

# Stochasticity-aware No-Reference Point Cloud Quality Assessment

Songlin Fan<sup>1,3</sup>, Wei Gao<sup>1\*</sup>, Zhineng Chen<sup>2</sup>, Ge Li<sup>1</sup>, Guoqing Liu<sup>4</sup> and Qicheng Wang<sup>4</sup>

<sup>1</sup>Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology, Shenzhen Graduate School, Peking University

<sup>2</sup>School of Computer Science, Fudan University

<sup>3</sup>China Mobile Shanghai ICT Co., Ltd

<sup>4</sup>Youiia Innov Tech Co., Ltd

{slfan, gaowei262, geli}@pku.edu.cn, zhincheng@fudan.edu.cn, {guoqing, wangqicheng}@minieye.cc

## Abstract

The evolution of point cloud processing algorithms necessitates an accurate assessment for their quality. Previous works consistently regard point cloud quality assessment (PCQA) as a MOS regression problem and devise a deterministic mapping, ignoring the stochasticity in generating MOS from subjective tests. This work presents the first probabilistic architecture for no-reference PCQA, motivated by the labeling process of existing datasets. The proposed method can model the quality judging stochasticity of subjects through a tailored conditional variational autoencoder (CVAE) and produces multiple intermediate quality ratings. These intermediate ratings simulate the judgments from different subjects and are then integrated into an accurate quality prediction, mimicking the generation process of a ground truth MOS. Specifically, our method incorporates a Prior Module, a Posterior Module, and a Quality Rating Generator, where the former two modules are introduced to model the judging stochasticity in subjective tests, while the latter is developed to generate diverse quality ratings. Extensive experiments indicate that our approach outperforms previous cutting-edge methods by a large margin and exhibits gratifying cross-dataset robustness. Codes are available at <https://git.openi.org.cn/OpenPointCloud/nrpcqa>.

## 1 Introduction

Point clouds are essential representations of 3D scenes and have found wide applications in various fields [Qi *et al.*, 2017; Fan *et al.*, 2024]. Due to the unaffordable complexity and size of raw point clouds, compression techniques are commonly used to reduce relevant processing and storage costs. However, compression can lead to visual quality degradation. Therefore, accurate measurement of point cloud quality is crucial to assess the fidelity of processed data and serves as a foundation for benchmarking and improving point cloud processing algorithms. Similar to the definitions for image quality assessment, point cloud quality assessment (PCQA)

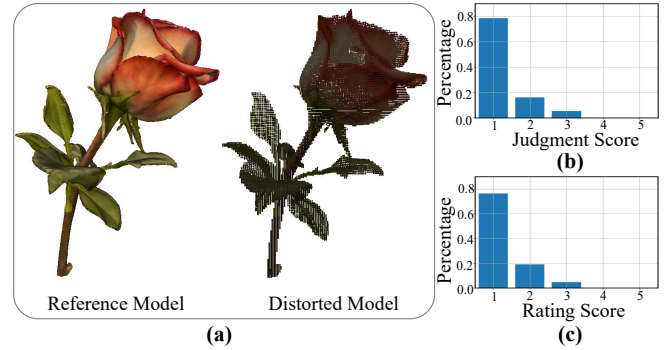


Figure 1: Stochasticity in standard dataset construction and predictions of our method. (a) Paired stimuli. (b) Distribution of 37 quality judgments about (a) in a recent subjective test. (c) Distribution of 37 intermediate quality ratings predicted from our model.

can be categorized into three factions: full-reference PCQA (FR-PCQA), reduced-reference PCQA (RR-PCQA), and no-reference PCQA (NR-PCQA), depending on the availability of reference data. As pristine data are often transparent in typical client applications, NR-PCQA has received significant attention and exploration efforts.

Conventional NR-PCQA methods [Wang *et al.*, 2023; Zhang *et al.*, 2022a; Liu *et al.*, 2021b; Zhang *et al.*, 2023; Chetouani *et al.*, 2021; Liu *et al.*, 2023] consistently adhere to a standardized framework that devises a deterministic fitting function or neural network mapping  $M = f(I; \omega)$ , where  $\omega$  denotes the model parameters/weights,  $I$  represents the input point cloud or its derivatives (*e.g.*, patches or projections), and  $M$  corresponds to the mean opinion score (MOS). MOSs, the most reliable visual quality description for point clouds, are usually obtained via subjective tests [Fan and Gao, 2023; Liu *et al.*, 2022a; Liu *et al.*, 2021a; Yang *et al.*, 2020a]. A subjective test involves recruiting several subjects to conduct laborious subjective experiments and gather human judgments for the viewed point cloud stimuli. Eventually, the discrepant judgment scores of a sample from different subjects are synthesized or directly averaged into the final MOS. We argue that though distortions in point clouds are deterministic, existing NR-PCQA solutions deviate from the generation of ground truth MOSs in subjective tests and fail to model the stochasticity in the human visual system (HVS) when giving

\*Corresponding author.

point cloud quality judgements.

Results of subjective tests [Fan and Gao, 2023; Liu *et al.*, 2023; Wu *et al.*, 2021; Javaheri *et al.*, 2020b] reveal that the HVS exhibits inherent volatility and stochasticity. That is, the quality judgments of different subjects may vary when evaluating the same point cloud stimuli due to the subjective nature of the HVS, which is influenced by complex factors such as individual visual sensitivity, cognitive biases, and perceptual preferences. Even a single subject may have distinct judgments for the same stimuli due to temporal fluctuations induced by prior context and mood. Figure 1 illustrates the stochasticity of a standard subjective test [Fan and Gao, 2023], showing that the actual quality judgments of a point cloud given by different subjects exhibit a specific distribution rather than a deterministic opinion. Herefore, existing overconfident methods that rely on deterministic mappings are susceptible to judgment biases. Furthermore, fitting a potentially biased MOS is one of the most critical factors that weaken the robustness of existing methods. Inspired by practical subjective tests, this work presents the first attempt at judgment distribution prediction and models the stochasticity in ground truth MOSs to produce multiple quality ratings for each point cloud. Multiple ratings are then integrated into a comprehensive quality prediction.

In specific, we propose a novel conditional variational autoencoder (CVAE) architecture for NR-PCQA, which leverages a latent variable conditioned on the projections of point clouds to accommodate the judgment variations of different subjects. Since the raw quality judgments in subjective tests of popular datasets are unavailable, which is necessary for training a CVAE to avoid posterior collapse, we utilize an effective KL annealing strategy [Sønderby *et al.*, 2016] to achieve diverse quality outputs for a point cloud. The proposed CVAE architecture comprises a Prior Module ( $PM_{prior}$ ), a Posterior Module ( $PM_{post}$ ), and a Quality Rating Generator (QRG). The  $PM_{prior}$  and  $PM_{post}$  are developed to predict the prior and posterior distributions of the latent variable, from which stochastic features can be sampled. Each sampled stochastic feature expresses the factors that may result in a possible judgment variant in subjective tests. Given the sampled stochastic feature, we further introduce the QRG to excavate the uncertain information in the stochastic feature and deterministic distortion information in point cloud projections and produce a stochastic quality rating. As the prior and posterior distribution mismatch of CVAEs can merely guarantee suboptimal performance [Sohn *et al.*, 2015], we utilize the Gaussian Stochastic Neural Network (GSNN) [Sohn *et al.*, 2015] to mitigate the disparity between training and testing. In the testing phase, we can sample the stochastic feature multiple times to predict diverse quality ratings of a point cloud. As shown in Figure 1, the diverse rating outputs of our method well capture the actual distribution of judgments derived from subjective tests, which can finally be averaged into an accurate quality prediction. Experimental results demonstrate that our method significantly outperforms all compared counterparts, including FR-PCQA and NR-PCQA methods.

We highlight the contributions of this work as follows:

- We are the first to probe stochasticity in dataset labeling for PCQA and propose a probabilistic perspective mim-

icking subjective tests and patterning the quality judgment distribution rather than a potentially biased MOS.

- We are the first to utilize a tailored CVAE in the quality assessment field, which can model both the uncertain factors in subjective tests and deterministic distortions in point clouds for accurate quality prediction.
- We verify the effectiveness of our method by achieving a new state-of-the-art performance and gratifying cross-dataset robustness, surpassing the existing best method by a significant margin across all datasets.

## 2 Related Work

### 2.1 Point Cloud Quality Assessment

The surging popularity of point clouds has sparked significant interest in PCQA. The development of PCQA has drawn inspiration from the advancements in image quality assessment, leading to categorizing methods into FR-PCQA, RR-PCQA, and NR-PCQA concerning their reliance on pristine point clouds. Early interest in FR-PCQA emerges from MPEG, with a primary focus on evaluating and optimizing point cloud compression algorithms. For example, point-to-point error [Mekuria *et al.*, 2016] and point-to-plane error [Tian *et al.*, 2017] are proposed to measure the geometry quality of point clouds and lay a solid foundation for subsequent investigations [Alexiou and Ebrahimi, 2018; Javaheri *et al.*, 2020a; Meynet *et al.*, 2019; Yang *et al.*, 2020a]. Inspired by the widely-used metrics in images (*i.e.*, PSNR and SSIM), Torlig *et al.* [Torlig *et al.*, 2018] and Alexiou *et al.* [Alexiou and Ebrahimi, 2020] design the point cloud-based versions, respectively, namely PSNR-yuv and PointSSIM. Yang *et al.* [Yang *et al.*, 2020b] infer the perceptual quality of point clouds using graph signal gradient and construct local graph representations. PCQM [Meynet *et al.*, 2020] considers both geometry and color distortions and utilizes a weighted combination to calculate a comprehensive quality description of point clouds. Recently, a novel distortion quantification method is devised by Yang *et al.* [Yang *et al.*, 2021] to model point cloud quality via multi-scale potential energy discrepancy. Apart from the aforementioned FR-PCQA works, some notable RR-PCQA methods [Liu *et al.*, 2021a; Liu *et al.*, 2022b] are also proposed and enrich the research on PCQA.

As considerable applications cannot observe reference point clouds, NR-PCQA [Wang *et al.*, 2023] has attracted lots of interest. PQA-Net [Liu *et al.*, 2021b] is the first NR-PCQA method that divides the PCQA problem into two meaningful sub-tasks, including distortion classification and quality prediction. These two sub-tasks cooperate to obtain gratifying results. Chetouani *et al.* [Chetouani *et al.*, 2021] adopt conventional deep learning routes on patch-level distortion characterizations to regress the quality scores of input samples. 3D-NSS [Zhang *et al.*, 2022b] utilizes support vector regression (SVR) to regress the quality-aware feature and obtain the visual quality score, which can address both point cloud and mesh quality assessment problems. Fan *et al.* [Fan *et al.*, 2022] first capture three videos by rotation operations and then conduct the PCQA task based on video features.

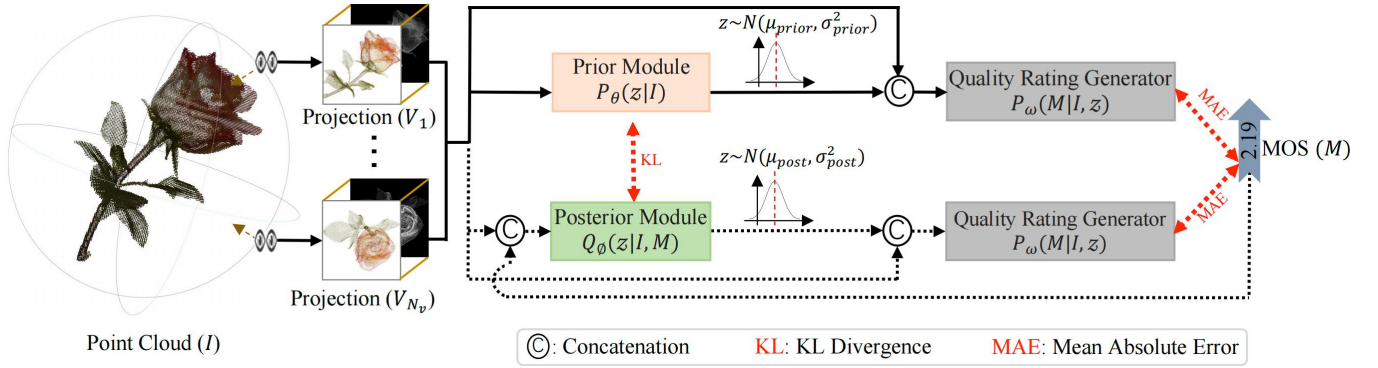


Figure 2: Overall architecture of the proposed method. Our method comprises a Prior Module ( $\text{PM}_{\text{prior}}$ ), a Posterior Module ( $\text{PM}_{\text{post}}$ ), and a Quality Rating Generator (QRG). Note that the two QRGs share network parameters. Solid lines indicate functionality in both the training and testing phases, while dashed lines indicate only working in the training stage.

Liu *et al.* [Liu *et al.*, 2023] present an effective quality metric, termed ResSCNN, which can accurately estimate MOSs of point clouds. Besides, they also construct the largest-scale dataset. To promote the robustness of existing NR-PCQA models, Yang *et al.* [Yang *et al.*, 2022] leverage existing abundant subjective scores of natural images and transfer relevant knowledge to help point cloud quality recognition. All these discussed NR-PCQA methods treat their tasks as a MOS fitting problem and fail to take the stochasticity in the MOS generation process (subjective tests) into account.

## 2.2 Generative Model

Variational autoencoders (VAE) [Kingma and Welling, 2013] and conditional variational autoencoders (CVAE) [Sohn *et al.*, 2015] are types of generative models that combine the principles of autoencoders and variational inference. VAEs extend traditional autoencoders by incorporating probabilistic modeling and typically map the input to a standard Gaussian distribution. The training process of VAEs involves maximizing a lower bound on data log-likelihood while minimizing the Kullback-Leibler (KL) divergence between the prior and posterior distributions of latent features. CVAEs are a further extension of VAEs, which modulate the Gaussian distribution of latent features by integrating conditional information. Due to their outstanding performance, the frameworks of VAEs and CVAEs have been widely used in various computer vision tasks. Liu *et al.* [Liu *et al.*, 2021c] design a novel RefVAE for reference-based image super-resolution, while Esser *et al.* [Esser *et al.*, 2018] apply the VAE to the image generation task. Recently, we also observed that some innovative works [Baumgartner *et al.*, 2019; Zhang *et al.*, 2021] introduce CVAEs into segmentation tasks and achieve visibly improved performance.

In the quality assessment field, generative adversarial networks (GAN) as another type of generative models have been exploited by some works [Lin and Wang, 2018; Ma *et al.*, 2020; Zhu *et al.*, 2021; Ren *et al.*, 2018; Yang *et al.*, 2020c] to assess image quality. Moreover, most of these works employ the framework of GANs to generate pseudo-reference images or for transfer learning purposes. To the best of our knowledge, no work explores the utilization of VAEs or CVAEs in

the context of quality assessment.

## 3 Proposed Method

As shown in Figure 1, our method aims to learn the conditional probability distribution of quality judgments in subjective tests instead of a solitary and potentially biased judgment. With a training dataset  $\mathcal{D} = \{I_i, M_i\}_{i=1}^{N_s}$  comprising  $N_s$  point clouds, we endeavor to estimate the distribution  $P_\omega(M|I, z)$ , where  $I$ ,  $M$ , and  $z$  represent the point cloud sample, corresponding MOS, and a low-dimensional latent variable, respectively. The expression  $P_\omega(M|I, z)$  indicates that the quality judgment from a subject depends on both the deterministic distortions in point cloud  $I$  and the uncertain factors  $z$  influenced by the HVS. Our proposed framework, illustrated in Figure 2, has three modules: a Prior Module ( $\text{PM}_{\text{prior}}$ ), a Posterior Module ( $\text{PM}_{\text{post}}$ ), and a Quality Rating Generator (QRG). During training, the  $\text{PM}_{\text{post}}$  approximates the posterior distribution of the latent variable  $Q_\phi(z|I, M)$  and aids the learning of  $\text{PM}_{\text{prior}}$  regarding the prior distribution  $P_\theta(z|I)$ . In the testing phase, stochastic features are sampled from the prior distribution, which, along with point cloud projections, are input to the QRG for predicting stochastic quality ratings.

As investigated by Liu *et al.* [Liu *et al.*, 2022a], most previous subjective tests for annotating PCQA datasets adopt an interactive 2D monitor as their display tool. To exactly agree with previous subjective tests, our model projects each point cloud into  $N_v$  four-channel RGB-D projections in the input stage, similar to the previous work [Yang *et al.*, 2020a]. We define the  $N_v$  four-channel projections from a point cloud as  $\{V_i\}_{i=1}^{N_v}$ . When computing projections, unlike previous methods [Yang *et al.*, 2020a; Liu *et al.*, 2021b; Fan *et al.*, 2022; Zhang *et al.*, 2023] that employ fixed viewpoints, our model allows for the random selection of viewpoints to imitate the interactive viewing pattern in subjective tests, where any view can be used to judge the quality of a point cloud. This random viewpoint selection strategy not only provides stochastic priors in the stochasticity modeling process, but also enhances the robustness of neural networks, since fixing viewpoint selection can be treated as a special case of our random view-

point selection strategy.

### 3.1 Problem Formulation

CVAEs often incorporate a conditioning variable, an output variable, and a latent variable. In the PCQA task, the conditioning and output variables correspond to the input point cloud  $I$  and the predicted quality rating  $M$ , respectively. The prior distribution of the latent variable  $z$  follows a modulated Gaussian distribution  $P_\theta(z|I)$ , with its parameters conditioned on the input point cloud  $I$ . The quality rating  $M$  is derived from  $P_\omega(M|I, z)$ , leading to the posterior distribution of  $z$  expressed as  $Q_\phi(z|I, M)$ . The primary objective of CVAEs is to approximate the posterior distribution of the latent variable, given the conditioning information. This is achieved by maximizing the evidence lower bound (ELBO). Specifically, the loss function of standard CVAEs consists of a reconstruction loss and a regularizer, which can be represented as:

$$\mathcal{L}_{CVAE} = \mathbb{E}_{z \sim Q_\phi(z|I, M)} [-\log P_\omega(M|I, z)] + D_{KL}(Q_\phi(z|I, M) || P_\theta(z|I)), \quad (1)$$

where the first term denotes the reconstruction loss, which quantifies the negative conditional log-likelihood of the output variable  $M$ , given the conditioning variable  $I$  and the latent variable  $z$  drawn from  $Q_\phi(z|I, M)$ . The second term represents the KL divergence of  $Q_\phi(z|I, M)$  and  $P_\theta(z|I)$ , acting as a regularizer to minimize the distribution discrepancy between the prior and posterior distributions.

However, training CVAEs to generate diverse quality ratings for a point cloud requires annotating training samples with multiple versions of quality judgments, as manifested by the varying judgments obtained from different subjects in subjective tests. Unfortunately, existing popular datasets [Yang *et al.*, 2020a; Liu *et al.*, 2022a; Liu *et al.*, 2021a] only provide a single MOS for each sample without disclosing the raw quality judgments obtained in subjective tests. Consequently, directly employing the conventional CVAE architecture with these datasets may lead to potential posterior collapse, rendering the latent variable invalid and making the training process unstable. To address the issue of posterior collapse, we adopt an effective KL annealing strategy [Sønderby *et al.*, 2016] during training, gradually increasing the weight of the regularizer in Equation (1). This strategy can be described as:

$$\mathcal{L}_{CVAE}^\lambda = \mathbb{E}_{z \sim Q_\phi(z|I, M)} [-\log P_\omega(M|I, z)] + \lambda * D_{KL}(Q_\phi(z|I, M) || P_\theta(z|I)), \quad (2)$$

where  $\lambda$  equals the ratio of the current training epoch to the total training epoch. As depicted in Figure 1, the model with our KL annealing strategy enables diverse quality rating outputs for a point cloud, despite being trained on samples with merely a single annotation.

### 3.2 Prior/Posterior Module

With the adapted objective of CVAEs for PCQA in Equation (2), we utilize two modules, the Prior Module (PM<sub>prior</sub>) and the Posterior Module (PM<sub>post</sub>), to model the prior  $P_\theta(z|I)$  and posterior  $Q_\phi(z|I, M)$  distributions of the latent variable

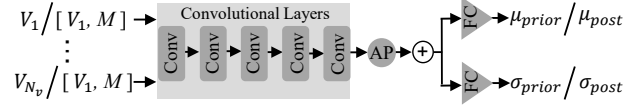


Figure 3: Network structure of the Prior Module and the Posterior Module. “[·]” denotes the concatenated input of the Posterior Module. “Conv” represents the convolutional layer. “ $\oplus$ ” is element-wise summation. “AP” means the global average pooling layer. “FC” stands for the fully connected layer.

$z$ . The two modules share an identical network structure but possess independent parameters,  $\theta$  and  $\phi$ , separately. The PM<sub>prior</sub> is designed to map the four-channel projections of a point cloud  $\{V_i\}_{i=1}^{N_v}$  to a low-dimensional prior statistic. To this end, we first employ five convolutional layers, as illustrated in Figure 3, to extract the individual latent feature from each projection. Subsequently, the latent feature is spatially averaged into a vector representation, and the vector representations of different projections are aggregated via element-wise summation. We map the aggregated vector representation to the Gaussian prior statistic  $P_\theta(z|I)$  using two fully connected layers, which is achieved by predicting the mean vector  $\mu_{prior} \in \mathbb{R}^{K_1}$  and standard deviation vector  $\sigma_{prior} \in \mathbb{R}^{K_1}$  of a Gaussian distribution. Differing from the PM<sub>prior</sub>, the PM<sub>post</sub> takes the five-channel concatenation of each point cloud projection and the spatially-expanded MOS as input, inferring the  $\mu_{post}, \sigma_{post} \in \mathbb{R}^{K_1}$  of the Gaussian posterior distribution  $Q_\phi(z|I, M)$ .

During training, the adapted CVAE maximizes the conditional log-likelihood of quality ratings while alleviating the distribution mismatch measured by  $D_{KL}(Q_\phi(z|I, M) || P_\theta(z|I))$ . As illustrated in Figure 2, we use the reparameterization trick to sample the stochastic feature  $F_s \in \mathbb{R}^{K_1 \times H \times W}$  from the posterior statistic  $Q_\phi(z|I, M)$ , where  $H$  and  $W$  are the height and width of point cloud projections. Specifically, we generate a  $K_1$ -dimensional independent random variable  $\epsilon$ , in which each element is drawn from a standard Gaussian distribution. The reparameterized feature vector  $z^s \in \mathbb{R}^{K_1}$  is then computed as  $z^s = \sigma_{post} * \epsilon + \mu_{post}$ . To obtain the stochastic feature  $F_s$ , we expand the spatial dimensions of  $z^s$  to match the point cloud projections. Ultimately, the stochastic feature  $F_s$  is combined with the projection set  $\{V_i\}_{i=1}^{N_v}$  to produce the quality rating prediction. In the testing phase, the stochastic feature is sampled from the prior statistic  $P_\theta(z|I)$  in the same way to predict a stochastic quality rating.

### 3.3 Quality Rating Generator

The stochastic feature  $F_s$  encodes uncertain variability in subjective tests, while the point cloud projections  $\{V_i\}_{i=1}^{N_v}$  incarnate the deterministic point cloud distortions. We further develop an effective Quality Rating Generator (QRG) to mimic a subject that associates the uncertain and deterministic information to give quality judgments in subjective tests. Our QRG (see Figure 4) is built on the ResNet-50 [He *et al.*, 2016] backbone, whose input is the concatenation of each point cloud projection and the stochastic feature. We



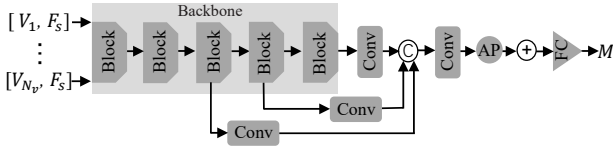


Figure 4: Illustration of our Quality Rating Generator. “Block” represents the convolutional block of the backbone.

use the last three levels of features to compute point cloud quality. Specifically, we first introduce three convolutional layers to reduce the channel size of these three high-level features to  $K_2$  and then fuse them into a multi-scale feature by the concatenation operation followed by convolution. The effectiveness of multi-scale representations has been validated in various computer vision tasks [Gao *et al.*, 2019; Gao *et al.*, 2023]. The multi-scale feature is spatially averaged into a vector representation, and we integrate the vector representations from different projections into the quality-aware feature using element-wise summation. The final stochastic quality rating can be predicted from the quality-aware feature via the fully connected layer. During testing, we can sample the stochastic feature multiple times from the prior statistic and produce diverse stochastic ratings for a point cloud. Similar to the calculation of ground truth MOSs, we treat the averaged score of multiple ratings as the final quality prediction for a point cloud.

### 3.4 Objective Function

Conventional CVAEs often sample the posterior statistic  $Q_\phi(z|I, M)$  for reconstruction during training while relying on the prior statistic  $P_\theta(z|I)$  during testing. However, it has been proved [Sohn *et al.*, 2015] that the distribution mismatch between  $Q_\phi(z|I, M)$  and  $P_\theta(z|I)$  may lead to sub-optimal performance in the testing phase. Although increasing the weight  $\lambda$  of the regularizer in Equation (2) can mitigate the issue of distribution mismatch, this solution cannot bring overall performance improvement due to the relatively insufficient emphasis on result reconstruction. Inspired by GSNN [Sohn *et al.*, 2015], we introduce the “testing branch” (see Figure 2) to reflect the testing circumstances during training. Concretely, we additionally sample the stochastic feature from  $P_\theta(z|I)$  for another reconstruction process in the training stage, which can be formulated as:

$$\mathcal{L}_{GSNN} = \mathbb{E}_{z \sim P_\theta(z|I)} [-\log P_\omega(M|I, z)]. \quad (3)$$

In this approach, we can force the  $PM_{prior}$  to learn an informative prior statistic and alleviate the distribution mismatch of the latent variable during the training and testing stages.

Consequently, the overall objective function of our PCQA-tailored CVAE architecture during training is composed of two components, including the adapted CAVE loss  $\mathcal{L}_{CVAE}^\lambda$  and the GSNN loss  $\mathcal{L}_{GSNN}$ , which can be expressed as:

$$\mathcal{L}_{overall} = \alpha * \mathcal{L}_{CVAE}^\lambda + (1 - \alpha) * \mathcal{L}_{GSNN}, \quad (4)$$

where  $\alpha$  is a weighting factor for balancing the two loss terms. Furthermore, we use the Mean Absolute Error (MAE) to measure the reconstruction loss between the predicted stochastic quality ratings and ground truth MOSs.

## 4 Experiments

### 4.1 Implementation Details

We implement our model on two NVIDIA RTX 3090 Ti GPUs with the PyTorch toolbox and initialize the ResNet-50 [He *et al.*, 2016] backbone in the QRG with parameters pre-trained on ImageNet, while other neural network parameters are randomly initialized. We use the Adam optimizer with an initial learning rate of  $2.5e-5$  and betas set to  $[0.5, 0.999]$ . Our model is trained for a total of 200 epochs, and the learning rate is reduced by a factor of 0.5 when the training process reaches the halfway mark. We set the training batch size to 8 while convincing ablations demonstrate that the weighting term  $\alpha = 0.4$  to emphasize the disparity reduction between the training and testing stages can obtain the best performance. The spatial resolution of point cloud projections is  $480 \times 480$ , and experiments reveal that the projection number  $N_v = 4$  can achieve the best balance between prediction efficiency and accuracy. We take the dimension of the latent variable  $K_1 = 3$  and the channel size of intermediate features  $K_2 = 32$  for a light computation overhead.

### 4.2 Datasets & Evaluation Metrics

**Datasets.** Our experiments utilize three widely-used PCQA datasets, including SJTU-PCQA [Yang *et al.*, 2020a], WPC [Liu *et al.*, 2022a], and WPC2.0 [Liu *et al.*, 2021a]. We strictly follow the operations of previous works [Fan *et al.*, 2022; Zhang *et al.*, 2023] and employ the k-fold cross-validation strategy to evaluate our method and other existing competitors. Specifically, we employ 9-fold, 5-fold, and 4-fold cross-validation for SJTU-PCQA, WPC, and WPC2.0, respectively, resulting in an approximate 8:2 split between training and testing sets. The final reported performance on a dataset is the average results across all folds. It is worth noting that all compared counterparts comply with the same dataset split to avoid any unfair comparisons.

**Evaluation Metrics.** We leverage four widely adopted evaluation metrics to quantify the level of agreement between predicted quality scores and ground truth MOSs. These metrics encompass the Spearman Rank Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC), Kendall’s Rank Correlation Coefficient (KRCC), and Root Mean Squared Error (RMSE). Since there can be value range misalignment between predicted results and ground truth MOSs, we follow the previous method [Zhang *et al.*, 2023] and introduce a common four-parameter logistic function to align their range. By employing the aforementioned four metrics, we can obtain convincing benchmarking results.

### 4.3 Comparisons with State-of-the-art Methods

We compare our method with 15 representative methods, including 8 FR-PCQA methods (*i.e.*, MSE-p2po [Mekuria *et al.*, 2016], HD-p2po [Mekuria *et al.*, 2016], MSE-p2pl [Tian *et al.*, 2017], HD-p2pl [Tian *et al.*, 2017], PSNR-yuv [Torlig *et al.*, 2018], PCQM [Meynet *et al.*, 2020], GraphSIM [Yang *et al.*, 2020b], and PointSSIM [Alexiou and Ebrahimi, 2020]) and 7 NR-PCQA methods (*i.e.*, BRISQUE [Mittal *et al.*, 2012a], NIQE [Mittal *et al.*, 2012b], IL-NIQE [Zhang *et al.*, 2015], ResSCNN [Liu *et al.*, 2023], PQA-Net [Liu

Method		SJTU-PCQA [Yang <i>et al.</i> , 2020a]				WPC [Liu <i>et al.</i> , 2022a]				WPC2.0 [Liu <i>et al.</i> , 2021a]			
		SRCC↑	PLCC↑	KRCC↑	RMSE↓	SRCC↑	PLCC↑	KRCC↑	RMSE↓	SRCC↑	PLCC↑	KRCC↑	RMSE↓
FR-PCQA	MSE-p2po	0.7294	0.8123	0.5617	0.1361	0.4558	0.4852	0.3182	0.1989	0.4315	0.4626	0.3082	0.1916
	HD-p2po	0.7157	0.7753	0.5447	0.1448	0.2786	0.3972	0.1943	0.2090	0.3587	0.4561	0.2641	0.1890
	MSE-p2pl	0.6277	0.5940	0.4825	0.2282	0.3281	0.2695	0.2249	0.2282	0.4136	0.4104	0.2965	0.2104
	HD-p2pl	0.6441	0.6874	0.4565	0.2126	0.2827	0.2753	0.1696	0.2199	0.4074	0.4402	0.3174	0.1952
	PSNR-yuv	0.7950	0.8170	0.6196	0.1315	0.4493	0.5304	0.3198	0.1931	0.3732	0.3557	0.2277	0.2015
	PCQM	0.8644	0.8853	0.7086	0.1086	0.7434	0.7499	0.5601	0.1516	0.6825	0.6923	0.4929	0.1563
	GraphSIM	0.8783	0.8449	0.6947	0.1032	0.5831	0.6163	0.4194	0.1719	0.7405	0.7512	0.5533	0.1499
NR-PCQA	PointSSIM	0.6867	0.7136	0.4964	0.1700	0.4542	0.4667	0.3278	0.2027	0.4810	0.4705	0.2978	0.1939
	BRISQUE	0.3975	0.4214	0.2966	0.2094	0.2614	0.3155	0.2088	0.2117	0.0820	0.3353	0.0487	0.2167
	NIQE	0.1379	0.2420	0.1009	0.2262	0.1136	0.2225	0.0953	0.2314	0.1865	0.2925	0.1335	0.2251
	IL-NIQE	0.0837	0.1603	0.0594	0.2338	0.0913	0.1422	0.0853	0.2401	0.0911	0.1233	0.0714	0.2400
	ResSCNN	0.8600	0.8100	-	-	-	-	-	-	0.7500	0.7200	-	-
	PQA-Net	0.8372	0.8586	0.6304	0.1072	0.7026	0.7122	0.4939	0.1508	0.6191	0.6426	0.4606	0.1698
	3D-NSS	0.7144	0.7382	0.5174	0.1769	0.6479	0.6514	0.4417	0.1657	0.5077	0.5699	0.3638	0.1772
	MM-PCQA	0.9103	0.9226	0.7838	0.0772	0.8414	0.8556	0.6513	0.1235	0.8023	0.8024	0.6202	0.1343
	Ours	0.9474	0.9636	0.8192	0.0628	0.8744	0.8766	0.6901	0.1089	0.8742	0.8880	0.6922	0.0979

Table 1: Benchmarking results of state-of-the-art methods on the SJTU-PCQA, WPC, and WPC2.0 datasets. “↑”/“↓” indicates that larger/smaller is better. The best method is highlighted in **bold**.

Method	WPC→ SJTU		WPC→ WPC2.0	
	SRCC↑	PLCC↑	SRCC↑	PLCC↑
PQA-Net	0.5411	0.6102	0.6006	0.6377
3D-NSS	0.1817	0.2344	0.4933	0.5613
MM-PCQA	0.7693	0.7779	0.7607	0.7753
<b>Ours</b>	<b>0.9133</b>	<b>0.9404</b>	<b>0.8300</b>	<b>0.8511</b>

Table 2: Result comparisons of cross-dataset generalization.

*et al.*, 2021b], 3D-NSS [Zhang *et al.*, 2022a], and MM-PCQA [Zhang *et al.*, 2023]). The performance comparisons of different methods are illustrated in Table 1. We can observe that: 1) Our method obtains the best performance among all NR-PCQA methods on all datasets, even surpassing the newly-proposed MM-PCQA [Zhang *et al.*, 2023] by a significant margin. For example, with the same experimental settings, our method outperforms the second-best MM-PCQA by 7.19% in terms of SRCC on the challenging WPC2.0 dataset. 2) Our method outperforms all FR-PCQA methods despite the additional reference information utilized by these methods. For instance, compared with the representative GraphSIM, our method shows 10.67% SRCC improvement on the WPC2.0 dataset. 3) Our method exhibits robust performance (fewer performance fluctuations) on both the relatively easy SJTU-PCQA dataset and the challenging WPC2.0 dataset.

#### 4.4 Generalization Analyses

Due to the data distribution variations and annotation biases in existing datasets, previous PCQA methods consistently suffer from poor cross-dataset generalization capabilities, which has been a persistent challenge for the practical application of existing PCQA algorithms. Our proposed method estimates point cloud quality by learning the conditional probability distribution of quality judgments in subjec-

tive tests rather than a solitary and potentially biased judgment, thus expected to have an excellent generalization capability. To compare the generalization performance of different methods, we select the relatively larger WPC (740 samples) as the training dataset and evaluate the models on the other two datasets, *i.e.*, SJTU-PCQA (378 samples) and WPC2.0 (400 samples). Besides, since there exist reference overlaps between the WPC and WPC2.0 datasets, we remove the samples in WPC with overlapped references when testing on WPC2.0, to ensure a considerable cross-dataset generalization difficulty.

The experimental results of cross-dataset generalization are listed in Table 2, from which we can learn that the proposed method obtains the best generalization performance and exceeds other counterparts by a noticeable margin. In concrete terms, our method surpasses the existing best method MM-PCQA [Zhang *et al.*, 2023] by around 14.4% and 6.9% in terms of SRCC on the SJTU-PCQA and WPC2.0 datasets, respectively. Moreover, the compelling performance of our method even exceeds all existing approaches directly trained on the SJTU-PCQA and WPC2.0 datasets in Table 1. Hence, this work offers a promising approach to addressing the enduring PCQA challenge on algorithm generalizability.

#### 4.5 Ablation Studies

To study the effectiveness of our designs, we tune our model on the WPC2.0 dataset and analyze the performance changes. **Stochastic versus Deterministic.** To highlight the superiority of our probabilistic architecture over previous deterministic mapping-based approaches, we build a deterministic model that solely comprises the QRG and takes point cloud projections without stochastic features as input. As shown in Table 3, the performance of the deterministic model, denoted as “w/o Stochastic,” exhibits noticeably inferior performance compared to our advanced probabilistic architecture. We attribute to the cause that traditional deterministic meth-

Method	SRCC $\uparrow$	PLCC $\uparrow$	KRCC $\uparrow$	RMSE $\downarrow$
MM-PCQA	0.8023	0.8024	0.6202	0.1343
<b>Ours</b>	<b>0.8742</b>	<b>0.8880</b>	<b>0.6922</b>	<b>0.0979</b>
w/o Stochastic	0.7756	0.7871	0.5892	0.1341
w/o KL Annealing	$\times$	$\times$	$\times$	$\times$
w/o $\mathcal{L}_{GSNN}$	0.8478	0.8643	0.6531	0.1092
only $\mathcal{L}_{GSNN}$	0.7229	0.7583	0.5356	0.1410
w/o Depth	0.8476	0.8589	0.6603	0.1083
Fixed Viewpoint	$\times$	$\times$	$\times$	$\times$
Early Average	0.8606	0.8735	0.6743	0.1028

Table 3: Ablations of our method on WPC2.0. “ $\times$ ” indicates an untrainable model due to gradient explosion.

Projection Number	SRCC $\uparrow$	PLCC $\uparrow$	KRCC $\uparrow$	RMSE $\downarrow$
2	0.8650	0.8837	0.6754	0.1002
<b>4</b>	<b>0.8742</b>	<b>0.8880</b>	<b>0.6922</b>	<b>0.0979</b>
6	0.8498	0.8696	0.6630	0.1039

Table 4: Ablation studies on the projection number.

ods overlook the uncertain factors in subjective tests and are susceptible to biased ground truth annotations.

**KL Annealing.** Training a CVAE capable of generating diverse outputs necessitates training samples with multiple annotation versions. We adopt the KL annealing strategy to overcome the absence of diverse annotations from existing datasets. To prove the validity of our scheme, we further conduct experiments without the KL annealing strategy and find this approach makes the model become untrainable, indicating the necessity of the KL annealing strategy.

**GSNN Loss.** To showcase the efficacy of our GSNN loss, we explore two additional variants. The first variant merely encompasses the  $\mathcal{L}_{CVAE}^{\lambda}$  term in Equation (4), referred to as “w/o  $\mathcal{L}_{GSNN}$ ”. The second variant exclusively incorporates the  $\mathcal{L}_{GSNN}$  term, denoted as “only  $\mathcal{L}_{GSNN}$ ”. Experiments in Table 3 show that both variants yield suboptimal performance. The first variant fails to mitigate the distribution mismatch during training and testing, while the second variant invalids the learning of the posterior distribution.

**Depth Information.** Previous projection-based NR-PCQA methods [Liu *et al.*, 2021b; Fan *et al.*, 2022; Zhang *et al.*, 2023] only utilize RGB colors in the point cloud projections, which cannot capture the spatial geometry distortions. To verify the benefits of additional depth information in our projections, we conduct a comparative experiment denoted as “w/o depth,” where the input projections only contain RGB colors. As shown in Table 3, the absence of depth information indeed degrades the performance of models.

**Random Viewpoint Selection.** To simulate the interactive viewing in subjective tests, we introduce random viewpoint selection when computing point cloud projections. To investigate the role of our random viewpoint selection strategy, we perform ablation experiments with the fixed viewpoint strategy adopted by the previous work [Zhang *et al.*, 2023]. As demonstrated in Table 3 “Fixed Viewpoint”, using fixed viewpoints leads to an untrainable model, revealing the im-

Value of $\alpha$	SRCC $\uparrow$	PLCC $\uparrow$	KRCC $\uparrow$	RMSE $\downarrow$
0.2	0.8686	0.8791	0.6832	0.0996
<b>0.4</b>	<b>0.8742</b>	<b>0.8880</b>	<b>0.6922</b>	<b>0.0979</b>
0.6	0.8726	0.8805	0.6892	0.0995

Table 5: Ablation experiments of different  $\alpha$  settings.

portance of our strategy in introducing stochastic priors and stabilizing the training process.

**Late Average versus Early Average.** During testing, our approach, marked as “Late Average,” samples the prior statistic multiple times to obtain stochastic features, each leading to a quality rating. We average all ratings of a sample to calculate the final quality prediction similar to the computation of MOSs. An alternative “Early Average” manner involves averaging the stochastic features first and then computing the final quality prediction on the averaged stochastic feature. As demonstrated in Table 3, our “Late Average” scheme reflecting practical subjective tests can obtain better performance.

**Projection Number.** To investigate the influence of the number of projections, as shown in Table 4, we conduct experiments with varying projection numbers, showing that the projection number  $N_v = 4$  can achieve the best performance. Too few or many projections may cause insufficient or redundant information, eventually hindering the performance promotion.

**Value of Hyperparameter  $\alpha$ .** Though both the adapted CVAE loss  $\mathcal{L}_{CVAE}^{\lambda}$  and the GSNN loss  $\mathcal{L}_{GSNN}$  are indispensable for our model to achieve its optimal performance, unbalanced ratio of these two terms in the objective function may discourage the performance. To find the most proper value of the weighting factor  $\alpha$ , we conduct ablation experiments on different values of  $\alpha$ . As shown in Table 5,  $\alpha = 0.4$  can achieve the best balance between  $\mathcal{L}_{CVAE}^{\lambda}$  and  $\mathcal{L}_{GSNN}$ .

## 5 Conclusion

This work presents the first probe of the stochasticity in dataset labeling for PCQA. To mimic the generation of ground truth MOSs, we propose a novel probabilistic architecture to model the conditional probability distribution of quality judgments in subjective tests. Specifically, our method utilizes a PCQA-tailored conditional variational autoencoder (CVAE) architecture to capture the uncertain variability in subjective tests and the deterministic distortions in point cloud projections, which is then integrated into diverse stochastic quality ratings for a point cloud. The stochastic ratings representing the labeling variants in subjective tests are finally averaged into an accurate quality prediction. Extensive experiments show that our method significantly outperforms previous approaches.

## Acknowledgments

This work was supported by The Major Key Project of PCL (PCL2024A02), Natural Science Foundation of China (62271013, 62031013), Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology (2024B1212010006), Guangdong Province Pearl

River Talent Program (2021QN020708), Shenzhen Science and Technology Program (JCYJ20240813160202004, JCYJ20230807120808017), Guangdong Basic and Applied Basic Research Foundation (2024A1515010155).

## References

- [Alexiou and Ebrahimi, 2018] Evangelos Alexiou and Touradj Ebrahimi. Point cloud quality assessment metric based on angular similarity. In *IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2018.
- [Alexiou and Ebrahimi, 2020] Evangelos Alexiou and Touradj Ebrahimi. Towards a point cloud structural similarity metric. In *IEEE International Conference on Multimedia & Expo Workshops*, pages 1–6. IEEE, 2020.
- [Baumgartner et al., 2019] Christian F Baumgartner, Kerem C Tezcan, Krishna Chaitanya, Andreas M Hötker, Urs J Muehlemaier, Khoschy Schawkat, Anton S Becker, Olivio Donati, and Ender Konukoglu. Phiseg: Capturing uncertainty in medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention*. Springer, 2019.
- [Chetouani et al., 2021] Aladine Chetouani, Maurice Quach, Giuseppe Valenzise, and Frédéric Dufaux. Deep learning-based quality assessment of 3d point clouds without reference. In *IEEE International Conference on Multimedia & Expo Workshops*, pages 1–6. IEEE, 2021.
- [Esser et al., 2018] Patrick Esser, Ekaterina Sutter, and Björn Ommer. A variational u-net for conditional appearance and shape generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8857–8866, 2018.
- [Fan and Gao, 2023] Songlin Fan and Wei Gao. Screen-based 3d subjective experiment software. In *ACM International Conference on Multimedia*, pages 9672–9675, 2023.
- [Fan et al., 2022] Yu Fan, Zicheng Zhang, Wei Sun, Xiongkuo Min, Ning Liu, Quan Zhou, Jun He, Qiyan Wang, and Guangtao Zhai. A no-reference quality assessment metric for point cloud based on captured video sequences. In *IEEE International Workshop on Multimedia Signal Processing*, pages 1–5. IEEE, 2022.
- [Fan et al., 2024] Songlin Fan, Wei Gao, and Ge Li. Point-mpp: Point cloud self-supervised learning from masked position prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
- [Gao et al., 2019] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [Gao et al., 2023] Wei Gao, Songlin Fan, Ge Li, and Weisi Lin. A thorough benchmark and a new model for light field saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [He et al., 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [Javaheri et al., 2020a] Alireza Javaheri, Catarina Brites, Fernando Pereira, and João Ascenso. A generalized hausdorff distance based quality metric for point cloud geometry. In *International Conference on Quality of Multimedia Experience*, pages 1–6. IEEE, 2020.
- [Javaheri et al., 2020b] Alireza Javaheri, Catarina Brites, Fernando Pereira, and Joao Ascenso. Point cloud rendering after coding: Impacts on subjective and objective quality. *IEEE Transactions on Multimedia*, 23:4049–4064, 2020.
- [Kingma and Welling, 2013] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [Lin and Wang, 2018] Kwan-Yee Lin and Guanxiang Wang. Hallucinated-iqa: No-reference image quality assessment via adversarial learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 732–741, 2018.
- [Liu et al., 2021a] Qi Liu, Hui Yuan, Raouf Hamzaoui, Honglei Su, Junhui Hou, and Huan Yang. Reduced reference perceptual quality model with application to rate control for video-based point cloud compression. *IEEE Transactions on Image Processing*, 30:6623–6636, 2021.
- [Liu et al., 2021b] Qi Liu, Hui Yuan, Honglei Su, Hao Liu, Yu Wang, Huan Yang, and Junhui Hou. Pqa-net: Deep no reference point cloud quality assessment via multi-view projection. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(12):4645–4660, 2021.
- [Liu et al., 2021c] Zhi-Song Liu, Wan-Chi Siu, and Li-Wen Wang. Variational autoencoder for reference based image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 516–525, 2021.
- [Liu et al., 2022a] Qi Liu, Honglei Su, Zhengfang Duanmu, Wentao Liu, and Zhou Wang. Perceptual quality assessment of colored 3d point clouds. *IEEE Transactions on Visualization and Computer Graphics*, 2022.
- [Liu et al., 2022b] Yipeng Liu, Qi Yang, and Yiling Xu. Reduced reference quality assessment for point cloud compression. In *IEEE International Conference on Visual Communications and Image Processing*, pages 1–5. IEEE, 2022.
- [Liu et al., 2023] Yipeng Liu, Qi Yang, Yiling Xu, and Le Yang. Point cloud quality assessment: Dataset construction and learning-based no-reference metric. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2s):1–26, 2023.
- [Ma et al., 2020] Jupo Ma, Jinjian Wu, Leida Li, Weisheng Dong, and Xuemei Xie. Active inference of gan for no-reference image quality assessment. In *IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2020.



- [Mekuria *et al.*, 2016] Rufael Mekuria, Zhu Li, Christian Tulvan, and Phil Chou. Evaluation criteria for point cloud compression. *ISO/IEC MPEG*, 2016.
- [Meynet *et al.*, 2019] Gabriel Meynet, Julie Digne, and Guillaume Lavoué. Pc-msdm: A quality metric for 3d point clouds. In *International Conference on Quality of Multimedia Experience*, pages 1–3. IEEE, 2019.
- [Meynet *et al.*, 2020] Gabriel Meynet, Yana Nehmé, Julie Digne, and Guillaume Lavoué. Pcpqm: A full-reference quality metric for colored 3d point clouds. In *International Conference on Quality of Multimedia Experience*, pages 1–6. IEEE, 2020.
- [Mittal *et al.*, 2012a] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.
- [Mittal *et al.*, 2012b] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012.
- [Qi *et al.*, 2017] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017.
- [Ren *et al.*, 2018] Hongyu Ren, Diqi Chen, and Yizhou Wang. Ran4iqa: Restorative adversarial nets for no-reference image quality assessment. In *AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [Sohn *et al.*, 2015] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. *Advances in Neural Information Processing Systems*, 28, 2015.
- [Sønderby *et al.*, 2016] Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. Ladder variational autoencoders. *Advances in Neural Information Processing Systems*, 29, 2016.
- [Tian *et al.*, 2017] Dong Tian, Hideaki Ochimizu, Chen Feng, Robert Cohen, and Anthony Vetro. Geometric distortion metrics for point cloud compression. In *IEEE International Conference on Image Processing*, pages 3460–3464. IEEE, 2017.
- [Torlig *et al.*, 2018] Eric M Torlig, Evangelos Alexiou, Tiago A Fonseca, Ricardo L de Queiroz, and Touradj Ebrahimi. A novel methodology for quality assessment of voxelized point clouds. In *Applications of Digital Image Processing XLI*, volume 10752, pages 174–190. SPIE, 2018.
- [Wang *et al.*, 2023] Jilong Wang, Wei Gao, and Ge Li. Applying collaborative adversarial learning to blind point cloud quality measurement. *IEEE Transactions on Instrumentation and Measurement*, 2023.
- [Wu *et al.*, 2021] Xinju Wu, Yun Zhang, Chunling Fan, Junhui Hou, and Sam Kwong. Subjective quality database and objective study of compressed point clouds with 6dof head-mounted display. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(12):4630–4644, 2021.
- [Yang *et al.*, 2020a] Qi Yang, Hao Chen, Zhan Ma, Yiling Xu, Rongjun Tang, and Jun Sun. Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration. *IEEE Transactions on Multimedia*, 23:3877–3891, 2020.
- [Yang *et al.*, 2020b] Qi Yang, Zhan Ma, Yiling Xu, Zhu Li, and Jun Sun. Inferring point cloud quality via graph similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3015–3029, 2020.
- [Yang *et al.*, 2020c] Xiaohan Yang, Fan Li, and Hantao Liu. Ttl-iqa: Transitive transfer learning based no-reference image quality assessment. *IEEE Transactions on Multimedia*, 23:4326–4340, 2020.
- [Yang *et al.*, 2021] Qi Yang, Siheng Chen, Yiling Xu, Jun Sun, M Salman Asif, and Zhan Ma. Point cloud distortion quantification based on potential energy for human and machine perception. *arXiv e-prints*, pages arXiv–2103, 2021.
- [Yang *et al.*, 2022] Qi Yang, Yipeng Liu, Siheng Chen, Yiling Xu, and Jun Sun. No-reference point cloud quality assessment via domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21179–21188, 2022.
- [Zhang *et al.*, 2015] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015.
- [Zhang *et al.*, 2021] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Saeed Anwar, Fatemeh Saleh, Sadegh Aliakbarian, and Nick Barnes. Uncertainty inspired rgb-d saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5761–5779, 2021.
- [Zhang *et al.*, 2022a] Zicheng Zhang, Wei Sun, Xiongkuo Min, Tao Wang, Wei Lu, and Guangtao Zhai. No-reference quality assessment for 3d colored point cloud and mesh models. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11):7618–7631, 2022.
- [Zhang *et al.*, 2022b] Zicheng Zhang, Wei Sun, Xiongkuo Min, Tao Wang, Wei Lu, and Guangtao Zhai. No-reference quality assessment for 3d colored point cloud and mesh models. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11):7618–7631, 2022.
- [Zhang *et al.*, 2023] Zicheng Zhang, Wei Sun, Xiongkuo Min, Quan Zhou, Jun He, Qiyuan Wang, and Guangtao Zhai. Mm-pcqa: Multi-modal learning for no-reference point cloud quality assessment. *International Joint Conference on Artificial Intelligence*, 2023.
- [Zhu *et al.*, 2021] Yunan Zhu, Haichuan Ma, Jialun Peng, Dong Liu, and Zhiwei Xiong. Recycling discriminator: Towards opinion-unaware image quality assessment using wasserstein gan. In *ACM International Conference on Multimedia*, pages 116–125, 2021.