Supplementary Materials to "An Image Quality Assessment Dataset for Portraits"

1. Portrait quality attributes

We ask more than 30 experts to annotate PIQ23 following carefully designed guidelines, on four attributes: face detail preservation, face target exposure, global color quality, and overall image quality. In this section, we elaborate on the guidelines for each attribute and give the remarks that were taken into consideration when developing the dataset.

Face detail preservation

We study the fidelity of face rendering in terms of detail preservation and skin smoothness. We tend to appreciate natural facial renderings over unnatural ones. For example, some over-sharpened faces can be considered worse than slightly blurred faces if the latter has a more natural and smooth texture rendering. Additionally, we have found that it is necessary to explicitly define the priorities in penalizing noise when the texture quality is similar. Hence, high-frequency noise is preferable to low-frequency noise, true random noise is better than patterned noise, and luminance noise is better than chromatic noise. Finally, a more general comparison of the skin texture and the facial details were considered, in particular beard, eyebrows, hair, etc.

Face target exposure

This attribute is used to evaluate the quality of the light rendering on the face. We have asked to find the balance between target exposure, contrast, and dynamic range on the face. This evaluation is sometimes hard since some of the attributes can be contradictory and finding a sweet spot between all the criteria is not always straightforward. In case of ambiguous comparisons (when it is not clear which image is better), we have left the choice to the observer and his preferences.

Global color quality *

Using the full image, we have asked to find a trade-off between the overall white balance and color rendering of the portrait image. The focus is usually on the subject since it constitutes the main element in the image, but we have not forced more specific guidelines for this attribute. We have also noted the following points:

- **Note on heavily under-exposed images:** In the case of a heavily under-exposed image, color was penalized, since we cannot really see any color.
- **Note on HDR scenes:** In case there was a failure in the dynamic range of the image, color also was penalized.
- **Note on skin tones:** Evaluating the skin tone rendering was not explicitly taken into consideration in the color attribute, since we need a reference image per person to evaluate it correctly. Defining good skin tone rendering is complicated and prone to subjectivity. The quality of skin tone rendering is hard to judge objectively without having a previous idea about the skin tone of the person. Also, skin reflectance (how much light the skin reflects) should be taken into consideration when analyzing skin tone renderings and when evaluating the target exposure on the face. Another type of analysis should be taken into consideration, such as comparing the quality of the skin tone rendering to a reference image of the person in multiple controlled lighting conditions.

Overall image quality

The overall attribute is considered a trade-off of all the main attributes in a portrait image. We have evaluated the overall quality of the image while prioritizing some aspects over others. We have defined a list of prioritized attributes that were only taken into consideration when the quality difference is ambiguous. Also, important failures were directly penalized. We have left the choice to the observer to estimate the severity of the quality difference in case multiple aspects came into play. We have tried to estimate the role of each attribute on three levels: Essential, important, and subjective. Essential means the attribute always plays a role in the overall quality of the image. Important represents the attributes that occasionally play a role in the overall quality. Finally, subjective attributes rarely play a role in the overall quality and their impact varies accordingly to the observer's opinion. Let us now list the attributes judged to play one of the three aforementioned roles in the overall quality analysis:

- **Face target exposure (essential):** we usually prioritize naturally well-exposed faces over underexposed or overexposed ones.

*This attribute was omitted from the final version of the dataset because of its annotation challenges.

- **Dynamic range, contrast and global target exposure (essential):** Some essential qualities of an image are how well it preserves the details in dark and bright areas, as well as the general exposure and contrast of the scene. Therefore, the dynamic range of the image, the target exposure, and the contrast are three of the first things to look at in the image. This is mostly interesting in portraits since maintaining a high dynamic of the subject and background is not straightforward.
- **White balance, color rendering, and skin tone (essential):** The quality of the color rendering between the portrait and the background can play a big role in image quality. A bit similar to the separate attribute, this one compromises the overall quality of a portrait image.
- **Sharpness and focus (important):** These attributes can be analyzed on single and group photos, with single and multi-planes. We try to analyze the capacity of the camera to focus on the main subject and the ability to separate planes.
- **Artifacts (subjective):** This attribute was be analyzed in case of an important artifact failure or when two images have very close quality levels. Several artifacts can play a subjective role in the overall quality of a portrait. We note some of these such as ghosting, halos, hue shift, ringing, flare, etc.

2. Examples of attribute quality

Face detail preservation

Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
					
-2.8	-0.81	-0.1	0.75	1.57	4.14

Figure 1. Examples of different levels of **face detail preservation quality** and their corresponding *JOD* scores illustrated on the face regions and on crops around the eyes.

Face target exposure

Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
					
-4.91	-2.21	-1.39	0.02	0.9	1.96

Figure 2. Examples of different levels of **face target exposure quality** and their corresponding *JOD* scores.

Global color quality

Level 0	Level 1	Level 2	Level 3
			
-3.97	-0.61	-0.09	1.96

Figure 3. Examples of different levels of **global color quality** and their corresponding *JOD* scores.

Overall image quality

Level 0	Level 1	Level 2	Level 3	Level 4
				
-4.7	-1.78	-0.36	0.7	2.36

Figure 4. Examples of different levels of **overall image quality** and their corresponding *JOD* scores.

3. Domain shift

Face detail preservation

(a) Outdoor scenes. (b) **Left:** outdoor scene; **right:** lowlight scene.

Figure 5. Examples of domain shift in **face detail preservation** annotations between different scenes of **(a)** same lighting condition and **(b)** different lighting conditions. All images have a *JOD* value close to 0.

Face target exposure

(a) Outdoor scene. (b) Indoor scene. (c) Lowlight scene.

Figure 6. Examples of domain shift in **face target exposure** annotations between different lighting conditions. All images have a *JOD* value close to 0.

Global color quality

(a) Outdoor scene. (b) Indoor scene. (c) Lowlight scene.

Figure 7. Examples of domain shift in **global color quality** annotations between different lighting conditions. All images have a *JOD* value close to 0.

Overall image quality

(a) Outdoor scene. (b) Indoor scene. (c) Lowlight scene.

Figure 8. Examples of domain shift in **overall image quality** annotations between different lighting conditions. All images have a *JOD* value close to 0.

4. PIQ23 Characteristics

Smartphone devices

PIQ23 was collected with over 100 smartphone devices and 14 brands from the last decade. Additionally, multiple camera modes were included across the scenes. Table 1 shows the distribution of the smartphone brands of PIQ23.

Brand	Nb devices
Samsung	20
Xiaomi	14
Oppo	14
Apple	13
Vivo	9
Huawei	7
OnePlus	6
Sony	5
Google	4
Asus	3
Realme	2
Motorola	2
Other	3

Table 1. Smartphone brand distribution in PIQ23

Ethnicities and genders

We have constructed PIQ23 with extra attention to gender and ethnic biases. We have tried to the best of our capabilities to minimize those biases through bias analysis. It should be noted that PIQ23 is the result of multiple years of engineering and photographic efforts and is not necessarily uniformly distributed through all characteristics. We have separated the skin tones, following the Fitzpatrick skin type (FST) ruler [1], into four categories: Fair, Asian, Medium, and Deep. We recognize the difficulties encountered by image enhancement algorithms on deep skin tones, hence we have also designed some of the PIQ23 scenes to include uniquely deep skin tones. Table 2 shows a rough estimation of the skin tone distribution in PIQ23. A more detailed analysis could be provided later throughout the life cycle of PIQ23.

Skin tone	Estimated %
Fair	50%
Deep	30%
Asian	15%
Medium	5%

Table 2. Skin tone estimated distribution in PIQ23

5. Annotation tool

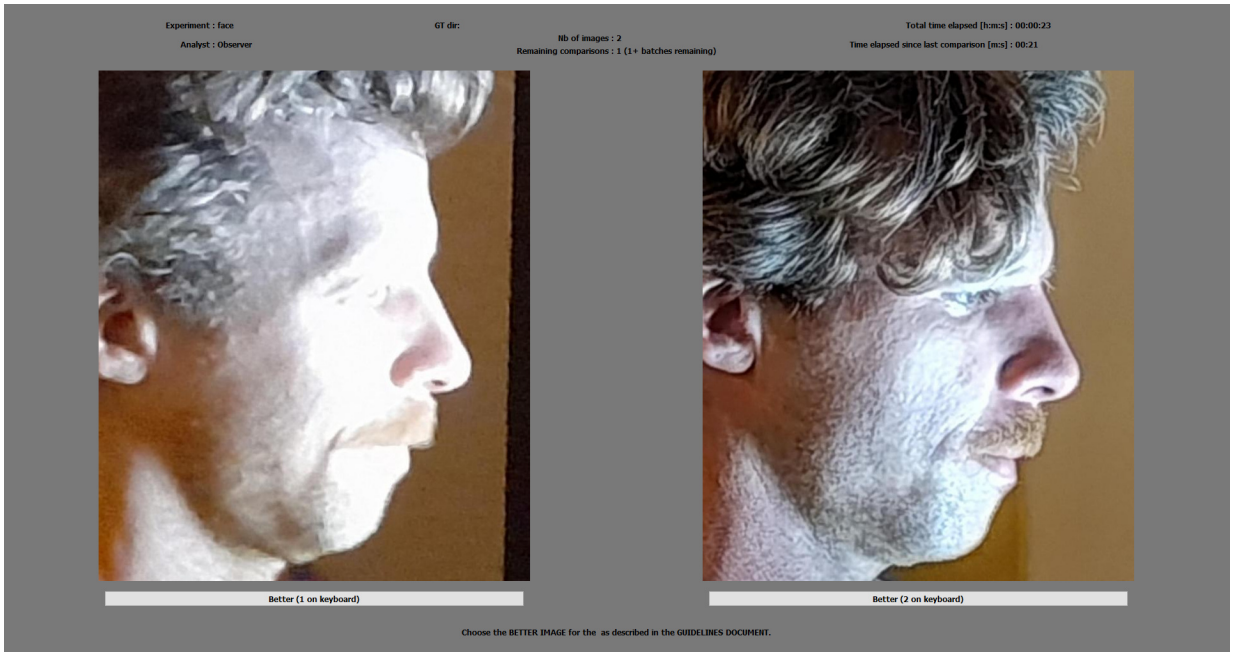

Figure 9. Example of the annotation tool used to acquire pairwise comparisons.

References

[1] Thomas B Fitzpatrick. “The validity and practicality of sun-reactive skin types I through VI”. In: *Archives of dermatology* 124.6 (1988), pp. 869–871.