Probabilistic Prompt Adaptation
for Unified Image Aesthetics and Quality Assessment

Supplementary Material

Overview

This document provides supplementary details and additional analyses for "Probabilistic Prompt Adaptation for Unified Image Aesthetics and Quality Assessment". It covers (i) prompt pool generation details, (ii) prompt pool coverage and generalization, (iii) computational efficiency and comparison with an off-the-shelf MLLM baseline, (iv) additional discussion on positioning, (v) detailed settings of the user study, and (vi) hyperparameter ablations. Example images shown in this paper and supplementary material are sampled from the Open Images Dataset [1], which provides per-image source and license metadata. For the Open Images images used in this paper and supplementary material, the per-image license is CC BY 2.0 (https://creativecommons.org/licenses/by/2.0/); images are only resized for presentation.

A. Details of Prompt Generation

A.1. Prompts used for generation

This section describes the details of how we generated the prompt samples $\mathcal{T}_{\text{samp}}$ used throughout all experiments for marginalizing text prompts in the prompt selection model (see Sec. 3.4 of the main paper). We employed the Large Language Model (LLM) GPT-5 to automatically generate $\mathcal{T}_{\text{samp}}$ by providing the following instructions.

Aesthetics (Attribute-Specific).

```
Create antonymous text prompt pairs that
describe positive and negative aspects of
image aesthetics. For example: "Stunning
photo." / "Dull photo." , "Vibrant photo."
/ "Faded photo." , "Elegant photo." /
"Clumsy photo." , and "Captivating photo."
/ "Uninspiring photo." Generate four
pairs of prompts in the form of "XXX
photo." and "YYY photo." for each of
the following 11 aesthetic attributes:
Interesting content, Subject emphasis,
Good lighting, Color harmony, Vivid
colors, Depth of field, Motion blur, Rule
of thirds, Balancing element, Repetition,
and Symmetry, resulting in a total of 44
pairs.
```

Image Quality (Attribute-Specific).

```
Similarly, create antonymous text
prompt pairs describing good and poor
image quality. Generate four pairs for
each of the following nine degradation
factors: Blur, Color-related, Contrast,
Compression, Noise, Over-exposure,
Quantization, Under-exposure, and
Spatially-localized, resulting in 36 pairs
in total.
```

Aesthetics (General).

```
Create 100 pairs of antonymous text
prompts in the form of "XXX photo."
and "YYY photo." that describe general
aesthetic quality, excluding the following
elements: Interesting content, Subject
emphasis, Good lighting, Color harmony,
Vivid colors, Depth of field, Motion
blur, Rule of thirds, Balancing element,
Repetition, and Symmetry.
```

Image Quality (General).

```
Create 80 pairs of antonymous text
prompts in the form of "XXX photo."
and "YYY photo." that describe
general image quality, excluding the
following degradation factors: Blur,
Color-related, Contrast, Compression,
Noise, Over-exposure, Quantization,
Under-exposure, and Spatially-localized.
```

A.2. Generated text pairs

The text pairs generated for each prompt are shown in Tabs. S6 to S9.

B. Prompt Pool Coverage and Generalization

This section presents results illustrating how PPA generalizes to textual variations and to prompts beyond those sampled during training. We include qualitative examples, a reliability analysis, and prompt replacement experiments. Our prompt pool is designed to cover the standard semantic space of IAA/IQA tasks, including 11 aesthetics and 9 quality attributes, as well as general descriptors beyond this predefined attribute taxonomy. All prompts in the pool—both attribute-specific and general—are treated identically by our training objective; the attribute taxonomy is only used to structure prompt generation. Attribute-specific concepts may nonetheless be better represented because each attribute is covered by multiple antonym pairs, increasing the local density of semantically similar prompts.

B.1. Qualitative Examples

We provide additional qualitative examples of prompt-specific scoring, complementing the samples shown in Fig. 1 of the main paper, in Figs. S1 and S2. Overall, PPA assigns scores that are broadly consistent with the intended prompt semantics across a wide range of prompts. All example images in this subsection are sampled from the Open Images Dataset [1].

As illustrated in Fig. S1, PPA produces prompt-specific scores that remain consistent under textual rephrasing within the training-time attribute categories, preserving the polarity between the positive and negative prompts. Fig. S2 shows that many prompts beyond the predefined attribute taxonomy still yield reasonable score ordering, while certain abstract or affective prompts (e.g., ”warm”/”cold”) can still exhibit weaker alignment between images and prompt semantics. These cases also serve as qualitative examples of failure or ambiguity in prompt suitability (see Fig. S2). This is consistent with our user study (see Sec. 4.3), where PPA does not always show clear superiority over CLIP [2, 3] or UniQA [4] for such abstract prompts.

B.2. Prompt Similarity and Reliability

To further assess reliability across the prompts used in our user study (see Sec. 4.3), we analyze how their semantic proximity to the training-time prompt set relates to PPA's relative performance. We embed each prompt using CLIP and construct a *difference vector* between its positive and negative forms. For each user-study prompt, we compute the maximum cosine similarity between its difference vector and the set of training-time difference vectors. We then correlate this similarity with PPA's relative performance against each baseline for that prompt in the user study (measured by win rate).

The resulting correlations are 0.220 for CLIP [2] and -0.034 for UniQA [4]. This suggests that prompts that are semantically closer to the training prompt set tend to yield higher relative performance against CLIP, whereas this relationship is weak for UniQA, possibly due to differences in its own training distribution.

B.3. Quantitative Evaluation via Task-Specific Scoring

To quantify robustness, we keep all learned model parameters fixed after training and modify only the prompt candidates used for computing prompt weights at inference time. Specifically, we either restrict the prompt candidates to a subset of the original pool ($\mathcal{T}_{\text{samp}}$) or replace them with newly constructed prompts, without updating any learned parameters. In this analysis, we also use a simple prompt-dropout regularization during training by randomly disabling 50% of prompts per iteration to reduce reliance on any fixed prompt set. We report SRCC/PLCC across 12 datasets. We consider: (i) attribute-only prompts extracted from the original set, (ii) rephrased prompts within the same attributes, (iii) new-attribute prompts, and (iv) general-purpose prompts constructed without specifying attributes. For fair comparison, each replacement setting uses 80 prompts. The results are summarized in Tab. S1.

These results show that PPA is robust to textual rephrasing within the same attribute categories, and retains moderate robustness under unseen attributes. Performance degrades mainly for prompts that are weakly grounded in perceptual cues or semantically distant from the training-time prompt set.

B.4. Limitations and Future Work

While this strategy ensures broad coverage, abstract or affective concepts (e.g., ”warm/cold”) can remain challenging due to weaker visual grounding and sparser neighborhoods of semantically similar prompts, even when included among general prompts. We leave investigating whether applying prompt-dropout regularization more broadly (e.g., into prompt-specific scoring) can further improve performance for future work. We do not compare alternative prompt construction strategies (e.g., human-curated prompts or dataset-driven prompt discovery) and leave such comparisons for future work.

C. Efficiency and Relation to MLLM Baselines

C.1. Computational Efficiency

PPA uses an LLM only *offline* to generate a pool of text prompts. At inference time, all prompt text embeddings are pre-computed and cached, and prompt marginalization is implemented as a lightweight matrix multiplication. As a result, inference cost is similar to fixed-prompt CLIP-IQA [3] while retaining prompt-controllable scoring.

C.2. Comparison with an MLLM Baseline

Recent MLLM-based approaches have demonstrated strong performance in unified image understanding. To clarify the positioning of PPA relative to such models, we compare prompt–image consistency against an off-the-shelf MLLM (GPT-5 mini) in a controlled human study following the same protocol as Sec. 4.3 of the main paper. We focus on eight prompts selected to cover diverse semantic properties, spanning both visually grounded attributes (e.g., surface homogeneity) and more abstract or affective concepts (e.g., ”warm/cold”).

Since existing MLLMs are not specifically fine-tuned for fine-grained score prediction conditioned on arbitrary prompts as required in our setting, we use the off-the-shelf

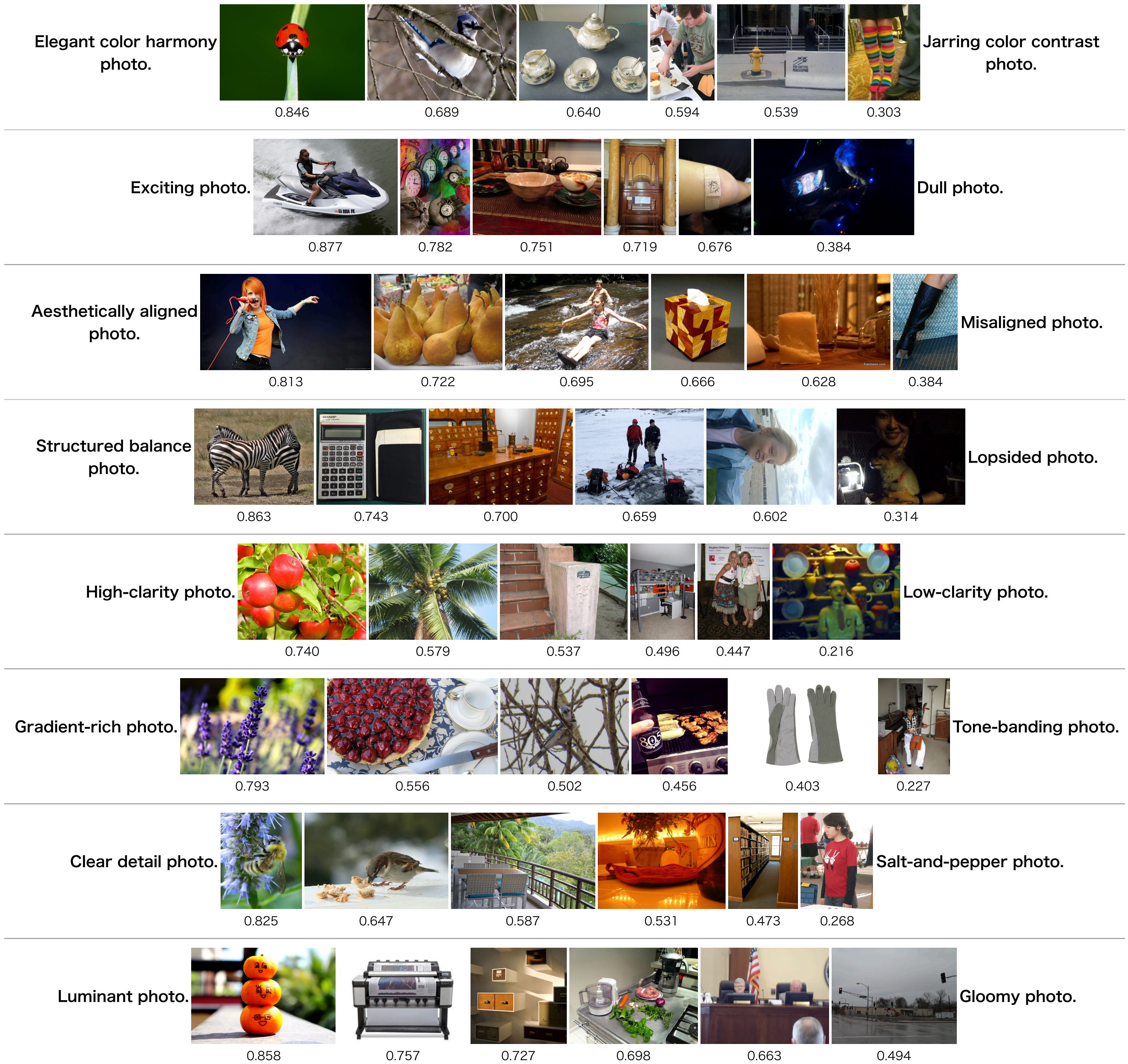


Figure S1. Examples of prompt-specific scoring with rephrased prompts within training-time attribute categories. The text on the left is the *positive prompt*, and the text on the right is the *negative prompt*. Images are arranged from higher to lower predicted scores, and the values below each image are PPA scores.

model as a reference baseline.

We report (i) computational efficiency and overall preference in Tab. S2, and (ii) prompt-wise prompt–image consistency in Tab. S3. The parameter count and FLOPs for GPT-5 mini in Tab. S2 are rough reference values and are provided only to indicate the order-of-magnitude difference in computational scale.

As shown in Tab. S2, despite using a lightweight CLIP-B/16 backbone, PPA achieves no statistically significant difference in overall preference compared to GPT-5 mini (1174 vs. 1226 wins; $p = 0.140$) while operating at a substantially smaller model and computational scale (151M params and 34G FLOPs vs. >8B params and >150G FLOPs).

We further provide a prompt-wise breakdown of prompt–image consistency in Tab. S3.

Figure S2. Examples of prompt-specific scoring using prompt pairs beyond the predefined attribute taxonomy. While many such prompts still yield reasonable ordering, certain abstract or affective prompts show weaker alignment.

Table S1. Comparison of SRCC and PLCC under different prompt replacement settings.

Prompts	#Prompts	SRCC	PLCC
Original Prompts	260	0.807	0.825
Attribute-Only Prompts	80	0.805	0.822
Rephrased Attribute Prompts	80	0.783	0.802
New-Attribute Prompts	80	0.756	0.772
General-Purpose Prompts	80	0.749	0.761

As seen in Tab. S3, PPA tends to perform better on visually grounded prompts (e.g., ”homogeneous surface”, $p < 0.001$), whereas GPT-5 mini shows advantages on certain abstract or affective concepts (e.g., ”cold and uninvit-

Table S2. Efficiency and overall preference comparison against GPT-5 mini in our human study.

Model	Params	FLOPs	Overall Preference
GPT-5 mini	>8B	>150G	51.1% (1226 wins)
PPA (Ours)	151M	34G	48.9% (1174 wins)

No statistically significant difference overall ($p = 0.140$).

Table S3. Comparison of prompt–image consistency between PPA and GPT-5 mini. Each cell shows the number of times each method was preferred in the format ”PPA / GPT-5 mini (p-value)”. Blue indicates that PPA is significantly superior at the 5% level, and pink indicates that GPT-5 mini is significantly superior.

Prompt	PPA / GPT-5 mini
photo where the subject stands out beautifully	140 / 160 (0.113)
photo where the subject is lost in the background	146 / 154 (0.302)
warm and beautiful photo	141 / 159 (0.136)
cold and uninviting photo	130 / 170 (0.009)
balanced contrast high-quality photo	156 / 144 (0.263)
uneven contrast low-quality photo	143 / 157 (0.193)
homogeneous surface high-quality photo	184 / 116 (< 0.001)
uneven quality low-quality photo	134 / 166 (0.028)
Total	1174 / 1226 (0.140)

ing”, $p = 0.009$).

For GPT-5 mini, scores were obtained by supplying an image and a text prompt through the API under standardized and deterministic settings (temperature=0):

- **System prompt:**

```
You are an image evaluation specialist.
```

- **Instruction prompt:**

```
Evaluate how well the input image matches
the concept of '{keyword}' on a scale
from 0.000 to 1.000.  Respond only with
the score as a decimal number with three
digits after the decimal point.
```

Here, {keyword} is replaced with either the *positive* or *negative* side of an antonymous prompt pair. Using these scores, we selected the top 10% and bottom 10% images for each prompt and asked participants to choose which image better matches the prompt semantics.

D. Method Positioning and Connections

PPA reformulates prompt-based IAA/IQA scoring by treating the text prompt as a *latent semantic variable*. Instead of optimizing a scorer for a fixed prompt, PPA learns a task-conditioned prompt distribution and optimizes the marginal likelihood of scores over a prompt pool under task-level supervision. This yields prompt-controllable scoring without prompt- or attribute-level annotations and provides interpretable prompt posterior weights. Conceptually, this is related to mixture-of-experts and attention-based selection in that the model learns to weight a set of experts, but PPA uses *explicit natural-language prompts* as the experts and marginalizes over them probabilistically. Unlike learned expert branches, the experts in PPA are semantically meaningful text prompts.

Table S4. Ablation study on the effect of task embedding dimension.

Dimension	SRCC	PLCC
2	0.785	0.805
4	0.808	**0.828**
8	**0.809**	**0.828**

E. User Study Settings

We provide detailed settings of the user study described in Sec. 4.3 of the main paper.

E.1. Crowdsourcing Platform and Evaluation Interface

The evaluation interface used in our user study is shown in Fig. S3. The user study was conducted through a web-based crowdsourcing platform (Yahoo! Japan crowd sourcing). Participants were free to choose their display device, with 50.6% using smartphones and 49.4% using personal computers. The experiment was conducted online with participants residing in Japan, and Japanese translations of the prompts were provided alongside the original text.

E.2. Construction of Evaluation Prompts

Candidate prompts were generated using GPT-5 to form antonymous text pairs in the form ”XXX photo.” / ”YYY photo.”. We removed those overlapping with the training-time prompt set and selected 21 diverse prompts for the user study. Example retrieval results used for the study, which complement Fig. 3 of the main paper, are shown in Fig. S4. The example images are sampled from the Open Images Dataset [1].

F. Hyperparameter Ablation

We report hyperparameter ablation results in Tabs. S4 and S5. In Tab. S4, we vary the dimension of the task embedding in the prompt selection model and observe that performance largely saturates at a dimension of four. In Tab. S5, we vary the standard deviation σ in the score prediction model and find that $\sigma = 0.1$ (our default setting) performs best.

References

[1] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Tom Duerig, and Vittorio Ferrari. The open

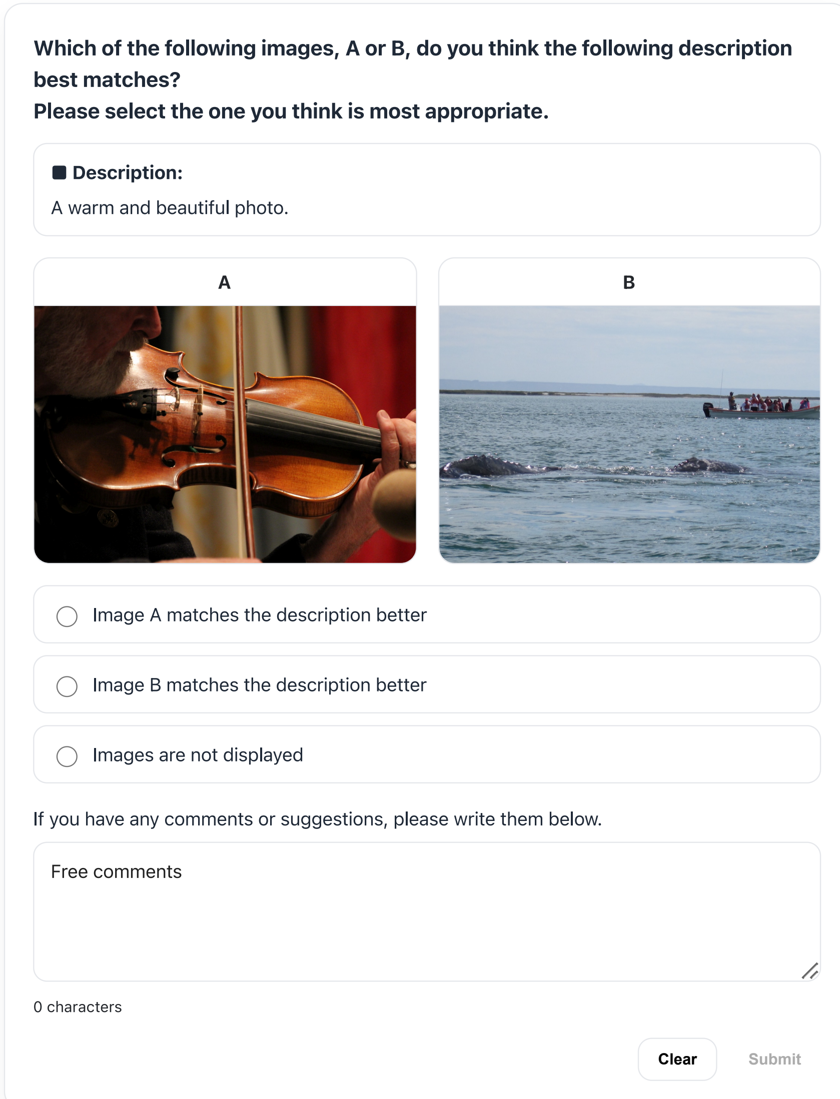


Figure S3. UI of the evaluation form used in the user study. Two images are displayed side-by-side, and participants select the one that best matches the given text description.

Table S5. Ablation study on the standard deviation σ of the score prediction model.

σ	SRCC	PLCC
0.05	0.805	0.825
0.1	**0.808**	**0.828**
0.2	0.790	0.807

images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *IJCV*, 128: 1956–1981, 2020.

[2] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

[3] Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, pages 2555–2563, 2023.

[4] Hantao Zhou, Longxiang Tang, Rui Yang, Guanyi Qin, Yan Zhang, Runze Hu, and Xiu Li. Uniqa: Unified vision-language pre-training for image quality and aesthetic assessment. *arXiv preprint arXiv:2406.01069*, 2024.

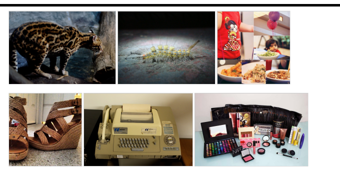
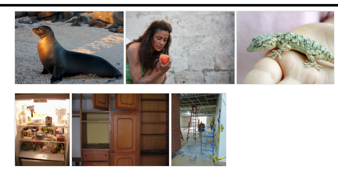
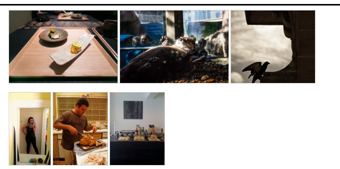
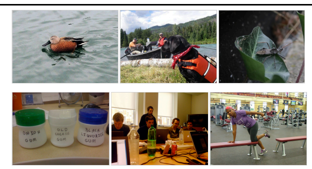
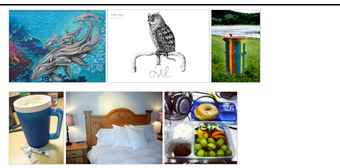
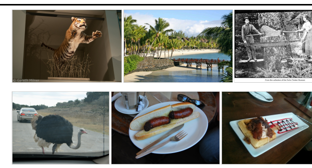
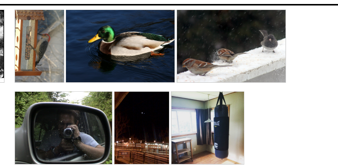
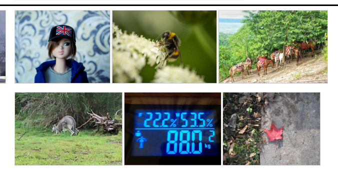
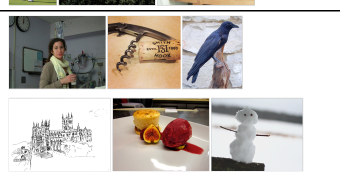
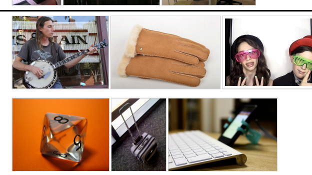
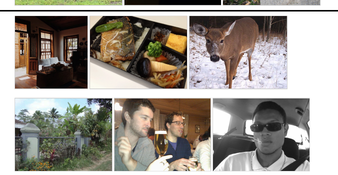
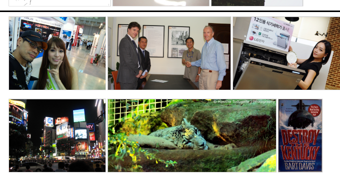
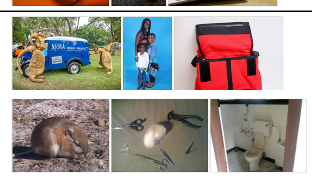
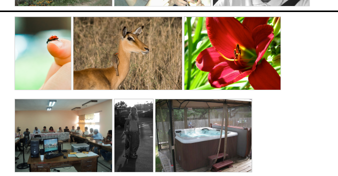
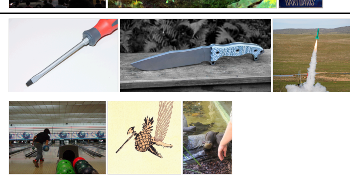
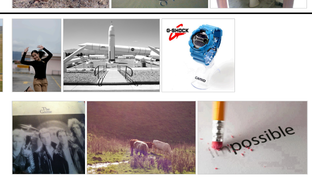
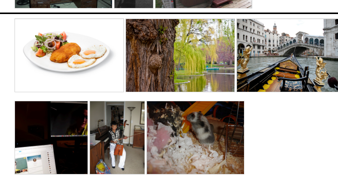
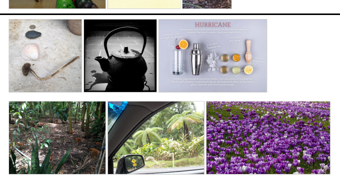
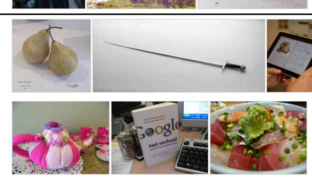
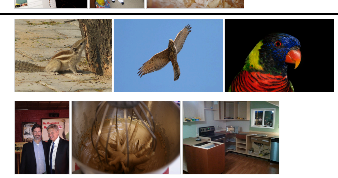
	CLIP	UniQA	PPA
beautifully composed photo poorly composed, unattractive photo			
naturally colored and beautiful photo unnaturally colored, unattractive photo			
beautiful photo with depth and dimensionality flat, unappealing photo			
beautifully detailed photo coarse, unattractive photo			
focused high-quality photo defocused low-quality photo			
clean color high-quality photo color-bleeding low-quality photo			
clean-edge high-quality photo pixelated low-quality photo			
pure-tone high-quality photo color-noisy low-quality photo			

Figure S4. Examples of retrieved images used in the user study. For each prompt, CLIP [2], UniQA [4], and PPA retrieve images corresponding to the positive or negative side of the antonymous prompt pair.

Table S6. Antonymous text prompt pairs for aesthetic attributes.

Attribute	Positive Prompt	Negative Prompt
Interesting Content	Fascinating photo. Engaging photo. Intriguing photo. Compelling photo.	Boring photo. Unremarkable photo. Tedious photo. Forgettable photo.
Subject Emphasis	Well-focused subject photo. Clearly emphasized subject photo. Strong focal point photo. Subject-centered photo.	Distracting composition photo. Unclear subject photo. Weak focal point photo. Background-dominated photo.
Good Lighting	Beautifully lit photo. Soft lighting photo. Well-exposed photo. Evenly illuminated photo.	Poorly lit photo. Harsh lighting photo. Underexposed photo. Uneven lighting photo.
Color Harmony	Harmonious color photo. Color-balanced photo. Pleasing tones photo. Well-blended color photo.	Clashing color photo. Unbalanced color photo. Harsh tones photo. Mismatched color photo.
Vivid Colors	Vibrant photo. Colorful photo. Rich tone photo. Bright color photo.	Faded photo. Washed-out photo. Pale tone photo. Dull color photo.
Depth of Field	Shallow depth photo. Beautiful bokeh photo. Well-isolated subject photo. Focused foreground photo.	Flat depth photo. Distracting background photo. Busy background photo. Blurred subject photo.
Motion & Blur	Smooth motion photo. Intentional blur photo. Dynamic motion photo. Graceful blur photo.	Shaky motion photo. Accidental blur photo. Frozen lifeless photo. Chaotic blur photo.
Rule of Thirds	Well-composed thirds photo. Balanced thirds photo. Dynamic thirds composition photo. Off-center focus photo.	Poorly composed photo. Centered photo. Static centered photo. Middle-weighted photo.
Balance & Symmetry	Well-balanced photo. Visually stable photo. Even composition photo. Symmetrically balanced photo.	Unbalanced photo. Top-heavy photo. Tilted composition photo. Asymmetrically awkward photo.
Repetition & Pattern	Pattern-rich photo. Rhythmic repetition photo. Structured pattern photo. Geometric repetition photo.	Patternless photo. Random repetition photo. Chaotic pattern photo. Disordered repetition photo.
Symmetry	Perfectly symmetrical photo. Mirror-like symmetry photo. Centered symmetry photo. Geometric symmetry photo.	Asymmetrical photo. Off-balance photo. Skewed photo. Distorted photo.

Table S7. Antonymous text prompt pairs for image quality attributes.

Attribute	Positive Prompt	Negative Prompt
Blur	Sharp photo.	Blurry photo.
	Crisp detail photo.	Out-of-focus photo.
	Well-defined photo.	Soft photo.
	Clear-edge photo.	Smudged-edge photo.
Color-related	Natural color photo.	Color-distorted photo.
	Accurate color photo.	Color-shifted photo.
	Vivid color photo.	Faded color photo.
	Balanced hue photo.	Uneven hue photo.
Contrast	High-contrast photo.	Low-contrast photo.
	Crisp tone photo.	Flat tone photo.
	Well-separated shadow and highlight photo.	Muddy contrast photo.
	Rich dynamic range photo.	Compressed dynamic range photo.
Compression	Uncompressed photo.	Compressed photo.
	Smooth texture photo.	Blocky texture photo.
	High-quality photo.	Heavily compressed photo.
	Artifact-free photo.	Artifacted photo.
Noise	Clean photo.	Noisy photo.
	Smooth surface photo.	Grainy photo.
	Noise-free photo.	Speckled photo.
	Low-noise photo.	High-noise photo.
Over-exposure	Properly exposed photo.	Overexposed photo.
	Balanced brightness photo.	Blown-out photo.
	Detailed highlight photo.	Clipped highlight photo.
	Natural light photo.	Too-bright photo.
Quantization	Smooth gradient photo.	Banding photo.
	Fine tonal transition photo.	Posterized photo.
	Continuous tone photo.	Stepwise tone photo.
	Soft shading photo.	Quantized shading photo.
Under-exposure	Properly lit photo.	Underexposed photo.
	Balanced exposure photo.	Too-dark photo.
	Visible detail photo.	Shadow-crushed photo.
	Bright-enough photo.	Dim photo.
Spatially-localized	Uniform quality photo.	Locally distorted photo.
	Even texture photo.	Patchy photo.
	Consistent detail photo.	Regionally blurred photo.
	Smooth area photo.	Localized artifact photo.

Table S8. Antonymous text prompt pairs for general aesthetics.

Positive	Negative	Positive	Negative
Beautiful photo.	Ugly photo.	Elegant photo.	Clumsy photo.
Refined photo.	Crude photo.	Graceful photo.	Awkward photo.
Artistic photo.	Uncreative photo.	Aesthetic photo.	Unaesthetic photo.
Stylish photo.	Tasteless photo.	Sophisticated photo.	Primitive photo.
Charming photo.	Bland photo.	Delicate photo.	Rough photo.
Emotional photo.	Emotionless photo.	Heartfelt photo.	Detached photo.
Romantic photo.	Unfeeling photo.	Peaceful photo.	Tense photo.
Serene photo.	Chaotic photo.	Dreamlike photo.	Literal photo.
Poetic photo.	Plain photo.	Calm photo.	Agitated photo.
Soothing photo.	Harsh photo.	Mystical photo.	Ordinary photo.
Creative photo.	Predictable photo.	Innovative photo.	Conventional photo.
Imaginative photo.	Literal photo.	Experimental photo.	Safe photo.
Conceptual photo.	Cliché photo.	Visionary photo.	Derivative photo.
Expressive photo.	Flat photo.	Inventive photo.	Unoriginal photo.
Unique photo.	Generic photo.	Original photo.	Copied photo.
Dynamic photo.	Static photo.	Vivid photo.	Faded photo.
Energetic photo.	Lifeless photo.	Dramatic photo.	Boring photo.
Powerful photo.	Weak photo.	Impactful photo.	Forgettable photo.
Exciting photo.	Dull photo.	Captivating photo.	Uninspiring photo.
Bold photo.	Timid photo.	Majestic photo.	Mundane photo.
Natural photo.	Artificial photo.	Organic photo.	Synthetic photo.
Authentic photo.	Fake photo.	Candid photo.	Staged photo.
Raw photo.	Overprocessed photo.	Genuine photo.	Manipulated photo.
Unfiltered photo.	Overedited photo.	Realistic photo.	Unnatural photo.
Pure photo.	Distorted photo.	Honest photo.	Deceptive photo.
Minimalist photo.	Busy photo.	Clean photo.	Cluttered photo.
Simple photo.	Overcomplicated photo.	Balanced composition photo.	Overloaded composition photo.
Subtle photo.	Gaudy photo.	Understated photo.	Excessive photo.
Neat photo.	Messy photo.	Orderly photo.	Chaotic photo.
Elegant simplicity photo.	Visual noise photo.	Clear layout photo.	Confusing layout photo.
Storytelling photo.	Meaningless photo.	Narrative photo.	Empty photo.
Symbolic photo.	Literal photo.	Metaphoric photo.	Flat photo.
Evocative photo.	Unmoving photo.	Thought-provoking photo.	Mindless photo.
Insightful photo.	Superficial photo.	Philosophical photo.	Trivial photo.
Cultural photo.	Contextless photo.	Reflective photo.	Shallow photo.
Timeless photo.	Outdated photo.	Classic photo.	Trendy photo.
Iconic photo.	Forgettable photo.	Mature photo.	Amateurish photo.
Professional photo.	Unskilled photo.	Polished photo.	Rough photo.
Elegant tone photo.	Harsh tone photo.	High-quality finish photo.	Unrefined finish photo.
Balanced tone photo.	Uneven tone photo.	Well-crafted photo.	Careless photo.
Inviting photo.	Repelling photo.	Warm photo.	Cold photo.
Comforting photo.	Disturbing photo.	Friendly photo.	Hostile photo.
Intimate photo.	Distant photo.	Welcoming photo.	Alienating photo.
Pleasant photo.	Unpleasant photo.	Gentle photo.	Aggressive photo.
Soft photo.	Harsh photo.	Tender photo.	Severe photo.
Majestic landscape photo.	Unremarkable landscape photo.	Charming portrait photo.	Unflattering portrait photo.
Atmospheric cityscape photo.	Flat cityscape photo.	Peaceful nature photo.	Chaotic nature photo.
Impressive architecture photo.	Clumsy architecture photo.	Graceful motion photo.	Awkward motion photo.
Elegant still-life photo.	Messy still-life photo.	Dreamy light photo.	Flat light photo.
Balanced perspective photo.	Distorted perspective photo.	Artful framing photo.	Unconsidered framing photo.

Table S9. Antonymous text prompt pairs for general image quality.

Positive	Negative	Positive	Negative
High-quality photo.	Low-quality photo.	Clear photo.	Hazy photo.
Sharp photo.	Soft photo.	Detailed photo.	Vague photo.
Well-defined photo.	Indistinct photo.	Clean photo.	Dirty photo.
Polished photo.	Rough photo.	Pristine photo.	Worn photo.
Crisp photo.	Smudged photo.	Refined photo.	Coarse photo.
Stable photo.	Shaky photo.	Consistent photo.	Inconsistent photo.
Accurate photo.	Inaccurate photo.	Reliable photo.	Unstable photo.
Uniform photo.	Uneven photo.	Well-rendered photo.	Poorly rendered photo.
Faithful photo.	Distorted photo.	Natural-looking photo.	Artificial-looking photo.
True-to-source photo.	Altered photo.	Unwarped photo.	Warped photo.
Smooth photo.	Rough-textured photo.	Even-toned photo.	Blotchy photo.
Gradual transition photo.	Harsh transition photo.	Fine-textured photo.	Grainy-textured photo.
Continuous-tone photo.	Patchy-tone photo.	Balanced texture photo.	Broken texture photo.
Clean edge photo.	Jagged edge photo.	Precise line photo.	Irregular line photo.
Consistent surface photo.	Uneven surface photo.	Well-blended photo.	Blocky photo.
Artifact-free photo.	Artifacted photo.	Undistorted photo.	Distorted photo.
Uncompressed photo.	Visibly compressed photo.	Seamless photo.	Fragmented photo.
Smooth gradient photo.	Posterized photo.	Natural tone photo.	Artificial tone photo.
Stable exposure photo.	Fluctuating exposure photo.	Clean gradient photo.	Stepped gradient photo.
Fine-detail preserved photo.	Detail-lost photo.	Integrity-preserved photo.	Integrity-damaged photo.
Properly focused photo.	Out-of-focus photo.	Evenly illuminated photo.	Unevenly illuminated photo.
Color-accurate photo.	Color-shifted photo.	Balanced lighting photo.	Harsh lighting photo.
Well-calibrated photo.	Mis-calibrated photo.	Proper white balance photo.	Wrong white balance photo.
Tone-consistent photo.	Tone-inconsistent photo.	Exposure-balanced photo.	Exposure-fluctuated photo.
Shadow-detailed photo.	Shadow-blocked photo.	Highlight-preserved photo.	Highlight-blown photo.
Undamaged photo.	Corrupted photo.	Stable structure photo.	Broken structure photo.
Uniform pixel distribution photo.	Irregular pixel distribution photo.	Aligned photo.	Misaligned photo.
Rectilinear photo.	Tilted photo.	Geometrically correct photo.	Geometrically distorted photo.
Perspective-correct photo.	Perspective-warped photo.	Undegraded photo.	Degraded photo.
Full-resolution photo.	Low-resolution photo.	High-fidelity photo.	Low-fidelity photo.
Balanced tone photo.	Overcontrasted photo.	Smooth highlight photo.	Clipped highlight photo.
Soft shadow photo.	Crushed shadow photo.	Controlled brightness photo.	Inconsistent brightness photo.
Neutral color balance photo.	Color-cast photo.	Consistent chroma photo.	Uneven chroma photo.
Natural gradation photo.	Abrupt gradation photo.	Well-tuned tone photo.	Poorly tuned tone photo.
Proper gamma photo.	Mismatched gamma photo.	Linear response photo.	Nonlinear response photo.
Stable frame photo.	Wobbly frame photo.	Uncropped photo.	Overcropped photo.
Properly scaled photo.	Improperly scaled photo.	Full-detail photo.	Detail-truncated photo.
Non-banding photo.	Banded photo.	Well-textured photo.	Textureless photo.
High dynamic range photo.	Limited dynamic range photo.	Balanced exposure photo.	Over/underexposed photo.
Unclipped photo.	Clipped photo.	Noise-free texture photo.	Noisy texture photo.