# HOLISTIC CORESET SELECTION FOR DATA EFFICIENT IMAGE QUALITY ASSESSMENT

## Supplementary Material

In our work, we proposed Perceptually Guided Coreset Selection (PGCS) for Image Quality Assessment (IQA). This supplementary document provides a detailed elaboration on the experimental findings presented in the main paper, offering additional insights into the datasets, baseline coreset selection methods, experimental settings, and hyperparameters.

## 1. EXTENDED RESULTS

In this section, we first illustrate the performance evaluation of our proposed PGCS method alongside other baseline methods on MUSIQ [1], followed by an ablation study for the PGCS image encoder $\mathcal{E}$ component.

### 1.1. Performance on IQA architecture - MUSIQ

This section discusses results on IQA architecture, MUSIQ [1]. Table 1 presents a performance comparison of coreset selection methods for the MUSIQ, trained on dataset fractions (1%, 5%, and the full dataset) and evaluated on their respective test sets. The coreset selection methods compared include Herding [2], K-center [3], Contextual Diversity (CD) [4], Moderate Coreset [5], and our proposed PGCS. Results are reported using two metrics: Spearman's Rank Correlation Coefficient (SRCC) and Pearson's Linear Correlation Coefficient (PLCC) for three datasets: KonIQ-10k [6], SPAQ [7], and AGIQA-3k [8]. A weighted average of SRCC and PLCC is also provided, calculated using the number of images in the test sets as weights. At 1% and 5% dataset fractions, PGCS performs better in terms of SRCC and PLCC metrics.

### 1.2. Ablation Study: PGCS Image Encoder

IQA relies on an image's semantic content and distortion to predict the quality score of an image. To validate the effectiveness of the image encoder used in PGCS, we conducted experiments to evaluate the suitability of the LIQE image encoder [10] for our approach. We experimented with extracting features from pre-trained networks, ResNet-50 [11] and VGG-19 [12] before their fully connected layers. Additionally, we performed an experiment using the logits from LIQE as the image representation. The results, presented in Table 2, validate our choice of including the LIQE image encoder, as it consistently outperforms the other configurations in terms of performance.

### 1.3. Ablation Study: Projection and Latent Space Partitioning

Table 3 presents the ablation study results evaluating the impact of non-linear projections and latent space partitioning on the proposed PGCS method. It reports the test set performance of the MANIQA architecture trained using a 5% dataset fraction as the coreset. The results highlight how different projection techniques and partitioning strategies influence model performance, demonstrating the effectiveness of our approach. Non-linear projections are particularly beneficial as they better capture complex distortions and feature relationships in IQA datasets compared to linear ones, leading to more informative coreset selection.

## 2. DATASET DESCRIPTION

We validated the performance of our proposed coreset selection method, PGCS, across five benchmark IQA datasets, each with distinct characteristics.

- **KADID-10k** [13]: This dataset consists of 10,125 images with 25 different types of synthetic distortions applied to 81 pristine images. These distortions include artifacts such as noise, blur, and compression, making it a comprehensive resource for evaluating IQA coreset selection methods on artificially altered content.
- **TID2013** [14]: Comprising 3,000 images with 24 reference images and 25 distortion types at 5 levels each, this dataset focuses on synthetic distortions, including additive noise, color quantization, and more complex transformations, providing a rich set of variations for IQA coreset selection methods evaluation.
- **KonIQ-10k** [6]: This dataset includes 10,073 images with authentic distortions collected from real-world sources. Captured under diverse conditions, it reflects realistic scenarios with distortions arising from natural photography errors, such as focus issues and lighting variations.

**Table 1**: Performance comparison of coreset selection methods for the MUSIQ [1], trained on dataset fractions selected by each method and evaluated on the respective test sets. **Bold** values mark the top-2 best performing methods for each dataset. The weighted average is calculated using the number of images in the test set as weights.

| Dataset Fraction | Methods | KonIQ-10k |  | SPAQ |  | AGIQA-3K |  | *Weighted Average* |  |
|---|---|---|---|---|---|---|---|---|---|
|  |  | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| 1% | Herding | 0.4041 | **0.4204** | 0.6919 | 0.6866 | 0.0083 | 0.1159 | **0.4876** | **0.5053** |
|  | K-center | **0.4204** | 0.4064 | 0.5011 | 0.5110 | 0.0422 | 0.1533 | 0.4108 | 0.4232 |
|  | CD | 0.4044 | 0.4111 | 0.6407 | 0.6315 | **0.0656** | **0.2257** | 0.4713 | 0.4896 |
|  | Moderate | 0.3661 | 0.3537 | **0.7054** | **0.7013** | **0.0496** | 0.1950 | 0.4831 | 0.4940 |
|  | PGCS | **0.4563** | **0.4671** | **0.7194** | **0.7062** | 0.0350 | **0.1963** | **0.5253** | **0.5436** |
| 5% | Herding | 0.4430 | 0.4351 | 0.6967 | 0.6860 | 0.0203 | 0.1513 | 0.5075 | 0.5155 |
|  | K-center | 0.4624 | 0.4514 | 0.7125 | 0.7007 | **0.0806** | **0.2374** | 0.5303 | 0.5397 |
|  | CD | 0.4655 | 0.4414 | **0.7344** | 0.7142 | 0.0631 | 0.1907 | 0.5395 | 0.5359 |
|  | Moderate | **0.5954** | **0.5803** | 0.7271 | **0.7189** | 0.0398 | 0.2003 | **0.5874** | **0.5971** |
|  | PGCS | **0.6075** | **0.6058** | **0.7703** | **0.7665** | **0.0749** | **0.2305** | **0.6167** | **0.6334** |
| FULL |  | 0.7471 | 0.7531 | 0.8262 | 0.8299 | 0.5133 | 0.5094 | 0.7546 | 0.7583 |

**Table 2**: Ablation study evaluating configurations for extracting image embeddings for the proposed PGCS method. The numerical values corresponding to the test set results for MANIQA [9] architecture trained using a 5% dataset fraction as the coreset.

| Image Embedding for PGCS | KADID-10K |  | TID2013 |  | KonIQ-10k |  | SPAQ |  | AGIQA-3K |  |
|---|---|---|---|---|---|---|---|---|---|---|
|  | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| Extracted from VGG-19 | 0.8068 | 0.8144 | 0.6013 | 0.6738 | 0.8374 | 0.8731 | 0.8923 | 0.8947 | 0.7784 | 0.8344 |
| Extracted from ResNet-50 | 0.8247 | 0.8320 | 0.5858 | 0.6366 | 0.8486 | 0.8736 | 0.8902 | 0.8909 | 0.7804 | 0.8262 |
| LIQE logits | 0.8340 | 0.8472 | 0.6293 | 0.7055 | 0.8468 | 0.8722 | 0.8895 | 0.8935 | 0.7905 | 0.8478 |
| LIQE Image Encoder | **0.8372** | **0.8473** | **0.6628** | **0.7061** | **0.8603** | **0.8776** | **0.8923** | **0.8956** | **0.8079** | **0.8598** |

- **SPAQ** [7]: The Smartphone Photography Attribute and Quality (SPAQ) dataset contains 11,125 images captured using various smartphone models. It emphasizes authentic distortions associated with mobile photography, such as noise and overexposure, making it suitable for evaluating IQA techniques tailored to handheld devices.

- **AGIQA-3K** [8]: This dataset features artificially generated distortions based on generative models. It includes content with unique and complex distortions not found in traditional datasets, offering a challenging testbed for assessing IQA methods in emerging scenarios involving synthetic imagery.

Table 4 provides details for each dataset, including Mean Opinion Score (MOS) range and the number of images corresponding to dataset fractions ranging from 1% to 95% for each dataset. This enables a direct comparison of performance results across different fractions. These datasets collectively provide a diverse evaluation benchmark, encompassing synthetic, authentic, and artificially generated distortions.

## 3. BASELINE CORESET SELECTION METHODS

We compared the performance of our proposed PGCS with four other coreset selection baselines, Herding [2], K-center greedy [3], and Contextual Diversity (CD) [4] and Moderate Coreset [5].

- **Herding** [2]: This method selects data points or generates pseudo-samples to approximate a target distribution efficiently. By iteratively minimizing the distance between the centers of the coreset and the original dataset in the feature space avoiding explicit model fitting. It draws inspiration from maximum entropy principles.

- **k-Center Greedy** [3]: This approach is designed to solve the minimax facility location problem, which seeks to select $k$ samples as coreset $CS$ from a full dataset $D$ such that the largest distance between any data point in $D \setminus CS$ and its closest data point in $CS$ is minimized. This problem is NP-hard, thus, greedy approximation is employed. To tackle the unlabeled coreset problem for CNNs, k-Center Greedy proposes a rigorous bound between the average loss over any given subset and the remaining data points using the geometry of the data. As an active learning algorithm, selecting a subset that minimizes this bound is objective.

- **Contextual Diversity (CD)** [4]: This approach was initially proposed to enhance active learning for CNNs. Unlike traditional methods that rely on visual diversity or prediction uncertainty, CD captures variations in spatial context by accounting for the confusion arising from spatially co-occurring classes.

- **Moderate Coreset** [5]: Traditional coreset selection

**Table 3**: Ablation study evaluating choice of non-linear projections and latent space partitioning for proposed PGCS method. The numerical values corresponding to the test set results for MANIQA [9] architecture trained using a 5% dataset fraction as the coreset.

| Non Linear Projection + Partitioning Method | KADID-10K | | TID2013 | | KonIQ-10k | | SPAQ | | AGIQA-3K | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC | PLCC |
| PCA + K-Means Clustering | 0.7877 | 0.7954 | 0.5818 | 0.6369 | 0.8012 | 0.8506 | 0.8668 | 0.8671 | 0.7527 | 0.8198 |
| t-SNE + K-Means Clustering | 0.8132 | 0.8134 | 0.6031 | 0.6403 | 0.8152 | 0.8599 | 0.8734 | 0.8764 | 0.7775 | 0.8233 |
| PCA + Latent Space Partioning ( Algorithm 2) | 0.8012 | 0.8044 | 0.6150 | 0.6498 | 0.8222 | 0.8418 | 0.8759 | 0.8854 | 0.7805 | 0.8354 |
| t-SNE + Latent Space Partioning ( Algorithm 2) | **0.8372** | **0.8473** | **0.6628** | **0.7061** | **0.8603** | **0.8776** | **0.8923** | **0.8956** | **0.8079** | **0.8598** |

**Table 4**: Details of the IQA datasets used in the experiments, including the number of images corresponding to each dataset fraction.

| Dataset | Distortion Category | MOS Range | #images | Coreset Size = #images corresponding to dataset fraction | | | | | | | | Full |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1.0% | 5.0% | 10.0% | 30.0% | 50.0% | 70.0% | 90.0% | 95.0% | |
| KADID-10K | Synthetic | [1,5] | 10125 | 70 | 354 | 708 | 2125 | 3543 | 4960 | 6377 | 6731 | 7086 |
| TID2013 | Synthetic | [0,9] | 3000 | 20 | 104 | 209 | 629 | 1049 | 1469 | 1889 | 1994 | 2099 |
| KonIQ-10k | Authentic | [1,5] | 10073 | 70 | 352 | 705 | 2115 | 3525 | 4935 | 6345 | 6698 | 7051 |
| SPAQ | Authentic | [0,100] | 11125 | 77 | 389 | 778 | 2336 | 3893 | 5450 | 7008 | 7397 | 7787 |
| AGIQA-3K | AI Generated | [0,5] | 2982 | 20 | 104 | 208 | 625 | 1043 | 1460 | 1877 | 1981 | 2086 |

methods rely on predefined score criteria to choose data points within a specific score range, which may not adapt well when the scenario changes, as the optimal range varies. Moderate coreset overcomes this by using the score median as a generalizable criterion. However, this approach applies the median score range as a global selection criterion for all classes, without accounting for the local distribution of each class, which may vary due to differences in class density.

## 4. EXPERIMENTAL SETTINGS

In this section, we describe the implementation details of the baselines and our proposed PGCS. The baselines include Herding [2], K-center greedy [3], and Contextual Diversity (CD) [4] and Moderate Coreset [5]. We also present the hyperparameter settings used for the IQA architecture to evaluate the performance of the baseline coreset selection methods and PGCS.

### 4.1. Implementation Details

PGCS is based on distortion, quality, and semantics-aware image embeddings extracted from LIQE. Thus, to ensure a fair comparison, we also employed embeddings from the LIQE image encoder incorporated with the following baselines: Herding [2], K-center greedy [3], and Contextual Diversity (CD) [4] and Moderate Coreset [5]. The other implementation details of all baselines and our proposed PGCS are provided below.

- **Herding** [2] and **k-Center Greedy** [3]: We employ the implementation given in publicly available GitHub repository DeepCore[1] with embeddings from LIQE image encoder as feature matrix, instead of VGG [12] or ResNet-50 [11].
- **Contextual Diversity (CD)** [4]: We utilize the implementation available in the publicly accessible GitHub repository[2]. To adapt CD for IQA, quality scores (or MOS) are discretized into bins, and logits representing quality levels from the LIQE [10] are employed to construct a co-occurrence matrix.
- **Moderate Coreset** [5]: We use the implementation given in publicly available GitHub repository[3]. Moderate coreset requires class labels; for IQA, we discretized quality scores (or MOS) into bins, with each bin representing a distinct class.
- **PGCS**: Our proposed approach involves several key components to optimize coreset selection. We first use a pre-trained image encoder $\mathcal{E}$ of LIQE [10] to extract image embeddings of dimension-512 for each image of the dataset $\mathcal{D}$. These embeddings are then projected to a lower-dimensional space using the projection operator $\mathcal{P}_m$, where t-SNE [15] is selected as $\mathcal{P}_m$ with $m = 3$. Next, we partition the projected embeddings using latent space partitioning algorithm with the number of partitions as $K = 10$. A dataset fraction $\alpha$ defines the percentage of the dataset to be selected as a coreset. For each partition, we calculate the median distance from the partition center to all the instances belonging to that partition. We choose $\alpha\%$ of instances from each partition to ensure diversity in the selected images. The

[1] https://github.com/PatrickZH/DeepCore

[2] https://github.com/sharat29ag/CDAL

[3] https://github.com/tmllab/2023_ICLR_Moderate-DS

dataset fraction $\alpha$ can range from 1% to 95%, allowing flexibility in the proportion of the dataset selected.

### 4.2. Hyperparameters Settings - IQA Architecture

We evaluated the performance of PGCS against other baseline methods by using the selected coreset to train the IQA architecture MANIQA [9] and MUSIQ [1]. This evaluation was conducted across dataset fractions $\alpha$ ranging from 1% to 95% for all datasets. We conducted experiments using PyTorch 2.4.0 and CUDA 12.2 for training and testing on an NVIDIA A100 GPU.

- **MANIQA**: MANIQA was trained and evaluated on coresets selected from all five datasets mentioned in Section 2. We utilize the MANIQA implementation publicly available on GitHub[4], and kept default configurations in our experiments. The training images are resized to a size of $224 \times 224$ and random horizontal flipping applied with a probability of 0.7. We kept the patch size $P$ set to 8. The MANIQA framework consists of two stages, each comprising two Transposed Attention Blocks (TAB) and one Scale Swin Transformer Block (SSTB). The embedding dimensions of the first and second SSTBs were set to $D_1 = 768$ and $D_2 = 384$, respectively. The Multi-Layer Perceptron (MLP) hidden layer dimension $D_m$, number of heads $H$, and window size were set to 768, 4, and 4 for each SSTB. The scaling factor $\alpha$ in the SSTB set to 0.80. The training was performed with a batch size $B$ of 2 and an initial learning rate $lr$ of $1 \times 10^{-5}$, using the ADAM optimizer with a weight decay of $1 \times 10^{-5}$. A cosine annealing learning rate scheduler was applied, with $T_{\text{max}} = 50$ and $\eta_{\text{min}} = 0$. The model was trained for 25 epochs. The Mean Squared Error (MSE) loss function was used as the training objective. The test set is kept consistent across all dataset fractions to ensure a concrete comparison between the baseline coreset methods and our proposed PGCS. During inference, we calculated metrics Spearman Rank Order Correlation (SRCC) and Pearson Linear Correlation Coefficient (PLCC) for test set to report the results.

- **MUSIQ**: We used the MUSIQ implementation publicly available on GitHub[5], and kept default configurations in our experiments. MUSIQ employs a Transformer-based architecture with a feed-forward hidden layer dimension of 384 and a multi-head attention mechanism consisting of 6 heads, each operating on a channel dimension of 384. The final MLP has a hidden layer size of 1152. Dropout is applied at a rate of 0.1 to both the attention and feed-forward layers, with an additional embedding dropout also set to 0.1. The model incorporates layer normalization with an epsilon value of $1 \times 10^{-12}$ to ensure numerical stability. The grid size of 10 is used for the spatial embedding. A cosine annealing learning rate scheduler was applied, with $T_{\text{max}} = 3 \times 10^4$ and $\eta_{\text{min}} = 0$. The training was done with a batch size $B$ of 2 using the SGD optimizer. The learning rate $lr$ and #epochs are tailored for each dataset, with specific values provided in Table 5.

**Table 5**: Learning rate $lr$ and #epochs for training on MUSIQ.

| Dataset →<br>Hyperparameter ↓ | KonIQ-10k | SPAQ | AGIQA-3K |
|---|---|---|---|
| $lr$ | $1 \times 10^{-4}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
| #epochs | 30 | 10 | 10 |

## 5. CORESET VISUALIZATION

In the main paper, we plotted t-SNE [15] embeddings for each dataset and coreset method for 5% dataset fraction. In this supplementary document, Figures 1 to 3 present the results for dataset fractions of 1%, 10%, and 30% to further demonstrate the performance of PGCS in selecting a representative and diverse coreset across a range of dataset sizes. The selected coreset points are marked in red to clearly indicate the chosen data points for each method. The saturated red regions indicate areas with a higher density of selected points. PGCS exhibits the widest distribution of coreset points with minimal overlap, as illustrated in the t-SNE plots. This distinct spread and reduced overlap highlight PGCS's capability to effectively select diverse and representative data points, minimizing redundancy and ensuring that the coreset used for training the IQA architecture effectively captures the key variations in the dataset.

## 6. REFERENCES

[1] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang, "Musiq: Multi-scale image quality transformer," in *IEEE CVPR*, 2021, pp. 5148–5157.

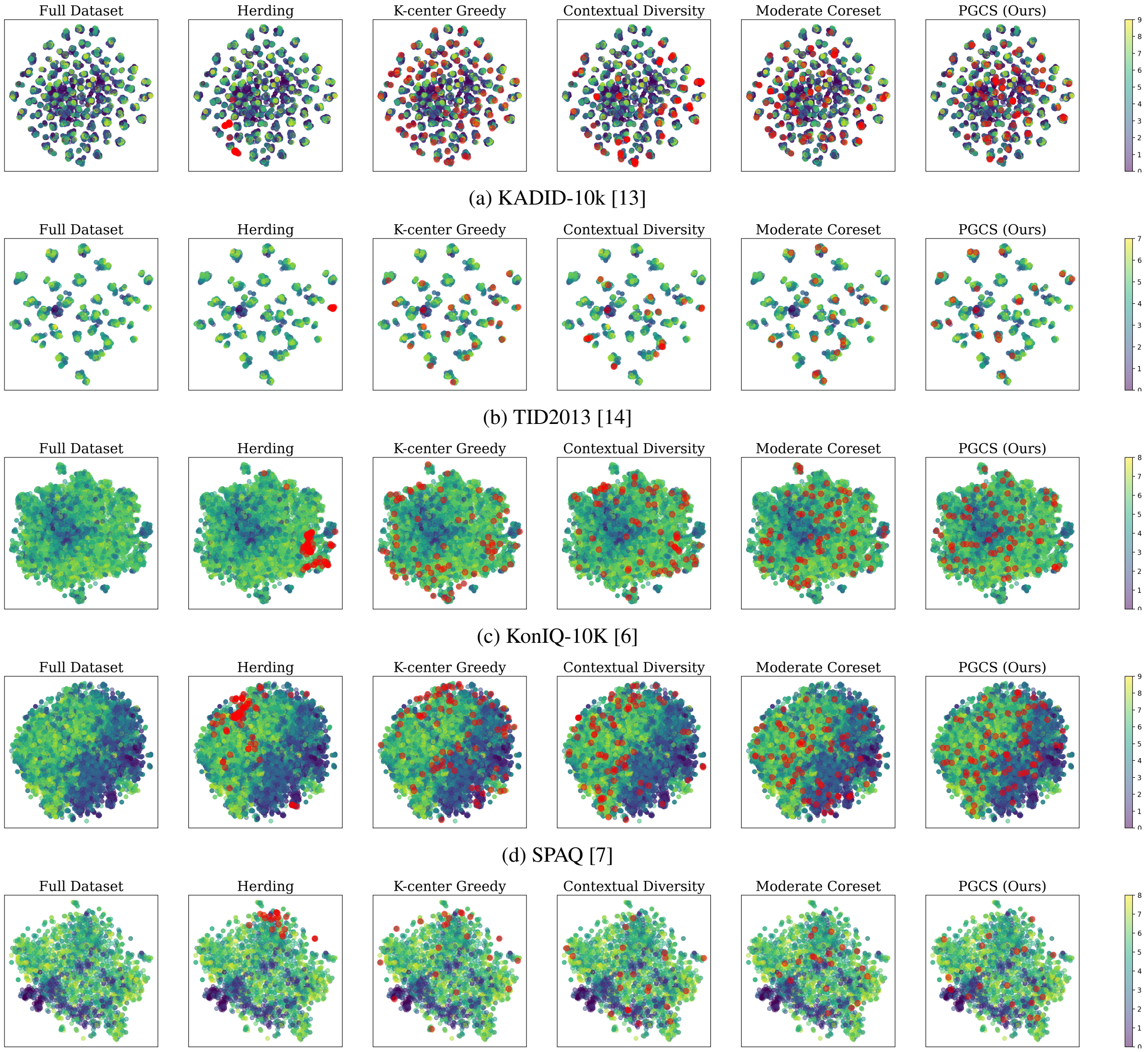
[2] Y. Chen, M. Welling, and A. Smola, "Super-samples from kernel herding," in *UAI*, 2010, pp. 109–116.

[3] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," in *ICLR*, 2018.

[4] S. Agarwal, H. Arora, S. Anand, and C. Arora, "Contextual diversity for active learning," in *ECCV*. Springer, 2020, pp. 137–153.

[4] https://github.com/IIGROUP/MANIQA

[5] https://github.com/anse3832/MUSIQ



(a) KADID-10k [13]

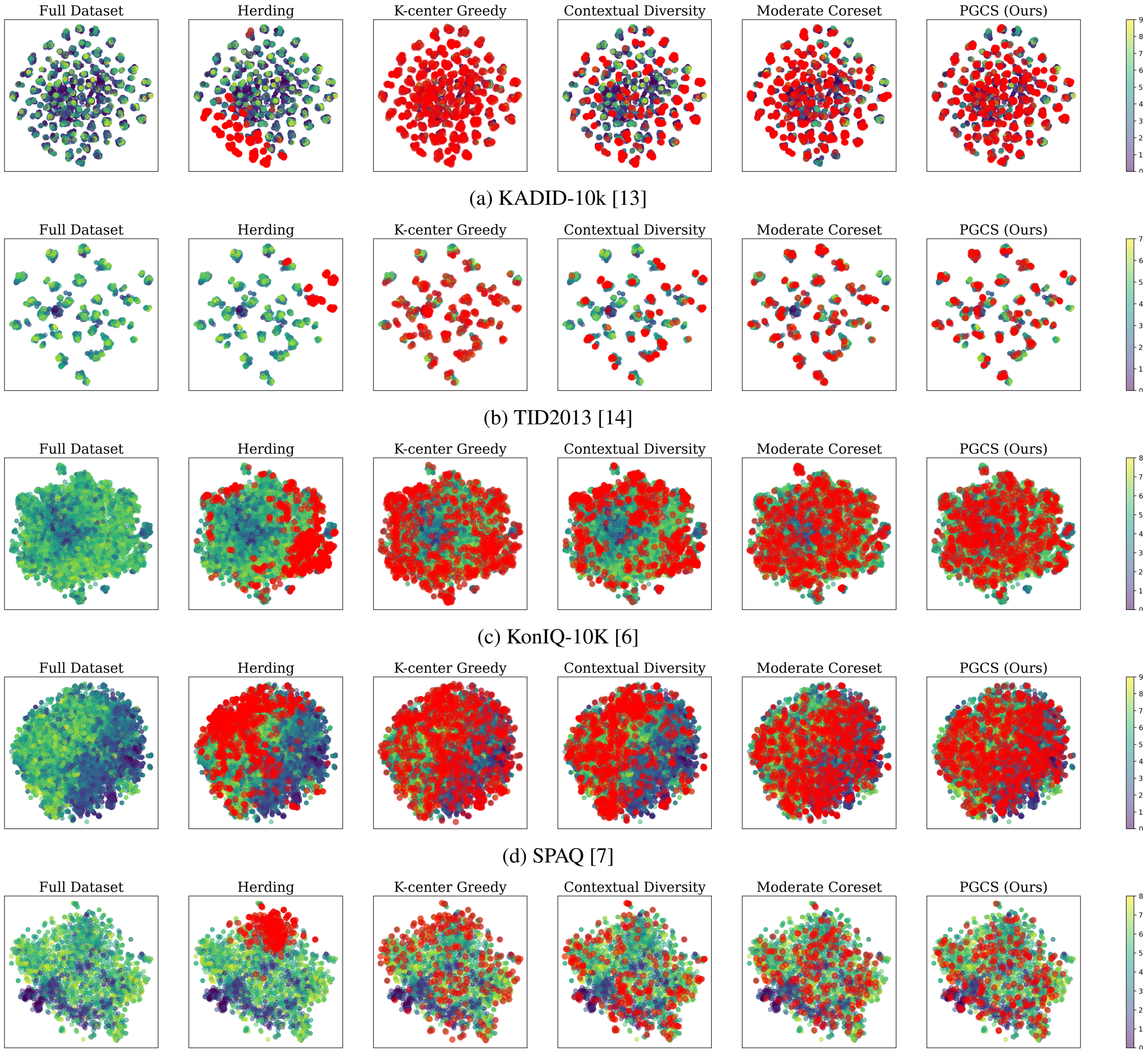
(b) TID2013 [14]

(c) KonIQ-10K [6]

(d) SPAQ [7]

(e) AGIQA-3K [8]

**Fig. 1**: t-SNE embeddings for each dataset and coreset method. Coreset (Dataset Fraction = 1% ) highlighted in red to illustrate selected data points. The saturated red regions indicate areas with a higher density of points, highlighting the concentration of selected points in the same location. The colorbar (on the right) shows the MOS score range for the dataset.



(a) KADID-10k [13]

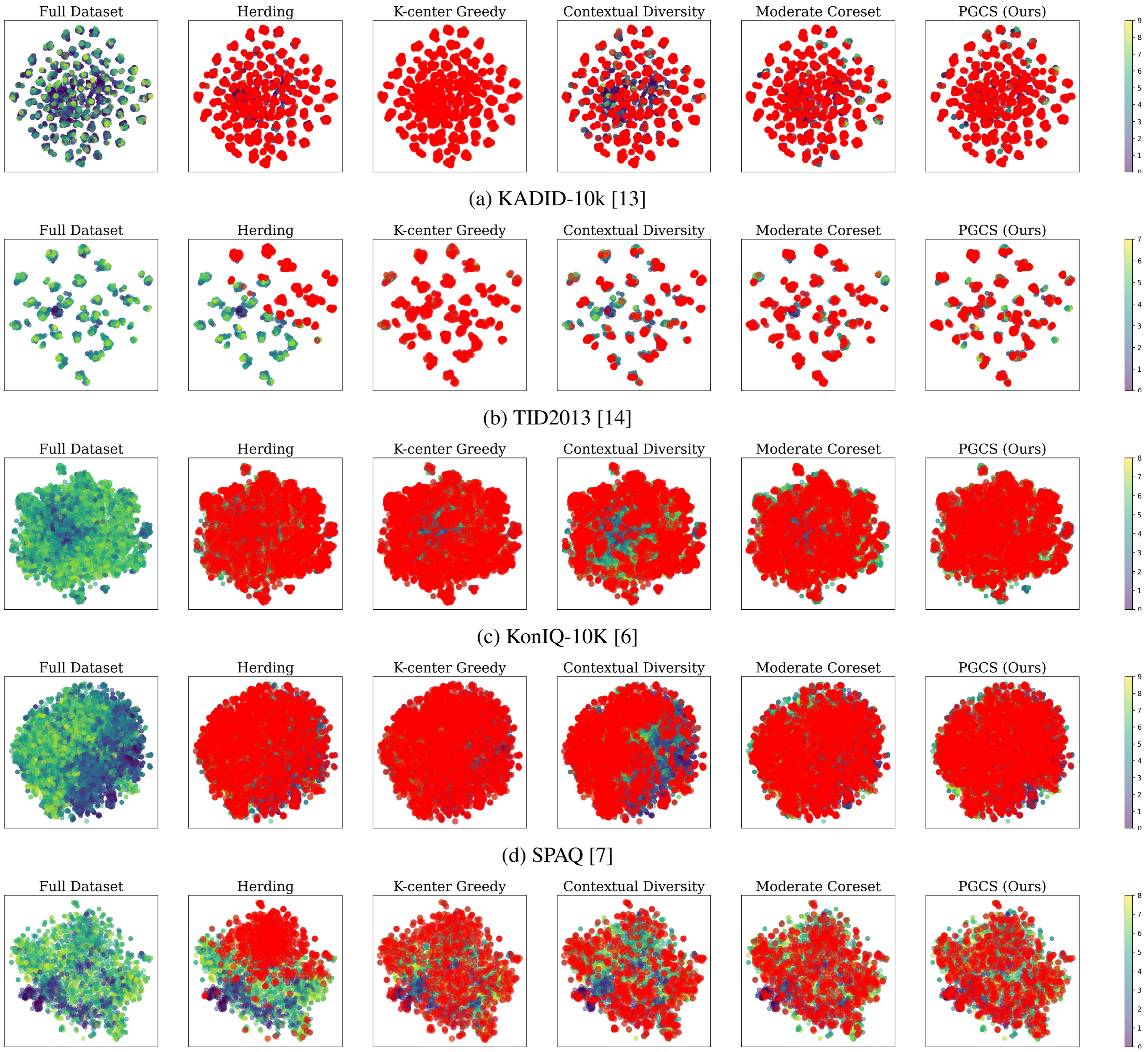
(b) TID2013 [14]

(c) KonIQ-10K [6]

(d) SPAQ [7]

(e) AGIQA-3K [8]

**Fig. 2**: t-SNE embeddings for each dataset and coreset method. Coreset (Dataset Fraction = 10% ) highlighted in red to illustrate selected data points. The saturated red regions indicate areas with a higher density of points, highlighting the concentration of selected points in the same location. The colorbar (on the right) shows the MOS score range for the dataset.



(a) KADID-10k [13]

(b) TID2013 [14]

(c) KonIQ-10K [6]

(d) SPAQ [7]

(e) AGIQA-3K [8]

**Fig. 3**: t-SNE embeddings for each dataset and coreset method. Coreset (Dataset Fraction = 30% ) highlighted in red to illustrate selected data points. The saturated red regions indicate areas with a higher density of points, highlighting the concentration of selected points in the same location. The colorbar (on the right) shows the MOS score range for the dataset.

[5] X. Xia, J. Liu, J. Yu, X. Shen, B. Han, and T. Liu, “Moderate coreset: A universal method of data selection for real-world data-efficient deep learning,” in *ICLR*, 2022.

[6] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, “Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment,” *IEEE TIP*, vol. 29, pp. 4041–4056, 2020.

[7] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, “Perceptual quality assessment of smartphone photography,” in *IEEE CVPR*, 2020, pp. 3677–3686.

[8] C. Li, Z. Zhang, H. Wu, W. Sun, X. Min, X. Liu, G. Zhai, and W. Lin, “Agiqa-3k: An open database for ai-generated image quality assessment,” *IEEE TCSVT*, 2023.

[9] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, and Y. Yang, “Maniqa: Multi-dimension attention network for no-reference image quality assessment,” in *IEEE CVPR*, 2022, pp. 1191–1200.

[10] W. Zhang, G. Zhai, Y. Wei, X. Yang, and K. Ma, “Blind image quality assessment via vision-language correspondence: A multitask learning perspective,” in *IEEE CVPR*, 2023, pp. 14071–14081.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE CVPR*, 2016, pp. 770–778.

[12] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *ICLR*, 2015.

[13] H. Lin, V. Hosu, and D. Saupe, “Kadid-10k: A large-scale artificially distorted iqa database,” in *IEEE QoMEX*, 2019, pp. 1–3.

[14] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., “Image database tid 2013: Peculiarities, results and perspectives,” *SPIC*, vol. 30, pp. 57–77, 2015.

[15] G. Hinton and S. Roweis, “Stochastic neighbor embedding,” *NeurIPS*, vol. 15, 2002.