# Few-Shot Image Quality Assessment via Adaptation of Vision-Language Models - Supplementary Material -

## 1. Appendix Overview

The supplementary material is organized as follows: Sec. 2 shows more training details in various BIQA settings. Sec. 3 provides more performance cooperation on the AI-generated BIQA dataset. Sec. 4 provides Zero-shot performance cooperation with other zero-shot BIQA methods. Sec. 5 provides more ablation and performance comparisons. Sec. 6 offers a detailed qualitative analysis of the gMAD competition.

## 2. Training Details

In the fully supervised setting, we provide additional details on the training preprocessing for various datasets—LIVE [22], CSIQ [11], TID2013 [20], KADID [14], LIVEC [3], KonIQ [8], LIVEFB [30], and SPAQ [2]—in Table 1, which were not included in the main paper. We used different training settings for each benchmark to ensure a fair comparison. In the data-efficient learning setup, no additional data enhancements were applied. The training images were simply resized to a resolution of 224×224, with the number of visual prompts set to 100 and the length of learnable vectors for text prompts set to 4.

| Dataset | Resolution | Resize | Batch Size | Label Range |
|---|---|---|---|---|
| LIVE | $768 \times 512$ | $512 \times 384$ | 12 | DMOS [0,100] |
| CSIQ | $512 \times 512$ | $512 \times 512$ | 12 | DMOS [0,1] |
| TID2013 | $512 \times 384$ | $512 \times 384$ | 48 | MOS [0,9] |
| KADID | $512 \times 384$ | $512 \times 384$ | 128 | MOS [1,5] |
| LIVEC | $500P \sim 640P$ | $500P \sim 640P$ | 16 | MOS [1,100] |
| KonIQ | $768P$ | $512 \times 384$ | 128 | MOS [0,5] |
| LIVEFB | $160P \sim 700P$ | $512 \times 512$ | 128 | MOS [0,100] |
| SPAQ | $1080P \sim 4368P$ | $512 \times 384$ | 128 | MOS [0,100] |

Table 1. Training preprocessing details of BIQA datasets.

## 3. Results on AI-generated BIQA dataset

### 3.1. Dataset Protocol

**AGIQA-1K:** The AGIQA-1K [33] dataset is the inaugural collection specifically created for assessing the perceptual quality of Artificial General Intelligence (AGI) and includes 1,080 images produced by diffusion models. Researchers gathered subjective quality ratings for these images through experimental evaluations and conducted benchmarks to test how well current image quality assessment models perform. The dataset features a diverse set of images depicting various subjects, including birds, cats, bats, children, and adults, showcasing its variety and complexity.

**AGIQA-3K:** The AGIQA-3K [12] dataset is an open database designed for the assessment of AI-generated image quality, comprising 2,982 images produced by six distinct models. The test model is divided into three groups: Loss-function models, SVR-based models, and DL-based models. These models use various methods to assess image quality. The DL-based group includes the latest deep learning metrics such as DBCNN [32], CLIPIQA [26], CNNIQA [9], and HyperNet [24], which characterize quality perception information by training deep neural networks. The dataset is randomly split into an 80/20 training and testing set, ensuring that images with the same object labels are grouped. The evaluation process is repeated 10 times to reduce performance bias. We report the average of SRCC, KRCC and PLCC to quantify the model's performance in terms of prediction accuracy and monotonicity.

### 3.2. Experiment Results

In this section, we showcase the performance of our model on the recent aigc-iqa dataset. Specifically, we chose the $\lambda$ value of 0.5 and conducted ten experiments to obtain the average while also reporting the performance of mainstream methods on the AGIQA-1K and AGIQA-3K subsets. For the comparison models, we either directly use the publicly available implementation results or re-train them on our datasets using the open-source training codes.

**Results on AGIQA-1K database.** Table 2 displays the performance of our method alongside existing methods on the AGIQA-1K database and two subsets. The handcrafted IQA methods do not perform satisfactorily in assessing AGIs due to the prior knowledge from NSIs being unsuitable for AI-generated BIQA datasets, highlighting the disparities between scenes. Our framework reaches SOTA performance under existing methods, demonstrating the superiority of our proposed meta-learning in rapidly adapting

| Metric | Database | ALL | | | stable-inpainting-v1 | | | stable-diffusion-v2 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Corr | SRCC | KRCC | PLCC | SRCC | KRCC | PLCC | SRCC | KRCC | PLCC |
| Hand crafted- based | BMPRI [17] | 0.0651 | 0.0400 | 0.1646 | 0.3746 | 0.2643 | 0.4094 | -0.0158 | -0.0112 | -0.0111 |
| | CEIQ [29] | 0.3069 | 0.2097 | 0.2836 | 0.2348 | 0.1607 | 0.2000 | 0.1314 | 0.0898 | 0.1392 |
| | DSIQA [19] | -0.3047 | -0.2148 | -0.0559 | 0.0428 | 0.0241 | 0.4106 | 0.0046 | 0.0041 | 0.0184 |
| | NIQE [18] | -0.5490 | -0.3824 | -0.5048 | 0.0414 | 0.0240 | 0.0712 | -0.2275 | -0.1564 | -0.2392 |
| Handcrafted& SVR-based | friquee [4] | 0.4938 | 0.3469 | 0.4192 | 0.4231 | 0.3024 | 0.3989 | 0.1783 | 0.1244 | 0.2069 |
| | GMLF [28] | 0.5575 | 0.4052 | 0.6356 | 0.5062 | 0.3649 | 0.6167 | 0.1501 | 0.1039 | 0.1713 |
| | HIGRADE [10] | 0.4056 | 0.2860 | 0.4425 | 0.2493 | 0.1732 | 0.2886 | 0.1358 | 0.0943 | 0.1308 |
| | NFERM [6] | 0.4540 | 0.3224 | 0.5396 | 0.3874 | 0.2743 | 0.4901 | 0.1193 | 0.0817 | 0.1474 |
| | NFSDM [5] | 0.4314 | 0.3055 | 0.4714 | 0.3840 | 0.2743 | 0.4576 | 0.1002 | 0.0690 | 0.0911 |
| Deep learning -based | ResNet50 [7] | 0.6365 | 0.4777 | 0.7323 | 0.6000 | 0.4485 | 0.7728 | 0.3961 | 0.2785 | 0.4739 |
| | StairIQA [25] | 0.5504 | 0.4039 | 0.6088 | 0.4669 | 0.2519 | 0.5050 | 0.3486 | 0.2519 | 0.4186 |
| | MGQA [27] | 0.6011 | 0.4456 | 0.6760 | 0.5618 | 0.4250 | 0.7206 | 0.3715 | 0.2584 | 0.3593 |
| | DEIQT [21] | 0.8309 | 0.6475 | 0.8651 | 0.7791 | 0.5975 | 0.8164 | 0.5129 | 0.3617 | 0.5345 |
| | GMRP-IQA (ours) | **0.8354** | **0.6614** | **0.8883** | **0.8168** | **0.6366** | **0.8386** | **0.6699** | **0.4848** | **0.6785** |

Table 2. Performance results on the AGIQA-1k database and two different generative model subsets. The bold entries indicate the best results, and underlines indicate the second-best.

| Metric | ALL | | | Bad Model | | | Medium Model | | | Good Model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SRCC | KRCC | PLCC | SRCC | KRCC | PLCC | SRCC | KRCC | PLCC | SRCC | KRCC | PLCC |
| DSIQA [19] | 0.4955 | 0.3403 | 0.5488 | 0.1908 | 0.1331 | 0.3139 | 0.2140 | 0.1469 | 0.3655 | 0.1665 | 0.1120 | 0.2520 |
| NIQE [18] | 0.5623 | 0.3876 | 0.5171 | 0.2031 | 0.1354 | 0.3309 | 0.2259 | 0.1483 | 0.2526 | 0.1750 | 0.1172 | 0.2533 |
| DBCNN [32] | 0.8207 | 0.6336 | 0.8759 | 0.5520 | 0.3958 | 0.6825 | 0.5011 | 0.3531 | 0.5575 | 0.4288 | 0.2975 | 0.4853 |
| CLIPIQA [26] | 0.8426 | 0.6468 | 0.8053 | 0.1882 | 0.1255 | 0.2549 | 0.6537 | 0.4693 | 0.6014 | 0.5038 | 0.3407 | 0.5081 |
| CNNIQA [9] | 0.7478 | 0.5580 | 0.8469 | 0.3233 | 0.2275 | 0.4547 | 0.4278 | 0.2807 | 0.4534 | 0.3952 | 0.2805 | 0.4517 |
| HyperNet [24] | 0.8355 | 0.6488 | 0.8903 | 0.5086 | 0.3628 | 0.5985 | 0.4687 | 0.3260 | 0.5480 | 0.5562 | 0.3927 | 0.6149 |
| DEIQT [21] | 0.8501 | 0.6684 | 0.9054 | 0.6449 | 0.4785 | 0.7363 | 0.5664 | 0.4033 | 0.7054 | 0.7886 | 0.5998 | 0.8420 |
| GRMP-IQA | **0.8799** | **0.7039** | **0.9202** | **0.6824** | **0.5129** | **0.7533** | **0.6742** | **0.4917** | **0.7909** | **0.8101** | **0.6273** | **0.8587** |

Table 3. Perceptual metric performance results on AGIQA-3k database and different subsets of different T2I AGI models. The bold entries indicate the best results, and underlines indicate the second-best.

CLIP to various IQA scenarios. Moreover, in the stable-diffusion-v2 subset, which utilizes more keywords, we substantially surpass the mainstream method DEIQT [21], indicating that our proposed gradient-regulated framework is better tailored for downstream IQA tasks rather than being confined to generalized semantic knowledge. Therefore, GRMP-IQA surpasses all other methods in terms of performance despite the wide range of image content and distortion types.

**Results on AGIQA-3K database.** For the AGIQA-3k dataset, to analyze the evaluation consistency of perception models, we categorize AI-generated images into three groups: bad, medium, and good models. The specific performances are presented in Table 3. Compared to traditional handcrafted feature-based methods such as DSIQA [19] and NIQE [18], deep learning models more closely mirror the human visual system, with our method outperforms all other deep learning models. Notably, CLIP-IQA [26] is challenged in performance when dealing with more distortion in the Bad Model subset, reflecting how CLIP's general knowledge can misclassify quality knowledge. DBCNN [32], CNNIQA [9], and HyperNet [24] all perform equally poorly in the Medium Model subset, indicating that traditional CNN-based feature extraction methods are ineffective at learning quality information from images with subtle quality traits. Although DEIQT extracts features more effectively, transformer-based methods are designed for classification rather than quality information. Overall, our proposed GMRP-IQA framework effectively addresses these issues.

## 4. Zero-Shot Performance

Zero-shot methods refer to training that occurs without human supervision. Consequently, we compared the performance of our GMRP-IQA with unsupervised BIQA

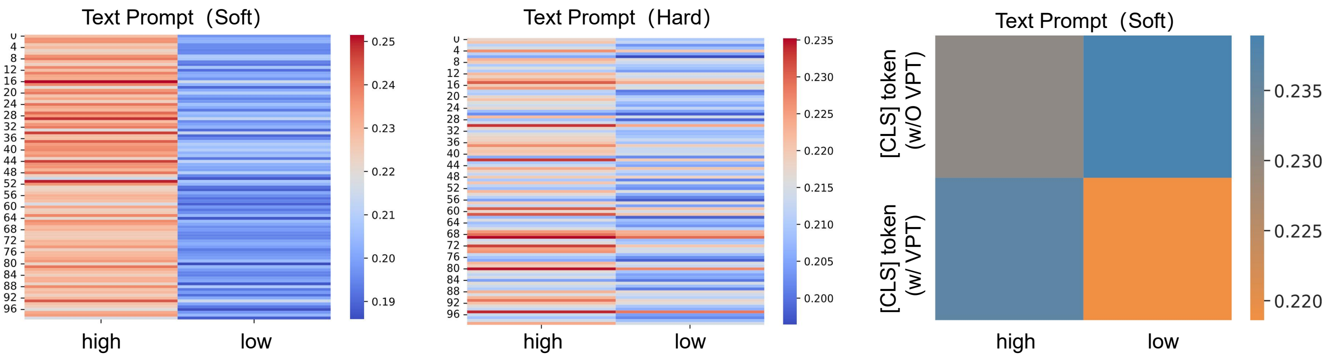

Figure 1. Cosine similarity between visual and textual prompts under different perspectives: (a) and (b) analyze from a visual-to-text perspective, and (c) from a text-to-visual perspective. After implementing visual-text prompts, there has been an enhancement in the model's ability to discriminate image quality, highlighting that the integration of Visual-Text Prompts is harmonious and efficient.

| Method | LIVEC | KonIQ | CSIQ | LIVE | PIPAL |
|---|---|---|---|---|---|
| NIQE [18] | 0.463 | 0.530 | 0.613 | 0.836 | 0.153 |
| IL-NIQE [31] | 0.440 | 0.507 | **0.814** | 0.847 | 0.282 |
| CL-MI [1] | 0.507 | 0.645 | 0.588 | 0.663 | 0.303 |
| CLIP-IQA [26] | 0.612 | 0.700 | 0.690 | 0.652 | 0.261 |
| GRepQ [23] | 0.740 | **0.768** | 0.693 | 0.741 | 0.436 |
| GRMP-IQA | **0.770** | 0.713 | 0.781 | **0.911** | **0.450** |

Table 4. SRCC performance comparison of GRMP-IQA (zero-shot) with other zero-shot methods on various IQA databases.

methods such as NIQE [18], IL-NIQE [31], contrastive learning based on mutual information (CL-MI) [1] and CLIP-IQA [26], alongside the self-supervised pre-trained GRepQ [23]. We evaluated all methods using the entire assessment database. As Table 4 indicates, GMRP-IQA ranked in the top two across five datasets, notably outperforming other methods by a significant margin on the LIVEC and LIVE datasets. Particularly, it demonstrated a notable performance lead over the similarly Vision-Language Model-based CLIP-IQA method, with improvements reaching up to **25.9** percentage points. GMRP-IQA also achieved state-of-the-art (SOTA) performance on the challenging PIPAL dataset, which includes various distortions, especially images that have undergone super-resolution and denoising through various restoration methods, including those based on GANs. It is important to note that the model performed slightly less effectively on the larger real-world dataset, KonIQ. We attribute this to the pre-trained CLIP model's inherent bias towards semantic information, which still affects its ability to assess image quality in a zero-shot setting. This underscores the value of our proposed quality-aware gradient regularization (QGR) during few-sample fine-tuning, with QGR ablation studies in our manuscript confirming its significant gains on KonIQ.

# 5. More Ablation Results

## 5.1. Effect of Visual-Text Prompt

Considering the domain differences between upstream pre-training tasks and downstream IQA tasks, the image encoder tends to encode semantic information rather than quality-related information. This inclination can potentially affect the alignment with the textual component, thereby impacting the model's final quality prediction, as the ultimate quality assessment is inferred collaboratively by the text and image branches. To address this, we introduce a visual prompt to complement the text prompt. In this section, we further analyze their synergistic relationship.

As illustrated in Figure 1, we analyze the cosine similarity between visual and textual prompts from two perspectives: (1) from a visual-to-text direction and (2) from a text-to-visual direction. For the first perspective (Fig. 1 (a) and (b)), we visualized the similarity matrix between the visual soft prompts and both soft and hard textual prompts. It is evident that the quality discrimination produced by the hard textual prompts in conjunction with learnable visual prompts is ambiguous, meaning that the prediction probabilities for high and low image quality are extremely close. In contrast, soft textual prompts, when used together with visual prompts, can yield more discriminative quality judgments. For the second perspective (Figure 1 (c)), we visualized the similarity between the [cls] token with and without visual soft prompts in response to soft text prompts. The observations suggest a similar phenomenon where, without the support of visual soft prompts, the [cls] token struggles to adapt to soft text prompts. This further corroborates the need for appropriate adjustments in the visual encoder to address the domain discrepancies between upstream and downstream tasks. In summary, our Visual-Text Prompt integration is harmonious and efficient, significantly enhancing the ability of the CLIP model to adapt to IQA tasks.

| Prompt Length | LIVEC | | KonIQ | |
|---|---|---|---|---|
| | PLCC | SRCC | PLCC | SRCC |
| M=4 | 0.867 | 0.836 | **0.880** | **0.853** |
| M=16 | **0.874** | **0.841** | 0.870 | 0.842 |

Table 5. Performance comparison for different prompt lengths.

| Method | Training Ratio | SPAQ | | KonIQ | |
|---|---|---|---|---|---|
| | | PLCC | SRCC | PLCC | SRCC |
| Q-Align | 20% | 0.911 | 0.909 | 0.901 | 0.903 |
| Ours | 20% | **0.925** | **0.920** | **0.931** | **0.915** |
| Q-Align | 80% | **0.933** | **0.930** | 0.941 | **0.940** |
| Ours | 80% | 0.932 | 0.927 | **0.945** | 0.934 |

Table 6. Performance comparison with Q-Align in terms of PLCC and SRCC under varying training data ratios.

### 5.2. Ablation of text prompt length

Drawing on empirical findings [13, 34], we uniformly set the prompt length to 4 (M=4) to balance performance and efficiency, as shown in Table 5. The results reveal that the optimal prompt length varies across datasets, indicating that increasing M from 4 to 16 improves performance on the LIVEC dataset, yet reduces KonIQ's PLCC. This divergence suggests dataset-specific adaptability of prompt length. While M=4 offers a balanced choice, adjusting it per dataset could optimize performance further.

### 5.3. Comparsion with the Q-Align.

Our method differs from Q-Align in two key aspects. First, while Q-Align fine-tunes numerous parameters (including the LLM), we use prompt tuning—this significantly reduces the data needed to adapt large models to IQA tasks. Second, we explicitly explore the connection between IQA and high-level vision tasks, which further alleviates semantic overfitting. Notably, GRMP-IQA outperforms Q-Align when using just 20% of the training data, and remains competitive even with the full dataset. Detailed results are reported in Table 6.

## 6. Qualitative Analysis

To further assess our framework's generalization, we trained models on the entire LIVE database and then tested them using the gMAD competition [16] on the Waterloo Database [15]. gMAD efficiently selects image pairs with maximum quality difference predicted by an attacking IQA model to challenge another defending model which considers them to be of the same level of quality. The selected pairs are shown to the observer to determine whether the attacker or the defender is robust. As shown in Fig. 2, In the first two columns, our model attacks the competing method DEIQT, where each column represents images chosen from the poorer and better quality levels predicted by the defender. In the last two columns, we fix our model as the defender, giving image pairs selected from poorer and better quality levels, respectively. From Fig. 2, it is evident that when our model serves as the defender, the image pairs chosen by the attacker show little perceptual quality change, whereas, as the attacker, our model selects image pairs with more significant quality differences in succession. This indicates that the model has strong defensive and offensive capabilities. Additionally, it is important to highlight that the image pairs in the second column, which share similar semantic information, misled the DEIQT into classifying them as similar quality. Conversely, our model effectively identified the quality differences between them. These findings underscore the strong generalization capability of our model in tackling complex distortions in real-world images.

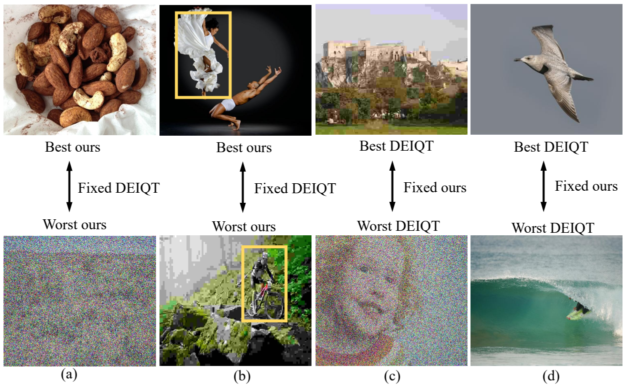

Figure 2. gMAD results between DEIQT [21] and ours. (a) Fixed DEIQT at low quality. (b) Fixed DEIQT at high quality. (c) Fixed ours at low quality. (d) Fixed ours at high quality.

## References

[1] Nithin C Babu, Vignesh Kannan, and Rajiv Soundararajan. No reference opinion unaware quality assessment of authentically distorted images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2459–2468, 2023. 3

[2] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. Perceptual quality assessment of smartphone photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3677–3686, 2020. 1

[3] Deepti Ghadiyaram and Alan C Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2015. 1

[4] Deepti Ghadiyaram and Alan C Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. *Journal of vision*, 17(1):32–32, 2017. 2

[5] Ke Gu, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Longfei Liang. No-reference image quality assessment metric by combining free energy theory and structural degradation model. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2013. 2

[6] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Using free energy principle for blind image quality assessment. *IEEE Transactions on Multimedia*, 17(1):50–63, 2014. 2

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[8] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 29:4041–4056, 2020. 1

[9] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1733–1740, 2014. 1, 2

[10] D Kundu, D Ghadiyaram, AC Bovik, and BL Evans. Large-scale crowdsourced study for high dynamic range images. *IEEE Trans. Image Process*, 26(10):4725–4740, 2017. 2

[11] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19 (1):011006, 2010. 1

[12] Chunyi Li, Zicheng Zhang, Haoning Wu, Wei Sun, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai, and Weisi Lin. Agiqa-3k: An open database for ai-generated image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 1

[13] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8094–8103, 2023. 4

[14] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2019. 1

[15] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26 (2):1004–1016, 2016. 4

[16] Kede Ma, Qingbo Wu, Zhou Wang, Zhengfang Duanmu, Hongwei Yong, Hongliang Li, and Lei Zhang. Group mad competition-a new methodology to compare objective image quality models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1664–1673, 2016. 4

[17] Xiongkuo Min, Guangtao Zhai, Ke Gu, Yutao Liu, and Xiaokang Yang. Blind image quality estimation via distortion aggravation. *IEEE Transactions on Broadcasting*, 64 (2):508–517, 2018. 2

[18] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 2, 3

[19] Niranjan D. Narvekar and Lina J. Karam. A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In *2009 International Workshop on Quality of Multimedia Experience*, pages 87–91, 2009. 2

[20] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Image database tid2013: Peculiarities, results and perspectives. *Signal processing: Image communication*, 30:57–77, 2015. 1

[21] Guanyi Qin, Runze Hu, Yutao Liu, Xiawu Zheng, Haotian Liu, Xiu Li, and Yan Zhang. Data-efficient image quality assessment with attention-panel decoder. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, 2023. 2, 4

[22] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006. 1

[23] Suhas Srinath, Shankhanil Mitra, Shika Rao, and Rajiv Soundararajan. Learning generalizable perceptual representations for data-efficient no-reference image quality assessment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 22–31, 2024. 3

[24] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3667–3676, 2020. 1, 2

[25] Wei Sun, Huiyu Duan, Xiongkuo Min, Li Chen, and Guangtao Zhai. Blind quality assessment for in-the-wild images via hierarchical feature fusion strategy. In *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 01–06. IEEE, 2022. 2

[26] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023. 1, 2, 3

[27] Tao Wang, Wei Sun, Xiongkuo Min, Wei Lu, Zicheng Zhang, and Guangtao Zhai. A multi-dimensional aesthetic quality assessment model for mobile game images. In *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–5. IEEE, 2021. 2

[28] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Transactions on Image Processing*, 23(11):4850–4862, 2014. 2

[29] Jia Yan, Jie Li, and Xin Fu. No-reference quality assessment of contrast-distorted images using contrast enhancement. *arXiv preprint arXiv:1904.08879*, 2019. 2

[30] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to

pictures (paq-2-piq): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3575–3585, 2020. 1

[31] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 3

[32] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1):36–47, 2018. 1, 2

[33] Zicheng Zhang, Chunyi Li, Wei Sun, Xiaohong Liu, Xiongkuo Min, and Guangtao Zhai. A perceptual quality assessment exploration for aigc images. In *2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 440–445. IEEE, 2023. 1

[34] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337–2348, 2022. 4