Towards Efficient Vision Transformers for Perceptual Quality Assessment of Diffusion-Generated Images

Shivam Bhardwaj and Tushar Shinde
Indian Institute of Technology Madras Zanzibar
shinde@iitmz.ac.in

Abstract

The rapid rise of Artificial Intelligence Generated Content (AIGC) demands scalable and efficient Image Quality Assessment (IQA) techniques, particularly for deployment in resource-constrained environments. Our approach leverages on the foundation models such as CLIP and DINO, but their computational overhead limits practical usage. In this work, we present a novel framework for efficient IQA by adaptively compressing Large Vision Models (LVMs) using layer-wise pruning and mixed-precision quantization guided by layer importance scores. Our method dynamically reduces model size while preserving perceptual sensitivity critical to assessing generative artifacts. We evaluate our approach on two dedicated AIGC IQA benchmark datasets: AGIQA-1K and AGIQA-3K and demonstrate up to 95% reduction in model size with minimal drop in human-perception correlation metrics such as SRCC and PLCC. Our method sets a new state-of-the-art in compact foundation models for generative content quality assessment and offers a scalable path toward real-time AIGC monitoring on edge platforms.

1. Introduction and Related Work

Recent advances in generative models, including Generative Adversarial Networks (GANs) [6] and diffusion-based approaches [4, 10], have enabled the synthesis of high-quality visual content. This progress has accelerated the emergence of AI-generated content (AIGC) across various modalities such as text, image, audio, and video [1, 30]. Within this domain, AI-generated images (AGIs) have gained particular attention due to their expressive potential and broad applicability. However, unlike natural scene images (NSIs) that conform to physical imaging constraints, AGIs often exhibit semantic anomalies and visually implausible textures [36], challenging traditional assumptions in computer vision.

Evaluating the perceptual quality of AGIs is crucial for both benchmarking generative models and ensuring visual consistency in downstream tasks. Conventional full-reference metrics such as PSNR and SSIM [20, 29] rely on pixel-wise fidelity and are inadequate for AGIs, which demand high-level assessments of semantic alignment, aesthetics, and generative artifacts [22]. In response, datasets like AGIQA-1K [36] and AGIQA-3K [16] have been curated with Mean Opinion Scores (MOS) to reflect human perceptions of AGI-specific distortions.

To leverage such annotations, deep learning-based IQA models have been developed, utilizing data-driven features for perceptual prediction [5, 11, 13]. These models typically extract features $\mathbf{f}(I)$ from images I and map them to perceptual scores $\hat{s} = g(\mathbf{f}(I))$. However, these methods often remain optimized for NSIs and fail to generalize to the unique distortions of AGIs. Classical no-reference methods like BRISQUE [18] and NIQE [19] also fall short, as AGIs frequently violate natural scene statistics due to stylistic and semantic deviations.

Recent works have started exploring dedicated IQA models for AGIs. For instance, Zhang et al. [36] proposed an MOS-correlated framework capturing AI-induced artifacts and perceptual unnaturalness. Other efforts [27, 33, 34] have employed task-specific architectures or foundation models, yet many still rely on handcrafted or pixel-based loss functions, limiting their semantic understanding in the absence of reference images.

Vision Transformers (ViTs), known for their global context modeling and superior performance in image classification [3], are underexplored in AGI IQA. While ViTs extract hierarchical features through self-attention and have shown promise in full-reference IQA, their adaptation to AGI scenarios is hindered by domain gaps and computational costs. Foundation models like CLIP [23] and DINO [2] offer semantically rich representations and remain promising yet underutilized for perceptual modeling in synthetic image domains.

Furthermore, the deployment of AGI models on resource-limited devices underscores the need for compact and efficient IQA models. Model compression techniques,

such as pruning and quantization [8], have become essential for reducing memory and computation overhead. Pruning eliminates less important parameters based on saliency, while quantization maps high-precision weights to lower-bit representations, enabling faster and smaller models without significant loss in performance.

Contributions. We propose a lightweight and semantically-aware IQA framework tailored for AGIs, with the following key contributions:

- We introduce an AGI-specific IQA framework that leverages frozen Vision Transformers (CLIP [23], DINO [2]) paired with a lightweight regression head, enabling semantically-informed quality prediction with minimal computational cost.
- We incorporate a layer-wise compression scheme combining pruning and quantization, guided by each layer's contribution to perceptual quality. This approach achieves substantial model compression with negligible performance loss.
- Extensive experiments on AGIQA-1K and AGIQA-3K demonstrate state-of-the-art correlation with human judgments. Our compressed models retain high accuracy while significantly reducing model size, making them ideal for practical AGI evaluation.

The rest of this paper is organized as follows: Section 2 introduces our IQA framework and adaptive compression technique. Section 3 details the experimental setup. Section 4 reports quantitative and qualitative results. Finally, Section 5 concludes the paper and outlines future research directions.

2. Methodology

We propose an efficient Image Quality Assessment (IQA) framework for AI-Generated Content (AIGC) by leveraging pre-trained Large Vision Models (LVMs) such as CLIP and DINO. These serve as fixed feature extractors, followed by a lightweight, trainable Multi-Layer Perceptron (MLP) regressor for perceptual quality prediction. To enable deployment on resource-constrained devices, we introduce a model compression approach inspired by [24] and driven by the *Quality-preserving Layer Importance Score* (QLIS), which guides both pruning and quantization in an adaptive, layer-wise manner.

2.1. Model Architecture

Given an input image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$, the frozen backbone f_θ (CLIP or DINO) extracts high-level features:

$$\mathbf{z} = f_\theta(\mathbf{I}) \in \mathbb{R}^d \tag{1}$$

where d is the output dimensionality of the backbone (e.g., $d = 768$ for CLIP). The features $\mathbf{z}$ are then passed through a trainable MLP regressor g_ϕ:

$$\hat{q} = g_\phi(\mathbf{z}) \tag{2}$$

producing a scalar quality prediction $\hat{q} \in \mathbb{R}$. The MLP is optimized using Mean Squared Error (MSE) loss against ground-truth subjective scores:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{q}_i - q_i^{\text{true}}\right)^2 \tag{3}$$

2.2. Quality-preserving Layer Importance Score

To facilitate perceptually-aware compression of the vision backbone, we introduce a Quality-preserving Layer Importance Score (QLIS) for each layer l. QLIS integrates structural, statistical, and information-theoretic features to estimate the importance of individual layers. Inspired by [24], this metric helps identify layers that are perceptually more significant and should therefore be retained with higher fidelity during compression.

The structural importance is captured by the parameter proportion P_l, which denotes the relative number of parameters in layer l compared to the entire model. Information diversity is quantified through normalized entropy E_l, derived from Shannon entropy over quantized weight distributions. Sparsity, indicating the compressibility of a layer, is captured using normalized sparsity S_l, which reflects the proportion of near-zero weights in the layer. To unify entropy and sparsity, we define the Entropy-Weighted Density Score (EWDS), which assigns higher scores to dense and information-rich layers. Finally, QLIS combines structural significance and compressibility using a tunable parameter $\beta \in [0, 1]$, balancing parameter proportion and density. A higher QLIS value implies greater perceptual relevance and guides the selection of pruning and quantization parameters for that layer. The equations defining the QLIS computation are as follows:

$$P_l = \frac{|\theta_l|}{\sum_j |\theta_j|} \tag{4}$$

$$E_l = \frac{H_l}{B} \tag{5}$$

$$S_l = \frac{|\theta_l^{\approx 0}|}{|\theta_l|} \tag{6}$$

$$\text{EWDS}_l = E_l \cdot (1 - S_l) \tag{7}$$

$$\text{QLIS}_l = \beta \cdot P_l + (1 - \beta) \cdot (1 - \text{EWDS}_l) \tag{8}$$

where $|\theta_l|$ is the number of parameters in layer l, $|\theta_l^{\approx 0}|$ is the count of weights satisfying $|\theta| < \epsilon$ for a small threshold ϵ, H_l is the entropy of layer weights, and B denotes the

full-precision bit-width (e.g., 32 for FP32). The QLIS metric effectively balances model compactness with perceptual quality preservation. Layers with high QLIS are considered more perceptually important.

2.3. Layer-wise Adaptive Compression Scheme

QLIS guides both pruning and quantization:

Layer-wise Adaptive Pruning (LAP): Each layer is pruned by thresholding weights below:

$$\epsilon_l = k_l \cdot \sigma_l, \tag{9}$$

where σ_l is the standard deviation and k_l is QLIS-guided.

Layer-wise Adaptive Quantization (LAQ): Each layer is assigned a bit-width b_l based on its QLIS. Higher QLIS implies higher precision. Huffman encoding is applied post-quantization for further compression.

This dual strategy achieves significant compression (up to 95% size reduction) with negligible perceptual quality degradation, enabling scalable IQA deployment for AIGC.

3. Experimental Setup

Datasets. We evaluate our framework on two benchmark datasets for AI-generated content (AIGC) image quality assessment: **AGIQA-1K** [36] and **AGIQA-3K** [15]. Both datasets provide AI-generated images annotated with human-assigned Mean Opinion Scores (MOS), which serve as ground truth for perceptual quality modeling. AGIQA-1K contains 1,080 images in *Anime* and *Realistic* styles, each with a MOS score. AGIQA-3K comprises 2,982 images generated using six diffusion-based AIGC models with diverse prompts and annotated MOS values, offering rich semantic and visual variability.

Model Training Strategy. We adopt pre-trained vision encoders (CLIP and DINO) combined with lightweight MLP regressors. In **CLIPIQA**, we use the CLIP ViT-L/14 encoder [23], outputting 768-dimensional features, followed by an MLP regressor:

$$\texttt{Linear}(768 \to 256) \to \texttt{ReLU} \to \texttt{Linear}(256 \to 1).$$

The CLIP encoder is frozen during training to retain semantic features, and only the regressor (approx. 200K parameters) is optimized. Input images are resized to 224×224 and normalized using CLIP's preprocessing.

In **DINOIQA**, we utilize the DINO ViT-Base/16 encoder [2], trained with self-supervised distillation. The final transformer block (`blocks.11`) and normalization layer (`norm`) are unfrozen and fine-tuned along with the appended MLP regressor. This selective tuning enhances perceptual adaptation while maintaining efficiency.

All models are trained using the Adam optimizer for 20 epochs with a learning rate of 1×10^{-4} and weight decay of 1×10^{-5}. The loss function is Mean Squared Error (MSE) between predicted and ground-truth MOS. We use an 80/20 train-test split with fixed seeds and a batch size of 16. Data loading is managed with PyTorch's `DataLoader`.

Model Compression Settings. To enable model compression, we apply both pruning and quantization. For pruning, we compare uniform pruning with our proposed Layer-wise Adaptive Pruning (LAP), which assigns pruning multipliers $k_l \in \{1.5, 1, 0.5, 0.25, 0\}$ based on layer importance derived from statistical properties like parameter variance and activation entropy. For quantization, we perform post-training quantization using bit-widths $b_l \in \{1, 2, 4, 8\}$, with 8-bit as the baseline. Our Layer-wise Adaptive Quantization (LAQ) selects precision levels based on perceptual sensitivity. Huffman coding is applied to the quantized weights to further compress the model using entropy coding.

Evaluation Metrics. Evaluation is conducted using standard IQA metrics: Spearman Rank Correlation Coefficient (SRCC), Kendall Rank Correlation Coefficient (KRCC), and Pearson Linear Correlation Coefficient (PLCC), which assess monotonic consistency, rank agreement, and linear correlation between predictions and ground-truth MOS, respectively. Compression efficiency is measured using the Compression Ratio (CR), defined as the ratio of original to compressed model size.

4. Results and Analysis

We evaluate our compression framework for AIGC-IQA on AGIQA-1K and AGIQA-3K datasets using DINO and CLIP-based features. The model is tested under three compression scenarios: (i) pruning-only, (ii) quantization-only, and (iii) pruning followed by quantization. Performance is measured using SRCC, KRCC, PLCC, and compression metrics: CR and Huffman CR (HCR).

4.1. Pruning-only Results

Pruning effectiveness is assessed by varying pruning levels (`P level`). Aggressive pruning (`P level 1.5`) severely degrades performance, particularly for CLIP (e.g., SRCC = -0.2134 on AGIQA-1K). DINO exhibits better resilience (SRCC = 0.4137). Moderate pruning (`P level 0.25`) improves results significantly (e.g., DINO SRCC = 0.7931, CLIP SRCC = 0.6185) with a $\sim$1.3$\times$ CR.

Our Layer-wise Adaptive Pruning (LAP) achieves superior results by assigning pruning ratios based on layer importance. On AGIQA-1K, LAP yields SRCC = 0.8250 (DINO), 0.8258 (CLIP), with CRs of $2.2\times$ and $1.9\times$. Similar performance is observed on AGIQA-3K.

4.2. Quantization-only Results

Uniform quantization with bit-widths of 1, 2, 4, and 8 shows that lower precision yields higher CRs but degrades accuracy. For instance, 1-bit quantization leads to SRCC =

Table 1. Performance comparison on the **AGIQA-1K** and **AGIQA-3K** datasets using DINO and CLIP features under various pruning (P) and quantization (Q) levels. Metrics include SRCC, KRCC, PLCC (correlation coefficients), Compression Ratio (CR), and Huffman Compression Ratio (HCR).

Method	AGIQA-1K - DINO					AGIQA-1K - CLIP					AGIQA-3K - DINO					AGIQA-3K - CLIP				
	SRCC	KRCC	PLCC	CR	HCR	SRCC	KRCC	PLCC	CR	HCR	SRCC	KRCC	PLCC	CR	HCR	SRCC	KRCC	PLCC	CR	HCR
P level 1.5	0.4137	0.2899	0.3490	7.8	7.8	-0.2134	-0.1471	-0.2810	7.9	7.9	-0.0620	-0.0397	-0.0416	7.8	7.8	-0.0274	-0.0188	-0.0286	7.9	7.9
P level 1.0	0.5441	0.3731	0.4785	3.5	3.5	0.2687	0.1858	0.2612	3.4	3.4	-0.0080	-0.0053	-0.0396	3.5	3.5	-0.2308	-0.1532	-0.2831	3.4	3.4
P level 0.5	-0.3216	-0.2161	-0.2873	1.8	1.8	-0.1564	-0.0985	-0.1170	1.7	1.7	-0.1971	-0.1333	-0.2907	1.8	1.8	0.4960	0.3368	0.4807	1.7	1.7
P level 0.25	0.7931	0.6009	0.8188	1.3	1.3	0.6185	0.4469	0.6590	1.3	1.3	0.7709	0.5861	0.8472	1.4	1.4	0.7156	0.5232	0.7237	1.3	1.3
Q 1-bit	0.2519	0.1670	0.1638	32.0	32.0	0.3615	0.2430	0.3480	32.0	32.0	-0.0918	-0.0609	-0.2938	32.0	32.0	0.2221	0.1519	0.3527	32.0	32.0
Q 2-bit	-0.1196	-0.0840	-0.1392	16.0	27.2	-0.0677	-0.0445	-0.0193	16.0	27.1	-0.0136	-0.0075	-0.2285	16.0	27.2	0.0389	0.0257	-0.0962	16.0	27.1
Q 4-bit	-0.2044	-0.1381	-0.1750	8.0	19.6	-0.3575	-0.2334	-0.3426	8.0	20.7	0.1191	0.0754	0.2012	8.0	19.7	-0.0062	0.0005	-0.0285	8.0	20.7
Q 8-bit	0.8285	**0.6468**	0.8517	4.0	6.2	0.8240	**0.6424**	**0.8513**	4.0	6.2	0.7992	0.6161	0.8678	4.0	6.2	**0.8216**	**0.6439**	0.8880	4.0	6.2
Baseline	0.8245	0.6426	0.8497	1.0	1.0	0.8217	0.6388	0.8493	1.0	1.0	0.8000	**0.6163**	**0.8680**	1.0	1.0	0.8196	0.6415	**0.8895**	1.0	1.0
Ours LAP	0.8250	0.6375	**0.8580**	2.2	2.2	**0.8258**	0.6272	0.8147	1.9	1.9	0.8001	0.6130	0.8500	1.7	1.7	0.8204	0.6218	0.8147	1.5	1.5
Ours LAQ	0.8268	0.6345	0.8368	6.3	11.4	0.8248	0.6212	0.7947	4.8	8.1	0.8002	0.6125	0.8573	5.7	10.2	0.8195	0.6273	0.8289	4.7	7.9
Ours LAPQ	**0.8326**	0.6365	0.8313	**11.7**	**22.2**	0.8226	0.6215	0.7979	**8.1**	**15.3**	**0.8003**	0.6117	0.8381	**7.8**	**14.7**	0.8200	0.6232	0.8150	**6.5**	**12.0**

0.2519 (DINO, AGIQA-1K), while 8-bit maintains near-baseline performance (SRCC ≈ 0.82) with $4\times$ CR and $6.2\times$ HCR.

Our Layer-wise Adaptive Quantization (LAQ) assigns bit-widths per layer, improving the trade-off. On AGIQA-1K, LAQ yields SRCC = 0.8268 (DINO), 0.8248 (CLIP), with CRs up to $6.3\times$ and HCR up to $11.4\times$. Similar trends are noted on AGIQA-3K.

4.3. Pruning + Quantization Results

The full compression pipeline (LAPQ) combines LAP and LAQ for maximum efficiency. On AGIQA-1K, LAPQ achieves SRCC = 0.8326 (DINO) and 0.8226 (CLIP), with CR = $11.7\times$ (HCR = $22.2\times$) and $8.1\times$ ($15.3\times$), respectively. On AGIQA-3K, it maintains high accuracy (DINO: 0.8003, CLIP: 0.8200) with substantial compression (HCRs: $14.7\times$, $12.0\times$).

These results confirm that LAPQ not only preserves model fidelity but also offers synergistic compression benefits, likely aided by structured sparsity and quantization-induced regularization.

4.4. Comparison with Existing Works

Unlike prior studies focusing individually on pruning or quantization, we present the first combined compression strategy tailored for AIGC-IQA. Existing handcrafted or SVR-based methods (e.g., CEIQ, NIQE, BMPRI, GMLF) achieve low SRCCs (<0.7). Deep models (e.g., ResNet50, DBCNN, HyperNet) perform better but are resource-intensive.

Our LAPQ framework outperforms or matches deep full-precision models (e.g., SRCC = 0.8326 on AGIQA-1K) while achieving up to $22.2\times$ compression. The adaptive nature of our approach enables high efficiency with minimal accuracy loss, making it ideal for edge deployment in perceptual quality assessment of AIGC.

5. Conclusion

We proposed LAPQ, a Layer-wise Adaptive Pruning and Quantization framework for compressing deep neural networks in the context of AI-generated content image quality assessment (AIGC-IQA). To the best of our knowledge, this is the first dedicated effort addressing model compression in this domain for efficient deployment on resource-constrained platforms.

Table 2. Performance comparison of existing methods and the proposed approach on the AGIQA-1K and AGIQA-3K datasets.

Method	AGIQA-1K				AGIQA-3K			
	SRCC	KRCC	PLCC	HCR	SRCC	KRCC	PLCC	HCR
Hand-crafted based								
CEIQ [32]	0.3069	0.2097	0.2836	1.0	0.3228	0.2220	0.4166	1.0
DSIQA [21]	-0.3047	-0.2148	-0.0559	1.0	0.4955	0.3403	0.5488	1.0
NIQE [19]	-0.5490	-0.3824	-0.5048	1.0	0.5623	0.3876	0.5171	1.0
SISBLIM [7]	-0.1309	-0.0889	-0.3575	1.0	0.5479	0.3788	0.6477	1.0
SVR-based								
BMPRI [17]	0.0651	0.0400	0.1646	1.0	0.6794	0.4976	0.7912	1.0
GMLF [31]	0.5575	0.4052	0.6356	1.0	0.6987	0.5119	0.8181	1.0
HIGRADE [14]	0.4056	0.2860	0.4425	1.0	0.6171	0.4410	0.7056	1.0
Deep Learning based								
ResNet50 [9]	0.6365	0.4777	0.7323	1.0	N/A	N/A	N/A	N/A
StairIQA [26]	0.5504	0.4039	0.6088	1.0	N/A	N/A	N/A	N/A
MGQA [28]	0.6011	0.4456	0.6760	1.0	N/A	N/A	N/A	N/A
DBCNN [35]	N/A	N/A	N/A	N/A	0.8207	0.6336	0.8759	1.0
CNNIQA [12]	N/A	N/A	N/A	N/A	0.7478	0.5580	0.8469	1.0
HyperNet [25]	N/A	N/A	N/A	N/A	**0.8355**	**0.6488**	**0.8903**	1.0
Ours: Efficient LAPQ + LVM backbone based								
LAPQ + DINOIQA	**0.8326**	**0.6365**	**0.8313**	**22.2**	0.8003	0.6117	0.8381	**14.7**
LAPQ + CLIPIQA	0.8226	0.6215	0.7979	15.3	0.8200	0.6232	0.8150	12.0

Unlike fixed pruning or uniform quantization methods, LAPQ adaptively selects layer-wise sparsity and precision based on importance, achieving up to 95% model size reduction with minimal performance loss across multiple AIGC-IQA benchmarks and large vision model backbones.

Our results demonstrate LAPQ's effectiveness as a scalable and practical solution for real-world AIGC-IQA applications. Future work will explore joint optimization strategies, hardware-aware compression, and extension to other generative modalities such as video and 3D content.

References

[1] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S Yu, and Lichao Sun. A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226*, 2023. 1

[2] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Pro-*

ceedings of the IEEE/CVF international conference on computer vision, pages 9650–9660, 2021. 1, 2, 3

[3] Manri Cheon, Sung-Jun Yoon, Byungyeon Kang, and Junwoo Lee. Perceptual image quality assessment with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 433–442, 2021. 1

[4] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869, 2023. 1

[5] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2567–2581, 2020. 1

[6] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 1

[7] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE Transactions on Broadcasting*, 60 (3):555–567, 2014. 4

[8] Song Han, Huizi Mao, and William J Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149*, 2015. 2

[9] 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. 4

[10] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1

[11] 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

[12] 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. 4

[13] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 1

[14] 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. 4

[15] 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*, 34(8):6833–6846, 2023. 3

[16] 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

[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. 4

[18] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12): 4695–4708, 2012. 1

[19] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 1, 4

[20] Pedram Mohammadi, Abbas Ebrahimi-Moghadam, and Shahram Shirani. Subjective and objective quality assessment of image: A survey. *arXiv preprint arXiv:1406.7799*, 2014. 1

[21] ND Narvekar and LJ Karam. A no-reference perceptual image sharpness metric based on a cumulative probability of blur. In *International Workshop on Quality of Multimedia Experience*, pages 87–91, 2009. 4

[22] Fei Peng, Huiyuan Fu, Anlong Ming, Chuanming Wang, Huadong Ma, Shuai He, Zifei Dou, and Shu Chen. Aigc image quality assessment via image-prompt correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6432–6441, 2024. 1

[23] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 1, 2, 3

[24] Tushar Shinde. Adaptive quantization and pruning of deep neural networks via layer importance estimation. In *Workshop on Machine Learning and Compression, NeurIPS 2024*, 2024. 2

[25] 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. 4

[26] 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. 4

[27] Jiarui Wang, Huiyu Duan, Jing Liu, Shi Chen, Xiongkuo Min, and Guangtao Zhai. Aigciqa2023: A large-scale image quality assessment database for ai generated images: from the perspectives of quality, authenticity and correspondence. In *CAAI International Conference on Artificial Intelligence*, pages 46–57. Springer, 2023. 1

[28] 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. 4

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 1

[30] Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and Hong Lin. Ai-generated content (aigc): A survey. *arXiv preprint arXiv:2304.06632*, 2023. 1

[31] 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. 4

[32] Jia Yan, Jie Li, and Xin Fu. No-reference quality assessment of contrast-distorted images using contrast enhancement. *arXiv preprint arXiv:1904.08879*, 2019. 4

[33] Zihao Yu, Fengbin Guan, Yiting Lu, Xin Li, and Zhibo Chen. Sf-iqa: Quality and similarity integration for ai generated image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6692–6701, 2024. 1

[34] Jiquan Yuan, Xinyan Cao, Changjin Li, Fanyi Yang, Jinlong Lin, and Xixin Cao. Pku-i2iqa: An image-to-image quality assessment database for ai generated images. *arXiv preprint arXiv:2311.15556*, 2023. 1

[35] 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. 4

[36] 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, 3