DSP-PCQA: Integrating Multiple Perception Preferences for Point Cloud Quality Assessment

Mingxuan Li, Fazhan Zhang, Zhenzhe Hou, Zihao Huang, Bohan Fu, Runze Hu, Xiaohui Chu*

School of Information and Electronics, Beijing Institute of Technology, Beijing, 100081, China
{limx1630, zfz63622, huangzihhhh, fubohan809, hrzlpk2015}@gmail.com, hzzrua@126.com, 3120225380@bit.edu.cn

Abstract

Point Cloud Quality Assessment (PCQA) faces a critical disconnect: existing methods operate on a flawed single-perception paradigm, while human observers evaluate quality through dual cognitive streams: **technical rationality** and **semantic sensibility**. This fundamental mismatch routinely produces assessment failures in real-world scenarios where technical and semantic signals conflict. To address this, we introduce Dual-Stream Perception PCQA (DSP-PCQA), the first framework that explicitly models this perceptual duality through parallel networks thoroughly mirroring the human cognitive pathway. DSP-PCQA introduces three key innovations: (1) a **Decoupled Focus Enhancer (DFE)** that surgically isolates technical and semantic information using two targeted transformations; (2) a **Context & Attribute Correlation Awareness (CACA)** module that captures the dynamic, non-linear relationships between different views and sub-models characteristic of human visual processing; and (3) an **Exchange-based Perceptual Injection (EPI)** module that strategically transfers information between perception streams, simulating how humans integrate multiple perceptual dimensions. Extensive evaluations show DSP-PCQA outperforms state-of-the-art methods across multiple benchmarks. Most importantly, our method resolves the perceptual discord that plagues existing approaches, maintaining high accuracy even in the challenging boundary cases where technical quality and semantic significance diverge, precisely where conventional methods often struggle.

Code — https://github.com/Limx1630/DSP-PCQA

Introduction

Point cloud technology has emerged as a cornerstone of 3D representation across virtual reality, autonomous driving, and robotic navigation (Guo et al. 2021; Weingarten, Gruener, and Siegwart 2004; Cui et al. 2022). However, a glaring deficiency plagues current PCQA approaches: the fundamental mismatch between how algorithms evaluate quality and how humans actually perceive it.

The crisis in PCQA stems from a profound misunderstanding of human perception (Li, Jiang, and Jiang 2019; Sheng et al. 2018). When evaluating point clouds, humans don't rely on a single assessment pathway—they simultaneously engage two distinct perceptual systems, a phenomenon well-established in biology (Van Essen and Maunsell 1983) and psychology (Bellini Leite 2022):

- **Technical Rationality**: An analytical evaluation that meticulously examines distortion artifacts, geometric precision, and color fidelity, representing our capacity to detect technical imperfections.
- **Semantic Sensibility**: An interpretive evaluation concerned with recognizing objects, understanding content, and appreciating aesthetic composition, reflecting our innate drive to extract meaning.

This perceptual duality creates assessment paradoxes that utterly confound current PCQA methods. A technically pristine point cloud of a meaningless object might receive perfect scores from algorithms while humans find it qualitatively poor. Conversely, a technically imperfect but semantically rich point cloud may be heavily penalized by algorithms despite high human appreciation. Fig.1 illustrates this disconnect: when technical and semantic signals align, assessment

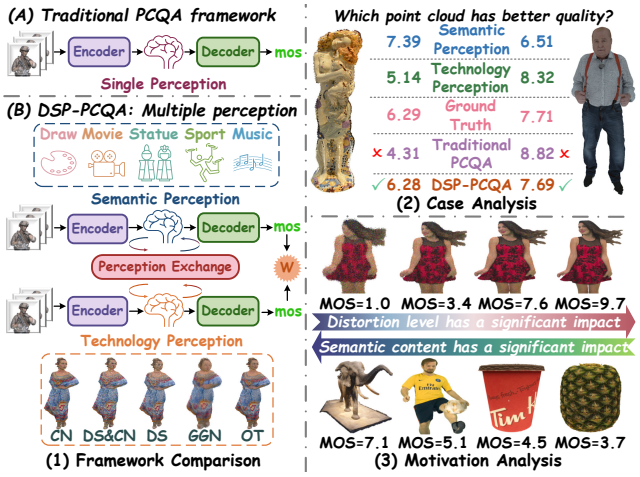

Figure 1: (1) Framework comparison between traditional PCQA method and DSP-PCQA. (2) Typical cases analysis in traditional framework and DSP-PCQA. (3) The source of the design motivation for DSP-PCQA.

*Corresponding author.

is straightforward, but when they conflict, current methods produce misaligned predictions. Moreover, it is known that different groups have different perceptual preferences: observers with technical rational inclination are more preoccupied with distortion detail, while users with semantic sensible mindset give precedence to content semantics and aesthetic value. This variation validates the practicality of considering multiple perception preferences in PCQA.

Existing methods to PCQA share this critical blind spot. By processing point clouds through a single perceptual lens, they fundamentally contradict the dual-stream nature of human assessment, creating an unbridgeable gap between algorithmic predictions and human judgments. The consequences of this mismatch extend beyond academic concern. In applications like autonomous driving, VR/AR experiences, and medical visualization, incorrect quality assessment can lead to serious performance degradation, compromised user experience (Liu et al. 2021a), or even safety risks when systems misallocate resources based on flawed quality metrics (Hu et al. 2025).

While recent research has begun exploring dual-perception approaches in 2D media(Chen et al. 2024a; Li et al. 2019), point cloud evaluation remains trapped in an outdated single-perception paradigm. This critical gap demands a fundamental reimagining of how we approach PCQA, aligning with actual human cognitive processes rather than continuing to force human perception into algorithmic convenience.

To address this crisis, we propose **Dual-Stream Perception Point Cloud Quality Assessment (DSP-PCQA)**, the first framework to model the technical-semantic duality in human point cloud perception. Our approach processes point clouds through parallel but interconnected streams, each specialized to extract information from a distinct perceptual dimension, and then integrates these perspectives in a manner that mirrors human cognitive processes.

The key innovations of DSP-PCQA include: (1) A **Decoupled Focus Enhancer (DFE)** that surgically separates technical and semantic information using targeted transformations: coordinate and grid confusion that preserves local distortions while disrupting global semantic structure for technical focus, and strategic application of random distortion that maintains semantic content while reducing distortion sensitivity for semantic focus. (2) A **Context & Attribute Correlation Awareness (CACA)** module that models the complex, non-linear dynamics of human vision by establishing comprehensive relationships between different views and sub-models using attention mechanisms that mirror the context and attribute integration occurring in human visual processing. (3) An **Exchange-based Perceptual Injection (EPI)** module that facilitates controlled information flow between perception streams by identifying quality-insensitive tokens and replacing them with complementary information, simulating the unconscious integration of technical and semantic assessments in human judgment formation.

Through extensive experiments on multiple benchmarks, we demonstrate that DSP-PCQA dramatically outperforms state-of-the-art methods. Most importantly, our approach maintains high accuracy even in challenging boundary cases where technical and semantic signals conflict, precisely where conventional methods fail catastrophically.

Related Work

No-Reference Point Cloud Quality Assessment

No-Reference PCQA (NR-PCQA) aims to evaluate the perceptual quality of distorted point clouds without available references. These methods can be categorized into three groups based on input data: projection-based, point-based, and multimodal methods. Xie et al. (Xie et al. 2023) projected point clouds onto diverse 2D maps to enhance representation. Liu et al. (Liu et al. 2023b) pioneered the first point-based methods that use an end-to-end voxel-based sparse CNN for raw 3D points without any 2D transformation. Beyond these, Zhang et al. proposed MM-PCQA(Zhang et al. 2023), a multimodal PCQA method that leverages multimodal fusion between multi-view images and point clouds.

While existing models (Zhou et al. 2024; Wang, Gao, and Li 2024; Tliba et al. 2023) demonstrate promising performance, they typically rely on single-dimensional perceptual for direct quality prediction, which overlooks the inherent complexity of human visual perception in PCQA.

Quality Assessment with Multiple Perception

Human quality judgments are influenced by multiple interacting perceptual factors(Wang et al. 2021). Recent studies leverage this by employing multi-network architectures to extract diverse perceptual features, achieving superior performance over single-network models. Min et al. (Min et al. 2017) first proposed a unified content-type-adaptive blind image quality assessment model suitable for various content types by utilizing a dual perception architecture. Yang et al. (Yang et al. 2024) proposed a novel mixture-of-experts approach for enhancing visual perception-driven and semantic-aware AIGC-image quality assessment. Xiang et al. (Xiang et al. 2024) proposed a semantic-aware and quality-aware interaction network for blind video quality assessment.

However, these advances remain confined to 2D media, e.g., images or videos(Chen et al. 2024a; Li et al. 2019; Liao et al. 2022). A significant research gap persists in integrating multi-perception cues for PCQA.

Method

In our study, we introduce the **DSP-PCQA** framework to model perceptual duality through parallel networks that mirror human cognitive pathways in PCQA. The detailed architecture is illustrated in Fig.2. We employ a specialized two-stage learning strategy to achieve our objectives. Before pretraining, we design a **Decoupled Focus Enhancer** to generate technical focus and semantic focus. These focuses and the original data are respectively used for pretraining process, enabling their corresponding branches to focus on extracting domain-specific information. Moreover, we present the pipeline for single perceptual information extraction, which includes a **Context & Attribute Correlation Awareness** module to construct context and attribute associations and introduce the procedure of pretraining using the technical and semantic focuses. Finally, we propose an **Exchange-based Perceptual Injection** module to facilitate a controlled flow of information between perception streams during the fine-tuning stage.

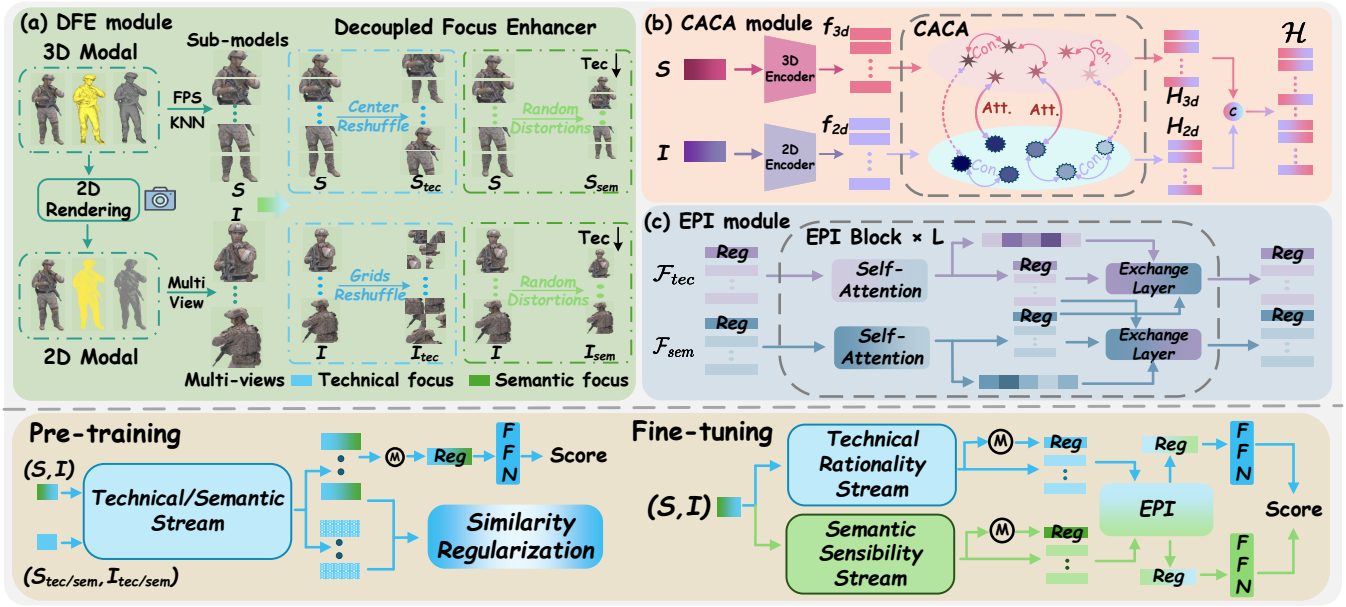

Figure 2: The proposed two-stage framework utilizes DFE, CACA, and EPI modules: 1)Pre-training: The DFE module generates Technical Focus and Semantic Focus data. This data, combined with Original data, pre-trains dual perception streams. Each stream incorporates a CACA module. 2)Fine-tuning: Original data is processed separately through each perception stream. The EPI module integrates the resulting perceptual information, enabling unconscious perceptual interaction.

Preliminaries

A point cloud with N points can be represented as $\mathcal{P}^{xyz} = \{p_i^{xyz}\}_{i=1}^{N}$, where p_i^{xyz} denotes the spatial coordinates. Following (Wang et al. 2023), we add additional attributes to enhance point cloud representation, obtaining color point cloud $\mathcal{P}^{rgb} = \{(p_i^{xyz}, p_i^{rgb})\}_{i=1}^{N}$ and normal point cloud $\mathcal{P}^{nor} = \{(p_i^{xyz}, p_i^{nor})\}_{i=1}^{N}$, where p_i^{rgb} denotes the color values and p_i^{nor} denotes the normal attribute. To simplify subsequent formulas, we define A as the 3D attribute set $\{xyz, rgb, nor\}$. To address the challenge of feature extraction from dense point clouds $\mathcal{P}^{\alpha}$, where $\alpha \in A$, we decompose it into L sub-models $S^{\alpha} = \{s_l^{\alpha} | l \in [1, 2, \cdots, L]\}$ using Farthest Point Sampling (FPS) and K-Nearest Neighbors (KNN) (Chen et al. 2024b). Each sub-model $s_l^{\alpha} \in \mathbb{R}^{K \times C}$, where K denotes the point count per sub-model and C denotes the feature dimension. Finally, we obtain the multi-attribute 3D modal data $\mathcal{S} = \{S^{\alpha} | \alpha \in A\}$.

Integrating multi-modal information can enhance feature representation and improve PCQA performance (Wang et al. 2024). Inspired by this, we project the point cloud into multi-view color maps, depth maps, and normal maps, denoted as $\mathcal{I} = \{I^{\beta} | \beta \in B\}$, where $B = \{rgb, dep, nor\}$ is the 2D attribute set. Each $I^{\beta} = \{i_m^{\beta} | m \in [1, 2, \cdots, M]\}$ represents M projected images with size $H \times W$.

Decoupled Focus Enhancer

To enable each stream to focus on extracting the corresponding perceptual features, we use both the original data and the focus data generated by our Decoupled Focus Enhancer (DFE) module in our pretraining process. In this section, we first introduce how the DFE module separates the original data $(\mathcal{S}, \mathcal{I})$ into two distinct focuses: the technical focus and the semantic focus.

Technical focus generation To ensure the raw data retains technical distortions while deliberately eliminating semantic information, we adopt a coordinate and patch confusion method(Wu et al. 2022). Specifically, given L 3D sub-models $\mathcal{S}$, we first calculate the center positions $C = \{c_1, c_2, \cdots, c_L\}$ of sub-model through spatial coordinates. Next, we define a bijection mapping function $\lambda(.)$ to perform a random permutation of original center coordinate indices. Specifically, given a set of integers from 1 to L, $\lambda(.)$ maps them to the same set, but in a random order, ensuring each index has a unique mapping. The rearranged center coordinates are defined as $\hat{C} = \{\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_L \mid \hat{c}_i = c_{\lambda(i)}\}$. Finally, we perform a center transformation on all points within each sub-model:

$$\begin{aligned} \hat{p}_{l,k}^{xyz} &= (p_{l,k}^{xyz} - c_l) + \hat{c}_l, \hat{s}_l^{\alpha} = \{(\hat{p}_{l,k}^{xyz}, p_{l,k}^{\alpha})\}_{k=1}^{K}, \\ \hat{S}^{\alpha} &= \{\hat{s}_l^{\alpha} | l \in \{1, 2, \cdots, L\}\}. \end{aligned} \tag{1}$$

Similarly, for the m-th view $i_m^{\beta} \in \mathbb{R}^{H \times W \times C}$, we first divide it into $\mathcal{K}$ small grids of $\mathcal{L} \times \mathcal{L}$ size, denoted as $i_m^{\beta} = \{g_1, g_2, \cdots, g_{\mathcal{K}}\}$, where $\mathcal{K} = \frac{H \times W}{\mathcal{L}^2}$. These mini-grids are subsequently rearranged to reconstruct a new image denoted as $\hat{i}_m^{\beta} = \{\hat{g}_1, \hat{g}_2, \cdots, \hat{g}_{\mathcal{K}} \mid \hat{g}_k = g_{\lambda(k)}\}$. Through the above process, we successfully establish the technical focus data $(\mathcal{S}_{tec}, \mathcal{I}_{tec})$. From an intuitive perspective, The confusion transformation preserves the relative positional relationships within individual sub-models or views while disrupting the

global correlations between sub-models or views, achieving the dual goal of preserving technical features and breaking down semantic information.

Semantic focus generation Image aesthetic assessment methods (Deng, Loy, and Tang 2017; Sheng et al. 2018) widely employ a downsampling strategy to degrade the technical information in original samples while maximally preserving semantic content and aesthetic information. Inspired by this approach, we apply random distortions of diverse types to multi-view inputs to generate 2D semantically-focused data. Notably, to prevent excessive distortions that would significantly compromise compositional integrity and semantic information, we randomly select only one distortion type from our predefined distortion categories and control distortion intensities during each training iteration.

For point cloud data, we similarly adopt the aforementioned strategy to degrade technical information while substantially preserving semantic content:

$$\bar{i}_m^{\beta} = \mathrm{RD}(i_m^{\beta}), \quad \bar{S_l}^{\alpha} = \mathrm{RD}(S_l^{\alpha}). \tag{2}$$

where $\mathrm{RD}(\cdot)$ denotes the random distortion operator, selected from within our predefined distortion set. Through this approach, we construct semantically-focused point cloud data $(\mathcal{S}_{sem}, \mathcal{I}_{sem})$. From an intuitive perspective, the random distortion method deliberately disrupts technical distortions level while preserving compositional relationships and aesthetic integrity among visual elements.

Context & Attribute Correlation Awareness

Existing PCQA methods typically process individual views or sub-models in isolation, overlooking the highly dynamic integrative nature of the human visual system. Therefore, we introduce our single perceptual stream pipeline, incorporating a CACA module to capture potential correlations across features from different views and sub-models. Additionally, in this section, we describe how our pretraining strategy enables the single perceptual network to extract domain-specific perceptual features effectively.

Feature Encoding We employ PointNet++ (Qi et al. 2017) and ResNet-50 (He et al. 2016) as feature encoders. Taking the raw data $(\mathcal{S}, \mathcal{I})$ as an example:

$$\begin{aligned} f_{3d}^{\alpha} &= \mathrm{PointNet++}(S^{\alpha}) \in \mathbb{R}^{L \times D}, \alpha \in A, \\ f_{2d}^{\beta} &= \text{ResNet-50}(I^{\beta}) \in \mathbb{R}^{M \times D}, \beta \in B. \end{aligned} \tag{3}$$

To this end, we extract multi-modal feature sets, denoted as $f_{3d} = \{f_{3d}^{\alpha} | \alpha \in A\}$ and $f_{2d} = \{f_{2d}^{\beta} | \beta \in B\}$ respectively.

The Structure of CACA After feature encoding, we first adopt an attention mechanism to model the contextual relationship within each attribute feature. Taking f_{3d}^{α} as an example, where $f_{3d}^{\alpha} = \{f_1^{\alpha}, f_2^{\alpha}, ..., f_L^{\alpha}\}$ and each f_l^{α} denotes a feature vector encoded by the l-th sub-model. First, we employ self-attention to model the contextual relationship between the different sub-models. Specifically, the query Q, key K, and value V are obtained through linear transformations. By computing weighted aggregations of different sub-models' features, each feature f_l^{α} can effectively capture enriched contextual information from other sub-models. The process is formulated as:

$$F_l^{\alpha} = \sum_{j=1}^{L} \frac{\exp(Q_l \cdot K_j^{\top})}{\sum_{k=1}^{L} \exp(Q_l \cdot K_k^{\top})} \cdot V_j. \tag{4}$$

f_{2d}^{β} undergoes identical processing to achieve the contextual relationship between different views. We denote the new features as $F_{3d} = \{F_{3d}^{\alpha} | \alpha \in A\}$ and $F_{2d} = \{F_{2d}^{\beta} | \beta \in B\}$.

To model the comprehensive relationships between different attribute features, we cross-attend each pair of attribute features in the combinatorial set $F = \{F_{3d}, F_{2d}\}$. We take F_{3d}^{rgb} as an example:

$$h_{3d}^{rgb} = \{\mathrm{Cross}(F_{3d}^{rgb}, F_j) \mid F_j \subseteq F, F_j \neq F_{3d}^{rgb}\}. \tag{5}$$

Furthermore, we compute a global 3D color feature by applying mean pooling over the sequence dimension of h_{3d}^{rgb}. The final color feature representation $H_{3d}^{rgb} \in \mathbb{R}^{(5L+1) \times D}$ is then constructed by concatenating this global feature vector with h_{3d}^{rgb} along the sequence dimension.

Following a similar procedure, we process the other five attribute features and denote the updated features as $\mathcal{H} = \{H_{3d}^{\alpha}, H_{2d}^{\beta}\}$, where $H_{3d} = \{H_{3d}^{\alpha} | \alpha \in A\}$ and $H_{2d} = \{H_{2d}^{\beta} | \beta \in B\}$. The entire process can be defined as:

$$\mathcal{H} = \mathcal{G}(\mathcal{S}, \mathcal{I}). \tag{6}$$

where $\mathcal{G}(\cdot)$ denotes the complete processing pipeline of single perceptual stream.

Pretraining The core objective of our pretraining is to enable the technical and semantic streams in DSP-PCQA to preserve quality-relevant information in their respective domains. Therefore, we optimize our model via feature similarity regularization loss.

Taking the technical stream as an example, we simultaneously process both the raw data $(\mathcal{S}, \mathcal{I})$ and technical focus data $(\mathcal{S}_{tec}, \mathcal{I}_{tec})$ through the technical stream. This yields features $\mathcal{H}_{tec}$ and $\mathcal{H}'_{tec}$, which are then aggregated via mean pooling:

$$\begin{aligned} \mathcal{H}_{tec} &= \mathcal{G}_{tec}(\mathcal{S}, \mathcal{I}), \quad \mathbb{H}_{tec} = \mathrm{Mean}(\mathcal{H}_{tec}), \\ \mathcal{H'}_{tec} &= \mathcal{G}_{tec}(\mathcal{S}_{tec}, \mathcal{I}_{tec}), \quad \mathbb{H'}_{tec} = \mathrm{Mean}(\mathcal{H'}_{tec}). \end{aligned} \tag{7}$$

In addition to the basic quality regression loss, we maximize the feature similarity between $\mathbb{H}_{tec}$ and $\mathbb{H'}_{tec}$ to reduce the impact of semantic factors in the technical feature space. The final overall loss for our pretraining is as follows:

$$\begin{aligned} \mathcal{L}_{reg}^{tec} &= \mathrm{MSE}(\mathrm{Reg}(\mathbb{H}_{tec}), Y_{true}), \\ \mathcal{L}_{similarity}^{tec} &= 1 - \frac{\mathbb{H}_{tec} \cdot \mathbb{H}'_{tec}}{\|\mathbb{H}_{tec}\| \cdot \|\mathbb{H}'_{tec}\|}, \\ \mathcal{L}_{total}^{tec} &= \lambda_1 \mathcal{L}_{similarity}^{tec} + \mathcal{L}_{reg}^{tec}. \end{aligned} \tag{8}$$

where $\mathrm{Reg}(\cdot)$ is a two-layer feed-forward network for quality regression, $\lambda_1 \in [0, 1]$ are used to control the proportion of loss. The same pre-training approach is applied to semantic stream. We successfully obtain two distinct perceptual streams that focus on the information in their respective fields.

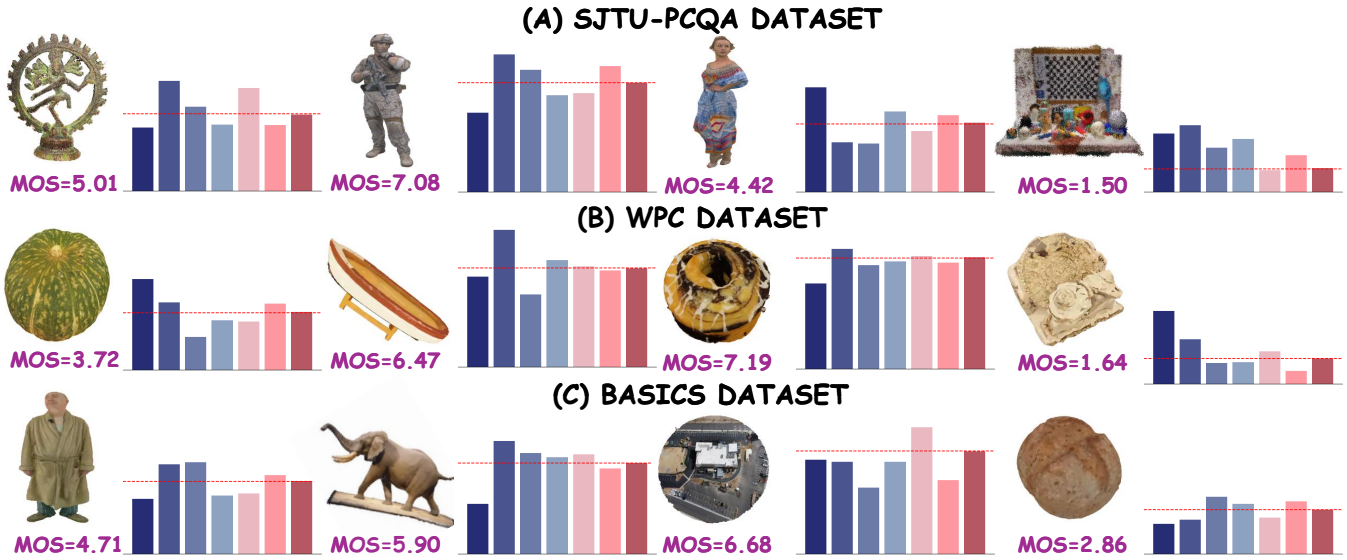

Figure 3: Examples of the point cloud evaluation results on SJTU, WPC, and BASICS datasets. The bar chart represents the predicted scores of the point cloud under different methods. From left to right, the methods are MM-PCQA, GMS-3DQA, CLIP-PCQA, LMM-PCQA, DSP-PCQA(Tec), DSP-PCQA(Sem), DSP-PCQA. The red dotted line indicates the MOS.

Exchange-based Perceptual Injection

Observers can't evaluate visual quality by focusing on a single dimension. Building on this insight, we propose an EPI module that exchanges perceptual details by exchanging quality-insensitive tokens during fine-tuning stage.

The Structure of EPI Technical features $\mathcal{H}_{tec}$ and semantic features $\mathcal{H}_{sem}$ are extracted via the technical rationality stream $\mathcal{G}_{tec}(\cdot)$ and the semantic sensibility stream $\mathcal{G}_{sem}(\cdot)$, respectively. These extracted features are then concatenated with the corresponding reg-tokens $\mathbb{H}_{tec}$ and $\mathbb{H}_{sem}$ to form the resultant features $\mathcal{F}_{tec}$ and $\mathcal{F}_{sem}$.

Our EPI module consists of multiple EPI blocks, where each block contains a self-attention layer and an exchange layer. At l-th block of EPI module, we first process $\mathcal{F}_{tec}^{l-1}$ and $\mathcal{F}_{sem}^{l-1}$ through separate self-attention layer, yielding attention score matrix A_T^l, A_S^l and updated $\mathcal{F}_{tec}^l$, $\mathcal{F}_{sem}^l$.

In particular, $A^l[0, 1:]$ represents the attention scores between the Reg-token and other tokens in the sequence. In the exchange layer, we select the subset of tokens corresponding to the bottom θ-percentile attention scores in $A^l[0, 1:]$ and replace their embedding vectors with the average embedding of all the tokens in another feature. Taking $\mathcal{F}_{tec}^l$ as an example, it can be updated by:

$$\begin{aligned} \mathcal{F}_{tec}^l[i,:] &= \frac{1}{n}\sum_{j=1}^{n}\mathcal{F}_{sem}^l[j,:] + \mathcal{F}_{tec}^l[i,:], \quad where \\ A_T^l[0,i] &\leq \mathcal{Q}\left(\{A_T^l[0,1:]\}, \theta\right). \end{aligned} \tag{9}$$

where $\mathcal{Q}$ denotes the function for calculating quantiles. Through this token-swapping mechanism between features, we simulate the perceptual exchange process involved in human quality judgment.

Fine-tuning The reg-tokens $\mathbb{H}_{tec}$ and $\mathbb{H}_{sem}$, extracted by the EPI module, are fed into separate quality regression heads, generating $Score_{tec}$ and $Score_{sem}$. The final quality score is computed as a weighted fusion of both scores, where learnable parameters adaptively determine their relative importance in different datasets:

$$Score = w \cdot Score_{tec} + (1-w) \cdot Score_{sem} \tag{10}$$

Following (Zhang et al. 2024b; Ding et al. 2024), we optimize the overall framework using a combined loss function comprising ranking loss and MSE loss.

$$\mathcal{L}_{total} = \mathcal{L}_{mse} + \lambda_2 \mathcal{L}_{rank}. \tag{11}$$

Here $\lambda_2 \in [0, 1]$ is used to control the proportion of loss.

Experimental

Datasets

We conduct experiments on three point cloud quality assessment benchmarks: SJTU-PCQA (9 reference point clouds, 7 distortions, 378 distorted point clouds) (Yang et al. 2021), WPC (20 reference point clouds, 4 distortion types, 740 distorted point clouds) (Liu et al. 2023a), and BASICS (75 reference point clouds, 4 distortion types, 1494 distorted point clouds) (Ak et al. 2024). Given the limited dataset sizes, we use k-fold cross-validation to evaluate performance: 9-fold for SJTU-PCQA; 5-fold for WPC. For BASICS, we split data into training, validation, and testing sets using an 8:1:1 ratio.

Implementation Details

Our experiments are performed using PyTorch on one NVIDIA 4090 GPU. Each sub-model has a cardinality K of 2048 and projected images have a resolution of 224×224. The training process is performed in two stages:

TYPE	METHODS	SJTU-PCQA			WPC			BASICS		
		SRCC↑	PLCC↑	RMSE↓	SRCC↑	PLCC↑	RMSE↓	SRCC↑	PLCC↑	RMSE↓
FR	MSE-P2PO(MEKURIA ET AL. 2016)	0.729	0.812	1.361	0.456	0.485	19.894	0.774	0.849	0.559
	MSE-P2PL(TIAN ET AL. 2017)	0.628	0.594	2.282	0.328	0.270	22.823	0.836	0.895	0.468
	PCQM(MEYNET ET AL. 2020)	0.864	0.885	1.086	0.743	0.750	15.164	0.808	0.891	0.478
	GRAHSIM(YANG ET AL. 2022B)	0.878	0.845	1.032	0.583	0.616	17.194	0.813	0.895	0.465
	TCDM(ZHANG ET AL. 2024A)	0.930	0.910	0.891	0.807	0.804	13.525	0.757	0.874	0.505
NR	IT-PCQA(YANG ET AL. 2022A)	0.865	0.828	1.166	0.487	0.433	19.896	0.310	0.302	1.024
	3D-NSS(ZHANG ET AL. 2022)	0.714	0.738	1.769	0.648	0.651	16.572	0.617	0.657	0.883
	RES-SCNN(LIU ET AL. 2023B)	0.880	0.889	0.878	0.435	0.429	23.270	0.352	0.391	0.975
	MM-PCQA(ZHANG ET AL. 2023)	0.910	0.923	0.772	0.841	0.856	12.351	0.738	0.793	0.628
	GMS-3DQA(ZHANG ET AL. 2024B)	0.911	0.918	0.787	0.831	0.834	12.229	0.807	0.895	0.472
	COPA (SHAN ET AL. 2024)	0.897	0.913	0.739	0.779	0.785	11.932	0.785	0.812	0.534
	CLIP-PCQA(LIU ET AL. 2025)	0.922	0.937	0.693	-	-	-	0.856	0.911	0.462
	LMM-PCQA(ZHANG ET AL. 2024C)	0.925	0.929	0.724	0.864	0.859	11.925	0.845	0.867	0.443
	DSP-PCQA(TEC)	0.936	0.941	0.689	0.861	0.849	12.013	0.842	0.889	0.484
	DSP-PCQA(SEM)	0.915	0.924	0.783	0.866	0.868	11.831	0.849	0.872	0.452
	DSP-PCQA	**0.951**	**0.964**	**0.637**	**0.892**	**0.889**	**11.572**	**0.875**	**0.946**	**0.347**

Table 1: Performance comparison between the proposed and state-of-the-art PCQA methods on the 3 datasets. "PC" and "I" stand for the method is based on the point cloud and image modality, respectively. DSP-PCQA(Tec) and DSP-PCQA(Sem) indicate using only single stream to assess quality. The best performance is marked with a gray background and with bold font; the second best is only marked with a gray background. FR: Full Reference, NR: No Reference

INDEX	DFE	CACA	EPI	SRCC↑	PLCC↑
①	×	✓	✓	0.912	0.923
②	TEC.	✓	✓	0.937	0.944
③	SEM.	✓	✓	0.929	0.935
④	✓	×	✓	0.911	0.934
⑤	✓	CON.	✓	0.928	0.940
⑥	✓	ATT.	✓	0.934	0.941
⑦	✓	✓	×	0.929	0.951
⑧	✓	✓	✓	**0.951**	**0.964**

Table 2: Ablation study of DSP-PCQA on key components.

TRAIN ON	SJTU		WPC	
TEST ON	WPC	BASICS	SJTU	BASICS
MM-PCQA	0.201	0.378	0.755	0.600
GMS-3DQA	0.396	0.268	0.684	0.469
CLIP-PCQA	0.324	0.470	—	—
LMM-PCQA	0.475	0.652	0.783	0.626
DSP-PCQA	**0.595**	**0.723**	**0.835**	**0.665**

Table 3: Cross-dataset validation. Both the training and testing are on the complete dataset.

Pretraining. The number of sub-models L and image projections M both equal to 6. In our DFE module, the size of small grids $\mathcal{L} \times \mathcal{L}$ is set to 32×32. The predefined distortion categories include: Downsampling, Color jittering, and Gaussian Noise. We employ the Adam optimizer with an initial learning rate of 5×10^{-5} and a weight decay of 10^{-4}. The hyper-parameter λ_1= 0.2. Each streams is trained for 100 epochs with a batch size of 4.

Fine-Tuning. The number of EPI blocks is set to 3, θ is set to 0.2, and the λ_2 is set to 1. The training details are kept consistent with those in the first stage.

Comparison Methods and Evaluation Criteria

In our study, we select 13 state-of-the-art quality assessment methods for comparison, consisting of 5 FR-PCQA and 8 NR-PCQA methods. The specific methods are listed in the "Method" column of Tab.1. We employ Spearman's Rank-Order Correlation Coefficient (SRCC), Pearson's Linear Correlation Coefficient (PLCC), and Root Mean Square Error (RMSE) to quantify model performance. Following established practice in quality assessment research (Zhu et al. 2024; Liu et al. 2021b), we employ a five-parameter nonlinear regression to compensate for scale discrepancies between algorithmic predictions and Mean Opinion Scores (MOS).

Overall Performance Discussion

We compare the performance of DSP-PCQA with other PCQA methods on the SJTU-PCQA, WPC, and BASICS databases. The results are illustrated in Tab. 1, and examples of the point cloud evaluation results are shown in Fig. 3. We can draw several conclusions: 1) DSP-PCQA outperforms all compared methods across the three databases. On SJTU-PCQA, it achieves an SRCC of 0.951, surpassing the best comparative method (vs. SRCC = 0.925). Similarly, it reaches an SRCC of 0.892 on WPC (vs. SRCC=0.864) and 0.875 on BASICS (vs. SRCC=0.856). 2) When independently evaluating the technical stream and the semantic stream, we find that the performance of these two streams is highly competitive compared with other PCQA methods. 3) As illustrated in Fig. 3, DSP-PCQA's predicted scores are significantly closer to the ground-truth MOS than those of other PCQA methods.

Ablation Study

To study the effectiveness of DSP-PCQA, we conducted thorough ablation experiments on the SJTU-PCQA database. **1)Contributions of Pre-training with the DFE module.** Comparing ①, ②, ③,and ⑧ in Tab. 2, it demonstrates the integration of both pre-training methods yields superior performance compared to employing a single one, indicating semantic and technical information play equally crucial role in DSP-PCQA. **2)Contributions of the CACA module.** To

Distortion	OT		CN		DS		DS + CN		DS + GGN		GGN		CN + GGN	
Method	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑
MM - PCQA	0.846	0.895	0.890	0.917	0.917	0.937	0.944	0.936	0.940	0.954	0.920	0.939	0.933	0.950
GMS - 3DQA	0.808	0.863	0.822	0.822	0.863	0.906	0.942	0.945	0.946	0.953	0.951	0.964	0.961	0.972
CLIP-PCQA	0.884	0.919	0.903	0.918	0.925	0.946	0.954	0.954	0.955	0.958	0.952	0.957	0.963	0.963
LMM - PCQA	0.915	0.936	0.863	0.900	0.853	0.905	0.929	0.933	0.931	0.954	0.922	0.940	0.953	0.956
DSP-PCQA(TEC)	0.857	0.875	0.882	0.931	0.876	0.926	0.949	0.958	0.958	0.969	0.967	0.971	0.922	0.935
DSP-PCQA(SEM)	0.922	0.938	0.917	0.946	**0.954**	**0.970**	0.961	0.967	0.943	0.953	**0.962**	**0.982**	0.968	0.973
DSP-PCQA	**0.938**	**0.951**	**0.950**	**0.973**	0.933	0.966	**0.974**	**0.981**	**0.967**	**0.981**	0.970	0.978	**0.979**	**0.987**

Table 4: Distortion-specific performance results on the SJTU-PCQA database. OT: Octree-based compression, CN: color noise, DS: down-sampling, DS+CN: down-sampling combined with color noise, DS+GGN: down-sampling combined with geometry Gaussian noise, GGN: geometry Gaussian noise, and CN+GGN: color noise combined with geometry Gaussian noise

Object	RB		RL		Loot		Soldier		ULB		LD		Statue		Shiva		Hhi	
Method	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑	SR↑	PL↑
MM - PCQA	0.946	0.951	0.895	0.918	0.944	0.940	0.962	0.967	0.936	0.952	0.955	0.954	0.907	0.937	0.834	0.870	0.936	0.940
GMS - 3DQA	0.900	0.924	0.899	0.923	0.924	0.934	0.963	0.968	0.911	0.955	0.947	0.955	0.843	0.909	0.868	0.906	0.912	0.923
CLIP-PCQA	0.942	0.955	0.903	0.913	0.943	0.923	0.954	0.959	0.952	0.960	0.942	0.934	0.920	0.949	**0.902**	**0.938**	0.952	0.957
LMM - PCQA	0.906	0.911	0.861	0.825	0.955	0.961	0.922	0.913	0.909	0.930	0.959	0.960	0.874	0.869	0.859	0.904	0.913	0.894
DSP-PCQA(TEC)	0.957	0.963	0.921	0.931	0.969	0.966	0.968	0.974	0.948	0.958	0.967	0.963	0.931	0.953	0.880	0.915	0.959	0.961
DSP-PCQA(SEM)	0.946	0.946	0.909	0.926	0.950	0.947	0.954	0.957	0.946	0.961	0.956	0.953	0.898	0.946	0.859	0.917	0.952	0.962
DSP-PCQA	**0.969**	**0.971**	**0.943**	**0.952**	**0.976**	**0.976**	**0.976**	**0.977**	**0.962**	**0.969**	**0.980**	**0.981**	**0.946**	**0.974**	0.898	0.924	**0.967**	**0.969**

Table 5: Object-specific performance results on the SJTU-PCQA database. LD:Longdress, RB: Redandblack, ULB: ULB-Unicorn, RL: Romanoillamp

evaluate the effectiveness of the CACA module, we remove the context correlation awareness and attribute correlation awareness modules, respectively. Comparing ④, ⑤, ⑥ and ⑧, We find that the improvement becomes more significant when both the context correlation awareness and attribute correlation awareness modules are included. **3) Contributions of EPI module.** Comparing ⑦, ⑧, we find that the EPI module can significantly improve the final performance of DSP-PCQA by exchanging different perceptual details.

Cross Dataset Validation

We perform cross dataset validation experiments as summarized in Tab. 3. We train on one dataset and evaluate on others, with the result corresponding to the minimal training loss recorded. As shown in Tab. 3, the performance of the cross-dataset evaluation is relatively low due to significant variations with respect to both the types and contents of distortion. Nevertheless, our DSP-PCQA still significantly outperforms all compared methods, which clearly demonstrates its outstanding generalizability.

Distortion-specific & Object-specific Evaluation

To validate that integrating diverse perceptual preferences is beneficial for PCQA, we conduct distortion-specific and object-specific evaluation experiments in this section.

Distortion-specific Evaluation We evaluate the quality of point clouds with the same distortion type but different objects and the results are summarized in Tab.4: 1) DSP-PCQA demonstrates the best performance across all distortion types, indicating that DSP-PCQA its high effectiveness in handling diverse distortion forms. 2) We find the semantic sensibility stream demonstrates superior performance over the technical rationality stream when evaluating the same distortion type. This result is expected because the semantic content usually plays a dominant role in PCQA when the distortion types of point clouds are similar. This finding aligns closely with the original intention of our model design and validates the rationality of our approach.

Object-specific Evaluation We evaluate the quality of point clouds with the same object type but different distortion types. The result is shown in Tab. 5: 1) DSP-PCQA achieves the best performance on most categories, indicating that DSP-PCQA can effectively conduct quality assessment for various point cloud contents, regardless of the specific content information presented in the point clouds. 2) We find the technical rationality stream demonstrates superior performance over the semantic sensibility stream. The underlying reason for this trend is that when point cloud content is similar, the associated semantic information exerts limited influence, allowing the technical rationality stream predominates.

Conclusion

In our study, we introduce the DSP-PCQA: the first framework that mirrors human dual perception preferences via parallel networks. Firstly, we employ a DFE module to further disentangle technical and semantic representations. Secondly, we introduce a CACA module to emulate the highly dynamic characteristics of the human visual system. Finally, dual perceptions are mutually injected by swapping quality-insensitive tokens through EPI module. Experimental results demonstrate that DSP-PCQA outperforms other methods, propelling groundbreaking advancements in PCQA.

Acknowledgments

This work was supported in part by the National Science Foundation of China under Grant 62301041, and in part by Beijing Institute of Technology Research Fund Program for Young Scholars.

References

Ak, A.; Zerman, E.; Quach, M.; Chetouani, A.; Smolic, A.; Valenzise, G.; and Le Callet, P. 2024. BASICS: Broad Quality Assessment of Static Point Clouds in a Compression Scenario. *IEEE Transactions on Multimedia*, 26: 6730–6742.

Bellini Leite, S. C. 2022. Dual Process Theory: Embodied and Predictive; Symbolic and Classical. *Frontiers in Psychology*, 13: 805386.

Chen, C.; Mo, J.; Hou, J.; Wu, H.; Liao, L.; Sun, W.; Yan, Q.; and Lin, W. 2024a. TOPIQ: A Top-Down Approach From Semantics to Distortions for Image Quality Assessment. *IEEE Transactions on Image Processing*, 33: 2404–2418.

Chen, W.; Jiang, Q.; Zhou, W.; Shao, F.; Zhai, G.; and Lin, W. 2024b. No-Reference Point Cloud Quality Assessment via Graph Convolutional Network. *IEEE Transactions on Multimedia*.

Cui, Y.; Chen, R.; Chu, W.; Chen, L.; Tian, D.; Li, Y.; and Cao, D. 2022. Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review. *IEEE Transactions on Intelligent Transportation Systems*, 23(2): 722–739.

Deng, Y.; Loy, C. C.; and Tang, X. 2017. Image Aesthetic Assessment: An experimental survey. *IEEE Signal Processing Magazine*, 34(4): 80–106.

Ding, X.; Luo, B.; Xiong, H.; Jin, D.; Wan, J.; and Jiang, L. 2024. Point Cloud Quality Assessment Based on Cross-Modal De-Redundancy. *IEEE Access*, 195516–195527.

Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; and Bennamoun, M. 2021. Deep Learning for 3D Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12): 4338–4364.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Hu, R.; Chu, X.; Dou, D.; Liu, X.; Liu, Y.; and Qi, B. 2025. Toward Real-World Applicability: Lightweight Underwater Acoustic Localization Model Through Knowledge Distillation. *IEEE Journal of Oceanic Engineering*, 50(2): 1429–1442.

Li, D.; Jiang, T.; and Jiang, M. 2019. Quality Assessment of In-the-Wild Videos. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, 2351–2359. New York, NY, USA: Association for Computing Machinery. ISBN 9781450368896.

Li, D.; Jiang, T.; Lin, W.; and Jiang, M. 2019. Which Has Better Visual Quality: The Clear Blue Sky or a Blurry Animal? *IEEE Transactions on Multimedia*, 21(5): 1221–1234.

Liao, L.; Xu, K.; Wu, H.; Chen, C.; Sun, W.; Yan, Q.; and Lin, W. 2022. Exploring the Effectiveness of Video Perceptual Representation in Blind Video Quality Assessment. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, 837–846. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392037.

Liu, Q.; Su, H.; Duanmu, Z.; Liu, W.; and Wang, Z. 2023a. Perceptual Quality Assessment of Colored 3D Point Clouds. *IEEE Transactions on Visualization and Computer Graphics*, 29(8): 3642–3655.

Liu, Q.; Yuan, H.; Hou, J.; Hamzaoui, R.; and Su, H. 2021a. Model-Based Joint Bit Allocation Between Geometry and Color for Video-Based 3D Point Cloud Compression. *IEEE Transactions on Multimedia*, 23: 3278–3291.

Liu, Q.; Yuan, H.; Su, H.; Liu, H.; Wang, Y.; Yang, H.; and Hou, J. 2021b. 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.

Liu, Y.; Yang, Q.; Xu, Y.; and Yang, L. 2023b. Point Cloud Quality Assessment: Dataset Construction and Learning-based No-reference Metric. *ACM Trans. Multimedia Comput. Commun. Appl.*, 19(2s).

Liu, Y.; Zhang, Y.; Shan, Z.; and Xu, Y. 2025. CLIP-PCQA: Exploring Subjective-Aligned Vision-Language Modeling for Point Cloud Quality Assessment. arXiv:2501.10071.

Mekuria, R.; Li, Z.; Tulvan, C.; and Chou, P. 2016. Evaluation criteria for point cloud compression. Technical report, ISO/IEC MPEG.

Meynet, G.; Nehmé, Y.; Digne, J.; and Lavoué, G. 2020. PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*.

Min, X.; Ma, K.; Gu, K.; Zhai, G.; Wang, Z.; and Lin, W. 2017. Unified Blind Quality Assessment of Compressed Natural, Graphic, and Screen Content Images. *IEEE Transactions on Image Processing*, 26(11): 5462–5474.

Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017. PointNet++: deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 5105–5114.

Shan, Z.; Zhang, Y.; Yang, Q.; Yang, H.; Xu, Y.; Hwang, J.-N.; Xu, X.; and Liu, S. 2024. Contrastive Pre-Training with Multi-View Fusion for No-Reference Point Cloud Quality Assessment. arXiv:2403.10066.

Sheng, K.; Dong, W.; Ma, C.; Mei, X.; Huang, F.; and Hu, B.-G. 2018. Attention-based Multi-Patch Aggregation for Image Aesthetic Assessment. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM '18, 879–886. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356657.

Tian, D.; Ochimizu, H.; Feng, C.; Cohen, R.; and Vetro, A. 2017. Geometric distortion metrics for point cloud compression. In *2017 IEEE International Conference on Image Processing (ICIP)*, 3460–3464.

Tliba, M.; Chetouani, A.; Valenzise, G.; and Dufaux, F. 2023. PCQA-Graphpoint: Efficient Deep-Based Graph Metric for Point Cloud Quality Assessment. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5.

Van Essen, D. C.; and Maunsell, J. H. 1983. Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6: 370–375.

Wang, J.; Chen, P.; Zheng, N.; Chen, B.; Principe, J. C.; and Wang, F.-Y. 2021. Associations between MSE and SSIM as cost functions in linear decomposition with application to bit allocation for sparse coding. *Neurocomputing*, 422: 139–149.

Wang, J.; Gao, W.; and Li, G. 2024. Zoom to Perceive Better: No-Reference Point Cloud Quality Assessment via Exploring Effective Multiscale Feature. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7): 6334–6346.

Wang, S.; Wang, X.; Gao, H.; and Xiong, J. 2023. Non-Local Geometry and Color Gradient Aggregation Graph Model for No-Reference Point Cloud Quality Assessment. In *Proceedings of the 31st ACM International Conference on Multimedia*, 6803–6810.

Wang, X.; Wang, X.; Liu, R.; and Huang, X. 2024. Rating-Augmented No-Reference Point Cloud Quality Assessment Using Multi-Task Learning. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4320–4324.

Weingarten, J.; Gruener, G.; and Siegwart, R. 2004. A state-of-the-art 3D sensor for robot navigation. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, volume 3, 2155–2160 vol.3.

Wu, H.; Chen, C.; Hou, J.; Liao, L.; Wang, A.; Sun, W.; Yan, Q.; and Lin, W. 2022. FAST-VQA: Efficient End-to-end Video Quality Assessment with Fragment Sampling. *Proceedings of European Conference of Computer Vision (ECCV)*.

Xiang, J.; Dang, Y.; Chen, P.; Liang, R.; Huan, R.; and Gao, N. 2024. Semantic-Aware and Quality-Aware Interaction Network for Blind Video Quality Assessment. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, 9970–9979. New York, NY, USA: Association for Computing Machinery. ISBN 9798400706868.

Xie, W.; Wang, K.; Ju, Y.; and Wang, M. 2023. pmBQA: Projection-based Blind Point Cloud Quality Assessment via Multimodal Learning. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3250–3258.

Yang, J.; Fu, J.; Zhang, W.; Cao, W.; Liu, L.; and Peng, H. 2024. MoE-AGIQA: Mixture-of-Experts Boosted Visual Perception-Driven and Semantic-Aware Quality Assessment for AI-Generated Images. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 6395–6404.

Yang, Q.; Chen, H.; Ma, Z.; Xu, Y.; Tang, R.; and Sun, J. 2021. Predicting the Perceptual Quality of Point Cloud: A 3D-to-2D Projection-Based Exploration. *IEEE Transactions on Multimedia*, 23: 3877–3891.

Yang, Q.; Liu, Y.; Chen, S.; Xu, Y.; and Sun, J. 2022a. No-Reference Point Cloud Quality Assessment via Domain Adaptation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21147–21156.

Yang, Q.; Ma, Z.; Xu, Y.; Li, Z.; and Sun, J. 2022b. Inferring Point Cloud Quality via Graph Similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6): 3015–3029.

Zhang, Y.; Yang, Q.; Zhou, Y.; Xu, X.; Yang, L.; and Xu, Y. 2024a. TCDM: Transformational Complexity Based Distortion Metric for Perceptual Point Cloud Quality Assessment. *IEEE Transactions on Visualization and Computer Graphics*, 30(10): 6707–6724.

Zhang, Z.; Sun, W.; Min, X.; Wang, T.; Lu, W.; and Zhai, G. 2022. 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.

Zhang, Z.; Sun, W.; Min, X.; Zhou, Q.; He, J.; Wang, Q.; and Zhai, G. 2023. MM-PCQA: Multi-Modal Learning for No-reference Point Cloud Quality Assessment. *IJCAI*.

Zhang, Z.; Sun, W.; Wu, H.; Zhou, Y.; Li, C.; Chen, Z.; Min, X.; Zhai, G.; and Lin, W. 2024b. GMS-3DQA: Projection-Based Grid Mini-patch Sampling for 3D Model Quality Assessment. *ACM Trans. Multimedia Comput. Commun. Appl.*, 20(6).

Zhang, Z.; Wu, H.; Zhou, Y.; Li, C.; Sun, W.; Chen, C.; Min, X.; Liu, X.; Lin, W.; and Zhai, G. 2024c. LMM-PCQA: Assisting Point Cloud Quality Assessment with LMM. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, 7783–7792. Association for Computing Machinery.

Zhou, W.; Yang, Q.; Chen, W.; Jiang, Q.; Zhai, G.; and Lin, W. 2024. Blind Quality Assessment of Dense 3D Point Clouds with Structure Guided Resampling. *ACM Trans. Multimedia Comput. Commun. Appl.*, 20(8).

Zhu, L.; Cheng, J.; Wang, X.; Su, H.; Yang, H.; Yuan, H.; and Korhonen, J. 2024. 3DTA: No-Reference 3D Point Cloud Quality Assessment With Twin Attention. *IEEE Transactions on Multimedia*, 26: 10489–10502.