https://doi.org/10.2352/EI.2026.38.12.GENAI-179

Adapting DeQA-Score for Attribute-Specific Portrait Quality Assessment

Yujin Cho[1,2], Minh Khang Tran[1], Benoit Pochon[1], Jean-Michel Morel[3], Gabriele Facciolo[2,4], Sira Ferradans[1]
[1] DXOMARK Image Labs, Paris, France
[2] Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 91190, Gif-sur-Yvette, France
[3] Lingnan University, China
[4] Institut Universitaire de France

Abstract

With the growing adoption of multimodal large language models (MLLMs) for image quality assessment, Vision–Language IQA systems such as DeQA-Score have demonstrated a strong correlation with human judgments on natural images. However, current MLLM-based quality predictors primarily provide global image quality scores and therefore lack the ability to quantitatively assess specific perceptual attributes such as noise, texture, contrast, and color factors that are essential for explainability and camera tuning. In this work, we extend DeQA-Score from global Mean Opinion Score (MOS)-based quality prediction to attribute-specific, Just Objectionable Difference (JOD)-based portrait assessment. Our study investigates how a MOS-trained model behaves when exposed to pairwise-annotated data and how lightweight adaptation can achieve perceptual alignment at the attribute level. Using a controlled mannequin dataset, we analyze the model's baseline behavior under different prompt strategies and spatial input configurations, revealing limited attribute sensitivity. We then apply LoRA fine-tuning on realistic portrait data annotated for texture and noise quality. The adapted model achieves correlations of SRCC $= 0.91/0.93$ *and PLCC* $= 0.91/0.90$ *with JOD scores for noise and texture, respectively. Subsequent analysis confirms that the vision encoder is the main contributor to perceptual learning. The proposed framework establishes an efficient path for converting global VLM-based Image Quality Assessment (IQA) models into attribute-aware, perceptually aligned assessors for real-world photography.*

Introduction

With the rapid progress of multimodal large language models (MLLMs) [1, 2, 3, 4], recent vision–language approaches have demonstrated strong performance in image quality assessment (IQA) by jointly reasoning over visual and textual representations to predict perceptual quality [5, 6, 7, 8, 9, 10]. However, current MLLM-based quality predictors primarily provide global image quality scores and therefore lack the ability to assess specific perceptual attributes—such as noise, texture, contrast, and color factors that are essential for explainability and camera tuning.

Among these, DeQA-Score [11] represents a state-of-the-art framework that takes both an image and a text prompt specifying the quality assessment task as inputs. The model is trained on large datasets annotated with Mean Opinion Scores (MOS), providing continuous and robust global quality prediction across natural images. However, such MOS-global quality models are inherently limited in their ability to capture attribute-specific perceptual factors, such as texture and detail preservation, which are critical to the perceived realism of portraits and other complex scenes. Furthermore, MOS values, obtained from absolute single-stimulus evaluations, differ fundamentally from Just Objectionable Difference (JOD) scales [12] derived from psychophysical pairwise comparisons. While MOS captures a global impression of quality, JOD scales specifically measure the smallest detectable perceptual distances between images, providing the fine-grained sensitivity required to isolate specific degradations like noise and texture. A JOD value represents the perceived quality difference where 1 JOD corresponds to approximately 75% observer agreement that one image is of higher quality than the other. This conceptual gap prevents direct transfer of MOS-trained IQA models to attribute-level, JOD-based applications. Bridging this gap is essential for developing perceptually grounded, attribute-aware image quality models that not only improve explainability and ISP optimization but also generalize to real-world portrait evaluation.

In this work, we focus on two perceptual quality attributes, texture and noise, that both affect the perceived realism of portrait images. Texture measures how well image details are preserved, contributing to the realistic appearance of surfaces and materials. In contrast, noise refers to unwanted grain-like or irregular patterns that do not originate from the scene content and can appear globally or locally, reducing overall visual quality when excessive. See Figure 1 for a visual example.

Together, these attributes form complementary aspects of perceptual quality: texture reflects desirable detail preservation, whereas noise represents undesired artifacts. Attribute-specific JOD annotations provide psychometric ground truth for these perceptual dimensions.

This work extends DeQA-Score from global image quality evaluation to the portrait domain by adapting it for attribute-specific, JOD-based quality assessment. The contributions of this paper are as follows:

- We analyze the behavior and limitations of a MOS-trained vision–language IQA model when applied to JOD-annotated data and attribute-specific prompts.
- We adapt a lightweight Low-Rank Adaptation (LoRA) fine-tuning [13] framework for attribute-specific learning, and we convert JOD-based annotations into a MOS-compatible training scheme for reliable estimation of perceptual texture and noise quality.

- We study the effects of fine-tuning configuration and data scale on perceptual correlation and model accuracy.

Methods

This section describes the experimental methodology used to analyze and adapt the DeQA-Score model for attribute-specific image quality assessment. We first evaluate the behavior of the pre-trained model on a controlled realistic mannequin dataset to study the effects of prompt formulation and spatial input on texture sensitivity. We then adapt the model to predict texture and noise quality using JOD-annotated real-world portrait data. Finally, we detail the training setup and evaluation protocol used for LoRA fine-tuning.

Baseline analysis on a controlled realistic mannequin dataset

A pre-trained DeQA-Score model was first evaluated on a controlled mannequin dataset [14] containing multiple scenes under varied illumination and acquisition conditions, without retraining. The original images in the dataset ranged in resolution from 4023×3024 to 1280×720 pixels. To evaluate the effect of spatial context, three input configurations were used: Original image (full frame, variable resolution), Face crops (546×772 pixels), and Eye crops (1500×1500 pixels) shown in Figure 2. These crops are obtained from the original images. Each scene was annotated with texture JOD scores [12] derived from pairwise comparisons, providing psychometric ground truth at the attribute level. The mannequin dataset comprised two annotation regimes: LQ (Face) for face-region texture quality, and HQ (Eye and Beard) for texture quality. In this experiment, only eye-region data from the HQ regime were used. The first experiment tested whether prompt engineering could guide a model trained for global image quality evaluation toward sensitivity to attribute-specific cues such as texture. Three prompt formulations were tested.

Figure 1. Examples illustrating the texture and noise attributes in portrait images. The left image exhibits stronger texture quality, with fine details in the skin and hat clearly preserved, though it also shows higher visible noise. In contrast, the right image displays reduced noise, resulting in a smoother appearance at the cost of losing natural fine details and surface texture.

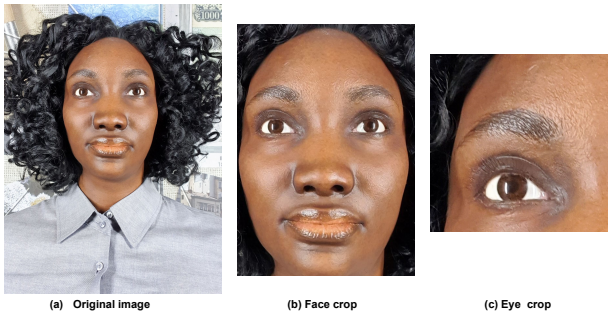


Figure 2. Examples of controlled mannequin dataset and spatial input configurations. The original image (left) is shown with its corresponding face crop (middle) and eye crop (right). Three mannequin identities (Asian, Black, and White) were used in the experiments, but only one is shown here for illustration.

- Prompt 1: "How would you rate the quality of this image?" representing global quality.
- Prompt 2: "Assess the image quality by analyzing the clarity, details, and rendering accuracy of the eye textures." focusing on the texture attribute within eye crop images.
- Prompt 3: "How would you evaluate the overall image quality based on the clarity, detail, and realism of the eye textures."

Each prompt was applied to the HQ eye-crop images across the entire mannequin dataset [14] to test consistency relative to prompt changes. Table 1 reports the mean DeQA-Scores and variance across prompts. The second experiment evaluates how the annotation regime and spatial input affect model performance. For each regime, DeQA-Score predictions were computed under three spatial input configurations (Original image, face crop, and eye crop) and fitted to the JOD scale using the VQEG logistic regression. [15, 16].

Table 1. Mean De-QA scores for eye-crop images across three prompt variations. Variance represents the across-prompt variability for each mannequin, showing the model's consistency to wording changes.

Mannequin	Prompt 1	Prompt 2	Prompt 3	Variance
Diana	3.28	3.29	3.27	0.07
Eugene	3.02	3.01	3.00	0.06
Sienna	3.63	3.63	3.65	0.03

Attribute-specific adaptation on real portrait data

The baseline experiments indicate that DeQA-Score's sensitivity to texture depends mainly on the spatial evaluation region, while the model is insensitive to different prompts. However, the mannequin dataset includes only texture-JOD annotations, limiting its scope to a single perceptual attribute. In the following experiment, we use full-face crops rather than small regions to ensure that local attributes such as noise and texture are assessed within the same perceptual context as defined during annotation. In the real portrait dataset, JOD annotations were defined at the full-face level because their perception is influenced by illuminance, skin tone, and camera processing across the entire face rather than isolated patches. To generalize the analysis and enable attribute-specific adaptation, we extend the study to a real-world

portraits annotated both for texture and noise quality under natural scene conditions.

Figure 1 illustrates the visual examples of the texture and noise attributes that influence perceived portrait realism. The portrait dataset consists of portraits captured in 15 different scenes with varied illumination and acquisition conditions. All images used in the realistic portrait dataset were collected with the written informed consent of the participants. This consent explicitly covers the use and publication of their photographs for research and scientific purposes. From these images, face crops are extracted and upscaled to 585×820 using bicubic interpolation. The dataset is annotated through pairwise comparisons and provides independent JOD labels for texture and noise separately, a sample is shown in Figure 3.

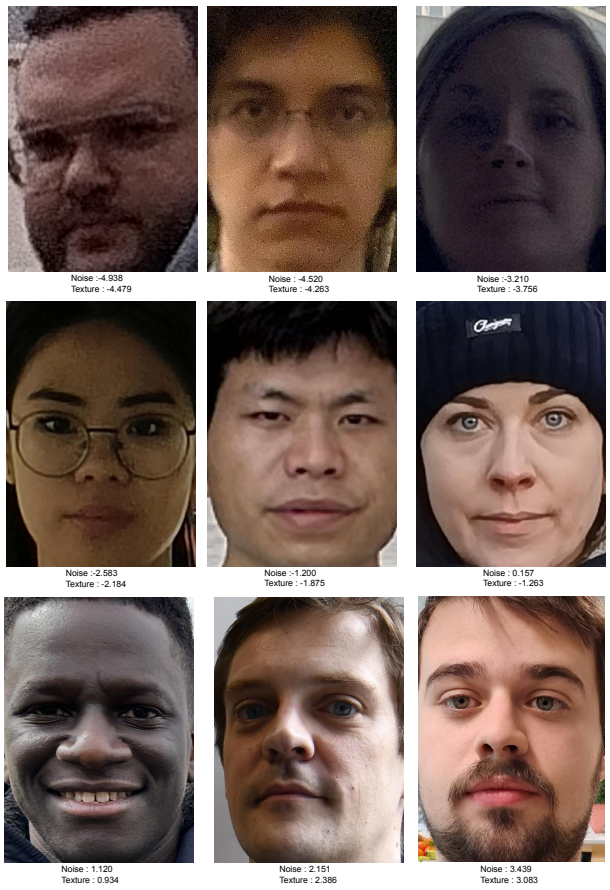


Figure 3. Sample images from realistic portrait datasets along with their texture and JOD scores. A JOD (Just Objectionable Difference) value represents the perceived quality difference derived from pairwise annotations, where 1 JOD corresponds to about 75% observer agreement that one image is of higher quality than the other.

Because large language models output discrete verbal quality levels rather than continuous values, the continuous JOD scores are normalized and discretized to match the text-based training format. Although JOD scores are derived from pairwise comparisons, they form an interval scale that can be linearly mapped, making them compatible with MOS-like discretization for regression-based learning. To align with the DeQA-Score model training scheme, JOD scores are normalized to the [1, 5] range and binned using three mapping strategies: uniform-range, uniform-range with outlier thresholding and uniform-sample (quantile-based), as shown in Figure 4. Since all binning methods produce similar correlations, we use the uniform-range method for simplicity. The attribute-specific prompts: "The texture quality is <level>" and "The noise quality is <level>" are used during training, following the original DeQA soft-label protocol.

Training Details

The DeQA-Score architecture consists of a visual encoder, a visual abstractor (which aligns visual features into the language embedding space), a tokenizer, and a language model that jointly learns from visual and textual information to estimate perceptual quality. Following the official DeQA-Score implementation, we LoRA-finetune the language model's attention layers while keeping both the visual encoder and visual abstractor fully trainable. This setup matches one of the configurations analyzed in the original DeQA-Score ablation study [11], where both visual components are fine-tuned, and the language model is LoRA-fine-tuned rather than fully fine-tuned. The LoRA rank was set to 128, following the default configuration of the original DeQA-Score paper to maintain architectural consistency and ensure sufficient capacity for visual-textual alignment. Training uses the AdamW optimizer (learning rate = 2×10^{-5}), a cosine decay schedule, batch size of 64, and three epochs, consistent with the authors' settings. To assess components' contributions, we also evaluate variants that selectively freeze the encoder and/or abstractor. For data efficiency, we vary the number of training samples (100–1,000) using stratified sampling across quality levels to preserve the JOD distribution. Performance is measured on a held-out test set for texture and noise using *SRCC*, *PLCC*, and *MAE*.

RESULTS

We evaluate the behavior and adaptation of the DeQA-Score model on two datasets. The controlled mannequin dataset is used to analyze the effects of prompt formulation and spatial input through ablation studies. The real portrait dataset is used to assess fine-tuning performance on noise and texture attributes under realistic conditions using JOD-based annotations.

Results on the Controlled Mannequin Dataset

On the Mannequin data [14], the DeQA-Score model exhibits insensitivity to different prompts, as shown in Table 1. Mean DeQA-Scores remain nearly identical across the three prompt formulations for all mannequins, indicating that the pre-trained model primarily captures global quality rather than attribute-specific texture cues. This lack of sensitivity stems from the initial training on global MOS datasets, which biased the model toward general quality features regardless of textual instructions. Consequently, without adaptation, the model treats attribute-specific prompts as synonyms for global quality evaluation. For the quantitative evaluation reported in Table 2, we use Prompt 1 ("How would you rate the quality of this image?") as a reference.

Following the VQEG and ITU-T J.149 recommendations [15, 16], we apply a four-parameter logistic function be-

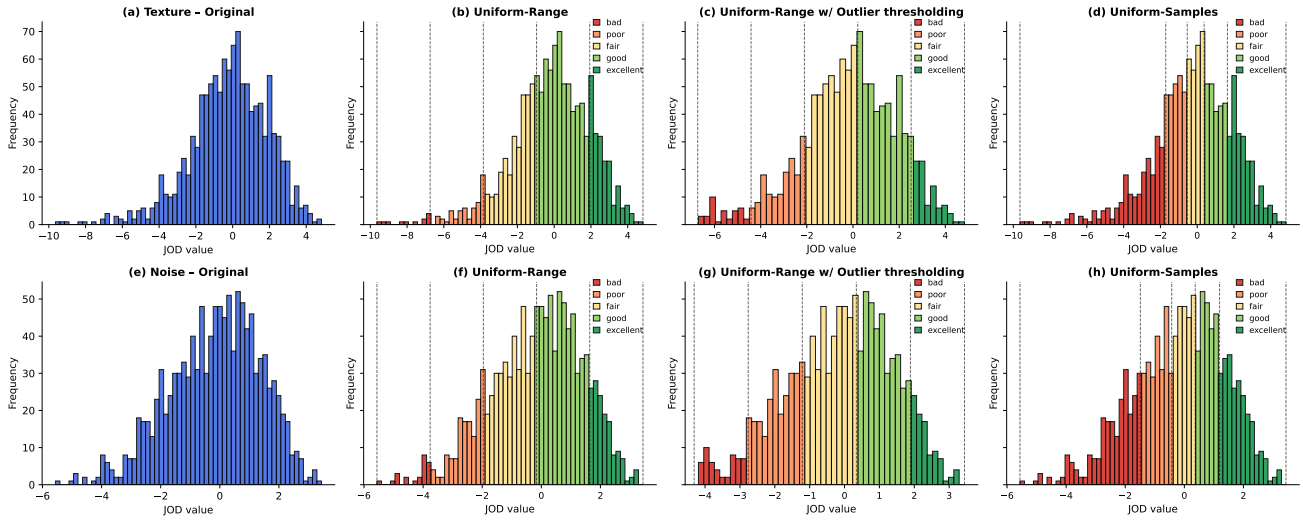


Figure 4. The original JOD distribution for Texture and Noise and a comparison of different data processing methods. Uniform-Range divides the JOD value range evenly; Uniform-Range w/ Outlier Thresholding combines uniform-range binning with outlier suppression; and Uniform-Samples uses quantile-based grouping. Each method ensures a balanced weight across quality classes to prevent model bias during training.

Table 2. Evaluation results (SRCC / PLCC / MAE) on the mannequin dataset for different annotation regimes (LQ and HQ) and spatial input configurations.

Annotation type	Input configuration	SRCC	PLCC	MAE
LQ (Face)	**Face crop**	**0.934**	**0.946**	**0.608**
	Original image	0.707	0.795	1.140
HQ (Eye)	**Eye crop**	**0.816**	**0.842**	**0.809**
	Original image	0.494	0.589	1.166

tween predicted and subjective JOD scores prior to computing correlation metrics. Table 2 summarizes results on the mannequin dataset, which includes eight scenes and three mannequin models (Sienna, Eugene, and Diana) with two annotation strategies: LQ (face-region texture quality) and HQ (eye-region texture quality). For each regime, predictions are evaluated using different spatial input configurations, including full images, face crops, and eye crops.

The results demonstrate that LQ annotations show the highest correlations with full-face inputs, while HQ annotations perform best with eye crops. In contrast, performance degrades when the spatial input does not match the annotated perceptual region. These findings indicate that the DeQA-Score model provides stable predictions only when the evaluation region is consistent with the perceptual scope of the annotation. To further investigate this calibration behavior beyond the controlled mannequin setup, the next section explores LoRA fine-tuning on real data so that the model can learn this JOD-to-MOS alignment directly without post-hoc fitting.

Fine-tuning on Real Portrait Data

After this controlled analysis, we fine-tune the DeQA-Score model on a realistic portrait dataset annotated with JOD scores for noise and texture. LoRA fine-tuning on this dataset leads to consistent improvements across all metrics (Figure 5). With

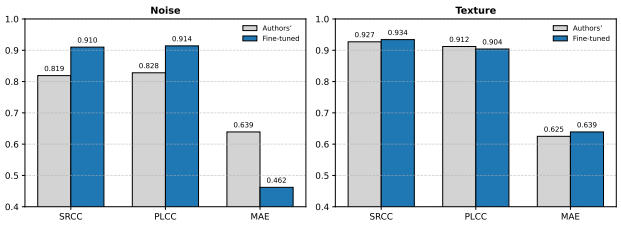


Figure 5. Comparison of baseline and LoRA fine-tuned model on Noise and Texture attributes using SRCC, PLCC, and MAE.

uniform-range discretization, the model achieves $SRCC = 0.910$, $PLCC = 0.914$, and $MAE = 0.462$ for noise attribute evaluation, and $SRCC = 0.934, PLCC = 0.904$, and $MAE = 0.639$ for texture. We also compare different binning strategies for discretizing JOD scores, such as uniform-range, uniform-range with outlier thresholding, and uniform-sample binning. All three approaches yield similar results, with differences in ($\Delta SRCC < 0.02$). This indicates the performance is stable regardless of the score-to-label mapping strategy in Table 3. For simplicity and consistency, uniform-range method is used in the main experiments.

Effect of Fine-tuning Components

To analyze how different adaptation strategies influence performance, we test four configurations of the DeQA-Score's components for fine-tuning: Tune/Tune, Tune/Freeze, Freeze/Tune and Freeze/Freeze, where the first term refers to the visual encoder and the second to the visual abstractor. As reported in Table 4, jointly fine-tuning both components achieves the best performance for both noise and texture attributes, indicating that the visual encoder is the main contributor to perceptual learning.

Freezing the encoder while tuning only the abstractor (Freeze/Tune) provides smaller gains, whereas freezing both

Table 3. Comparison of mapping strategies for JOD normalization in DeQA-Score fine-tuning. Uniform-Range divides the JOD value range evenly, Uniform-Range w/Outlier Thresholding combines uniform-range binning with outlier suppression, and Uniform-Sample uses quantile-based grouping. Bold values indicate the best performance and underlined values indicate the second best for each metric. Higher SRCC/PLCC and lower MAE indicate better results.

Attribute	Method	SRCC	PLCC	MAE
Noise	Uniform-Range	**0.910**	**0.914**	**0.462**
	Uniform-Range w/ Outlier Thresholding	0.897	0.901	0.485
	Uniform-Sample	0.907	0.907	0.475
Texture	Uniform-Range	0.934	0.904	0.639
	Uniform-Range w/ Outlier Thresholding	0.934	0.902	0.623
	Uniform-Sample	**0.941**	**0.909**	**0.620**

(Freeze/Freeze) results in negligible performance change. These findings show that while the visual encoder is key to overall performance, it cannot distinguish specific perceptual attributes out of the box. Effective adaptation therefore requires fine-tuning the encoder jointly with the prompts to learn attribute-aware representations.

Impact of Standard Deviation

The original DeQA-Score models perceptual uncertainty using the standard deviation of Mean Opinion Score (MOS) annotations. Since the PIPAL dataset does not include standard deviation values, the DeQA-Score authors [11] estimated a pseudo σ by measuring the average ratio between the standard deviation and score range in other MOS-based IQA datasets, and then scaling this ratio to the PIPAL dataset [17]. To examine how this choice of σ affects performance on our JOD-based data, we compared the pseudo σ method with several fixed σ values ranging from 0.2 to 0.5. As summarized in Table 5, all configurations produced comparable results, with a small fixed $\sigma = 0.2$ giving slightly higher correlations and more stable training. This behavior can be attributed to the higher consistency of JOD annotations compared to MOS scores. Since JOD data exhibit lower perceptual variability, explicitly modeling uncertainty through a pseudo σ provides limited benefit, while a small fixed value yields more uniform weighting that better matches the characteristics of the data.

Table 5. Comparison of pseudo and fixed standard deviation settings in DeQA training for Noise and Texture attributes. The fixed σ = 0.2 yields the most stable performance across metrics.

Standard deviation (σ)	SRCC (Noise / Texture)	PLCC (Noise / Texture)	MAE (Noise / Texture)
Pseudo σ	0.902 / 0.933	0.891 / 0.878	0.503 / 0.642
Fixed $\sigma = 0.2$	**0.907** / 0.934	**0.914** / **0.904**	**0.468** / **0.639**
Fixed $\sigma = 0.3$	0.895 / **0.935**	0.899 / 0.896	0.488 / 0.627
Fixed $\sigma = 0.4$	0.894 / 0.932	0.898 / 0.894	0.499 / 0.643
Fixed $\sigma = 0.5$	0.900 / 0.935	0.894 / 0.890	0.502 / 0.637

Data Efficiency of LoRA Fine-tuning

For the data-efficiency analysis, we fine-tune the model using progressively larger subsets of the training data (100-1043 samples). To avoid bias toward any specific quality level, each subset was sampled evenly across the five quality groups [1-5] and evaluated on a fixed test set of 153 images. As illustrated in Figure 6, correlations increase steadily with training size for both noise and texture. For noise, *SRCC* improves from 0.84 at 100 samples to 0.91 at 1,043 samples while *MAE* decreased from 0.69 to 0.46. For texture, *SRCC* remained consistently high around 0.92-0.93 with gradual improvement as data increased. This smaller gain reflects that the MOS baseline model already captured global texture information well, whereas noise required additional adaptation to align with perceptual JOD judgments. No clear saturation was observed even at the largest dataset size, indicating that LoRA fine-tuning continues to benefit from additional supervision. The steady performance improvement even at smaller sample sizes suggests that stratified sampling effectively preserves the JOD distribution, allowing the model to learn quality gradients efficiently with limited data. Overall, the adapted model successfully bridges the gap between MOS-based pretraining and JOD-based attribute prediction.

Conclusions

This study presents a systematic adaptation of a vision–language model trained on Mean Opinion Score (MOS) based datasets to Just Objectionable Difference (JOD)-based, attribute-specific IQA for portraits. The proposed LoRA framework enables efficient fine-tuning that achieves correlations above 0.9 for texture and noise using 1,043 annotated samples. Unlike conventional MOS-trained IQA models that predict only an overall quality score, the adapted model provides separate perceptual quality scores for noise and texture, each aligned with the underlying psychophysical JOD scale.

Because large language models produce discrete textual outputs rather than continuous values, the continuous JOD annotations must be discretized and expressed as textual levels during training. We also compare different ways of converting JOD data into MOS-like training targets, including uniform-range, uniform-range with outlier thresholding mappings and uniform-sample (quantile-based). All three strategies yield similar performance, indicating that the proposed framework is robust to the specific choice of label discretization. For clarity and simplicity, uniform-range normalization was used for the main results. Our experiments reveal a high correlation between global MOS and texture JOD, indicating that texture perception is a major component of how observers judge portrait quality. This strong overlap supports the need for attribute-specific adaptation, which enables the model to disentangle and separately evaluate the distinct effects of texture and noise on overall quality.

Finally, our ablations show that perceptual understanding resides mainly in the visual encoder, while the text encoder enhances interpretability and stability. These findings highlight a scalable and principled path for adapting vision–language IQA models into perceptually aligned, attribute-aware tools. As future research, it would be highly valuable to extend this framework to include generated portraits. Since many generated images contain unique texture and noise artifacts for which no established quality metric currently exists, evaluating how MLLM-based IQA models respond to these artifacts will be interesting for modern portrait assessment and future camera-quality evaluation.

Table 4. Effect of visual encoder and visual abstractor fine-tuning on Noise and Texture performance. Tuning both modules yields the highest correlations, while partial or frozen configurations reduce perceptual alignment. (Higher is better for SRCC/PLCC and lower for MAE.)

Encoder / Abstractor	SRCC (Noise)	SRCC (Texture)	PLCC (Noise)	PLCC (Texture)	MAE (Noise) ↓	MAE (Texture) ↓
Tune / Tune	**0.907**	0.934	**0.914**	0.904	**0.462**	0.639
Tune / Freeze	0.897	**0.937**	0.909	**0.919**	0.484	**0.610**
Freeze / Tune	0.842	0.932	0.841	0.900	0.638	0.683
Freeze / Freeze	0.837	0.929	0.838	0.901	0.637	0.680

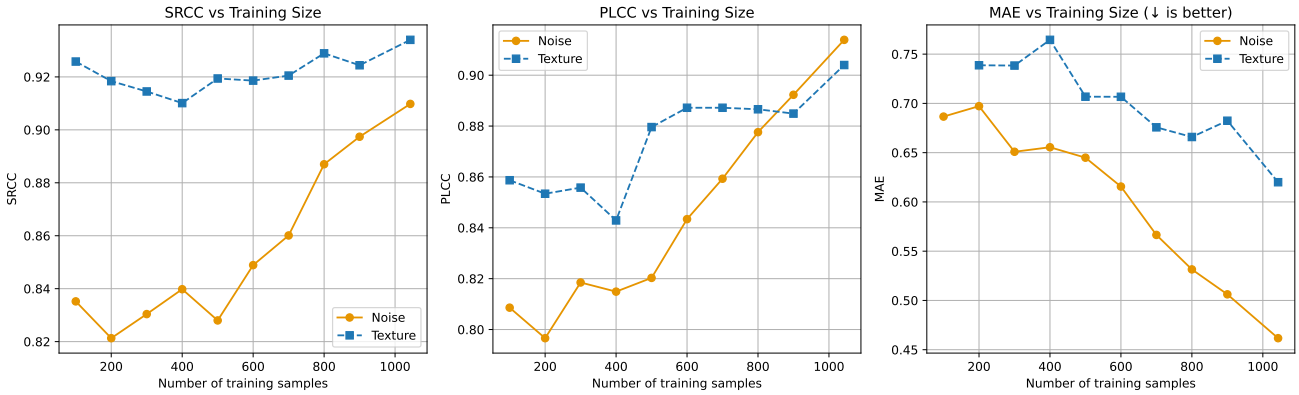


Figure 6. Effect of training data size on LoRA fine-tuning for Noise and Texture attributes. Each point represents a model fine-tuned on a subset of training samples (uniformly drawn across quality levels bad–excellent) and evaluated on the same 153 test images. Performance improves steadily with increasing data size, especially for the Noise attribute.

Acknowledgments

This project was provided with computer and storage resources by GENCI at IDRIS thanks to the grant AD011014305R1 on the supercomputer Jean Zay.

References

[1] H. Liu, C. Li, Q. Wu, and Y. J. Lee, "Visual instruction tuning," *Advances in neural information processing systems*, vol. 36, pp. 34892–34916, 2023.

[2] G. OpenAI, "4v (ision) system card https://cdn. openai. com/papers," *GPTV_System_Card. pdf*, 2023.

[3] Q. Ye, H. Xu, J. Ye, M. Yan, A. Hu, H. Liu, Q. Qian, J. Zhang, and F. Huang, "mplug-owl2: Revolutionizing multi-modal large language model with modality collaboration," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pp. 13040–13051, 2024.

[4] D. Zhu, J. Chen, X. Shen, X. Li, and M. Elhoseiny, "Minigpt-4: Enhancing vision-language understanding with advanced large language models," *arXiv preprint arXiv:2304.10592*, 2023.

[5] H. Wu, Z. Zhang, W. Zhang, C. Chen, L. Liao, C. Li, Y. Gao, A. Wang, E. Zhang, W. Sun, *et al.*, "Q-align: Teaching lmms for visual scoring via discrete text-defined levels," *arXiv preprint arXiv:2312.17090*, 2023.

[6] C. Chen, S. Yang, H. Wu, L. Liao, Z. Zhang, A. Wang, W. Sun, Q. Yan, and W. Lin, "Q-ground: Image quality grounding with large multi-modality models," in *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 486–495, 2024.

[7] H. Wu, Z. Zhang, E. Zhang, C. Chen, L. Liao, A. Wang, K. Xu, C. Li, J. Hou, G. Zhai, *et al.*, "Q-instruct: Improving low-level visual abilities for multi-modality foundation models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25490–25500, 2024.

[8] H. Wu, H. Zhu, Z. Zhang, E. Zhang, C. Chen, L. Liao, C. Li, A. Wang, W. Sun, Q. Yan, *et al.*, "Towards open-ended visual quality comparison," in *European Conference on Computer Vision*, pp. 360–377, Springer, 2024.

[9] Z. You, J. Gu, Z. Li, X. Cai, K. Zhu, C. Dong, and T. Xue, "Descriptive image quality assessment in the wild," *arXiv preprint arXiv:2405.18842*, 2024.

[10] Z. You, Z. Li, J. Gu, Z. Yin, T. Xue, and C. Dong, "Depicting beyond scores: Advancing image quality assessment through multi-modal language models," in *European Conference on Computer Vision*, pp. 259–276, Springer, 2024.

[11] Z. You, X. Cai, J. Gu, T. Xue, and C. Dong, "Teaching large language models to regress accurate image quality scores using score distribution," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2025.

[12] M. Perez-Ortiz, A. Mikhailiuk, E. Zerman, V. Hulusic, G. Valenzise, and R. K. Mantiuk, "From pairwise comparisons and rating to a unified quality scale," *IEEE Transactions on Image Processing*, vol. 29, pp. 1139–1151, 2019.

[13] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, *et al.*, "Lora: Low-rank adaptation of large language models.," *ICLR*, vol. 1, no. 2, p. 3, 2022.

[14] D. C. Ventura, G. P. Gouveia, A. Calarasanu, V. Tosel, N. Chahine, and S. Ferradans, "From video conferences to dslrs: An in-depth texture evaluation with realistic mannequins," *Electronic Imaging*, vol. 36, pp. 1–6, 2024.

[15] J. Recommendation, "149: Methdsod for specifying accuracy and cross-calibration of video quality metrics (vqm)," *ITU-T. March*, 2004.

[16] K. Brunnstrom, D. Hands, F. Speranza, and A. Webster, "Vqeg validation and itu standardization of objective perceptual video quality metrics [standards in a nutshell]," *IEEE Signal processing magazine*, vol. 26, no. 3, pp. 96–101, 2009.

[17] G. Jinjin, C. Haoming, C. Haoyu, Y. Xiaoxing, J. S. Ren, and D. Chao, "Pipal: a large-scale image quality assessment dataset for perceptual image restoration," in *European conference on computer vision*, pp. 633–651, Springer, 2020.

Author Biography

Yujin Cho *is a PhD student at ENS Paris-Saclay and DXOMARK. She holds a Master's degree in applied mathematics from Université Paris-Saclay. Her research interests include image quality assessment and image enhancement.*

Minh Khang Tran *is a Master's student at NTNU Gjøvik. He completed an internship at DXOMARK in 2025. His research interests include color science and image processing.*

Benoit Pochon *received his Master's degree in engineering from Centrale Supelec (2001) and his Master's degree in Electrical Engineering from GeorgiaTech University (2001). After several years working in the signal processing domain, he joined DXOMARK Image labs in 2017, as image science director.*

Jean-Michel Morel *started his career as teaching assistant in mathematics at University of Marseille, and obtained his doctorate at Sorbonne University in 1985. He then held chair professor positions in applied mathematics at Paris Sciences & Lettres University, Paris-Saclay University, City University of Hong Kong, and is currently chair professor in data science at Lingnan University. Trained as a pure mathematician, Jean-Michel Morel took an early interest in computer vision and AI and specialized in the mathematical formalization of visual perception. This led him to conceive image processing and image analysis algorithms, many of which have been implemented in imaging systems, including earth observation satellites, medical imaging devices, industrial cameras and mobile phones.*

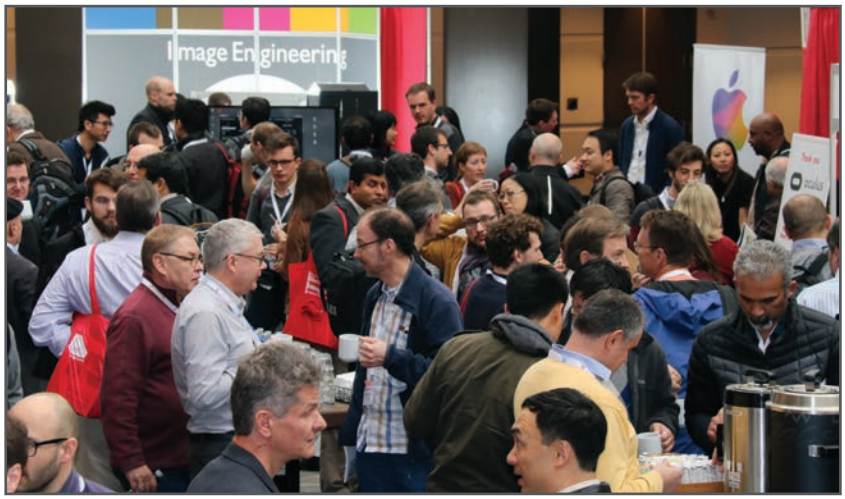
Gabriele Facciolo *is a Full Professor at École-Normale Supérieure Paris-Saclay (ENSPS) and senior member of the Institute Universitaire de France. He leads the image-processing group at Centre Borelli, researching mathematical modeling and machine learning for computer vision, image/video restoration, and remote sensing. He earned his Ph.D. (2011) from Universitat Pompeu Fabra, Barcelona. After post-doctoral research at ENSPS and École des Ponts ParisTech, he became senior researcher at DxO in 2016, and since 2018 he is Professor of Applied Mathematics at the ENSPS. He is a founding editor, and current Editor-in-Chief, of IPOL (ipol.im), the first journal to publish articles accompanied by executable code and online demonstrators.*

Sira Ferradans *is currently the AI Director at DXOMARK. She earned her PhD in Computer Vision from the Universitat Pompeu Fabra (Barcelona, Spain), and worked as a researcher at Duke University (North Carolina, US) and École Normale Supérieure (ENS Paris, France). Since 2016, she has worked in industry bridging the gap between research and product in the machine-learning domain.*

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org