Not All Distortions Are Created Equal: Distortion-Selective Domain Adaptation for Point Cloud Quality Assessment

Yangwei Li, Xiaochuan Wang*, Xin Shang, Haisheng Li

School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing, China
2431062120@st.btbu.edu.cn, wangxc@btbu.edu.cn, 2330702038@st.btbu.edu.cn, lihsh@btbu.edu.cn

Abstract

Point cloud quality assessment (PCQA) has advanced significantly with synthetic datasets offering diverse distortion coverage for model training. However, when applied to new application scenarios, models often suffer from performance drops due to mismatched distortion characteristics between source and target domains. Most current methods use all available synthetic distortions, which may introduce irrelevant features and hinder generalization. To address this, we propose **DST-PCQA**, a distortion-selective training framework for PCQA. Unlike previous approaches that treat all distortions equally, DST-PCQA identifies and selects distortion types most relevant to a target domain by analyzing inter-domain distortion similarity. This selective strategy reduces negative transfer and enables efficient domain-specific training. To fully leverage the selected distortions for both classification and quality prediction, we adopt a dual-branch architecture that fuses 2D visual cues and 3D geometric structure via cross-modal attention. This design supports multi-level feature alignment across modalities and enables fine-grained distortion understanding. Extensive evaluations across three target domains have verified the effectiveness of DST-PCQA over full-set training baselines. Moreover, its distortion-selective strategy is orthogonal to existing model-based PCQA methods, enabling improved cross-domain performance and reduced training costs across a wide range of architectures.

Introduction

Point cloud quality assessment (PCQA) plays a critical role in safeguarding visual fidelity across various applications. However, the demands of different applications, e.g., immersive VR systems that need to process and stream massive 3D asset libraries (Liang et al. 2024), and autonomous driving systems encountering unpredictable environmental degradations (Zang et al. 2019), necessitate a quality assessment method that is not just accurate, but universally adaptable. The development of such robust, task-driven models is fundamentally hampered by the profound lack of annotated data. Unlike natural image analysis, which benefits from massive datasets, PCQA databases typically contain only hundreds to thousands of samples. This severe data limitation creates a critical **generalization gap**, i.e., a model trained on one dataset performs poorly on another, a clear sign of overfitting to source-specific biases.

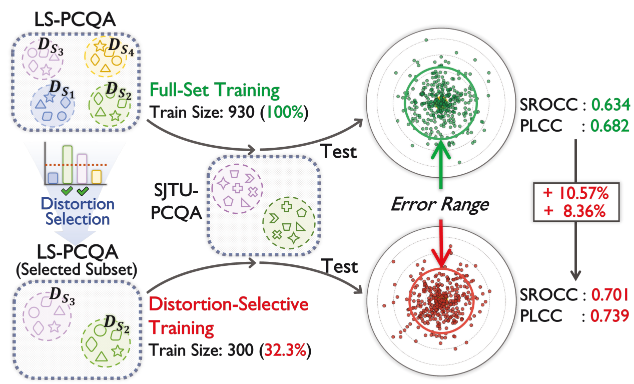

Figure 1: Comparison of Full-Set Training and our Distortion-Selective Training (DST). The full-set training was trained on LS-PCQA and tested on SJTU-PCQA, while DST was trained on the subset of LS-PCQA via distortion selection and tested on SJTU-PCQA. The error range indicates the variance of errors between the predicted scores and the corresponding MOS. Note that DST achieves superior SROCC and PLCC with tighten error range.

Confronted with the stark disparity in dataset scales, researchers tried to create increasingly diverse synthetic distortion datasets to mitigate the cross-dataset performance degradation, with the assumption that more distortion variety leads to better cross-domain generalization (Liu et al. 2023). Moreover, they adopted **full-set training**, which mechanically aggregates all references with all distortions. However, training on the full source dataset LS-PCQA yields suboptimal performance in practice.

As shown in Fig. 1, the model trained with all the 930 samples on LS-PCQA and tested on SJTU-PCQA is inferior to ours training with only 300 samples. Particularly, the cross-dataset performance is increased by 10.5% of SROCC and 8.36% of PLCC when the training scale is dropped to 32.3%. The observation indicates that *not all distortions are created equal*. We therefore hypothesized that a **distortion-**

*Corresponding author.

selective training strategy, i.e., using only a small, highly relevant subset of distortions, could paradoxically lead to superior performance.

Based on these, we proposed a **D**istortion-**S**elective **T**raining framework for PCQA, namely **DST-PCQA**. Specifically, we carefully design a distortion-aware selection mechanism to mitigate the negative transfer effects meanwhile containing the training samples. Our method begins by analyzing the similarity between the target domain and each type of synthetic distortion, which then guides the construction of a curated, target-relevant training subset. This selected data is processed using a dual-branch multi-modal architecture, where the 2D branch captures visual distortion patterns from multi-view projections, while the 3D branch extracts geometric features using a Point Transformer (Zhao et al. 2021). Finally, a cross-modal attention mechanism integrates the complementary information from both branches, resulting in a comprehensive model that effectively captures both appearance and structure-based degradations.

In summary, our salient contributions include:

- A novel **Distortion-Selective Training (DST-PCQA)** framework, powered by a **Domain-Adaptive Distortion Selection (DADS)** module that prunes irrelevant source data to mitigate negative transfer and yield superior performance with substantially less data.
- A unified dual-branch architecture that serves a dual role. Its versatile feature encoders first enable domain selection via classification, and form the backbone of a high-fidelity regression model using cross-modal attention.
- Extensive experimental validation demonstrating that DST-PCQA consistently and significantly outperforms full-set training baselines and existing domain adaptation methods across challenging cross-domain scenarios.

Related Work

Point Cloud Quality Assessment

Previous Point Cloud Quality Assessment (PCQA) has shifted from full-reference (FR) like GraphSIM (Yang et al. 2020) towards no-reference (NR). Recently, no-reference methods became spotlight, since they do not rely on pristine reference point clouds, therefore boosting the performance in real-world applications. Due to the development of deep learning, current NR methods can be further categorized into projection-based, point-wise, and hybrid ways.

Projection-based methods firstly project the point cloud onto 2D images, and then learn to regress to the subjective scores (Zhang et al. 2024a; Wang et al. 2024; Wang, Gao, and Li 2024). **Point-wise** methods directly learn to predict the quality in the 3D space (Neri and Battisti 2025). **Hybrid** methods incorporate both the 2D image and point-wise features, where the two modalities are integrated (Zhang et al. 2023; Chen et al. 2024).

Recently, researchers utilize Large Language Models (LLMs) to further underscore the distortion-aware description (Zhang et al. 2024b; Xie et al. 2024). Although these methods achieve impressive performance on single datasets, they merely explore the cross-dataset generalization.

Domain Adaptation for Quality Assessment

To bridge the generalization gap, Domain Adaptation (DA) has emerged as a key strategy following two paths. **Feature-level alignment** aims to learn a domain-invariant representation space, which has been introduced into PCQA. It extends from image quality assessment (IQA), where methods like StyleAM (Lu et al. 2025) align style-modulated features, to the more complex PCQA domain. Pioneering works include IT-PCQA (Yang et al. 2022) and Liu (Liu et al. 2025). IT-PCQA firstly investigated the external knowledge transfer from IQA dataset, where the features from 2D and 3D space are directly aligned. Liu (Liu et al. 2025) employed generative models to mitigate the modality gap from IQA to PCQA. However, these methods sacrifice target specificity for universal representation, i.e., they rely on the assumption that greater source diversity can benefit the transfer performance, ignoring that many distortions are irrelevant or detrimental to the target's profile. It becomes especially acute in cross-modality settings, where the semantic gap between 2D and 3D distortions makes feature alignment highly prone to negative transfer.

A complementary and more direct paradigm is **data-level selection**. The powerful insight that distortion-selective training can significantly outperform brute-force full-set training was compellingly demonstrated in IQA (Li et al. 2024). Our work is the first to introduce this paradigm to PCQA. We posit that algorithmically pruning irrelevant source distortions before training constitutes a more direct, robust, and principled path towards robust cross-dataset adaptation.

Preliminary

Cross-dataset PCQA generalization presents a challenging domain adaptation problem. We consider multiple labeled source domains $\{\mathcal{S}_i\}_{i=1}^{k}$, each corresponding to a synthetic distortion type, and an unlabeled target domain $\mathcal{T}$ representing a specific application scenario. Following the multi-source domain adaptation generalization bound (Zhao et al. 2018), the target risk $\varepsilon_T(h)$ is bounded by:

$$\varepsilon_T(h) \leq \sum_{i=1}^{k} \alpha_i \left(\hat{\varepsilon}_{S_i}(h) + \frac{1}{2} d_{\mathcal{H}}(\mathcal{S}_i, \mathcal{T}) \right) + C \qquad (1)$$

where $\hat{\varepsilon}_{S_i}(h)$ is the empirical source risk, $d_{\mathcal{H}}(\mathcal{S}_i, \mathcal{T})$ measures domain discrepancy, α_i are source domain weights satisfying $\sum_{i=1}^{k} \alpha_i = 1$, and C represents complexity terms.

This bound formally reveals that high discrepancy $d_{\mathcal{H}}(\mathcal{S}_i, \mathcal{T})$ from even a single irrelevant source domain can inflate the error bound, causing negative transfer. The prevalent full-set training paradigm, by aggregating all source data, implicitly assumes uniform weighting ($\alpha_i = 1/k$) across all domains. This non-selective approach makes the model highly susceptible to performance degradation from a few highly dissimilar source domains, demonstrating the critical need for an adaptive selection mechanism. In PCQA, this is particularly acute as datasets often differ fundamentally in both content and distortion patterns. For quality assessment, however, the model should be distortion-sensitive

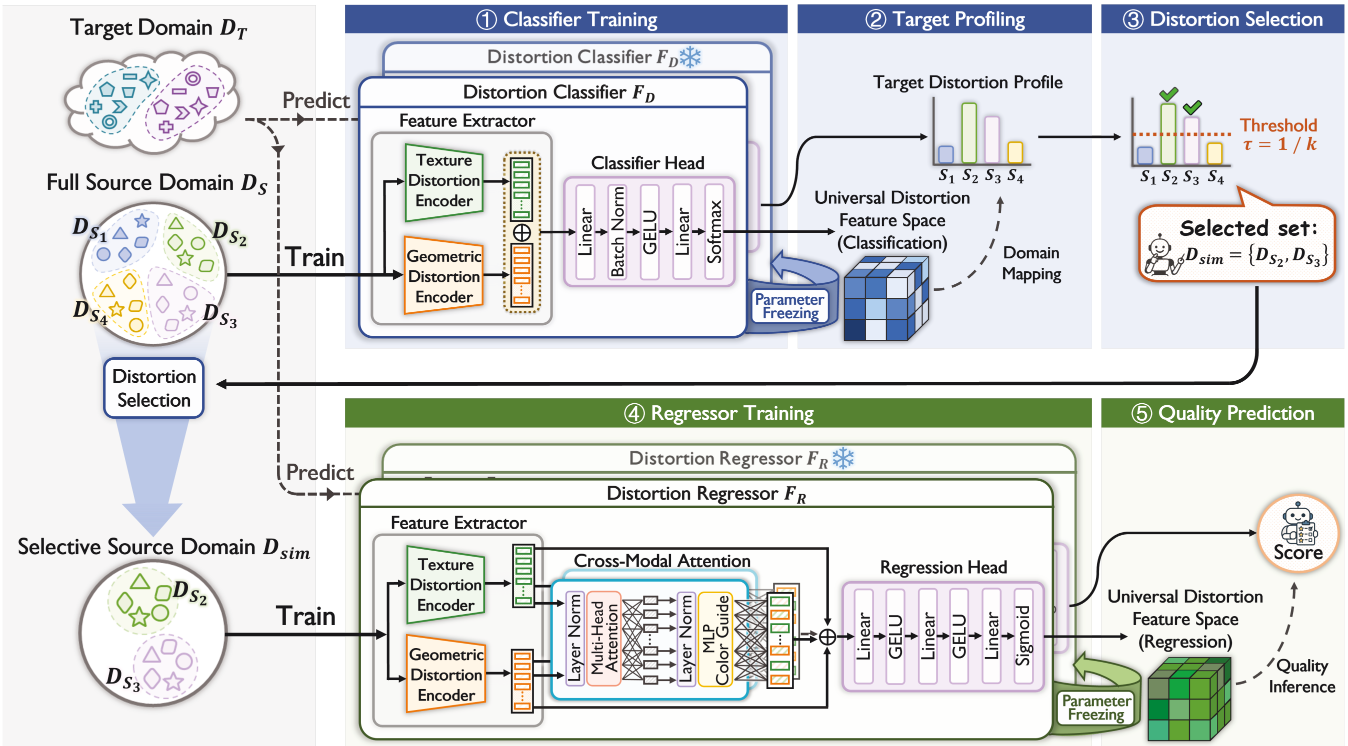

Figure 2: Overview of the proposed DST-PCQA framework, consisting of a two-stage, five-step pipeline. Stage I (Steps 1–3) implements the Domain-Adaptive Distortion Selection (DADS) module, while Stage II (Steps 4–5) leverages the selected distortion subset for targeted quality assessment.

but content-invariant. This allows the domain discrepancy to be conceptually decomposed:

$$d_{\mathcal{H}}(\mathcal{S}_i, \mathcal{T}) \approx \underbrace{d_{\text{content}}(\mathcal{S}_i, \mathcal{T})}_{\text{Content Gap}} + \underbrace{d_{\text{distortion}}(\mathcal{S}_i, \mathcal{T})}_{\text{Distortion Gap}} \tag{2}$$

Instead of reconciling the challenging Content Gap via full-set strategy, our distortion-selective training focuses on minimizing the more tractable Distortion Gap. This is formalized as finding a sparse weighting scheme $\boldsymbol{\alpha}^*$:

$$\boldsymbol{\alpha}^* = \arg \min_{\boldsymbol{\alpha} \in \{0, 1/|\mathcal{A}|\}^k} \sum_{i=1}^{k} \alpha_i \cdot \hat{d}_{\mathcal{H}}(\mathcal{D}_{d_i}, \mathcal{D}_T) \tag{3}$$

where $\mathcal{A} = \{i : \alpha_i > 0\}$ is the set of selected domains, and $\hat{d}_{\mathcal{H}}(\mathcal{D}_{d_i}, \mathcal{D}_T)$ is a practical proxy for the discrepancy between the source distortion distribution $\mathcal{D}_{d_i}$ and the target's distortion profile $\mathcal{D}_T$.

However, directly computing the domain discrepancy $\hat{d}_{\mathcal{H}}$ is intractable, which necessitates a practical proxy to estimate the relevance between source and target domains. To implement this, our DADS module leverages the key insight that distortion patterns (e.g., noise, compression) are largely content-agnostic (Liu et al. 2023). We operationalize our strategy by training a distortion classifier and using its predictions to estimate similarity:

$$\text{sim}(\mathcal{T}, \mathcal{S}_i) \propto \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} F_D(\mathbf{x})[i] \tag{4}$$

where $F_D(\mathbf{x})[i]$ is the classifier's predicted probability that a target sample $\mathbf{x}$ belongs to the i-th source distortion class. This allows us to select the most relevant source domains, effectively optimizing the generalization bound for PCQA.

Proposed Method

To mitigate negative transfer in cross-domain scenarios, we introduce Distortion-Selective Training for Point Cloud Quality Assessment (DST-PCQA) framework. As illustrated in Fig. 2, our method is instantiated through a five-step pipeline organized into two main stages. The first stage, **Domain-Adaptive Distortion Selection**, algorithmically prunes the full source domain to identify a target-relevant subset (*Step 1 to Step 3*). The second stage, **Cross-Domain Quality Assessment**, successively trains a dedicated regression model exclusively on this selected data to ensure robust performance on the target domain (*Step 4 to Step 5*).

Domain-Adaptive Distortion Selection

The cornerstone of our framework is the Domain-Adaptive Distortion Selection (DADS) module, which systematically identifies relevant source distortions for target-specific training. This process unfolds in three sequential steps: training a universal distortion expert, profiling the target domain, and selecting the most relevant source distortions.

Classifier Training. We begin by training a universal distortion classifier $F_D : \mathcal{X} \rightarrow \Delta^{k-1}$ on the complete source domain $\mathcal{D}_S$ to distinguish among all k synthetic distortion types. Here $\mathcal{X}$ indicates the input domain, Δ^{k-1} the $(k-1)$-dimensional probability simplex, i.e., the space of all possible k-dimensional probability vectors. The network employs multimodal feature processing to capture comprehensive distortion signatures, as point cloud degradations often manifest simultaneously in both geometry and appearance.

For geometric distortion encoding, we process the raw point cloud sample $\mathbf{x}$ using Point Transformer (Zhao et al. 2021), obtaining the geometry feature,

$$\mathbf{F}_{3D} = PointTransformer(\mathbf{x}) \tag{5}$$

For texture distortion encoding, we extend hierarchical feature extraction (Nguyen et al. 2025) to handle multi-view projections. Specifically, the point cloud is firstly projected onto six orthogonal views, and then processed through ResNet-50 to extract four level feature maps $\{\mathbf{H}_{ij}\}$, respectively. Here $i = 1, 2, \cdots, 6$ indicates the projection view, while $j = 1, 2, \cdots, 4$ indicates the feature level. The followed by cross-view aggregation:

$$\mathbf{F}_{2D} = \oplus_{i=1}^{6}(Max(Head_i(\mathbf{H}_{ji}))) \tag{6}$$

where each Head_i extracts distortion-aware features at different scales. $\oplus(\cdot)$ stands for the concatenation operation, and Max indicates the max-pooling operation.

The final classification combines both modalities:

$$\mathbf{p} = Softmax(\text{MLP}(\mathbf{F}_{2D} \oplus \mathbf{F}_{3D})) \tag{7}$$

where $\mathbf{p}$ indicates the final predicted probability distribution over the k distortion classes.

The training objective employs standard Cross-Entropy loss to produce well-calibrated probability distributions:

$$\mathcal{L}_{\text{Stage1}} = -\sum_{i=1}^{N} \log p(y_i|\mathbf{p}_i) \tag{8}$$

where N is the number of trained samples, y_i is the corresponding ground truth distortion label.

Target Profiling Once trained, the frozen classifier F_D is used to analyze the characteristics of the unlabeled target domain $\mathcal{T}$. We pass all target samples through F_D and average their output probability vectors to compute a Target Affinity Profile, $\boldsymbol{\mu}_T \in \mathbb{R}^k$. This process is defined as:

$$\boldsymbol{\mu}_T = \mathbb{E}_{\mathbf{x}\sim\mathcal{T}}[F_D(\mathbf{x})] \tag{9}$$

Here, the expectation $\mathbb{E}_{\mathbf{x}\sim\mathcal{T}}$ is computed by averaging the classifier's output probability vectors $F_D(\mathbf{x})$ over all samples $\mathbf{x}$ in the target domain $\mathcal{T}$. The resulting vector $\boldsymbol{\mu}_T$ quantitatively represents the target domain's affinity to each of the k source distortion types, effectively providing a practical estimate of the Distortion Gap (Eq. 2).

Distortion Selection With the Target Affinity Profile, we then select source domains that exhibit an above-average affinity to the target. This is achieved by applying a principled threshold $\tau = 1/k$, which corresponds to a uniform probability distribution. The set of selected distortion indices, $\mathcal{I}^*$, is thus formed by:

$$\mathcal{I}^* = \{i \mid \boldsymbol{\mu}_{T,i} > \tau\} \tag{10}$$

In this expression, $\boldsymbol{\mu}_{T,i}$ denotes the i-th element of the Target Affinity Profile vector, representing the affinity score for the i-th distortion type. This selection creates the final curated data subset $\mathcal{D}_{\text{sim}} = \bigcup_{i\in\mathcal{I}^*} \mathcal{S}_i$ for subsequent training.

Cross-Domain Quality Assessment

The quality prediction stage continues to use the feature processing structure built earlier, but adjusts it for cross-dataset regression. It includes two final steps that use the selected distortion subset to perform targeted quality assessment.

Regressor Training The quality assessment model employs the same multimodal feature extractors $(\mathbf{F}_{2D}, \mathbf{F}_{3D})$ but incorporates enhanced fusion mechanisms for regression. Unlike classification, which mainly relies on distinguishing features, the more subtle task of quality regression gains from a deeper fusion of information across different modalities. We adopt a Cross-Modal Attention (CMA) mechanism that allows each modality to be enhanced by complementary information from the other. A multi-head attention block, $\Gamma(\cdot)$, is applied symmetrically:

$$\mathbf{F}_{2D}^{enhanced} = \Gamma(Q = \hat{\mathbf{F}}_{3D}, K = \hat{\mathbf{F}}_{2D}, V = \hat{\mathbf{F}}_{2D}) \tag{11}$$

$$\mathbf{F}_{3D}^{enhanced} = \Gamma(Q = \hat{\mathbf{F}}_{2D}, K = \hat{\mathbf{F}}_{3D}, V = \hat{\mathbf{F}}_{3D}) \tag{12}$$

where $\hat{\mathbf{F}}$ denotes features projected into a common embedding space. The final quality-aware representation preserves both modality-specific and enhanced information:

$$\mathbf{F}_{quality} = \hat{\mathbf{F}}_{2D} \oplus \hat{\mathbf{F}}_{3D} \oplus \mathbf{F}_{2D}^{enhanced} \oplus \mathbf{F}_{3D}^{enhanced} \tag{13}$$

To concurrently optimize for prediction accuracy and ranking consistency, a critical aspect for perceptual quality assessment, we employ a hybrid loss function $\mathcal{L}_{\text{Stage2}}$, which is inspired by (Zhang et al. 2023). It is a weighted combination of the standard mean squared error and a pairwise ranking loss ($\mathcal{L}_{\text{rank}}$). The comprehensive formulation is expressed as:

$$\mathcal{L}_{\text{Stage2}} = \mathcal{L}_{\text{mse}} + \lambda \cdot \mathcal{L}_{\text{rank}} \tag{14}$$

where λ is a hyperparameter that balances the contribution of each loss component.

Quality Prediction The trained quality regressor performs cross-dataset prediction by processing target domain samples through the same feature extraction pipeline, producing final quality scores that reflect the learned distortion-specific patterns from the curated training subset.

Overall Training Pipeline

Our DST-PCQA framework operates through a two-stage pipeline. As illustrated in Figure 2: **Stage I (Steps 1-3)** trains a universal dual-branch distortion classifier on complete source domain $\mathcal{D}_S$, then profiles target domain $\mathcal{D}_T$ to compute distortion affinity scores and select relevant subset $\mathcal{D}_{\text{sim}}$ via threshold $\tau = 1/k$. **Stage II (Steps 4-5)** reuses the feature extractors but incorporates cross-modal attention, training the quality model exclusively on $\mathcal{D}_{\text{sim}}$. This decoupled design enables domain-agnostic selection and clean target-relevant training that mitigates negative transfer.

Experiments

This section systematically evaluates the DST-PCQA framework. We first detail the experimental setup, then analyze the selection behavior of our DADS module and compare performance against state-of-the-art methods. Finally, ablation studies confirm our key design choices and demonstrate that our selective strategy is orthogonal to existing model-based architectures, underscoring its broad applicability.

Databases and Evaluation Metrics

Datasets. To validate the adaptability of our framework from a general synthetic source to task-specific synthetic targets, we conduct experiments across multiple PCQA datasets. The source domain is **LS-PCQA** (Liu et al. 2023), a large-scale dataset containing 930 distorted point clouds generated from 30 pristine models under 31 diverse synthetic distortions. It serves as a universal degradation pool for learning generalizable representations. As target domains, we choose three widely-used synthetic PCQA datasets, each reflecting a distinct application scenario:

- *Perceptual Quality Assessment:* **SJTU-PCQA** (Yang et al. 2021) comprises 420 samples with seven distortion types that are most perceptually relevant to human observers.
- *Compression-Oriented Assessment:* **WPC** (Liu et al. 2022) contains 740 samples with distortions caused by downsampling, noise, and standardized codecs (G-PCC, V-PCC). **WPC2.0** (Liu et al. 2021), with 400 samples, focuses exclusively on V-PCC artifacts under varying quantization parameters, simulating rate-distortion optimization scenarios.

Evaluation Metrics. We primarily use the Spearman Rank-Order Correlation Coefficient (SROCC) and the Pearson Linear Correlation Coefficient (PLCC) to measure prediction monotonicity and linearity, respectively, with higher values being better.

Implementation Details

Our framework is implemented in PyTorch and trained on four NVIDIA RTX 3090 GPUs. The training follows a two-stage, from-scratch protocol. **(1) DADS Module:** A dual-branch distortion classifier (Point Transformer + ResNet-50) is trained for up to 100 epochs on the entire LS-PCQA source domain (where the number of distortion types $k = 31$) using a standard Cross-Entropy loss. We use the Adam optimizer (lr=1×10^{-4}, weight decay=1×10^{-5}) with a ReduceLROnPlateau scheduler and early stopping (patience=15). The effective batch size is 32. **(2) Quality Regression:** Following the Distortion-Selective Training (DST) strategy, a separate regression model, which shares the same backbone but incorporates a Cross-Modal Attention (CMA) module, is trained from scratch for up to 100 epochs exclusively on the DADS-selected subset ($\mathcal{D}_{\text{sim}}$). It is optimized with Adam (lr=5×10^{-5}, weight decay=1×10^{-4}) and a hybrid L2-Rank loss (λ_{l2}=1.0, λ_{rank}=1.0), using a batch size of 8 and early stopping (patience=30). No weights are transferred between stages, adhering to a strict unsupervised domain adaptation protocol.

Selected Distortion	**SJTU-PCQA**	**WPC**	**WPC2.0**
`GaussShift`	✓	✓	✓
`C2AI_Lossy`	✓	✓	✓
`LocalLoss`	✓	✓	✓
`ColorNoise`	✓		
`GaussNoise`	✓		
`UniformNoise`	✓	✓	
`HighFreqNoise`	✓		
`UniformShift`	✓		
`LossyG_LossyA`	✓		
`CorrGaussNoise`	✓	✓	
`Reconstruction`		✓	✓
`LosslessG_NearLossA`		✓	✓
`LosslessG_LossyA`		✓	✓
`Saturation`		✓	✓
`LocalOffset`		✓	✓
`MeanShift`		✓	✓
`Octree`		✓	
`DownSample`		✓	
Total Selected	**10**	**13**	**9**
Selection Ratio	**32.3%**	**41.9%**	**29.0%**

Table 1: DADS selection results for each target scenario.

Domain Selection Analysis

Our distortion-selective strategy operates by first training a universal distortion classifier on the full source domain, then repurposing it as a frozen feature extractor to profile the target domain. This process yields a distortion affinity vector that guides selection via a simple $\tau = 1/k$ threshold. The effectiveness of this strategy is evident in Table 1, which reveals highly specialized selection patterns. For the perception-focused SJTU-PCQA, DADS correctly prioritizes noise-related artifacts (e.g., `GaussShift`, `ColorNoise`), mirroring human visual sensitivity. Conversely, for the compression-centric WPC and WPC2.0 domains, it intelligently pivots to codec-induced artifacts (e.g., `Reconstruction`, `Octree`), adapting to the underlying technical requirements. Most tellingly, the universal selection of only a tiny core set of three distortions underscores a crucial insight: naive full-set training forces the model to learn from substantial irrelevant information. The targeted adaptation by DADS is therefore not merely beneficial, but essential for effective knowledge transfer. Guided by the specific selections for each target domain identified in Table 1, we now proceed with our main cross-domain experiments to demonstrate the direct impact of this tailored training on model generalization and performance.

Performance Comparison

Comparison with State-of-the-Art Methods. We first evaluate our framework against a comprehensive suite of state-of-the-art deep learning-based PCQA methods. As detailed in Table 2, all models are trained on the full LS-

Methods	LS-PCQA		SJTU-PCQA		WPC		WPC2.0	
	SROCC ↑	PLCC ↑	SROCC ↑	PLCC ↑	SROCC ↑	PLCC ↑	SROCC ↑	PLCC ↑
DHCN (Chen et al. 2024)	0.7556	0.7344	0.5337	0.5779	0.5458	0.5468	0.4468	0.4721
GMS-3DQA (Zhang et al. 2024a)	0.7354	0.7425	<u>0.5893</u>	<u>0.6240</u>	0.5872	0.5918	0.3638	0.3701
AKS-Net (Wang et al. 2024)	0.6465	0.6474	0.5496	0.6127	0.3658	0.4199	0.2233	0.3074
MM-PCQA (Zhang et al. 2023)	<u>0.8668</u>	<u>0.8560</u>	0.5384	0.5316	<u>0.6106</u>	<u>0.6231</u>	<u>0.4733</u>	<u>0.4789</u>
MOD-PCQA (Wang, Gao, and Li 2024)	0.6425	0.7004	0.5420	0.5785	0.2417	0.2843	0.1724	0.2374
PST-PCQA (Neri and Battisti 2025)	0.5806	0.5700	0.4625	0.4839	0.2713	0.3453	0.0067	0.0274
DST-PCQA (Ours)	**0.8760**	**0.8611**	**0.7009**	**0.7397**	**0.6535**	**0.6608**	**0.5576**	**0.5793**

Table 2: Performance comparison against state-of-the-art deep learning-based PCQA methods. All models are trained on the LS-PCQA source and evaluated on its in-domain test set and three unseen synthetic target domains. Our DST-PCQA demonstrates superior generalization. Best results are in **bold**, second-best are <u>underlined</u>.

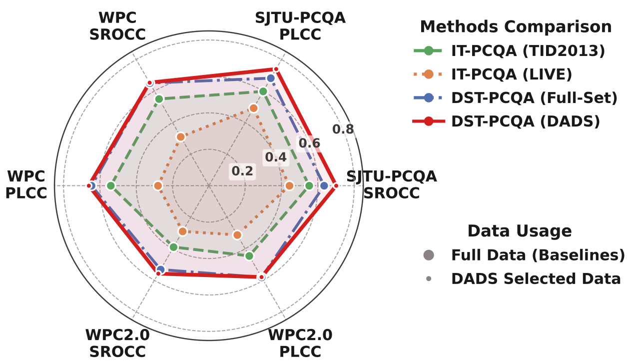

Figure 3: Quantitative comparison of our method and IT-PCQA on performance and data efficiency.

PCQA source dataset and then tested for generalization on the three unseen target domains. The results reveal a clear pattern: while many leading architectures, particularly the powerful MM-PCQA, achieve excellent in-domain scores, their performance sharply degrades when transferred. This drop highlights a critical generalization gap, confirming our hypothesis that forcing a model to learn from a full, uncurated set of distortions induces significant negative transfer. In stark contrast, our DST-PCQA not only remains competitive within the source domain but consistently and significantly outperforms all competitors in every cross-domain scenario. This robust performance provides compelling evidence that target-aware distortion selection is the key to effective generalization.

Cross-Modal Domain Adaptation Comparison. We benchmark DST-PCQA against IT-PCQA (Yang et al. 2022), a pioneering, reproducible, and open-sourced cross-modal adaptation framework. As shown in Figure 3, our method's superiority highlights a key strategic difference. Their cross-modal alignment is vulnerable to the large 2D-to-3D semantic gap, whereas our intra-modality selection avoids this issue, proving that optimizing distortion relevance within the same modality is a more direct path to generalization.

Universal Applicability Across Backbone Architectures. A key strength of our data-selection strategy is its model-agnostic design, functioning as a data-centric plug-and-play module. To validate its universal applicability, we applied our Distortion-Selective Training (DST) paradigm, powered by the DADS module, to several SOTA backbones. As shown in Table 3, the results offer compelling evidence of its effectiveness. By training selectively, the DST approach consistently yields substantial performance gains or maintains competitive results while drastically reducing data usage. For instance, applying our selective strategy to MM-PCQA on SJTU-PCQA boosts SROCC by a significant 19.2% while using only 32% of source distortions. This pattern demonstrates that DADS effectively mitigates the negative transfer inherent in full-set training, establishing a more efficient and fundamentally robust principle for cross-domain PCQA.

Ablation Studies and Analysis

Architecture Components. To validate our multimodal feature extractor design for the domain classifier, we ablate its core components. As shown in Table 4, the fused architecture (90.86% accuracy) significantly outperforms either modality in isolation. While the 2D visual branch provides a strong baseline (86.02%), the 3D geometric branch is less effective on its own due to the fundamental trade-off where necessary point downsampling for computational feasibility obscures fine-grained details critical for distortion analysis. The superior performance of the fused model confirms that 2D visual cues and 3D geometric structure are highly complementary, with the former being indispensable for capturing subtle artifacts that the latter can miss.

Selection Strategy Validation. To confirm that our approach performs distortion-aware selection rather than simply benefiting from a reduced dataset size, we benchmark it against a robust random selection baseline. This baseline mirrors the selection ratio for each target domain but chooses the distortions randomly, with results averaged over 5 runs for stability. Table 5 clearly shows that our Distortion-Selective strategy consistently and significantly outperforms this random approach across all scenarios. This result provides strong evidence that the performance gain is a direct result of selecting the most relevant distortions, not an incidental effect of training set reduction.

Backbone	Strategy	SJTU-PCQA		WPC		WPC2.0	
		SROCC ↑	PLCC ↑	SROCC ↑	PLCC ↑	SROCC ↑	PLCC ↑
MM-PCQA	Full-Set	0.5384	0.5316	0.6106	**0.6231**	0.4733	0.4789
	Distortion-Selective	**0.6418**	**0.6556**	**0.6374**	0.6219	**0.5044**	**0.4987**
MOD-PCQA	Full-Set	0.5421	0.5785	0.2417	0.2843	0.1724	0.2374
	Distortion-Selective	**0.5739**	**0.6087**	**0.2661**	**0.3115**	**0.2387**	**0.2768**
PST-PCQA	Full-Set	0.4625	0.4839	0.2713	0.3453	0.0067	0.0274
	Distortion-Selective	**0.4752**	**0.4919**	**0.2740**	**0.3746**	**0.1523**	**0.1129**
DST-PCQA (Ours)	Full-Set	0.6346	0.6819	0.6510	0.6461	0.5317	**0.5808**
	Distortion-Selective	**0.7009**	**0.7397**	**0.6535**	**0.6608**	**0.5576**	0.5793

Table 3: Model-agnostic generalizability of the Distortion-Selective strategy. The results validate that our selection strategy is highly effective as a plug-and-play module. When applied to various SOTA backbones, it consistently yields substantial performance gains or maintains competitive results, all while using only a fraction of the training data.

Method	Accuracy	Precision	F1-Macro
2D Branch Only	86.02%	87.60%	87.11%
3D Branch Only	39.25%	41.80%	36.37%
Dual-Branch (Ours)	**90.86%**	**92.70%**	**91.40%**

Table 4: Ablation study on classifier architecture. Results are reported on the LS-PCQA in-domain test set.

Target Domain	Strategy	SROCC ↑	PLCC ↑
SJTU-PCQA	Full-Set	0.6346	0.6819
	Random	0.5940	0.6257
	Distortion-Selective	**0.7009**	**0.7397**
WPC	Full-Set	0.6510	0.6461
	Random	0.6085	0.5898
	Distortion-Selective	**0.6535**	**0.6608**
WPC2.0	Full-Set	0.5317	**0.5808**
	Random	0.4745	0.5121
	Distortion-Selective	**0.5576**	0.5793

Table 5: Comparison between our DADS and a Random Selection baseline. Results are averaged over 5 random runs.

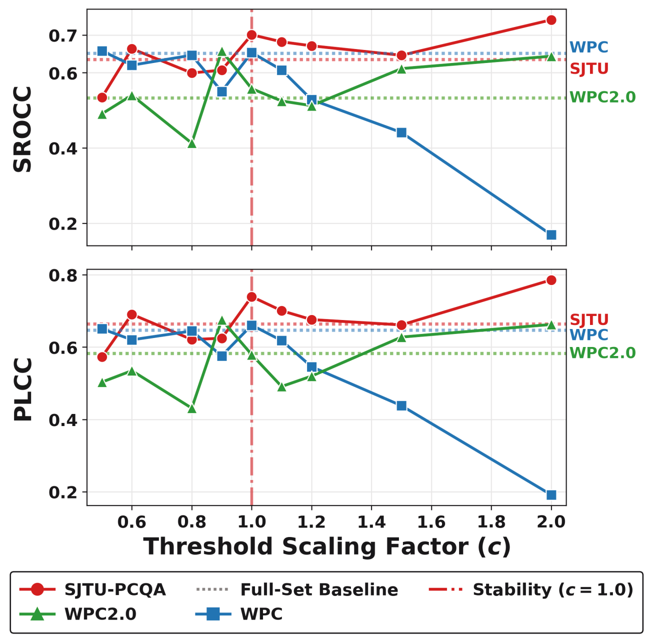

Figure 4: Impact of the threshold scaling factor c on SROCC and PLCC across target domains. Horizontal dotted lines indicate Full-Set baselines.

Hyperparameter Sensitivity. We analyze the sensitivity of our selection threshold, controlled by the scaling factor c where $\tau = c/k$. Figure 4 reveals that performance exhibits nuanced sensitivity to this hyperparameter. Our default choice of $c = 1.0$, marked by the dashed line, proves to be a robust and empirically-grounded selection, as performance for most target domains peaks at or near this value. Notably, performance for the WPC domain degrades sharply for $c > 1.0$, highlighting that an overly aggressive selection threshold can be detrimental. This confirms that our principled default of $c = 1.0$ provides a strong and reliable balance, consistently matching or outperforming the Full-Set baselines across diverse scenarios.

Conclusion

In this paper, we introduced DST-PCQA, a framework that effectively mitigates negative transfer in cross-dataset PCQA. Our distortion-selective training strategy achieves competitive performance with significantly less training data and computation, confirming that for effective knowledge transfer, *not all distortions are created equal*. Experimental results demonstrate that the proposed method can not only achieve better cross-dataset generalization, but also boost the performance of existing backbones. Consequently, our work opens a path toward future research on complete content-distortion disentanglement, which can benefit the generalization of PCQA in real-world applications.

Acknowledgments

This research was supported by the Beijing Municipal Natural Science Foundation (4232020), National Natural Science Foundation of China (62201017) and R&D Program of Beijing Municipal Education Commission (KM202310011014).

References

Chen, W.; Jiang, Q.; Zhou, W.; Xu, L.; and Lin, W. 2024. Dynamic Hypergraph Convolutional Network for No-Reference Point Cloud Quality Assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10): 10479–10493.

Li, A.; Wu, J.; Liu, Y.; and Li, L. 2024. Bridging the synthetic-to-authentic gap: Distortion-guided unsupervised domain adaptation for blind image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 28422–28431.

Liang, Z.; Liu, J.; Dasari, M.; and Wang, F. 2024. Fumos: Neural Compression and Progressive Refinement for Continuous Point Cloud Video Streaming. *IEEE Transactions on Visualization and Computer Graphics*, 30(5): 2849–2859.

Liu, Q.; Su, H.; Duanmu, Z.; Liu, W.; and Wang, Z. 2022. Perceptual quality assessment of colored 3D point clouds. *IEEE Transactions on Visualization and Computer Graphics*, 29(8): 3642–3655.

Liu, Q.; Yuan, H.; Hamzaoui, R.; Su, H.; Hou, J.; and Yang, H. 2021. Reduced Reference Perceptual Quality Model With Application to Rate Control for Video-Based Point Cloud Compression. *IEEE Transactions on Image Processing*, 30: 6623–6636.

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

Liu, Y.; Yang, Q.; Zhang, Y.; Xu, Y.; Yang, L.; and Li, Z. 2025. From Images to Point Clouds: An Efficient Solution for Cross-media Blind Quality Assessment without Annotated Training. *arXiv preprint arXiv:2501.13387*.

Lu, Y.; Li, X.; Liu, J.; and Chen, Z. 2025. StyleAM: Perception-Oriented Unsupervised Domain Adaption for No-Reference Image Quality Assessment. *IEEE Transactions on Multimedia*, 27: 2043–2058.

Neri, M.; and Battisti, F. 2025. Low-Complexity Patch-Based No-Reference Point Cloud Quality Metric Exploiting Weighted Structure and Texture Features. *IEEE Transactions on Broadcasting*, 71(2): 631–640.

Nguyen, H. T.; De Moor, K.; Larabi, M.-C.; and Amirshahi, S. A. 2025. A Distortion Aware Image Quality Assessment Model. In *Proceedings of the Winter Conference on Applications of Computer Vision*, 207–216.

Wang, H.; Wang, X.; Yuan, X.; Chen, X.; and Li, H. 2024. No-Reference Point Cloud Quality Assessment with Adaptive Keyframe Selection. In *2024 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 1–5.

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.

Xie, W.; Liu, Y.; Wang, K.; and Wang, M. 2024. LLM-Guided Cross-Modal Point Cloud Quality Assessment: A Graph Learning Approach. *IEEE Signal Processing Letters*, 31: 2250–2254.

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. 2022. No-reference point cloud quality assessment via domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 21179–21188.

Yang, Q.; Ma, Z.; Xu, Y.; Li, Z.; and Sun, J. 2020. Inferring point cloud quality via graph similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(6): 3015–3029.

Zang, S.; Ding, M.; Smith, D.; Tyler, P.; Rakotoarivelo, T.; and Kaafar, M. A. 2019. The Impact of Adverse Weather Conditions on Autonomous Vehicles: How Rain, Snow, Fog, and Hail Affect the Performance of a Self-Driving Car. *IEEE Vehicular Technology Magazine*, 14(2): 103–111.

Zhang, Z.; Sun, W.; Min, X.; Wang, Q.; He, J.; Zhou, Q.; and Zhai, G. 2023. MM-PCQA: Multi-Modal Learning for No-reference Point Cloud Quality Assessment. In Elkind, E., ed., *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, 1759–1767. International Joint Conferences on Artificial Intelligence Organization. Main Track.

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

Zhang, Z.; Wu, H.; Zhou, Y.; Li, C.; Sun, W.; Chen, C.; Min, X.; Liu, X.; Lin, W.; and Zhai, G. 2024b. LMM-PCQA: Assisting Point Cloud Quality Assessment with LMM. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, 7783–7792. New York, NY, USA: Association for Computing Machinery. ISBN 9798400706868.

Zhao, H.; Jiang, L.; Jia, J.; Torr, P. H.; and Koltun, V. 2021. Point Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 16259–16268.

Zhao, H.; Zhang, S.; Wu, G.; Moura, J. M. F.; Costeira, J. P.; and Gordon, G. J. 2018. Adversarial Multiple Source Domain Adaptation. In Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.