefficient α . Lines 6-8 numerically compute integrals used to calculate the intensity from the soft target. This is done as the expression doesn't possess a closed form and is done for all distances $R \in \{0.1, \dots, R_0\}$.

Next, lines 9-11 compute the intensity from the soft target by taking the largest value from the precomputed integrals i_{tmp} and multiplying that with $C_A P_0$ and β . C_A is a system constant, P_0 the pulse's peak power, and β is the backscattering coefficient. On line 12, the algorithm checks if $i_{\text{soft}} > i_{\text{hard}}$. If this is the case, then the point is shifted using a scaling factor, and its intensity is changed to i_{soft} . Otherwise, the point's coordinates remain intact while the point's intensity value is changed to i_{hard} . Lastly, a linear scaling of the intensity values is performed.



(a) $\alpha = 0.176$



(b) $\alpha = 0.354$



(c) $\alpha = 0.531$

Figure 3.5: The same image from Munkkivuori with different increasing amounts of foggy noise.

Algorithm 1 LiDAR fog simulation

```

1: procedure FOGGIFY( $\mathbf{p}, i, \alpha, \beta, \beta_0, \tau_H$ )
2:    $R_0 \leftarrow \|\mathbf{p}\|$ 
3:    $x, y, z \leftarrow \mathbf{p}$  ▷  $i = P_{R, \text{clear}}$ 
4:    $C_A P_0 \leftarrow i \frac{R_0^2}{\beta_0}$  ▷ follows from Eq. (12)
5:    $i_{\text{hard}} \leftarrow i \times \exp(-2\alpha R_0)$  ▷ see Eq. (17)
6:   for  $R$  in  $(0, 0.1, \dots, R_0)$  do ▷ 10cm accuracy
7:      $I_R \leftarrow \text{SIMPSON}(I(R, R_0, \alpha, \tau_H))$  ▷ see Eq. (18)
8:   end for
9:    $i_{\text{tmp}} \leftarrow \max(I_R)$ 
10:   $R_{\text{tmp}} \leftarrow \arg \max(I_R)$ 
11:   $i_{\text{soft}} \leftarrow C_A P_0 \beta \times i_{\text{tmp}}$  ▷ see again Eq. (18)
12:  if  $i_{\text{soft}} > i_{\text{hard}}$  then
13:     $s \leftarrow \frac{R_{\text{tmp}}}{R_0}$  ▷ scaling factor  $s$ 
14:     $p \leftarrow \text{RANDOM.UNIFORM.FLOAT}(-1, 1)$ 
15:     $n \leftarrow 2^p$  ▷ noise factor  $n \in (\frac{1}{2}, 2)$ 
16:     $x \leftarrow s \times n \times x$ 
17:     $y \leftarrow s \times n \times y$ 
18:     $z \leftarrow s \times n \times z$ 
19:     $i \leftarrow i_{\text{soft}}$ 
20:  else ▷ keep original location
21:     $i \leftarrow i_{\text{hard}}$  ▷ only modify intensity
22:  end if
23:  return  $x, y, z, i$ 
24: end procedure

```

Figure 3.6: The algorithm used to apply fog distortion to a particle p with intensity i [43]

Parameter	Explanation	Value(s)
fog_particle_share	Proportion of pixels where a fog particle is placed. As particles are larger than a single pixel, the total proportion of pixels is greater.	0.5
α	Factor determining the visual strength of the fog applied to the image, between 0 and 1.	[0.0, 0.7]
fog_particle_size	Size of each fog particle in pixels.	25
fog_particle_positions	Coordinates where fog particles are placed. Randomized and seeded for reproducibility.	5
fog_particle_radiuses	Radius of fog particles in pixels.	25

Table 3.2: Parameter values used for synthetic fog distortion.

3.6.1 Point cloud distortion implementation

Adding synthetic fog to point clouds first involves precomputing the integrals that are described on line 7 in figure 3.6. The integrals are computed using Simpson’s 1/3 rule, a step size of 0.1 m, $\tau_h = 20$ ns, and a maximum distance $R_0 = 200$ m. Each of the precomputed integral values is the same as in the original paper describing the method [43].

The precomputed integrals are subsequently used to add synthetic fog distortions to the point clouds. A set of 40 point clouds is selected from the REHEARSE dataset, and each one is distorted with the parameters described in table 3.3. For each point in each point cloud, the algorithm shown in Figure 3.6 is applied. All parameters, except α , are kept consistent with the original paper to ensure alignment with the proposed fog simulation model. α is progressively increased to simulate denser fog in the same linear manner used in the synthetic image distortions. A visualization of a point cloud distorted using three different α values can be seen in Figure 3.7.

Both MM-PCQA and MS-PCQE input point clouds of the form $P = \{g_m, c_m\}_{m=1}^N$ where $g_m \in \mathbb{R}^{1 \times 3}$ are the geometric coordinates and $c_m \in \mathbb{R}^{1 \times 3}$ consists of the RGB color information. As the point clouds in the dataset are of the form (x, y, z, i) and do not contain color information, the color is inferred. This is done by taking the reflectance value i , an integer between 0 and 255, and setting $c_m = (i, i, i)$.

3.7 Selection of NR-IQA techniques

The number of NR-IQA techniques is too extensive to cover within the scope of this thesis. Therefore, it is necessary to select a representative set of NR-IQA methods. The decision is made to include one method from each of the following paradigms in NR-IQA: NSS, CNN, attention, CLIP, and LMM. This decision is driven by the aim of understanding how each paradigm performs in the context of NR-IQA on synthetically weather-distorted images.

IL-NIQE [19] is selected as the NSS-based approach because of its proven effectiveness on NR-IQA benchmarks and because it builds upon and enhances earlier NSS-based NR-IQA techniques. DBCNN [45] is selected as the CNN-based approach because, in a comparative study of deep learning techniques, it achieved the second-highest performance among CNN-based methods [20]. The top-performing method, BLINDER [46], is rejected as no existing implementation of the algorithm is found. Topiq [47] is selected as the attention-based method. It is chosen because it has demonstrated better

Parameter name	Explanation	Value
α	Attenuation coefficient which reduces the signal power.	$[0.0, 0.25]$
β	Coefficient for backscattering, which quantifies the portion of the LiDAR signal that is reflected directly back toward the sensor by fog particles. The value of β is the same as in the original paper.	$0.046\alpha / \ln(20)$
β_0	Differential reflectivity of the target.	$10^{-6} / \pi$
τ_h	Half-power pulse width	20 ns

Table 3.3: Parameters for the synthetic fog distortion of point clouds.

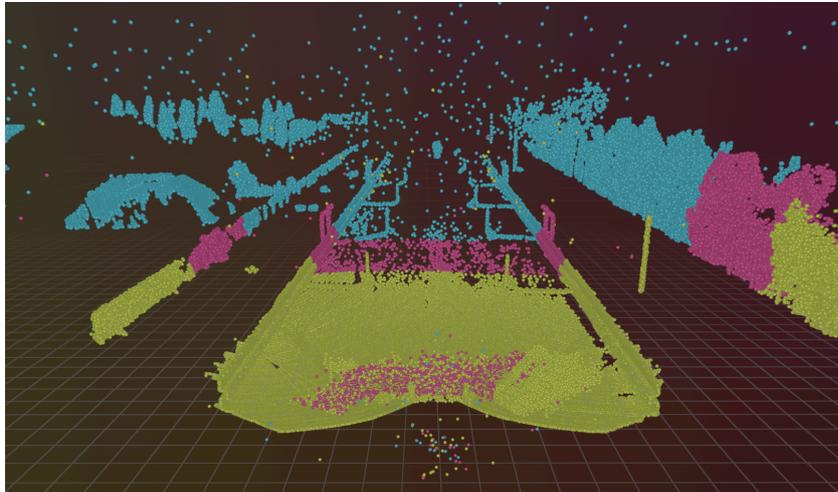


Figure 3.7: Visualization of three point clouds subjected to increasing distortion levels. The blue point cloud corresponds to $\alpha = 0.03$, the pink to $\alpha = 0.06$, and the yellow to $\alpha = 0.09$.

performance on NR-IQA benchmarks compared to similar methods, such as TReS [48] and Musiq [26]. QualiCLIP [39] is chosen as the CLIP-based approach as it outperforms other CLIP-based methods like CLIP-IQA [28] and LIQE [49]. Finally, Q-Align is chosen as the LMM-based model as it has achieved excellent performance on NR-IQA benchmarks [29]. A summary of each of the chosen models can be found in Table 3.4. In this table, the *Dataset* column refers to the dataset used during training. This information was ascertained for all methods but IL-NIQE and DB-CNN. The *Lower Better* column refers to whether a method interprets lower scores as corresponding to higher image quality

All of the chosen NR-IQA techniques are sourced from the pyiqa library [50], a Python toolbox for image quality assessment that implements both NR and FR-IQA techniques. pyiqa is used because it implements all of the selected NR-IQA methods. It also simplifies the evaluation process by eliminating the need to locate and compile individual NR-IQA implementations.

3.8 Selection of NR-PCQA techniques

This thesis evaluates two NR-PCQA methods, and the reason for limiting the scope to two methods is that each method requires significant configuration and setup. Adding additional methods to the evaluation would be unfeasible given the time constraints.

The selected NR-PCQA metrics are MS-PCQE and MM-PCQA. MS-PCQE is selected because it has been shown in previous research to have the best performance on several datasets. Porcu et al. [30] found that MS-PCQE achieved the best performance among all tested NR-PCQA methods on five NR-PCQA datasets. The authors of MS-PCQE developed multiple models, each trained on a different dataset. In this thesis, the model trained on the LS-PCQA dataset [31] is used. LS-PCQA is a large-scale NR-PCQA dataset consisting of 104 reference point clouds, 31 distortion types, and 22568 distorted point clouds. This model is selected because LS-PCQA is by far the largest and most diverse NR-PCQA dataset available [30].

MM-PCQA is chosen since it combines the strengths of projection-based and model-based methods. Additionally, Porcu et al. [30] found that MM-PCQA had similar performance to MS-PCQE and, for some distortion types, even outperformed it. The specific MM-PCQA model used in this thesis is the one trained on the WPC dataset [51, 52]. The WPC dataset comprises 20 high-quality reference point clouds, along with 740 distorted versions. The point clouds were distorted using five types of distortions. The model trained on the

WPC dataset is selected because it is the only predefined model provided by the authors. While it is possible to train the model on a different dataset, this falls outside the scope of the thesis.

3.9 Evaluation methods

NR-QA methods are typically assessed by comparing their predicted scores, also called objective scores, with human evaluations of the same images or point clouds. Human evaluations, also known as subjective scores, are gathered by asking humans to score images or point clouds, typically on a five-point scale [53, 30].

Using the subjective scores, NR-QA methods are typically assessed using the following four metrics: Pearson linear correlation coefficient (PLCC), Spearman’s rank correlation coefficient (SRCC), Kendall rank correlation coefficient (KRCC), and Root mean square error (RMSE) [16, 30, 14]. 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. Since the point clouds and images used in this thesis do not have subjective scores, PLCC and RMSE can not be used as evaluation metrics. However, SRCC and KRCC remain applicable, as the distortion method described in sections 3.5 and 3.6 generates ground-truth rankings of the images and point clouds.

SRCC measures the prediction monotonicity of the NR-QA by comparing the predicted scores with the ground-truth quality ranks. The formula is expressed as:

$$\text{SRCC} = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)} \quad (3.1)$$

where d_i is the difference between the objective and ground-truth quality rankings for image or point cloud i .

To illustrate the use of SRCC, consider the following example. Suppose a point cloud has been distorted four times, with each version having an increasing level of noise. These versions can be ranked in order of increasing distortion as $O = [1, 2, 3, 4]$. An NR-PCQA metric then scores each of the point cloud versions, and the scores are ranked from best to worst. This results in the ranking $P = [1, 3, 4, 2]$. Calculating the SRCC is then done using the

following calculation:

$$\text{SRCC} = 1 - \frac{6((O_1 - P_1)^2 + (O_2 - P_2)^2 + \dots + (O_n - P_n)^2)}{N(N^2 - 1)} \quad (3.2)$$

$$= 1 - \frac{6((1 - 1)^2 + (2 - 3)^2 + (3 - 4)^2 + (4 - 2)^2)}{4(4^2 - 1)} = 0.4 \quad (3.3)$$

KRCC, like SRCC, uses ranks as opposed to absolute scores in its evaluation [54]. It is computed by comparing all possible pairs of observations and classifying them as concordant, discordant, or tied. For two variables $X = \{X_1, \dots, X_n\}$ and $Y = \{Y_1, \dots, Y_n\}$, we consider all pairs $\{(i, j) : i < j \leq n\}$. A pair is concordant, discordant, or tied depending on the following conditions:

$$\begin{cases} \text{Concordant} & \text{if } (X_i - X_j)(Y_i - Y_j) > 0 \\ \text{Discordant} & \text{if } (X_i - X_j)(Y_i - Y_j) < 0 \\ \text{Tie} & \text{if } (X_i = X_j) \oplus (Y_i = Y_j) \end{cases}$$

With these definitions, the formula of KRCC (tau-b) is as follows:

$$\tau = \frac{C - D}{\sqrt{(C + D + T_X) * (C + D + T_Y)}} \quad (3.4)$$

where C is the number of concordant pairs, D is the number of discordant pairs, T_X is the number of ties in X , and T_Y is the number of ties in Y .

Using the same example as before, with ground-truth ranking of [1, 2, 3, 4] and predicted rankings of [1, 3, 4, 2]. The concordant pairs (by index) are [(0, 1), (0, 2), (0, 3), (1, 2)], the discordant ones are [(1, 3), (2, 3)], and there are no ties. This means that the KRCC is calculated as:

$$\text{KRCC} = \frac{4 - 2}{\sqrt{(4 + 2 + 0) * (4 + 2 + 0)}} = \frac{1}{3}$$

3.10 Scoring images and point clouds

Each set of distorted images is evaluated by the five selected NR-IQA methods, which assign a score to every image within the set. These scores are subsequently ranked and compared to the true quality ranking of each image set by calculating the SRCC and KRCC, as outlined in Section 3.9. This entire process is repeated for each of the distorted point cloud sets and NR-PCQA methods.

3.11 Analysis

A two-tailed permutation test is conducted for each NR-QA method to evaluate whether its predicted rankings are correlated with the ground-truth rankings. The hypotheses are defined as:

$$\begin{aligned} H_0 &: \text{SRCC}_{\text{mean}} = 0 \\ H_A &: \text{SRCC}_{\text{mean}} \neq 0 \end{aligned}$$

where $\text{SRCC}_{\text{mean}}$ refers to the mean SRCC value across all image or point cloud sets, depending on whether the method is an NR-IQA or NR-PCQA metric. The null hypothesis H_0 states that the mean SRCC is 0, implying that there is no correlation between the NR-QA's scores and the ground-truth ranking. The alternative hypothesis H_A states that there is a statistically significant correlation.

To estimate the p-value under the assumptions of the null hypothesis, the rankings of the NR-QA method are randomly shuffled within each image or point cloud set. For each permutation, the SRCC is computed using the shuffled rankings and the ground-truth rankings. The resulting SRCCs are then averaged across all sets, resulting in a permuted mean SRCC. This process is repeated 5000 times, and the empirical p-value \hat{p} is computed as:

$$\hat{p} = \frac{1}{5000} \sum_{i=1}^{5000} I(|\text{SRCC}_{\text{mean}}^{(i)}| \geq |\text{SRCC}_{\text{observed}}|)$$

where $\text{SRCC}_{\text{mean}}^{(i)}$ denotes the mean SRCC from the i -th permutation and $\text{SRCC}_{\text{observed}}$ is the mean SRCC obtained from the experimental data. $I()$ is the indicator function that is equal to 1 if $|\text{SRCC}_{\text{mean}}^{(i)}| \geq |\text{SRCC}_{\text{observed}}|$ and 0 otherwise. 5000 is chosen as the number of permutations following the recommendation of Marozzi [55]. As multiple hypotheses are tested, a Holm-Bonferroni correction is performed to minimize the risk of Type 1 errors.

A Friedman test is applied to identify significant differences between two or more of the selected NR-IQA methods. It is a non-parametric statistical test suitable for comparing multiple methods across multiple datasets [56]. In this evaluation, each dataset corresponds to an image set. The null hypothesis states that the methods are equivalent, while the alternative hypothesis states that at least one method differs significantly from at least one other method. The chi-square distribution approximates the Friedman test statistic when the number of methods or image sets is sufficiently large, typically greater than 5 [57]. In the evaluation, five methods and 40 image sets are used, indicating

that the chi-square approximation is reasonable.

If the null hypothesis of the Friedman test is rejected, pairwise Wilcoxon signed-rank tests are conducted to identify which NR-IQA methods differ significantly. Additionally, for the NR-PCQA methods, a Wilcoxon signed-rank test is performed to assess whether MM-PCQA and MS-PCQE differ significantly.

A Wilcoxon signed-rank test including two methods, A and B , is performed by first computing the differences $D = A - B$. The absolute values $|D|$ are then ranked, and zero differences are excluded. Then, the original signs of the differences are reintroduced and summed up into two separate sums. Positive differences are summed up into the sum $\sum R^+$, and negative differences are summed up into the sum $\sum R^-$. A low p-value suggests a statistically significant difference between the methods A and B . If $\sum R^+ > \sum R^-$ and p is sufficiently small, it suggests that A outperforms B [58]. By contrast, if $\sum R^- > \sum R^+$ and p is sufficiently small, it suggests that B outperforms A . Since multiple hypotheses are tested in the pairwise Wilcoxon signed-rank tests for the NR-IQA methods, a Holm-Bonferroni correction is applied to reduce the risk of Type I error.

Method	Approach	Dataset	Lower better	Description
IL-NIQE	NSS	–	True	NSS approach using a multivariate Gaussian model of image patches, each assessed with a Bhattacharyya-like distance. Final quality score is obtained by average pooling [19].
DB-CNN	CNN	–	False	Deep bilinear model with two CNNs for authentic distortions. Their outputs are pooled for final quality prediction [45].
TOPIQ	Attention-based	KonIQ-10k	False	Top-down network leveraging high-level semantic features to guide attention to perceptually relevant image regions. Uses coarse-to-fine and cross-scale attention [47].
QualiCLIP	CLIP	KonIQ-10k	False	Self-supervised method using CLIP to rank synthetically degraded images, removing the need for subjective scores [39].
Q-Align	LMM	KonIQ-10k, SPAQ, KADID-10k	False	LMM-based model trained with text-defined quality levels rather than numerical scores for better alignment with human ratings [29].

Table 3.4: NR-IQA methods included in the study

Method	Approach	Dataset	Lower better	Description
MM-PCQA	Multimodal	WPC	False	Multimodal approach that utilizes 2D projections and 3D submodels, fuses them using cross-modal attention, and regresses the result to a quality score.
MS-PCQE	Projection	WPC	False	Projection-based approach that merges projections of two focal lengths into a ConvGRU module, followed by a dual-branch transformer, and an MLP for score prediction. [45].

Table 3.5: Table of NR-PCQA methods included in the study

Chapter 4

Results and Analysis

Method	SRCC (SD)	KRCC (SD)
TOPIQ	-0.496** (0.685)	-0.443** (0.622)
DB-CNN	0.131** (0.872)	0.160** (0.805)
QualiCLIP	-0.062** (0.756)	-0.070** (0.696)
Q-Align	0.998** (0.002)	0.979** (0.015)
IL-NIQE	0.946** (0.094)	0.868** (0.156)

* $p < 0.05$ ** $p < 0.001$

Table 4.1: Mean SRCC and KRCC values of the NR-IQA techniques across both distortion types and locations, with standard deviations shown in parentheses. The p-values are obtained from permutation tests evaluating whether the mean SRCC and KRCC differ significantly from 0

4.1 NR-IQA results

With 20 base images selected and two distortion types employed, a total of 40 image sets are generated, resulting in 40 SRCC and KRCC values per NR-IQA technique. Table 4.1 presents the mean SRCC and KRCC values for the NR-IQA techniques under evaluation. Notably, Q-Align achieves near-perfect agreement with ground truth rankings, with mean SRCC and KRCC values of 0.998 and 0.970, respectively. In comparison, IL-NIQE also performs well, with corresponding values of 0.946 and 0.868.

In contrast, the performance of TOPIQ, DB-CNN, and QualiCLIP is considerably lower. QualiCLIP has a mean SRCC of -0.062 and a standard deviation of 0.756, suggesting that the SRCC values are widely dispersed

in both the positive and negative directions. DB-CNN has a higher mean SRCC value of 0.131, but like QualiCLIP, its values vary widely, with a standard deviation of 0.872. Interestingly, TOPIQ has a mean SRCC value of -0.496 and a mean KRCC value of -0.443, indicating a strong negative correlation between its scores and the ground truth scores. In other words, it performs worse than random, consistently assigning higher scores to lower-quality images than to higher-quality ones. However, TOPIQ also shows high variability, with standard deviations of 0.685 for the mean SRCC and 0.622 for the mean KRCC.

The exact SRCC values of each NR-IQA method on each image set can be seen in Figure 4.1. The points in the figure are grouped based on location (Otaniemi vs Munkkivuori) and distortion (rainy vs foggy). Q-Align has consistently high SRCC scores across the 40 image sets, regardless of distortion type or location. Notably, it is not possible to discern any subset of image sets where Q-Align performs worse. IL-NIQE also achieves consistently high SRCC scores but performs slightly better on foggy image sets than rainy ones. On the foggy image sets, IL-NIQE achieves an astonishing mean SRCC of 0.999 with a standard deviation of 0.002. However, on the rainy image sets, the mean SRCC is lower at 0.892 with a standard deviation of 0.110. Thus, IL-NIQE seems to struggle more with rain degradation than with fog degradation. Interestingly, QualiCLIP achieves excellent performance on the fog-distorted images from Otaniemi, achieving a mean SRCC of 0.966. However, on the foggy image sets from Munkkivuori, the opposite is true, as it achieves a mean SRCC of -0.977. Furthermore, the standard deviations of these calculations are pretty minor, at 0.050 and 0.023, indicating that QualiCLIP exhibits a strong consistency in its ranking of foggy images. This suggests that location is the determining factor for QualiCLIP when assessing foggy images. On the rainy image sets, the results are less consistent, with a mean SRCC value of -0.118 and a relatively high standard deviation of 0.436.

Differences in performance based on location are also evident for TOPIQ, which achieved a mean SRCC of 0.577 on the foggy Otaniemi image sets and -0.928 on the foggy Munkkivuori image sets. The standard deviation is relatively high in both cases. For the Otaniemi image sets, it is 0.494, and for the Munkkivuori drive, it is 0.169. On the rainy image sets, TOPIQ consistently exhibited clear inverse performance, often rating lower-quality images as higher quality. It achieved a mean SRCC value of -0.815 and a standard deviation of 0.177.

Inverse performance is also observed for DB-CNN on the rainy image sets. It achieved a mean SRCC of -0.601 across all rainy image sets with a high

standard deviation of 0.496. However, it achieved much better performance on the foggy image sets across the two locations. On the foggy image sets from Otaniemi, it achieved a high mean SRCC of 0.998, whereas on the foggy image sets from Munkkivuori, the mean SRCC dropped to 0.730. This indicates that DB-CNN, similar to TOPIQ and QualiCLIP, performed worse on images from Munkkivuori than on those from Otaniemi.

Figure 4.2, 4.3, 4.4, 4.5, and 4.6 show a boxplot of the scores of the NR-IQA methods across both locations and weather distortions. Every fifth image index is considered, and each boxplot represents the distribution of scores corresponding to the same image index. Figure 4.2 illustrates a general downward trend in DB-CNN scores for the foggy image sets from Otaniemi, and to a lesser but still notable extent, for those from Munkkivuori. These trends suggest good performance as DB-CNN, like all other NR-IQA metrics apart from IL-NIQE, assigns higher scores to images it considers to be of higher quality. For the rainy image sets at both locations, the trend is slightly upward, with SRCC values being negative in 16 out of 20 cases. Another interesting result visible in Figure 4.2 is the sharp divide between the absolute scores of the foggy and rainy images. Notably, not a single rainy image is deemed as higher in quality than any of the foggy images.

The scores of TOPIQ are found in Figure 4.3, where it is noticeable that the predicted scores of the foggy image sets from Otaneimi are rather consistently decreasing, with a mean SRCC score of 0.577. Meanwhile, the performance on foggy image sets from Munkkivuori is significantly worse, with a mean SRCC value of -0.927. This is also evident in Figure 4.3 where the Munkkivuori-foggy plot shows an upward trend across the image sets. TOPIQ's poor performance on rainy image sets is also evident from the slightly upward trend on the Otaniemi-rainy and Munkkivuori-rainy image sets.

Turning to Figure 4.4, Q-Align demonstrates remarkable consistency in its scoring ability across locations and weather distortions. A uniform downward trend is observed across all conditions, without any outliers. The lowest mean SRCC observed across all subsets occurs for the rainy image sets from Munkkivuori, and even in this instance, Q-Align achieved an astonishing mean SRCC of 0.996. Additionally, the small sizes of the boxplots indicate slow variability in the scores among images within the same image index.

Moving on to Figure 4.5, it is clear that QualiCLIP is highly inconsistent in its scoring. Figure 4.5 shows a distinct downward trend in scores for Otaneimi-foggy, indicating good performance. However, the opposite is true for the Munkkivuori-foggy image sets, where a striking upward trend in scores is observed. On these image sets, QualiCLIP, like TOPIQ, exhibits a clear

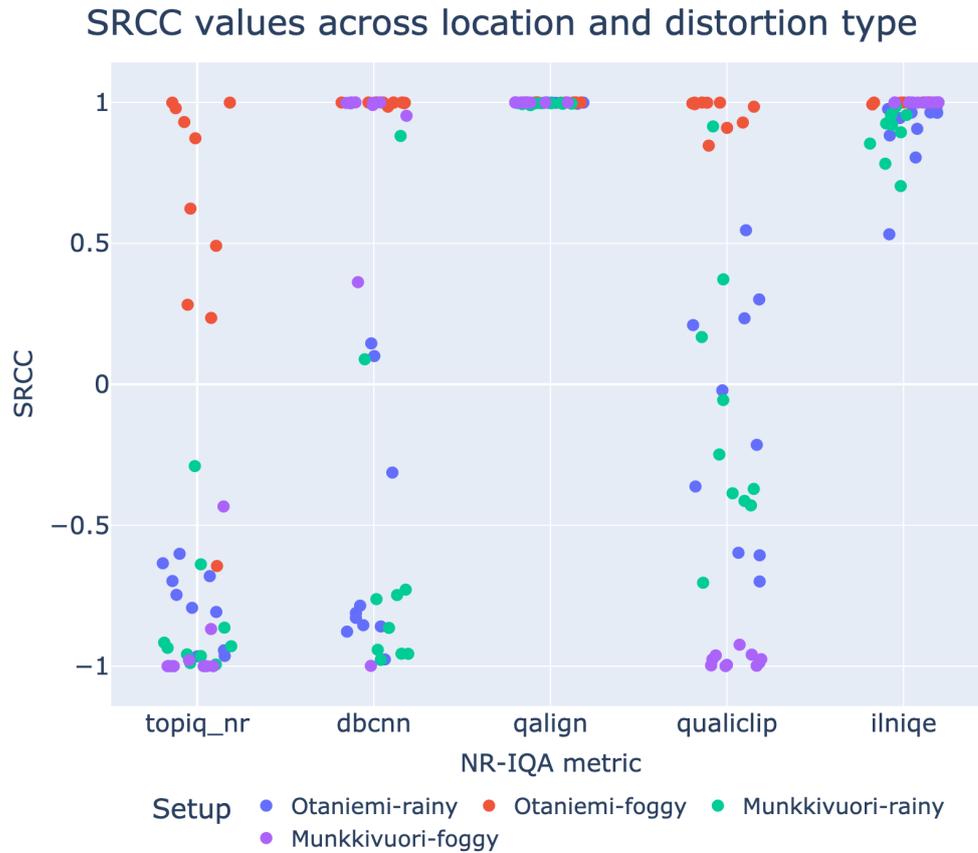


Figure 4.1: Distribution of SRCC values for the surveyed NR-IQA methods.

tendency towards inverse ranking, assigning progressively increasing scores to images with progressively decreasing quality. The inverse rankings are less pronounced but still present on the rainy image sets. QualiCLIP achieved mean SRCCs of -0.121 and -0.115 on the rainy image sets from Otaniemi and Munkkivuori, respectively.

Finally, the assigned scores of IL-NIQE are shown in Figure 4.6. Unlike the other NR-IQA metrics, IL-NIQE assigns lower scores to images it deems to have higher quality and higher scores to images it deems as having lower quality. Figure 4.6 shows a clear upward scoring trend across both locations and distortion types. The trend is more pronounced for the foggy image sets, as reflected in the higher mean SRCC of 0.999. In contrast, the trend is less evident for the rainy image sets, which have a lower mean SRCC of 0.892.

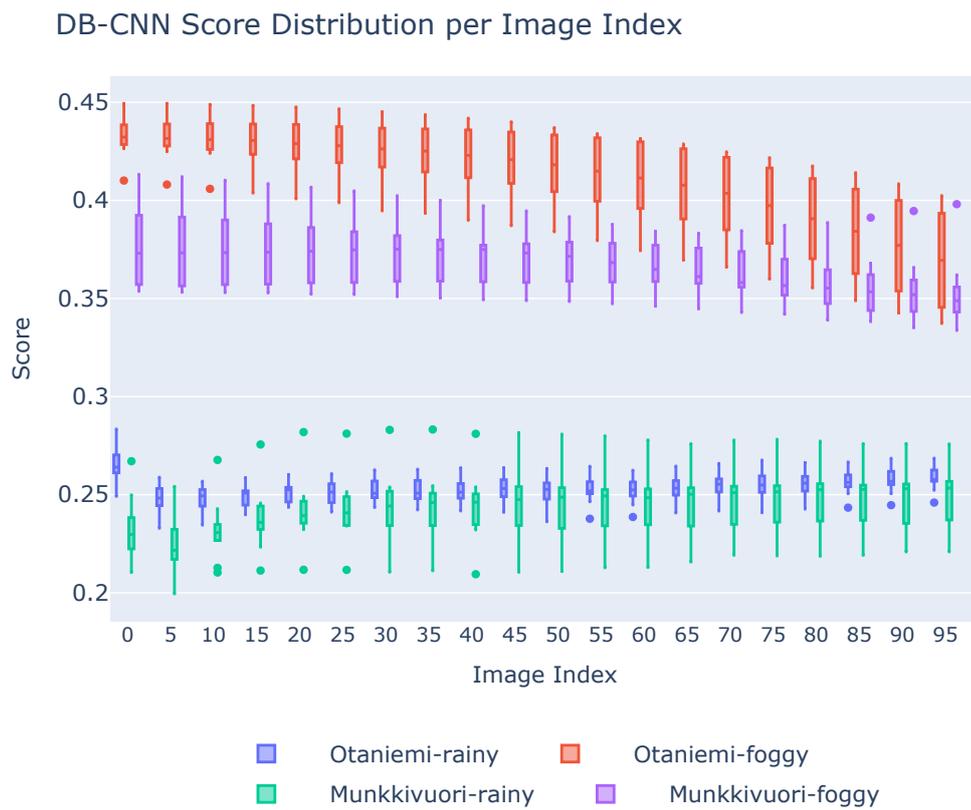


Figure 4.2: Boxplot of DB-CNN scores for every fifth image index, showing score distributions across weather and location variations

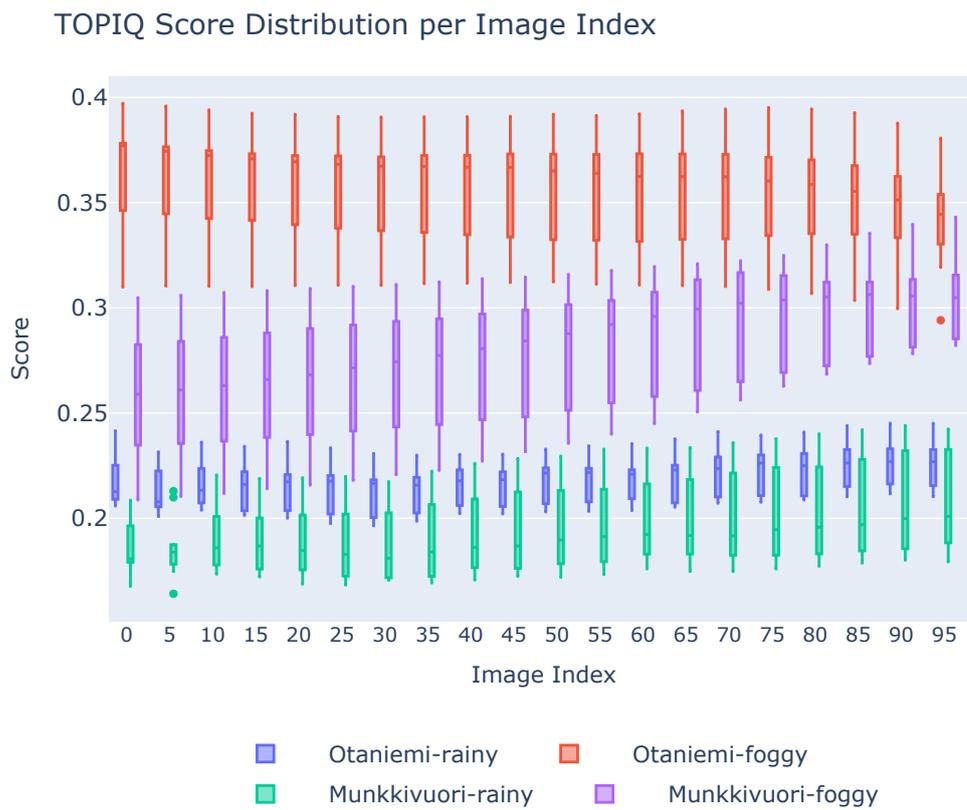


Figure 4.3: Boxplot of TOPIQ scores for every fifth image index, showing score distributions across weather and location variations

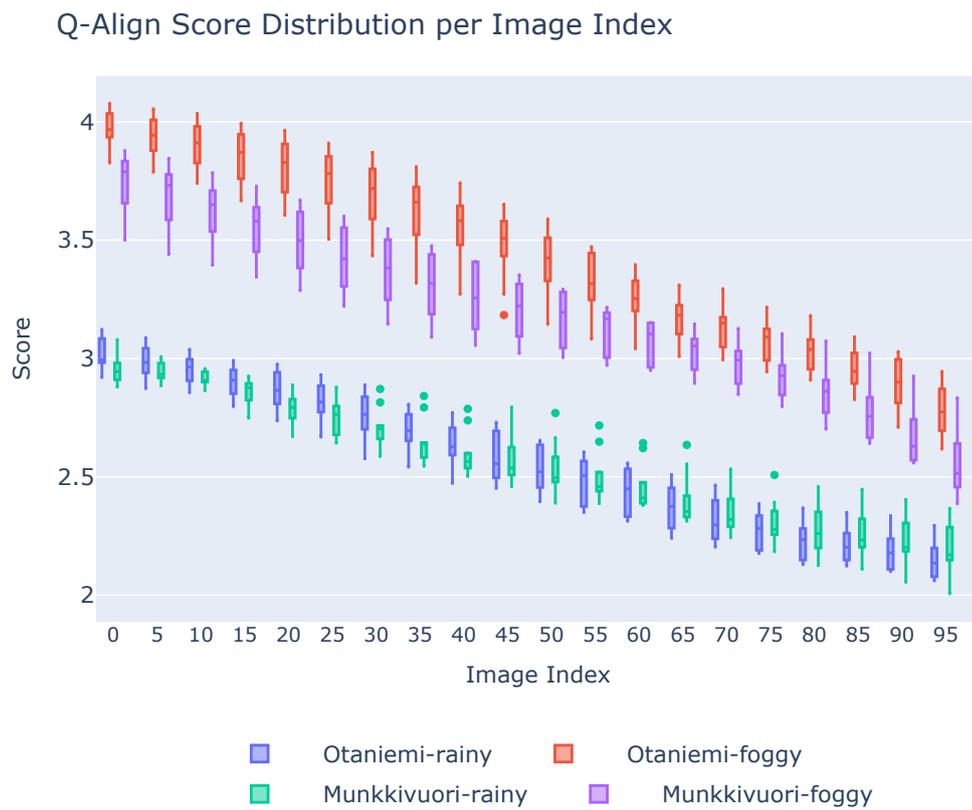


Figure 4.4: Boxplot of Q-Align scores for every fifth image index, showing score distributions across weather and location variations

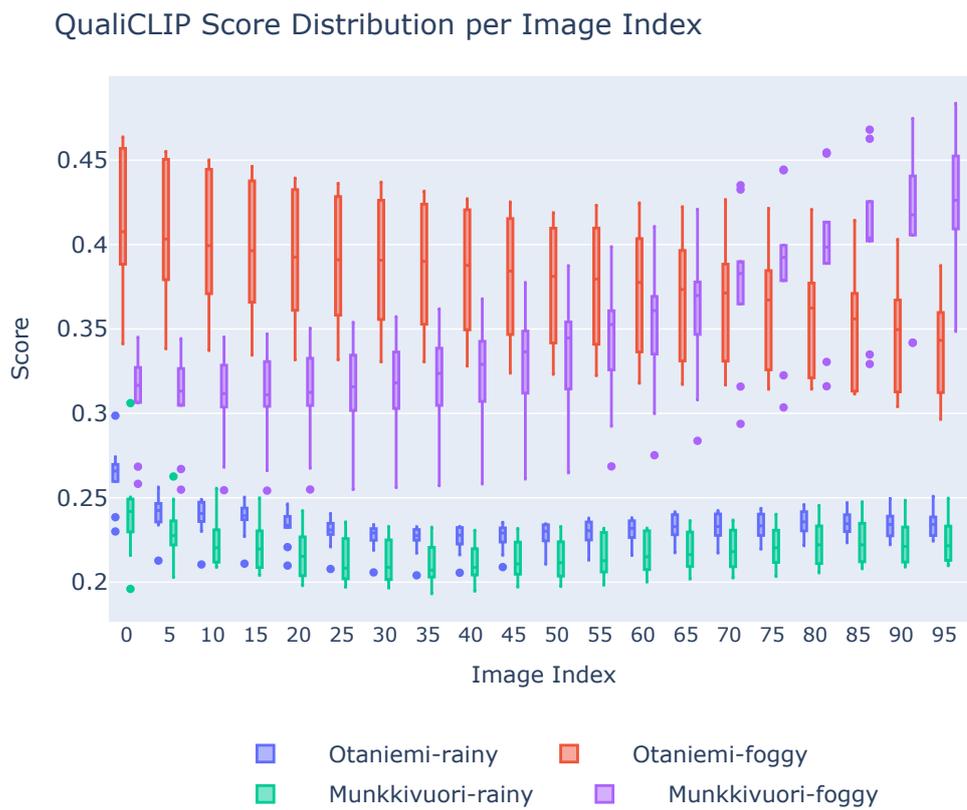


Figure 4.5: Boxplot of QualiCLIP scores for every fifth image index, showing score distributions across weather and location variations

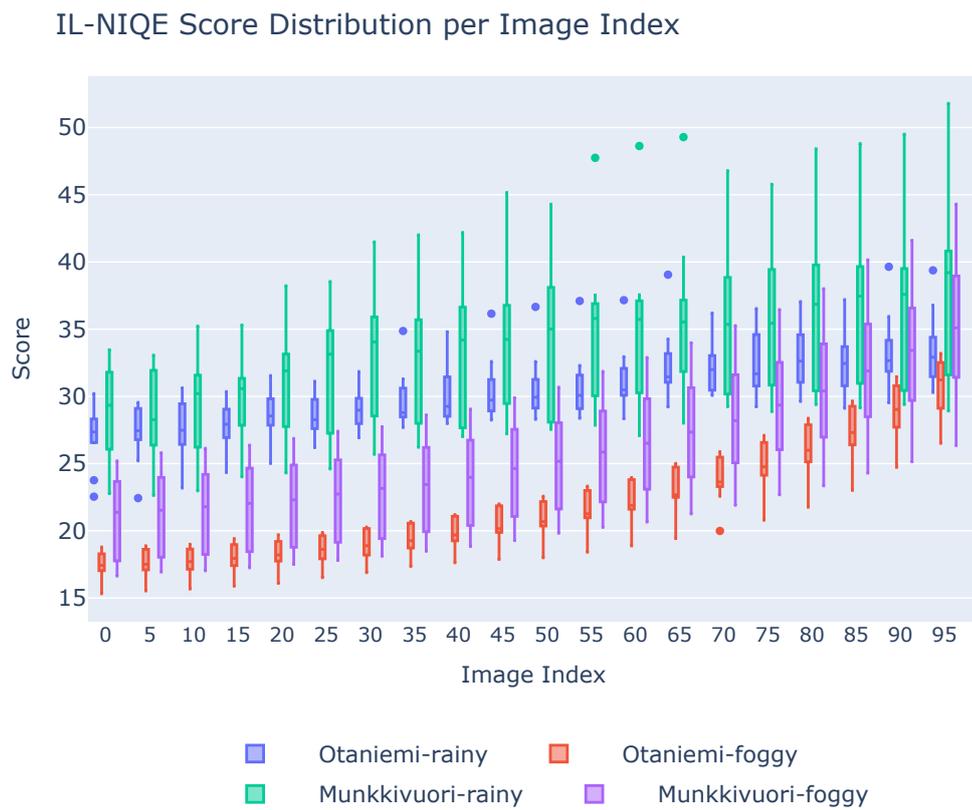


Figure 4.6: Boxplot of IL-NIQE scores for every fifth image index, showing score distributions across weather and location variations

4.1.1 Statistical tests

The result of the Friedman test is 109.11, with a p-value of 1.126×10^{-22} , meaning that the null hypothesis of no significant differences between the methods can be rejected at a 1% significance level. Thus, it can be demonstrated with strong statistical significance that at least one method consistently outperforms the others. Following the Friedman test, the results from the pairwise Wilcoxon signed-rank tests are found in Table 4.2. In this table, the model pair column lists each pairwise combination between the models, where the difference is computed as the first method minus the second. The statistic column lists the Wilcoxon signed-rank statistics, $\sum R^+$ lists the sums of the positive ranks, $\sum R^-$ lists the sums of the negative ranks, and the corrected p-value column lists the p-values after the Holm-Bonferroni correction.

The column of corrected p-values shows that all values, except for DB-CNN - QualiCLIP, are below 0.01. This means that, at the 1% significance level, the null hypothesis of equal performance can be rejected for all model pairs except DB-CNN and QualiCLIP. In the first four rows of Table 4.2, the negative rank sum $\sum R^-$ is substantially greater than the positive rank sum $\sum R^+$ for all four comparisons. Therefore, it can be concluded that TOPIQ is outperformed by all other models. The fifth row shows that there is no significant difference between DB-CNN and QualiCLIP. The next two rows show that DB-CNN is outperformed by both Q-Align and IL-NIQE. Similarly, the two rows that follow demonstrate that QualiCLIP is outperformed by both Q-Align and IL-NIQE. Finally, the last row indicates that Q-Align outperforms IL-NIQE.

Model Pair	Statistic	$\sum R^+$	$\sum R^-$	Corrected p-value
TOPIQ – DB-CNN	96	96	-724	0.000
TOPIQ – QualiCLIP	74	74	-746	0.000
TOPIQ – Q-Align	0	0	-820	0.000
TOPIQ – IL-NIQE	0	0	-820	0.000
DB-CNN – QualiCLIP	374	445	-374	0.646
DB-CNN – Q-Align	32	32	-788	0.000
DB-CNN – IL-NIQE	39	39	-741	0.000
QualiCLIP – Q-Align	4	4	-816	0.000
QualiCLIP – IL-NIQE	6	6	-814	0.000
Q-Align – IL-NIQE	96	724	-96	0.000

Table 4.2: Results of the Wilcoxon signed-rank tests after Holm-Bonferroni corrections.

4.2 NR-PCQA results

The mean SRCC and KRCC values of MM-PCQA and MS-PCQE are found in Table 4.3. The standard deviations of the various metrics are found in the parentheses. Furthermore, each p-value is from a two-tailed permutation test evaluating whether the mean differs significantly from 0. All the SRCC and KRCC values in Table 4.3 differ from 0 at the 1% significance level. Moreover, MM-PCQA has a mean SRCC value of 0.172 with a standard deviation of 0.401. This suggests a small and positive correlation between its assigned scores and the ground-truth qualities of the point clouds. Additionally, a standard deviation of 0.401 indicates that MM-PCQA exhibits inconsistent performance across the 40 point cloud sets. Interestingly, MS-PCQE exhibits a stronger negative correlation with the ground-truth qualities. It has a mean SRCC value of -0.294 with a standard deviation of 0.343. This shows that MS-PCQE, similar to some NR-IQA metrics mentioned earlier, consistently rates lower-quality point clouds as having higher quality than those of better quality. However, MS-PCQE is also inconsistent across the image set, with a standard deviation of 0.343.

Table 4.3 shows the mean SRCC and KRCC of MM-PCQA and MS-PCQE on the fog-distorted point clouds. MM-PCQA yielded a very weak positive correlation, with a mean SRCC of 0.172 and a mean KRCC value of 0.129. However, the high standard deviations of 0.401 and 0.312 suggest inconsistent values across point cloud sets. MS-PCQE achieved a slight negative correlation, with a mean SRCC of -0.294 and a mean KRCC of

Metric	SRCC (SD)	KRCC (SD)
MM-PCQA	0.172** (0.401)	0.129** (0.312)
MS-PCQE	-0.294** (0.343)	-0.220** (0.275)

* $p < 0.05$ ** $p < 0.001$

Table 4.3: Mean SRCC and KRCC values of the NR-PCQA techniques on the set of fog-distorted point clouds. The p-values are obtained from permutation tests evaluating whether the mean SRCC and KRCC differ significantly from 0.

-0.220. Similar to MM-PCQA, MS-PCQE exhibited high variability with standard deviations of 0.343 and 0.275, respectively.

Figure 4.7 plots the distribution of SRCC scores for both MM-PCQA and MS-PCQE across weather distortions and time of day. MM-PCQA generally achieved higher SRCC scores on point cloud sets recorded in clear weather and during daytime. The mean SRCC for point clouds captured in clear weather was 0.316, while for those captured during daytime, it was 0.237. However, both cases had high variability, with standard deviations of 0.325 and 0.395, respectively. MM-PCQA exhibited near-random performance on point clouds captured in rainy and nighttime conditions, with mean SRCC values of just 0.028 and 0.051, respectively. As with clear weather and daytime conditions, there was also high variability, with the standard deviation of 0.418 for rainy sets and 0.384 for nighttime sets.

In contrast, MS-PCQE demonstrates greater consistency in its quality predictions across varying weather conditions and between daytime and nighttime settings. On point clouds captured in clear weather, MS-PCQE achieved a mean SRCC of -0.298 (STD = 0.321), while in rainy weather, it achieved a mean SRCC of -0.290 (STD = 0.364). Similarly, the point clouds captured in daytime yielded a mean SRCC of -0.293 (STD = 0.337), and for point clouds captured in nighttime, it achieved -0.297 (0.354). These results suggest that MS-PCQE performs equally poorly across different weather conditions and between day and night scenarios.

Figures 4.8 and 4.9 show boxplots of the scores of the NR-PCQA methods across weather and time of day. Both plots contain 10 x-values, corresponding to the 10 distortion severity levels within each point cloud set. Figures 4.8 and 4.8 show no clear upward or downward trend for MM-PCQA or MS-PCQE. Interestingly, MM-PCQA assigns negative scores to many point clouds in the dataset. This is noteworthy, as it does not occur for any point cloud in the MM-PCQA training set. This result could be due to the point clouds in the

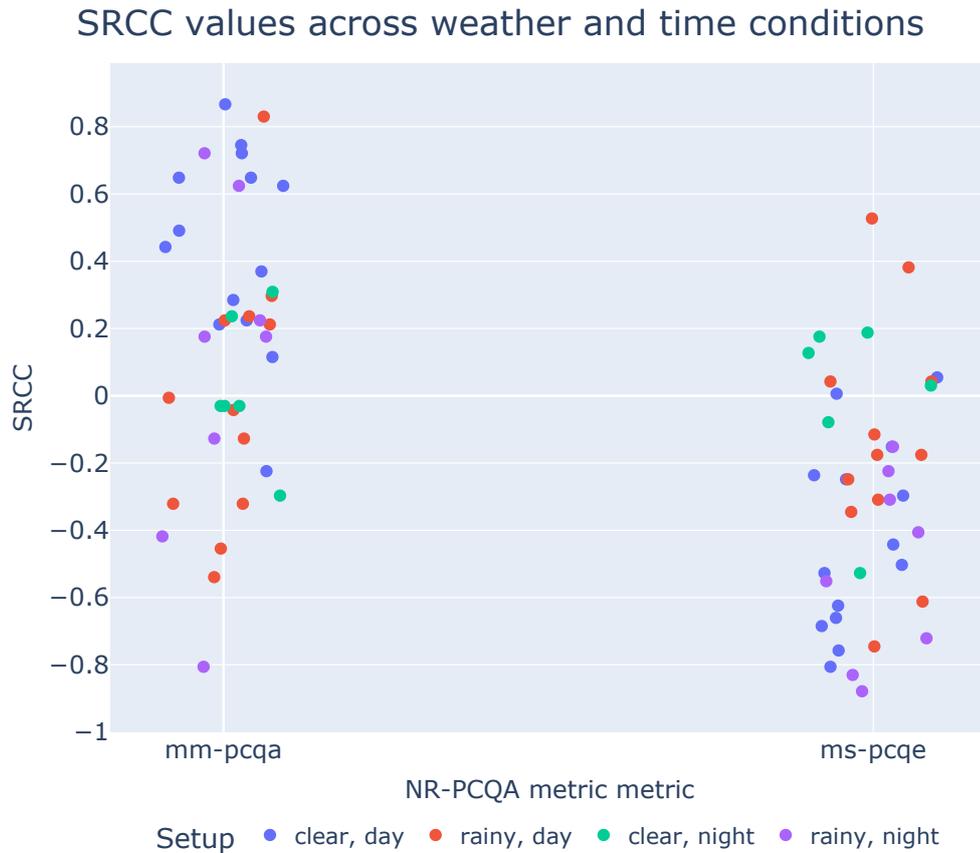


Figure 4.7: Distribution of SRCC values for the surveyed NR-PCQA methods.

dataset being significantly different from those used to train MM-PCQA.

4.2.1 Statistical tests

A two-sided Wilcoxon signed-rank test is performed to determine whether MM-PCQA and MS-PCQE differ significantly in performance. For this test, the SRCC values of each method across the 40 image sets are used. The resulting statistic is 90.5 with a p-value of $1.750153066180124 \times 10^{-5}$, meaning that the null hypothesis that the methods perform equally well can be rejected at the 1% significance level. In the test, the SRCC values of MM-PCQA are subtracted from those of MS-PCQE. The resulting positive rank sum $\sum R^+$ is 90.5 while the negative rank sum $\sum R^-$ is -729.5, indicating that MM-PCQA significantly outperforms MS-PCQE.

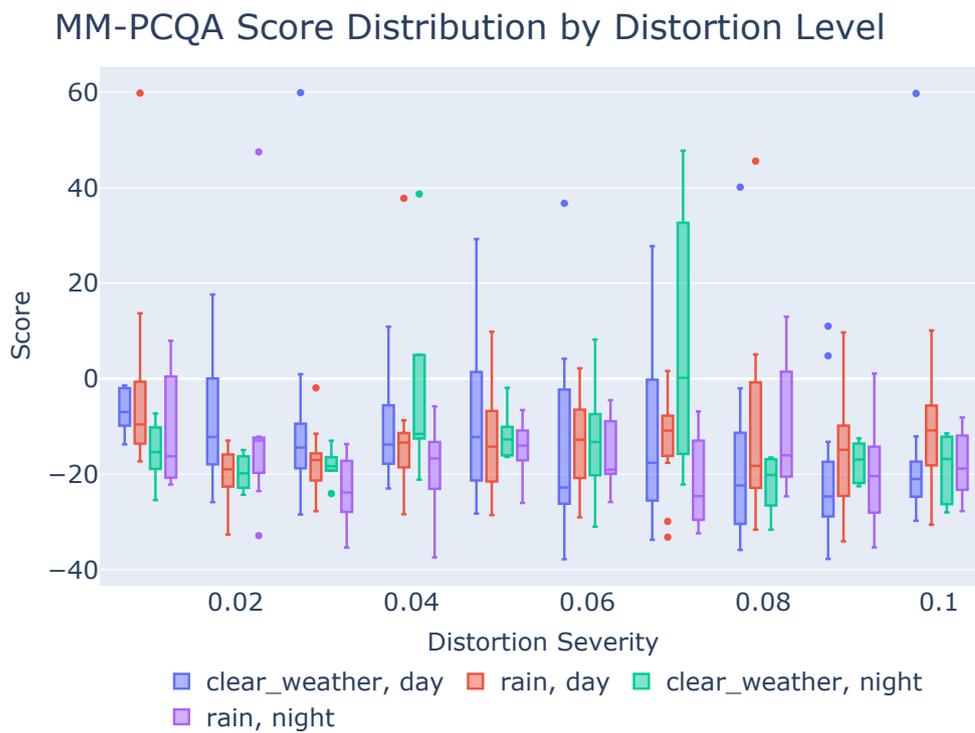


Figure 4.8: MM-PCQA score distributions across distortion levels and conditions.

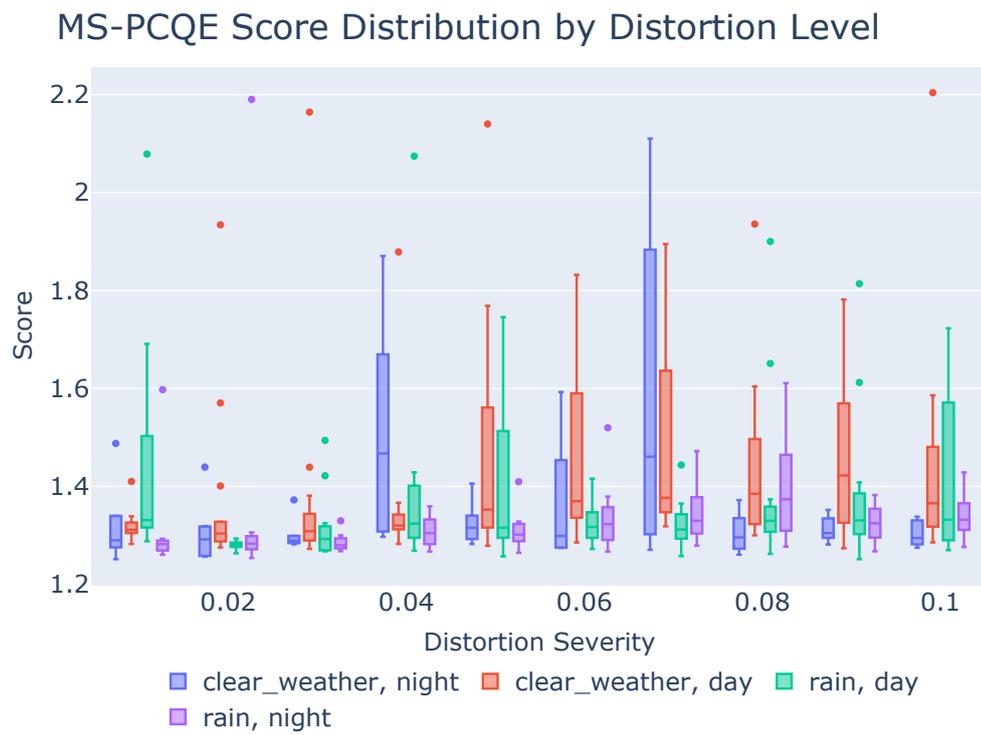


Figure 4.9: MS-PCQE score distributions across distortion levels and conditions,

Chapter 5

Discussion

Regarding the NR-IQA results, it is not surprising that 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. Q-Align was also trained on twelve NR-IQA datasets that contained both authentic and synthetic distortions. While training on both authentic and synthetic datasets typically reduces the accuracy of an NR-IQA model, 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 in the study.

What is more surprising is the poor performance of another state-of-the-art NR-IQA method, namely QualiCLIP. As mentioned earlier, QualiCLIP exhibited slight inverse performance on the rainy image sets, near-perfect performance on the foggy Otaniemi image sets, and near-perfect inverse performance on the foggy Munkkivuori image sets. Overall, the lack of performance is surprising given that QualiCLIP has achieved impressive results on traditional NR-IQA benchmark datasets. Additionally, the training process of QualiCLIP much resembles the evaluation process used in this thesis. QualiCLIP was trained by distorting images with increasing levels of noise and learning by ranking the distorted images. This is quite similar to the evaluation in this thesis, namely, distorting images and testing the NR-IQA's ability to rank them. One possible reason why QualiCLIP performed poorly is that the degradations used in this study differ from those used to train QualiCLIP. QualiCLIP was trained using 24 distortion types, none of which included weather-related noise. This could have had a negative impact on performance.

However, the fact that QualiCLIP was not trained on weather-related

distortions cannot explain the significant difference in performance between the foggy Otaniemi image sets and the foggy Munkkivuori image sets. This difference in performance also occurs for TOPIQ and means that on foggy images from Munkkivuori, both TOPIQ and QualiCLIP consistently assign higher scores to images with worse perceptual quality. An example of this can be seen in Figure 5.1. This discrepancy in performance is most likely attributed to differences in scene and lighting between the Otaniemi and Munkkivuori images. The images from Otaniemi are generally brighter and mainly consist of urban settings, while the Munkkivuori images are captured in rural environments and are darker. However, why these scene and lighting differences result in such dramatic changes in performance is unclear. It is also unclear why this does not occur for IL-NIQE, Q-Align, or DB-CNN.

Another notable result is the consistent performance of IL-NIQE. Despite relying on NSS, a somewhat antiquated technique in the field of NR-IQA, IL-NIQE outperforms QualiCLIP, DB-CNN, and TOPIQ. This is noteworthy as QualiCLIP and TOPIQ, both clearly outperform IL-NIQE on conventional NR-IQA benchmarks [47, 39]. One possible reason for IL-NIQE's strong performance is that synthetic rain and fog introduce consistent and predictable deviations in luminance and contrast compared to natural high-quality images. IL-NIQE may effectively detect these deviations by computing MSCN coefficients, which capture local variations in luminance and contrast.

A critical methodological decision in the thesis was to avoid using human subjective scores as the ground truth when evaluating images and point clouds. This decision was necessary, as there are no existing datasets of images and point clouds in the AV domain with subjective scores. However, it also has the consequence that it is not possible to evaluate whether a method's assigned scores are reliable. It is therefore possible that one of the surveyed NR-IQA metrics performs well in ranking images or point clouds, but that its absolute scores do not align with human perception.

The choice to use synthetic rather than authentic weather distortions also had its advantages and disadvantages. The main advantage was that it enabled distorting images with precise severity levels, which was necessary to establish ground truth rankings of the image and point cloud sets. However, the main disadvantage of this approach is that real-world weather phenomena are often more complex than what can be modeled using synthetic noise. For instance, rainy conditions can lead to droplets on the camera lens, which cause total or near-total blockages. The artificial rain generation cannot account for this. Furthermore, foggy images often display a non-uniform distribution of fog. Areas of an image with close objects may appear clearer, whereas areas that



(a) QualiCLIP score of 0.3047 and Topiq score of 0.2247



(b) QualiCLIP score of 0.4407 and Topiq score of 0.2861

Figure 5.1: Example where QualiCLIP and Topiq assign a higher score to a perceptually worse-quality image

depict distant backgrounds are more affected by fog. This is also something that the synthetic noise generation does not consider. The fact that synthetic noise does not fully model real-weather phenomena can be problematic, as a model can perform well on synthetic data but still struggle with data captured in authentic weather conditions. Therefore, it is crucial that the use of synthetic distortions is complemented with testing on authentic data to ensure that these systems perform reliably under real-world conditions.

Given the large number of NR-IQA methods in the literature, a selection of methods was necessary to keep the evaluation manageable. The goal with the selection was to cover a broad range of techniques that used different paradigms for NR-IQA. This meant that only one method per paradigm was included in the thesis. While this covers a broad range of techniques, it does not cover the nuances within each paradigm. Therefore, it is essential to recognize that alternative methods within these paradigms may yield significantly different results. Different results could possibly also have been obtained by using the same methods, but with different training datasets.

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 of varying distortion levels. Furthermore, while MM-PCQA achieved a statistically significant positive correlation, its mean SRCC value of 0.172 is too low to indicate practical usefulness. Additionally, the negative scores generated by MM-PCQA indicate that the AV point clouds fall outside the distribution of point clouds for which MM-PCQA was trained. This further highlights the limited usefulness of MM-PCQA on AV point clouds.

The reason for the NR-PCQA model's inability to rank point clouds is most likely that NR-PCQA methods are primarily designed to assess the quality of highly detailed colored point clouds. This is indicative by the fact that the NR-PCQA benchmark datasets all contain highly rich and detailed point clouds. For instance, in the WPC dataset [59], which was used to train MM-PCQA, the mean number of points in each point cloud is 1359007. In contrast, the number of points per point cloud in the AV point clouds used in the thesis is in the tens of thousands. The loss of color information in the AV point clouds is probably also a factor that explains the lack of performance.

Chapter 6

Conclusions and Future work

6.1 Conclusions

This thesis evaluated a set of NR-IQA and NR-PCQA methods on weather-distorted data used in the field of autonomous vehicles. The evaluation centered on synthetic weather distortions of images and point clouds, where the relative quality rankings were known. In the case of images, synthetic rain and synthetic fog were employed, whereas in the case of point clouds, only synthetic fog was utilized. Among the five evaluated NR-IQA methods, Q-Align and IL-NIQE demonstrated strong performance across all distortion types, suggesting that large multimodality models and NSS models are suitable for the task at hand. While both performed well, a Wilcoxon signed-rank test revealed that Q-Align ultimately outperformed IL-NIQE. The NR-PCQA evaluation showed that MM-PCQA outperformed MS-PCQE and achieved a weak, statistically significant correlation. However, this correlation is too weak to have any practical usefulness. Overall, the findings suggest that NR-IQA is more mature and reliable for applications in autonomous vehicles than NR-PCQA, which requires further research to be viable.

6.2 Limitations

The primary limitation of this thesis is the reliance on synthetic distortions rather than authentic distortions. While synthetic noise simplifies evaluation, it does not fully capture the complexity of real-world conditions. Furthermore, the lack of human-labelled subjective quality scores meant that only rank-based quality metrics, such as SRCC and KRCC, could be used in evaluating the NR-IQA and NR-PCQA methods. Typically, NR-IQA and NR-PCQA

methods are evaluated using both rank-based metrics and score-related metrics, such as the Pearson linear correlation coefficient. This means that relying solely on rank-based metrics is a significant limitation. Another limitation is that only five NR-IQA and two NR-PCQA methods were evaluated. Both NR-IQA and NR-PCQA are extensive fields of research containing numerous methods. Therefore, this thesis does not provide a comprehensive evaluation of how all NR-IQA or NR-PCQA methods perform in relation to the problem at hand. Finally, the AV point clouds used in the NR-PCQA evaluation differ significantly from those used to train the NR-PCQA methods. The NR-PCQA methods are trained on point clouds containing significantly more points and color information, which is not present in the AV point clouds.

6.3 Future work

One way to extend this work is to create a benchmark NR-IQA dataset of authentically distorted AV images with corresponding subjective scores. This would enable the evaluation of NR-IQA metrics not only using rank-based metrics, such as SRCC and KRCC, but also accuracy-based metrics, including the Pearson linear correlation coefficient and root mean square error, which are standard in the NR-IQA field. Using naturally occurring and authentic distortions would also provide insight into whether the results in this thesis generalize to the real world.

Another interesting research area is investigating the feasibility of online NR-IQA, which involves evaluating image quality in real-time. This approach has been used by Zhang and Eskandarian [60] and was missed during the literature review of this study. Zhang and Eskandarian [60] explored online NR-IQA of autonomous vehicles by creating the detection quality index (DQI), which scores AV images based on saliency maps and the performance of object detection algorithms. Based on this metric, they trained a neural network, SPA-NET, to predict the DQI. Future research in online NR-IQA could explore combining the DQI metric with one or more of the NR-IQA methods evaluated in this study.

To enhance the performance of NR-PCQA metrics, future research could utilize fused LiDAR and image data to generate colored point clouds. This process involves calibrating the LiDAR and camera, projecting LiDAR points onto the 2D image plane, and then recording the RGB values of each 3D point based on the corresponding pixel in the image. Given that both NR-PCQA metrics evaluated in the thesis require colored point clouds, their

performance would likely be improved with authentic color information rather than simulated color information, as used in this thesis.

References

- [1] B. Crisafulli, R. Guesalaga, and R. Dimitriu, “Consumers’ adoption of autonomous cars as a personal values-directed behavior,” *Journal of Business Research*, vol. 189, p. 115106, 2025. [Page 1.]
- [2] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixao, F. Mutz *et al.*, “Self-driving cars: A survey,” *Expert systems with applications*, vol. 165, p. 113816, 2021. [Page 1.]
- [3] Y. Dong, C. Kang, J. Zhang, Z. Zhu, Y. Wang, X. Yang, H. Su, X. Wei, and J. Zhu, “Benchmarking robustness of 3d object detection to common corruptions,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1022–1032. [Page 1.]
- [4] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, “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*, 2020, pp. 11 682–11 692. [Page 1.]
- [5] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge, and W. Brendel, “Benchmarking robustness in object detection: Autonomous driving when winter is coming,” *arXiv preprint arXiv:1907.07484*, 2019. [Page 1.]
- [6] Y. Suh, Y. Chung, and Y. Park, “Ai training data management for reliable autonomous vehicles using hashgraph,” *Applied Sciences*, vol. 15, no. 11, p. 6123, 2025. [Page 1.]
- [7] “The ROADVIEW Project - Developing robust automated driving in extreme weather conditions,” <https://roadview-project.eu/>, accessed: 2024-11-28. [Page 2.]

- [8] S. Sonawane and A. Deshpande, “Image quality assessment techniques: An overview,” *International Journal of Engineering Research*, vol. 3, no. 4, 2014. [Page 2.]
- [9] Y. ZHOU, Z. ZHANG, W. SUN, X. MIN, and G. ZHAI, “Perceptual quality assessment for point clouds: A survey,” *ZTE Communications*, vol. 21, no. 4, p. 3, 2023. [Pages 2, 12, and 17.]
- [10] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, “Perception and sensing for autonomous vehicles under adverse weather conditions: A survey,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 146–177, 2023. doi: <https://doi.org/10.1016/j.isprsjprs.2022.12.021>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271622003367> [Page 2.]
- [11] Y. Li and J. Ibanez-Guzman, “Lidar for Autonomous Driving: The principles, challenges, and trends for automotive lidar and perception systems,” *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, Jul. 2020. doi: 10.1109/MSP.2020.2973615 ArXiv:2004.08467 [cs]. [Online]. Available: <http://arxiv.org/abs/2004.08467> [Page 5.]
- [12] M. Liu, E. Yurtsever, J. Fossaert, X. Zhou, W. Zimmer, Y. Cui, B. L. Zagar, and A. C. Knoll, “A survey on autonomous driving datasets: Statistics, annotation quality, and a future outlook,” *IEEE Transactions on Intelligent Vehicles*, 2024. [Page 6.]
- [13] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, “Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2020. [Page 6.]
- [14] G. Zhai and X. Min, “Perceptual image quality assessment: a survey,” *Science China Information Sciences*, vol. 63, pp. 1–52, 2020. [Pages 7 and 31.]
- [15] Z. Wang and A. C. Bovik, “Modern image quality assessment,” PhD Thesis, Springer, 2006. [Page 7.]
- [16] 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. [Pages 7, 12, 17, and 31.]

- [17] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012. [Page 7.]
- [18] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012. [Page 7.]
- [19] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, 2015. [Pages 8, 28, and 35.]
- [20] X. Yang, F. Li, and H. Liu, “A survey of dnn methods for blind image quality assessment,” *IEEE Access*, vol. 7, pp. 123 788–123 806, 2019. [Pages 8, 12, 17, and 28.]
- [21] 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. [Page 8.]
- [22] L. Kang, P. Ye, Y. Li, and D. Doermann, “Convolutional neural networks for no-reference image quality assessment,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1733–1740. [Page 8.]
- [23] X. Liu, J. Van De Weijer, and A. D. Bagdanov, “Rankiqa: Learning from rankings for no-reference image quality assessment,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1040–1049. [Pages 8 and 18.]
- [24] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, “Deep neural networks for no-reference and full-reference image quality assessment,” *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 206–219, 2017. [Page 8.]
- [25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014. [Page 8.]
- [26] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, “Musiq: Multi-scale image quality transformer,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 5148–5157. [Pages 8 and 30.]

- [27] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, “Learning transferable visual models from natural language supervision,” in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763. [Page 9.]
- [28] J. Wang, K. C. Chan, and C. C. Loy, “Exploring clip for assessing the look and feel of images,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 2, 2023, pp. 2555–2563. [Pages 9 and 30.]
- [29] 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 lms for visual scoring via discrete text-defined levels,” *arXiv preprint arXiv:2312.17090*, 2023. [Pages 9, 30, and 35.]
- [30] 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. [Pages 10, 13, 17, 30, and 31.]
- [31] Y. Liu, Q. Yang, Y. Xu, and L. Yang, “Point cloud quality assessment: Dataset construction and learning-based no-reference metric,” *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 2s, pp. 1–26, 2023. [Pages 10 and 30.]
- [32] B. Liu, M. Wang, H. Foroosh, M. Tappen, and M. Pensky, “Sparse convolutional neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 806–814. [Page 10.]
- [33] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, “No-reference quality assessment for 3d colored point cloud and mesh models,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7618–7631, 2022. [Page 10.]
- [34] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, “Emerging mpeg standards for point cloud compression,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2018. [Page 11.]

- [35] Q. Liu, H. Yuan, H. Su, H. Liu, Y. Wang, H. Yang, and J. Hou, "Pqa-net: Deep no reference point cloud quality assessment via multi-view projection," *IEEE transactions on circuits and systems for video technology*, vol. 31, no. 12, pp. 4645–4660, 2021. [Page 11.]
- [36] X. Chai and F. Shao, "Ms-pcqe: Efficient no-reference point cloud quality evaluation via multi-scale interaction module in immersive communications," *IEEE Transactions on Consumer Electronics*, 2024. [Pages 11 and 13.]
- [37] 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. [Page 11.]
- [38] R. A. Manap and L. Shao, "Non-distortion-specific no-reference image quality assessment: A survey," *Information Sciences*, vol. 301, pp. 141–160, 2015. [Page 12.]
- [39] L. Agnolucci, L. Galteri, and M. Bertini, "Quality-aware image-text alignment for real-world image quality assessment," *arXiv preprint arXiv:2403.11176*, 2024. [Pages 18, 30, 35, and 53.]
- [40] 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, online; accessed 16 April 2025. [Page 18.]
- [41] 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. [Pages 18 and 19.]
- [42] 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. doi: 10.3390/info11020125. [Online]. Available: <https://www.mdpi.com/2078-2489/11/2/125> [Pages 20 and 23.]
- [43] M. Hahner, C. Sakaridis, D. Dai, and L. Van Gool, "Fog simulation on real lidar point clouds for 3d object detection in adverse weather," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15 283–15 292. [Pages ix, 24, 26, and 28.]

- [44] R. H. Rasshofer, M. Spies, and H. Spies, “Influences of weather phenomena on automotive laser radar systems,” *Advances in radio science*, vol. 9, pp. 49–60, 2011. [Page 24.]
- [45] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, “Blind image quality assessment using a deep bilinear convolutional neural network,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, 2020. doi: 10.1109/TCSVT.2018.2886771 [Pages 28, 35, and 36.]
- [46] F. Gao, J. Yu, S. Zhu, Q. Huang, and Q. Tian, “Blind image quality prediction by exploiting multi-level deep representations,” *Pattern Recognition*, vol. 81, pp. 432–442, 2018. [Page 28.]
- [47] C. Chen, J. Mo, J. Hou, H. Wu, L. Liao, W. Sun, Q. Yan, and W. Lin, “Topiq: A top-down approach from semantics to distortions for image quality assessment,” *IEEE Transactions on Image Processing*, 2024. [Pages 28, 35, and 53.]
- [48] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, “No-reference image quality assessment via transformers, relative ranking, and self-consistency,” in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 1220–1230. [Page 30.]
- [49] W. Zhang, G. Zhai, Y. Wei, X. Yang, and K. Ma, “Blind image quality assessment via vision-language correspondence: A multitask learning perspective,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 14 071–14 081. [Page 30.]
- [50] C. Chen and J. Mo, “IQA-PyTorch: Pytorch toolbox for image quality assessment,” [Online]. Available: <https://github.com/chaofengc/IQA-PyTorch>, 2022. [Page 30.]
- [51] Q. Liu, H. Su, Z. Duanmu, W. Liu, and Z. Wang, “Perceptual quality assessment of colored 3d point clouds,” *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2022. doi: 10.1109/TVCG.2022.3167151 [Page 30.]
- [52] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, “Perceptual quality assessment of 3d point clouds,” in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 3182–3186. [Page 30.]

- [53] D. C. Lepcha, B. Goyal, A. Dogra, and V. Goyal, “Image super-resolution: A comprehensive review, recent trends, challenges and applications,” *Information Fusion*, vol. 91, pp. 230–260, 2023. [Page 31.]
- [54] S. Xu, S. Jiang, and W. Min, “No-reference/blind image quality assessment: a survey,” *IETE Technical Review*, vol. 34, no. 3, pp. 223–245, 2017. [Page 32.]
- [55] M. Marozzi, “Some remarks about the number of permutations one should consider to perform a permutation test,” *Statistica*, vol. 64, no. 1, pp. 193–201, 2004. [Page 33.]
- [56] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine learning research*, vol. 7, no. Jan, pp. 1–30, 2006. [Page 33.]
- [57] “Interpret all statistics for Friedman Test.” [Online]. Available: <https://support.minitab.com/en-us/minitab/help-and-how-to/statistics/non-parametrics/how-to/friedman-test/interpret-the-results/all-statistics/> [Page 33.]
- [58] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures David J. Sheskin*, fifth edition ed. Chapman Hall/CRC, 2020. [Page 34.]
- [59] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, “Perceptual quality assessment of 3d point clouds,” in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 3182–3186. [Page 55.]
- [60] C. Zhang and A. Eskandarian, “A quality index metric and method for online self-assessment of autonomous vehicles sensory perception,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 13 801–13 812, 2023. [Page 58.]

€€€€ For DIVA €€€€

```
{
  "Author1": { "Last name": "Stenmark",
    "First name": "Victor",
    "Local User Id": "u100001",
    "E-mail": "vstenm@kth.se",
    "organisation": { "L1": "School of Electrical Engineering and Computer Science",
    }
  },
  "Cycle": "2",
  "Course code": "DA231X",
  "Credits": "30.0",
  "Degree1": { "Educational program": "Bachelor's Programme in Information and Communication Technology",
    "programcode": "TCOMK",
    "Degree": "Master's degree",
    "subjectArea": "Technology"
  },
  "Title": {
    "Main title": "Evaluating NR-IQA and NR-PCQA Methods on Weather-Distorted Data in Autonomous Driving",
    "Subtitle": "A subtitle in the language of the thesis",
    "Language": "eng",
    "Alternative title": {
      "Main title": "Detta är den svenska översättningen av titeln",
      "Subtitle": "Detta är den svenska översättningen av undertiteln",
      "Language": "swe"
    }
  },
  "Supervisor1": { "Last name": "Veronika",
    "First name": "Domova",
    "Local User Id": "u100003",
    "E-mail": "veronica.domova@gmail.com",
    "organisation": { "L1": "School of Electrical Engineering and Computer Science",
      "L2": "Computer Science" }
  },
  "Supervisor2": { "Last name": "Marsh",
    "First name": "Ian",
    "E-mail": "ian.marsh@ri.se",
    "Other organisation": "RISE"
  },
  "Examiner1": { "Last name": "Li",
    "First name": "Haibo",
    "Local User Id": "u1d13i2c",
    "E-mail": "haiboli@kth.se",
    "organisation": { "L1": "School of Electrical Engineering and Computer Science",
      "L2": "Computer Science" }
  },
  "Cooperation": { "Partner_name": "Företaget AB",
    "National Subject Categories": "10201, 10206",
    "Other information": { "Year": "2025", "Number of pages": "1,??",
    "Copyrightleft": "copyright",
    "Series": { "Title of series": "TRITA – EECS-EX", "No. in series": "2024:0000" },
    "Opponents": { "Name": "A. B. Normal & A. X. E. Normalè",
    "Presentation": { "Date": "2022-03-15 13:00",
      "Language": "eng",
      "Room": "via Zoom https://kth-se.zoom.us/j/ddddddddd",
      "Address": "Isafjordsgatan 22 (Kistagången 16)",
      "City": "Stockholm",
      "Number of lang instances": "2",
      "Abstract[eng ]": €€€€

€€€€,
"Keywords[eng ]": €€€€
Canvas Learning Management System, Docker containers, Performance tuning €€€€,
"Abstract[swe ]": €€€€
€€€€,
"Keywords[swe ]": €€€€
Canvas Lärplattform, Dockerbehållare, Prestandajustering €€€€,
}
```

acronyms.tex

```
%%% Local Variables:
%%% mode: latex
%%% TeX-master: t
%%% End:
% The following command is used with glossaries-extra
\setabbreviationstyle[acronym]{long-short}
% The form of the entries in this file is \newacronym[label]{acronym}{phrase}
% or \newacronym[options]{label}{acronym}{phrase}
% see "User Manual for glossaries.sty" for the details about the options, one example is shown below
% note the specification of the long form plural in the line below
%\newacronym[longplural={Debugging Information Entities}]{DIE}{DIE}{Debugging Information Entity}
%
% The following example also uses options
%\newacronym[shortplural={OSes}, firstplural={operating systems (OSes)}]{OS}{OS}{operating system}

% note the use of a non-breaking dash in long text for the following acronym
%\newacronym{IQL}{IQL}{Independent -QLearning}

% example of putting in a trademark on first expansion
%\newacronym[first={NVIDIA OpenSHMEM Library (NVSHMEM\texttrademark)}]{NVSHMEM}{NVSHMEM}{NVIDIA OpenSHMEM Library}

%\newacronym{KTH}{KTH}{KTH Royal Institute of Technology}

%\newacronym{LAN}{LAN}{Local Area Network}
%\newacronym{VM}{VM}{virtual machine}
% note the use of a non-breaking dash in the following acronym
%\newacronym{WiFi}{-WiFi}{Wireless Fidelity}

%\newacronym{WLAN}{WLAN}{Wireless Local Area Network}
%\newacronym{UN}{UN}{United Nations}
%\newacronym{SDG}{SDG}{Sustainable Development Goal}

\newacronym{IQA}{IQA}{Image quality assessment}
```