Supplementary Materials to "PPR10K: A Large-Scale Portrait Photo Retouching Dataset with Human-Region Mask and Group-Level Consistency"

Jie Liang[1,2,*], **Hui Zeng**[1,2,*], **Miaomiao Cui**[2], **Xuansong Xie**[2] and **Lei Zhang**[1,2]
[1]The HongKong Polytechnic University, [2]DAMO Academy, Alibaba Group

1. Study on the GLC Measure

As mentioned in the main paper, the GLC measure is defined based on the statistics of colors. In this section, we conduct ablation studies to find the most reliable color channels for computing the GLC measure.

We randomly select 50 groups of photos from the PPR10K dataset, and denoted by G their ground truths, which have good GLC since they are elaborately adjusted by human experts. We then randomly distort the photos in G by changing five attributes (including temperature, tint, exposure, contrast and saturation) via CameraRaw to obtain 100 new group sets $[G_1, G_2, \cdots, G_{100}]$ with decreasing GLC. Specifically, we set a range $[-a_i, a_i]$ for the i-th attribute A_i. To generate G_n, $n = 1, 2, \cdots, 100$, we randomly distort each attribute A_i in a restricted range $\frac{n}{100} \times [-a_i, a_i]$. Naturally, larger n results in lower GLC in G_n, as shown in Figure 1. We can set the ground-truth GLC of G_n as $\mathcal{M}_{GLC}^{real}(G_n) = 1 - \frac{n}{100}$. We then calculate the predicted GLC measure $\mathcal{M}_{GLC}^{predict}(G_n^c)$ on different color channels G_n^c using Eq. (3) in the main paper. Finally, we calculate the Pearson Correlation Coefficients (PCC) between $\mathcal{M}_{GLC}^{predict}(G_n^c)$ and $\mathcal{M}_{GLC}^{real}$ among the 100 distorted group sets (each set contains 50 groups of photos) to find the most reliable color channels for computing GLC. The results on the different color channels are reported in Table 1.

As shown in Table 1, the $\mathcal{M}_{GLC}$ calculated on the a and b channels in the Lab color space correlates better with the ground-truth GLC than L and other channels in the RGB space. This is natural since the statistics of a and b channels are more robust to content changes in a group of photos than the other channels. Combining a and b channels achieves the highest PCC among all competitors. As a result, we compute the $\mathcal{M}_{GLC}$ using a and b channels.

2. More Information about the PPR10K Dataset

This section provides more information about the PPR10K dataset. In Figure 2, we provide more example groups and their ground truths retouched by the three experts with consideration of the requirements of HRP and GLC in the PPR task. In Figure 3, we show the high-resolution human-region masks provided by our PPR10K dataset, which are very useful when the subjects and backgrounds need different renditions. In Figure 4, we summarize the statistical information about the PPR10K dataset, including the number of photos per group, shooting time and major adjustments on nine visual attributes in CameraRaw during retouching.

3. Data Augmentation Details

As illustrated in Section 5.1 of the main paper, we augment the training images using 6 major visual attributes, including temperature, tint, exposure, highlights, contrasts and saturation, to enrich the lighting and color distributions of the training set. Specifically, we randomly modify the 6 attributes of each source photo in CameraRaw and get 5 augmented ones. The ranges of the random modifications of each attributes are: Temperature: [-500, 500]; Tint: [-10, 10]; Exposure: [-1, 1]; Highlights: [-35, 35]; Contrasts: [-30, 0]; Saturation: [-30, 0]. Note that the setting of random modification ranges is based on the adjustment ranges of expert retouchers to approximate real illuminations or colors. After the offline tonal augmentation, we have $54, 174$ training pairs in total.

4. More Visual Results

This section provides more qualitative comparisons to demonstrate the effectiveness of the proposed dataset and learning strategies. Figure 5 and Figure 6 show comparisons between models trained on the FiveK dataset and our PPR10K dataset,

while Figure 7 and Figure 8 illustrate the effectiveness of the HRP and GLC learning strategies, respectively.

Table 1: Pearson correlation coefficients between $\mathcal{M}_{GLC}^{predict}$ and $\mathcal{M}_{GLC}^{real}$ on different channels and their combinations with different randomness settings. The best PCC in each setting are shown in **bold**.

#	Ranges	L	a	b	R	G	B	ab	Lab	RGB
#1	Temp ±200; Tint ±10; Expo ±0.5; Cont ±10; Satu ±10	0.8237	0.9188	0.9356	0.8673	0.8745	0.8693	**0.9469**	0.9170	0.8902
#2	Temp ±400; Tint ±30; Expo ±0.75; Cont ±20; Satu ±20	0.8512	0.9230	0.9408	0.8874	0.8933	0.8816	**0.9519**	0.9237	0.9003
#3	Temp ±600; Tint ±50; Expo ±1; Cont ±30; Satu ±30	0.8785	0.9279	0.9485	0.9025	0.9134	0.9082	**0.9591**	0.9280	0.9188

Figure 1: Example photos in a group that is distorted in terms of visual attributes with different randomness. Ranges #1, #2 and #3 indicate the three ranges in Table 1, while n indicates different levels of randomness within each range.

|(a) Input|(b) PPR10K-a|(c) PPR10K-b|(d) PPR10K-c|

Figure 2: Visual examples of (a) source photos and (b-d) their ground truths retouched by three experts. For the top group, the source photos have inconsistent exposure and temperature. For the middle group, the backgrounds are over-exposed. For the bottom group, the human regions are under-exposed. The ground truths retouched by the three experts fulfill both the HRP and GLC requirements while maintaining different styles.
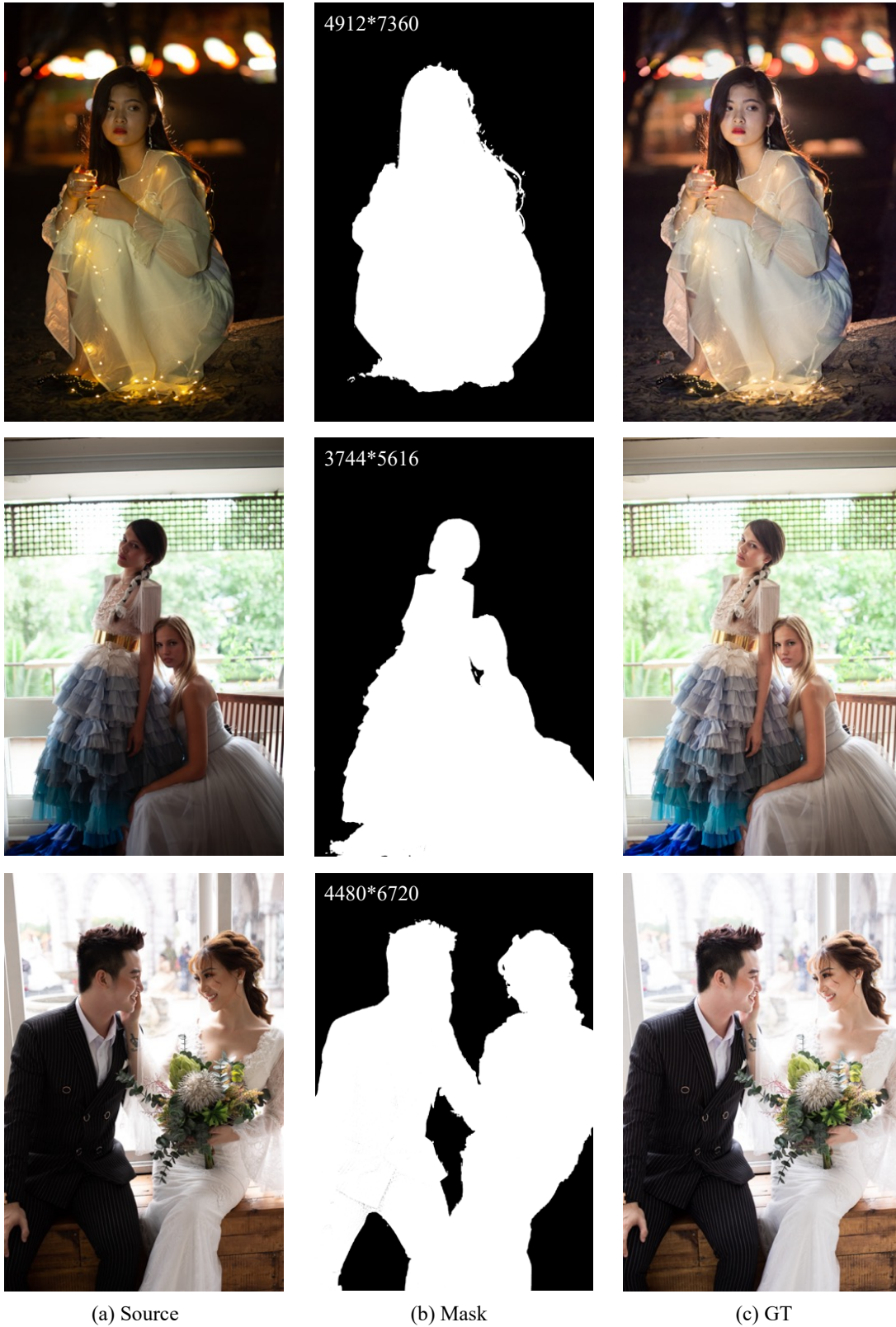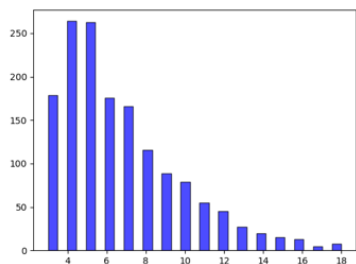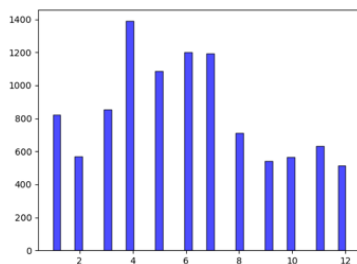
4912*7360

3744*5616

4480*6720

(a) Source   (b) Mask   (c) GT

Figure 3: Examples of (a) source photos and their (c) retouched ground truths with (b) full-resolution human-region masks. The masks help the human regions to receive proper retouching with higher priority than background regions.
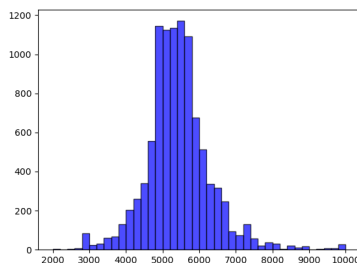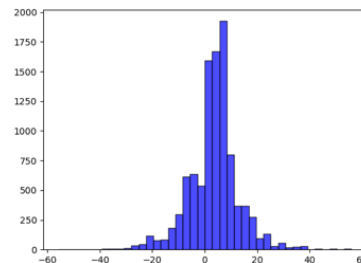
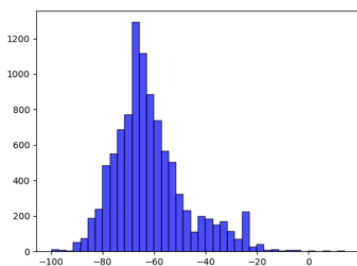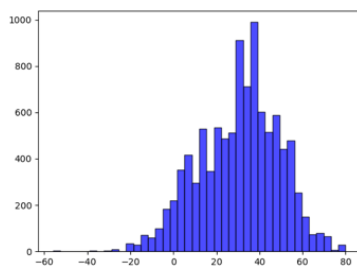(a) Number of photos per group (b) Shooting time, in month (c) Shooting time, in hour
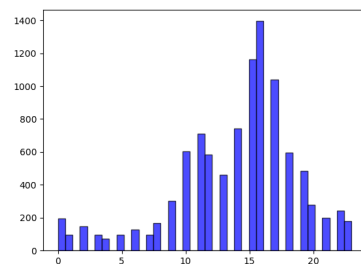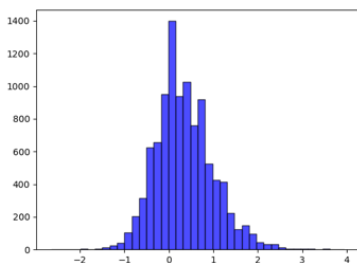
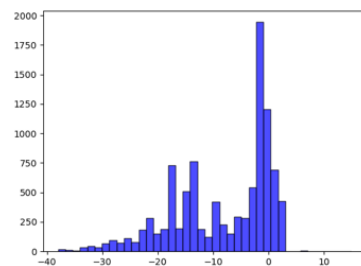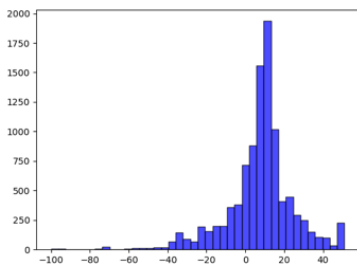(d) Exposure (e) Temperature (f) Tint
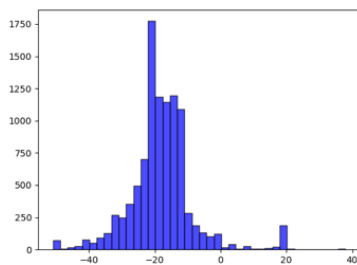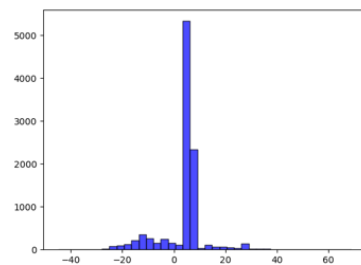
(g) Highlight (h) Shadow (i) Saturation

(j) White (k) Black (l) Contrast

Figure 4: Histograms of basic attributes of the PPR10K dataset, including (a) the number of photos per group, (b, c) the shooting time of each photo (in month and hour) and (d-l) the adjustments of nine major visual attributes in CameraRaw from one of the three retouchers (the other two are similar).

(a) Input   (b) Target   (c) HDRNet (FiveK)   (d) HDRNet (PPR10K)

Figure 5: Visual comparisons between the HDRNet models trained on (c) the FiveK dataset and (d) the proposed PPR10K dataset. The inputs are shown in (a) and the targets in (b) are from PPR10K-a.

(e) CSRNet (FiveK)   (f) CSRNet (PPR10K)   (g) 3DLUT (FiveK)   (h) 3DLUT (PPR10K)

Figure 6: Visual comparisons between the CSRNet/3D LUT models trained on (a, c) the FiveK dataset and (b, d) the proposed PPR10K dataset. The inputs and targets are shown in Figure 5.

| (a) Input | (b) 3DLUT | (c) 3DLUT+HRP |

Figure 7: More visual evaluation examples of the HRP learning strategy by using 3D LUT. The human skins are brighter when the HRP learning strategy is employed.

| (a) Input | (b) 3D LUT | (c) 3D LUT + GLC | (d) 3D LUT + GLC + HRP |

Figure 8: More visual evaluation examples of the GLC learning strategy by using 3D LUT. The lighting and color of human skin are more consistent when the GLC learning strategy is employed.