Time-Aware Auto White Balance in Mobile Photography Supplementary Material

Mahmoud Afifi* Luxi Zhao* Abhijith Punnappurath
Mohammed A. Abdelsalam Ran Zhang Michael S. Brown
AI Center–Toronto, Samsung Electronics
{m.afifi1, lucy.zhao, abhijith.p, m.abdelsalam, ran.zhang, michael.b1}@samsung.com

This supplementary material provides additional details on the experiments presented in the main paper (Sec. 1), additional details on the contextual information, ablation studies, and results (Sec. 2), and detailed information about our dataset (Sec. 3). Lastly, we provide additional ground-truth data to broaden the dataset's impact for other applications (Sec. 4).

1. Additional comparison details

In the main paper and this supplementary material, we report the results of various methods evaluated on our proposed dataset. We benchmark several approaches, including statistical-based (learning-free) methods, camera-specific learning-based methods, and cross-camera learning-based methods.

For the cross-camera learning-based methods (specifically, SIIE [2] and C5 [6]), we report results for three versions of each method:

- A model trained on the Cube++ [19] and NUS [17] datasets (results reported without any postfix). Since the Cube++ and NUS datasets lack user-preference ground truth, we only report results on neutral ground truth using our dataset.
- A model trained on the Cube++ and NUS datasets, as well as our proposed dataset (results reported with the postfix (tuned)). Similarly, due to the absence of user-preference ground truth in the Cube++ and NUS datasets, we only report results on neutral ground truth using our dataset.
- A model trained exclusively on our dataset (results reported with the postfix (tuned-CS), where 'CS' stands for camera-specific).

This approach ensures a fair comparison, as the generic models (trained on Cube++ and NUS datasets) lack exposure to the diverse lighting conditions present in our dataset.

For the C5 method [6], when training the tuned-CS model, we used only the input histogram and excluded the additional histograms proposed in the original method. These additional histograms were mainly intended to assist the model in calibrating for new cameras. Since our dataset uses a single camera, we removed the extra encoders from the C5 model for the tuned-CS version.

For FFCC [8], we first performed tuning to identify optimal hyperparameters before training the model. The model was tuned and trained on our dataset. When reporting FFCC with capture metadata—denoted as FFCC (capture info) in the tables—we used a vector of $[log(\text{shutter-speed}), log(\text{ISO}), 1]$ instead of the original metadata vector $[log(\text{shutter-speed}), log(\text{f-number}), 1] \times [\text{cam-1}, \text{cam-2}, 1]$, for the following reasons: we only have a single camera (the original method was designed to handle two different cameras), and our dataset focuses on smartphone cameras, which have a fixed aperture (no change in f-number per scene).

For the classification-CC method [35], we re-implemented the approach as the original code was unavailable. In our implementation, we set the number of clusters to 50, matching the value used in the original paper [35] for the NUS dataset [17].

For the KNN method [3], which was initially proposed for white-balance correction in the post-capture stage, we followed the adjustments used in the evaluation presented in [2]. Specifically, we replaced the polynomial function used in the original method with the ground-truth 3D illuminant vectors from the training data. The nearest-neighbor process was performed as described in the original paper, but the final output was an illuminant color, rather than a polynomial function.

For the quasi-unsupervised CC method [10], we report the results of both the unsupervised model and the tuned model on our dataset. We used the gray-world (GW) method [12] as the initial estimation for APAP [4].

For the TLCC method [40], we used the official checkpoint released by the authors, trained on the sRGB dataset [51] and raw datasets [17, 21]. We then finetuned the model on our proposed dataset to leverage transfer learning from

*Equal contribution.

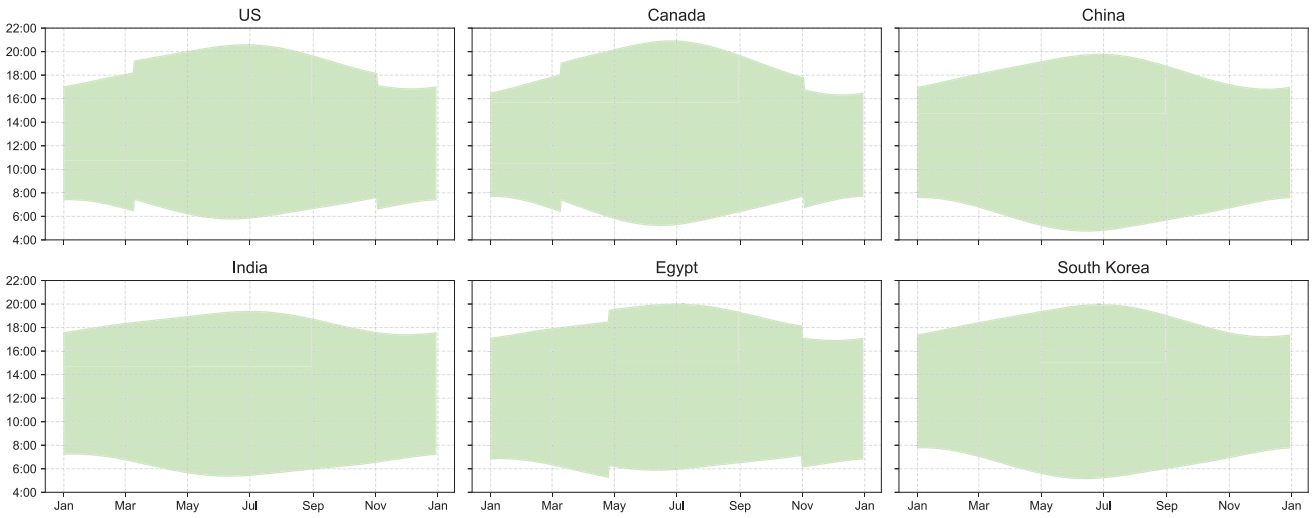


Figure 1. This figure illustrates the daily variations in sunrise and sunset times across different countries throughout the year 2024. The x-axis represents the date, while the y-axis denotes the time of day (4 AM – 10 PM). The light green shaded regions indicate the duration of daylight for each country, highlighting seasonal variations due to differences in latitude and geographical location.

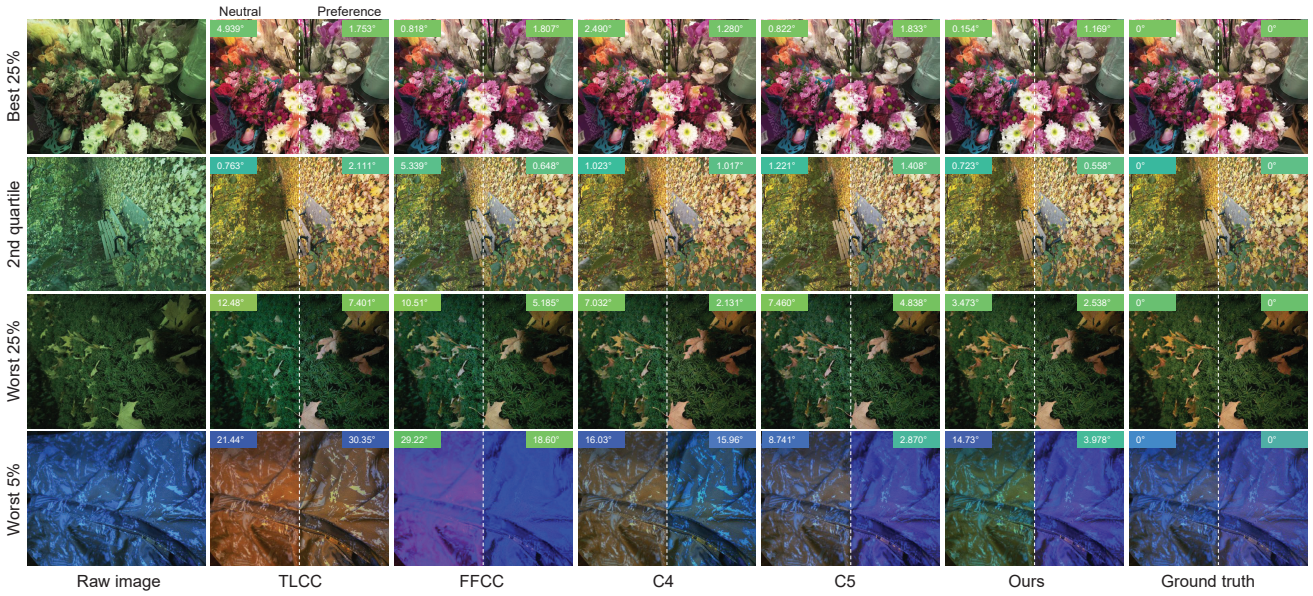


Figure 2. Qualitative examples from the best 25% (first quartile), second quartile, third quartile, and the worst 5% of our results. Shown images are white-balanced using the illuminant estimates from TLCC [40], FFCC [8], C4 [47], C5 [6], and our method. We show results of both types of white-balance corrections: 1) neutral (on the left side of each white-balanced image) and 2) user-preference (on the right side). All images are gamma-corrected to enhance visualization.

sRGB to raw, as described in the TLCC paper [40].

For the gamut method [23], we present the results for three canonical gamuts: edges, pixel colors, and the 1st-degree gradient. For the TECC method [9], we report the results with the 2nd-order gray-edge [41]. Lastly, for the gray-edge (GE) method [41], we provide the results based on both the first and second gradients of the images.

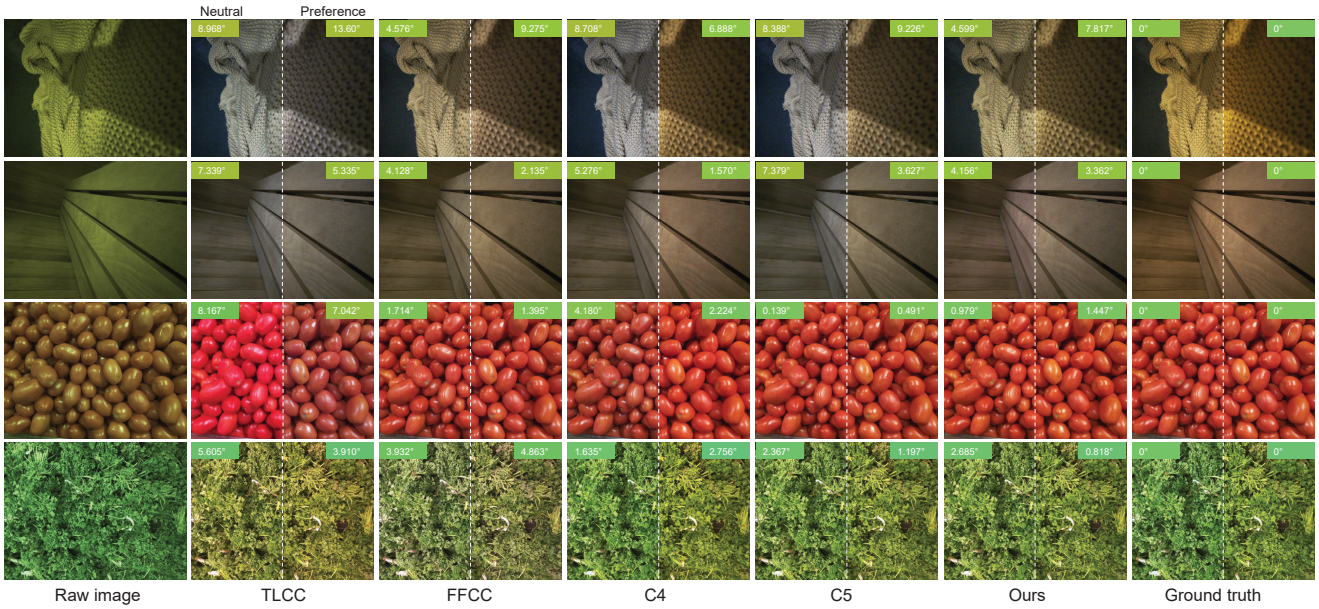


Figure 3. Additional qualitative examples of scenes with limited color variations, which are particularly challenging for illuminant estimation. Images are white-balanced using illuminant estimates from TLCC [40], FFCC [8], C4 [47], C5 [6], and our method. For reference, we also include results corrected using the ground-truth illuminant. We show results of both types of white-balance corrections: 1) neutral (on the left side of each white-balanced image) and 2) user-preference (on the right side). All images are gamma-corrected for better visualization.

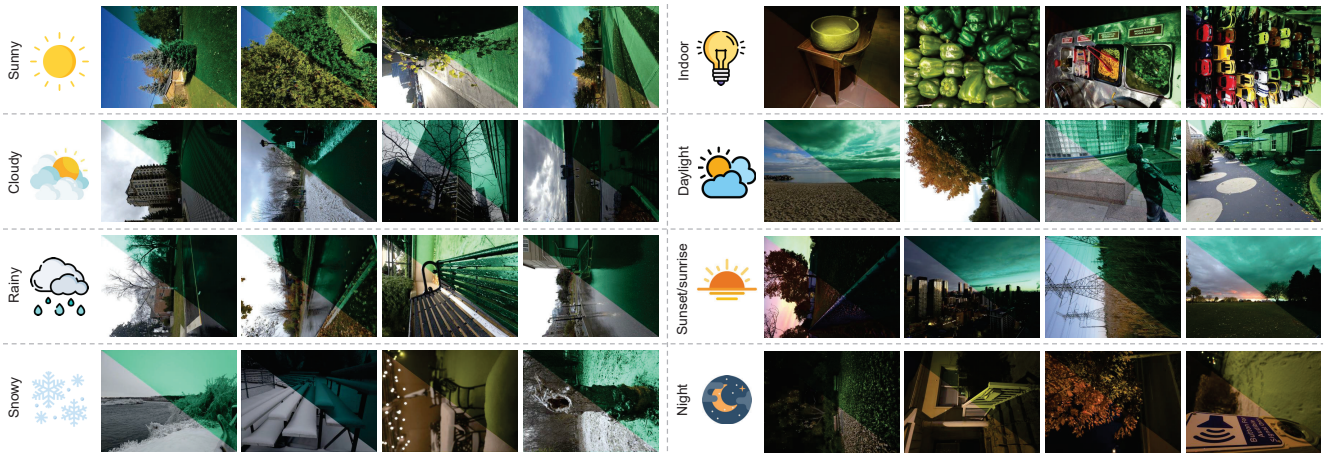


Figure 4. Our dataset includes diverse scenes captured under various weather conditions (sunny, cloudy, rainy, and snowy) and lighting conditions (indoor, daylight, sunset/sunrise, and night). For each example, we show raw images (gamma-corrected for better visualization) alongside their sRGB counterparts.

2. Additional details and results

2.1. Additional details on contextual information

In the main paper, we presented our method, which relies on the "probability" of an image being captured during one of the key solar events (e.g., sunrise, noon, sunset) along with additional capture metadata and color information represented as histograms.

Our method leverages the probability of the time of day, allowing the model to rely on an absolute time reference rather than being affected by location-specific time zones, thereby improving generalization.

Solar event times (i.e., dawn, sunrise, noon, sunset, dusk, and midnight) vary significantly based on the time of year and geographical location. For instance, locations near the equator experience relatively small variation in day length,

whereas higher-latitude regions exhibit more pronounced seasonal differences. In Fig. 1, we illustrate the average length of daylight across different countries and continents. As is well known, sunset/sunrise times vary depending on both location and date.

If we were to use the raw clock timestamp without geolocation, the information would be highly location-dependent and would not generalize well to regions with different solar event timings. An alternative approach would be to provide both geolocation and timestamp, allowing the model to learn their relationship with solar event timings. However, this would require a diverse dataset with images captured across different locations worldwide to ensure robust learning, which may be impractical due to the extensive data collection required. Our method is simpler and more effective–instead of relying on learned patterns, we use traditional astronomical methods [34, 42, 43] to compute solar event times for a given location. This allows us to represent time in an absolute manner, using the probability of an image being captured at each solar event rather than relying on location-specific timestamps.

2.2. Additional ablation studies

In the main paper, we presented a set of ablation studies to analyze the impact of different input features on our method. Here, we provide additional ablation studies on the validation set with masks applied, as shown in Table 1. By default, our results in Table 1 use the histogram feature, $\mathbf{H}$, and the time-capture feature, $\mathbf{c}$, with noise stats, $\mathbf{n}$.

In this additional set of ablation studies, we show results when using only the time feature, $\mathbf{p}$, without the histogram feature (w/o $\mathbf{H}$). We then added capture information, including ISO (i), shutter speed (s), flash status (f), and noise stats ($\mathbf{n}$), one at a time, in addition to the time feature, $\mathbf{p}$. Additionally, we examine the effect of using the histogram feature in combination with only ISO (i), shutter speed (s), flash status (f), and noise stats ($\mathbf{n}$).

Furthermore, we present results using the complete time feature with noise stats, $\mathbf{n}$, under the following conditions:

- Without the edge histogram, $\mathbf{H}_e$.
- Without the additional u and v positional encoding channels (u/v coord.).
- Using the log-uv histogram from prior work [6–8] instead of our R/G and B/G chromaticity histogram.
- Using the R/G, B/G chromaticity image, $\mathbf{I}_{\texttt{chroma}}$, instead of our histogram feature.
- Without pre-processing and normalization of the time-capture feature.
- Using a smaller histogram feature with 24 bins.
- Using a lower-resolution image of 64×48.
- Without the time feature, $\mathbf{p}$.
- Various combinations of noise stats, $\mathbf{n}$, and SNR stats, $\mathbf{r}$.

2.3. Analysis on outdoor vs. indoor scenes

For outdoor scenes, time-of-day information provides strong cues about the likely range of illuminants. However, it is not as informative for indoor scenes, which are typically illuminated by artificial lights. Therefore, for indoor scenes, image colors and additional capture information are necessary for accurate illuminant estimation.

In Table 2, we analyze the angular error of our model when trained using 1) the time feature $\mathbf{p}$ only, 2) the complete time-capture feature without the color histogram, and 3) the complete time-capture feature with the color histogram, across different scene types. The models are trained on all scene types and tested separately on outdoor and indoor scenes, with the indoor and outdoor scenes manually labeled (see Sec. 3.2 for details).

The first row of Table 2 presents results for using only the time feature $\mathbf{p}$, without any color information from the scene provided by the histogram $\mathbf{H}$. The results for outdoor scenes are significantly better than those for indoor scenes, indicating that the time feature $\mathbf{p}$ provides valuable cues for estimating illuminants in outdoor scenes.

The second row shows results for using the complete time-capture feature, without scene color information. The outdoor/indoor gap is smaller, suggesting that other capture data, such as ISO and noise stats, provide additional insights into the capturing environment. The third row shows results for using the complete time-capture feature along with the image histogram (our proposed method). This further reduces the outdoor/indoor gap, as the histogram provides color information for both indoor and outdoor scenes.

2.4. Additional quantitative results

In the main paper, we reported results on our testing set without masking out regions illuminated by light sources different from the dominant one used to obtain the ground truth. This setup mimics realistic scenarios where a single illuminant is not always present. In Table 3, we report results after masking out regions in the testing set that are illuminated by different light sources than the ground truth. Table 4 shows comparisons with other methods on the validation set without masking out regions lit by different illuminations than the dominant light color in the scene.

2.5. Qualitative results

Figure 2 presents qualitative results from our method alongside other illuminant estimation methods, namely TLCC [40], FFCC [8], C4 [47], and C5 [6]. We include randomly selected examples representing the top 25%, second quartile, third quartile, and bottom 5% of our results. For FFCC [8] and C5 [6], we show the best result from each method for every example shown in the figure, as we used multiple models for each method—FFCC includes models with and

Table 1. **Results on the validation set with masking**. We report the mean, median, best 25%, worst 25%, tri-mean, and maximum angular errors for each method on both neutral and user-preference white-balance ground truth, presented in the format (neutral / user-preference). Results for our method with various configurations are included, where $\mathbf{p}$, i, s, f, $\mathbf{n}$, and $\mathbf{r}$ represent the time feature, ISO, shutter speed, flash status, noise stats, and SNR stats, respectively. Symbols $\mathbf{H}$ and $\mathbf{c}$ denote histogram and time-capture feature. $\mathbf{H} \rightarrow \mathbf{I}_{\texttt{chroma}}$ indicates using R/G and B/G images instead of histograms, while $\mathbf{H}_e$ refers to histogram of image edges. log-$\mathbf{H}$ refers to the histogram used in [6, 7] and $\mathbf{c}$-raw refers to using time-capture features without any pre-processing or normalization. Additional configurations include h (number of histogram bins), and uv coord. (additional histogram channels of the u/v coordinates in histogram space). The number of parameters required by each method is reported. The best and second-best results are highlighted.

Method	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max	#params (K)
GW [12]	5.86 / 5.41	4.90 / 3.97	0.78 / 0.95	12.74 / 11.92	19.86 / 19.06	4.98 / 4.42	30.05 / 30.53	-
SoG [20]	4.38 / 3.95	3.28 / 2.71	0.57 / 0.69	9.87 / 9.03	15.63 / 14.53	3.62 / 3.09	31.28 / 31.66	-
GE-1st [41]	4.02 / 3.62	2.82 / 2.43	0.60 / 0.64	9.09 / 8.46	14.99 / 14.10	3.18 / 2.73	32.57 / 32.91	-
GE-2nd [41]	3.71 / 3.29	2.67 / 2.29	0.62 / 0.60	8.29 / 7.56	14.15 / 13.17	2.96 / 2.52	31.66 / 32.02	-
Max-RGB [11]	3.54 / 2.57	2.75 / 1.84	0.78 / 0.88	7.61 / 5.52	11.66 / 9.22	3.02 / 1.98	19.81 / 16.78	-
wGE [24]	3.96 / 3.55	2.62 / 2.19	0.56 / 0.61	9.20 / 8.62	15.69 / 14.66	3.03 / 2.63	33.91 / 34.32	-
PCA [17]	4.33 / 3.90	3.16 / 2.41	0.54 / 0.57	10.02 / 9.49	16.75 / 15.87	3.46 / 2.94	32.40 / 32.84	-
MSGP [37]	6.41 / 5.88	5.48 / 4.01	0.84 / 1.03	14.20 / 13.16	23.99 / 23.26	5.38 / 4.59	34.23 / 35.95	-
GI [36]	4.30 / 4.50	2.78 / 2.74	0.43 / 0.75	11.14 / 11.11	21.42 / 20.61	2.94 / 3.13	32.52 / 32.96	-
TECC [9]	3.78 / 3.30	2.69 / 2.23	0.62 / 0.59	8.49 / 7.66	14.16 / 13.18	2.99 / 2.57	31.41 / 31.75	-
Gamut (pixels) [23]	3.72 / 2.54	2.83 / 1.61	0.73 / 0.61	8.10 / 5.99	13.66 / 10.52	3.01 / 1.83	21.81 / 17.82	0.636
Gamut (edges) [23]	4.43 / 3.92	3.34 / 3.04	1.04 / 1.13	9.52 / 8.13	15.13 / 13.50	3.62 / 3.22	19.26 / 15.91	324
Gamut (1st) [23]	4.33 / 3.85	3.34 / 2.58	0.68 / 0.98	9.87 / 8.77	15.61 / 13.86	3.52 / 2.83	22.15 / 19.03	279
NIS [22]	4.28 / 3.80	3.75 / 2.73	0.72 / 0.85	9.14 / 8.21	14.57 / 13.33	3.80 / 3.09	31.72 / 31.87	0.078
Classification-CC [35]	2.55 / 1.64	2.15 / 1.15	0.63 / 0.36	5.37 / 3.57	8.88 / 5.47	2.17 / 1.32	17.82 / 7.89	58,384
FFCC [8]	2.21 / 1.54	1.33 / 0.90	0.42 / 0.24	5.46 / 3.92	10.38 / 8.03	1.55 / 1.00	17.18 / 14.53	12
FFCC (capture info) [8]	1.97 / 1.43	1.25 / 0.80	0.39 / 0.24	4.72 / 3.65	8.56 / 7.13	1.43 / 0.93	14.16 / 12.30	36.9
FC4 [26]	4.88 / 5.49	2.97 / 3.66	0.90 / 1.32	12.41 / 13.00	31.38 / 31.36	3.15 / 3.85	44.29 / 42.46	1,705
APAP (GW) [4]	3.43 / 1.99	2.66 / 1.53	0.86 / 0.52	7.25 / 4.11	10.91 / 6.02	2.92 / 1.66	15.11 / 7.10	0.289
SIIE [2]	3.67 / -	3.16 / -	0.88 / -	7.44 / -	10.75 / -	3.24 / -	16.91 / -	1,008
SIIE (tuned) [2]	2.90 / -	2.23 / -	0.45 / -	6.44 / -	10.27 / -	2.38 / -	13.97 / -	1,008
SIIE (tuned-CS) [2]	2.65 / 1.61	1.91 / 1.27	0.45 / 0.32	6.13 / 3.58	10.76 / 5.34	2.05 / 1.35	20.53 / 8.12	1,008
KNN (raw) [3]	2.43 / 1.42	1.52 / 0.99	0.34 / 0.25	6.08 / 3.27	11.70 / 5.90	1.63 / 1.06	22.00 / 8.96	757
Quasi-U-CC [10]	3.60 / 3.27	2.71 / 2.10	0.50 / 0.55	8.19 / 7.87	13.18 / 12.68	2.89 / 2.41	22.60 / 23.17	54,421
Quasi-U-CC (tuned) [10]	2.70 / 2.46	1.92 / 1.53	0.50 / 0.51	6.21 / 5.88	9.90 / 9.24	2.12 / 1.76	15.20 / 16.99	54,421
BoCF [30]	3.12 / 1.94	2.55 / 1.37	0.85 / 0.50	6.28 / 4.19	8.91 / 6.90	2.67 / 1.57	11.97 / 10.81	59
C4 [47]	1.63 / 1.46	1.04 / 0.94	0.30 / 0.29	3.87 / 3.49	6.73 / 5.68	1.17 / 1.07	9.89 / 10.59	5,116
CWCC [31]	3.21 / 2.21	2.44 / 1.79	0.83 / 0.73	6.84 / 4.48	10.67 / 7.89	2.68 / 1.83	12.99 / 14.34	101
C5 [6]	2.90 / -	2.34 / -	0.78 / -	5.95 / -	9.89 / -	2.44 / -	19.78 / -	412
C5 (tuned) [6]	1.87 / -	1.14 / -	0.29 / -	4.74 / -	8.69 / -	1.26 / -	12.44 / -	412
C5 (tuned-CS) [6]	1.80 / 1.44	1.24 / 0.92	0.33 / 0.27	4.23 / 3.45	7.44 / 5.69	1.37 / 1.05	13.70 / 8.17	172
TLCC [40]	2.69 / 2.51	2.09 / 1.77	0.63 / 0.57	5.75 / 5.60	9.22 / 9.63	2.21 / 1.98	13.51 / 21.24	32,910
PCC [48]	3.06 / 1.89	1.92 / 1.39	0.46 / 0.40	7.29 / 4.18	11.99 / 6.91	2.28 / 1.48	24.28 / 9.33	0.378
RGP [16]	4.31 / 4.39	2.92 / 2.81	0.39 / 0.69	10.84 / 10.82	20.28 / 18.68	3.06 / 3.26	33.93 / 34.36	-
CFCC [13]	2.74 / 1.57	2.04 / 1.25	0.57 / 0.42	6.17 / 3.34	10.45 / 5.82	2.18 / 1.29	14.53 / 9.35	0.283
Ours (w/o $\mathbf{H}$, $h = 0$)	2.24 / 1.60	1.68 / 1.20	0.47 / 0.38	4.93 / 3.61	8.83 / 6.18	1.77 / 1.25	19.44 / 8.52	2.1
Ours (w/o $\mathbf{c}$)	2.28 / 1.35	1.51 / 0.89	0.35 / 0.29	5.64 / 3.12	12.01 / 5.86	1.61 / 0.93	23.51 / 10.53	4.07
Ours (w/o $\mathbf{H}$, $\mathbf{c} = \mathbf{p}$)	5.28 / 4.20	3.15 / 2.28	0.61 / 0.55	13.39 / 10.83	19.41 / 15.72	3.83 / 2.84	27.27 / 19.60	1.95
Ours (w/o $\mathbf{H}$, $\mathbf{c} = \left[\mathbf{p}^T, i\right]^T$)	4.22 / 3.41	2.15 / 2.23	0.43 / 0.50	10.86 / 8.30	18.10 / 13.93	2.93 / 2.57	27.53 / 18.70	1.97
Ours (w/o $\mathbf{H}$, $\mathbf{c} = \left[\mathbf{p}^T, s\right]^T$)	4.61 / 3.79	2.34 / 2.23	0.59 / 0.52	12.14 / 9.61	18.89 / 14.74	3.06 / 2.61	26.83 / 19.58	1.97
Ours (w/o $\mathbf{H}$, $\mathbf{c} = \left[\mathbf{p}^T, f\right]^T$)	5.24 / 4.14	2.67 / 2.30	0.52 / 0.55	13.62 / 10.69	19.67 / 15.73	3.60 / 2.72	26.57 / 19.67	1.97
Ours (w/o $\mathbf{H}$, $\mathbf{c} = \left[\mathbf{p}^T, \mathbf{n}^T\right]^T$)	2.34 / 1.66	1.66 / 1.14	0.41 / 0.38	5.50 / 3.83	9.74 / 6.89	1.73 / 1.27	19.47 / 9.22	2.05
Ours ($\mathbf{c} = \left[i\right]^T$)	1.99 / 1.26	1.35 / 0.82	0.36 / 0.28	4.78 / 3.01	9.25 / 5.88	1.51 / 0.92	16.89 / 9.82	4.61
Ours ($\mathbf{c} = \left[s\right]^T$)	2.11 / 1.27	1.57 / 0.81	0.38 / 0.25	4.90 / 3.10	9.19 / 5.98	1.61 / 0.89	20.73 / 9.65	4.61
Ours ($\mathbf{c} = \left[f\right]^T$)	2.06 / 1.33	1.24 / 0.80	0.36 / 0.24	5.11 / 3.34	10.34 / 6.04	1.38 / 0.91	20.66 / 11.15	4.61
Ours ($\mathbf{c} = \mathbf{n}$)	1.94 / 1.26	1.20 / 0.88	0.38 / 0.32	4.66 / 2.93	8.85 / 5.77	1.36 / 0.88	23.85 / 9.55	4.69
Ours ($\mathbf{c} = \mathbf{n}$)	1.93 / 1.28	1.27 / 0.83	0.36 / 0.23	4.71 / 3.26	9.36 / 5.96	1.35 / 0.86	20.59 / 9.76	4.79
Ours (w/o $\mathbf{H}_e$)	1.96 / 1.30	1.37 / 0.87	0.35 / 0.23	4.73 / 3.03	10.29 / 4.65	1.41 / 0.96	26.11 / 5.84	4.86
Ours (w/o u/v coord.)	1.69 / 1.34	1.25 / 0.94	0.28 / 0.31	3.88 / 3.02	7.03 / 4.89	1.32 / 1.06	19.58 / 6.84	4.79
Ours (w/ log-$\mathbf{H}$ [6, 7])	1.88 / 1.27	1.37 / 0.89	0.40 / 0.26	4.24 / 2.86	7.28 / 4.92	1.43 / 1.00	23.83 / 6.96	4.93
Ours ($\mathbf{H} \rightarrow \mathbf{I}_{\texttt{chroma}}$)	2.17 / 1.32	1.46 / 0.85	0.41 / 0.25	5.20 / 3.19	10.29 / 5.65	1.58 / 0.93	17.97 / 7.61	4.79
Ours (w/ $\mathbf{c}$-raw)	2.11 / 1.46	1.45 / 1.10	0.46 / 0.38	4.79 / 3.23	9.00 / 5.51	1.60 / 1.17	21.59 / 10.09	4.93
Ours ($h = 24$)	1.74 / 1.14	1.24 / 0.74	0.29 / 0.23	4.12 / 2.74	7.80 / 4.84	1.32 / 0.84	21.94 / 7.24	4.93
Ours ($\mathbf{I} \in \mathbb{R}^{(64\times48)\times3}$)	1.86 / 1.16	1.26 / 0.73	0.34 / 0.24	4.35 / 2.81	8.36 / 4.91	1.41 / 0.83	18.76 / 8.33	4.93
Ours (w/o $\mathbf{p}$)	1.83 / 1.13	1.25 / 0.75	0.31 / 0.20	4.39 / 2.72	8.87 / 4.90	1.34 / 0.83	18.52 / 7.33	4.83
Ours (w/o $\mathbf{n}$, w/o $\mathbf{r}$)	1.79 / 1.24	1.20 / 0.86	0.34 / 0.24	4.23 / 2.95	7.77 / 5.53	1.35 / 0.89	17.82 / 10.33	4.83
Ours (w/o $\mathbf{n}$, w/ $\mathbf{r}$)	1.70 / 1.12	1.15 / 0.85	0.29 / 0.25	4.13 / 2.56	8.84 / 4.65	1.23 / 0.87	21.55 / 7.79	4.93
Ours (w/ $\mathbf{n}$, w/ $\mathbf{r}$)	1.63 / 1.12	1.14 / 0.71	0.33 / 0.25	3.78 / 2.67	7.19 / 4.98	1.25 / 0.83	19.96 / 9.28	5.03
Ours (w/ $\mathbf{n}$, w/o $\mathbf{r}$)	1.63 / 1.09	1.05 / 0.71	0.29 / 0.24	3.92 / 2.62	8.12 / 4.89	1.18 / 0.77	22.22 / 8.45	4.93

Table 2. **Results on outdoor vs. indoor scenes**. We report the mean, median, best 25%, worst 25%, tri-mean, and maximum angular errors for each experiment setting on the testing set (without masking). Models are trained and tested on the neutral ground-truth illuminants. $\mathbf{c}$ denotes the time-capture feature. $\mathbf{p}$ represents the time feature. '$\mathbf{c}$ = all' indicates that the full time-capture feature is used.

Method	Mean		Med.		Best 25%		Worst 25%		Worst 5%		Tri.		Max	
	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor
Ours (w/o $\mathbf{H}$, $\mathbf{c}$ = $\mathbf{p}$)	3.47	9.39	1.96	8.83	0.40	1.98	9.24	17.73	16.19	25.51	2.28	8.97	24.31	36.23
Ours (w/o $\mathbf{H}$, $\mathbf{c}$ = all)	2.12	3.05	1.41	2.41	0.38	1.13	5.05	6.00	8.58	10.00	1.53	2.53	17.55	11.25
Ours (w/ $\mathbf{H}$, $\mathbf{c}$ = all)	1.77	1.97	1.20	1.26	0.33	0.42	4.18	4.83	8.32	10.15	1.31	1.33	18.75	35.42

Table 3. **Results on the testing set with masking**. We report the mean, median, best 25%, worst 25%, tri-mean, and maximum angular errors for each method on both neutral and user-preference white-balance ground truth, presented in the format (neutral / user-preference). Symbols $\mathbf{n}$ and $\mathbf{r}$ represent noise stats and SNR stats, respectively. The best and second-best results are highlighted.

Method	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max
GW [12]	6.38 / 5.68	6.29 / 4.67	1.05 / 1.11	12.55 / 12.05	18.15 / 19.89	5.94 / 4.86	28.09 / 31.89
SoG [20]	4.36 / 3.78	3.44 / 2.19	0.66 / 0.70	9.52 / 9.27	14.53 / 16.36	3.68 / 2.69	23.22 / 32.06
GE-1st [41]	4.02 / 3.58	3.13 / 2.27	0.63 / 0.59	9.11 / 8.86	14.37 / 16.09	3.32 / 2.59	21.38 / 32.21
GE-2nd [41]	3.90 / 3.38	2.87 / 1.97	0.65 / 0.59	8.82 / 8.39	14.02 / 15.65	3.11 / 2.32	22.24 / 31.73
Max-RGB [11]	3.70 / 2.73	2.80 / 1.87	0.90 / 0.91	7.90 / 6.19	12.32 / 12.76	3.05 / 1.95	21.84 / 24.86
wGE [24]	3.76 / 3.33	2.68 / 2.01	0.57 / 0.53	8.66 / 8.45	13.85 / 16.01	2.98 / 2.33	21.39 / 31.94
PCA [17]	4.34 / 3.75	3.52 / 1.97	0.62 / 0.64	9.54 / 9.47	14.51 / 16.65	3.67 / 2.58	22.87 / 32.31
MSGP [37]	6.62 / 5.87	5.91 / 4.56	0.98 / 1.05	13.56 / 12.87	20.62 / 22.30	5.85 / 4.85	37.25 / 36.92
GI [36]	4.76 / 4.85	3.24 / 2.83	0.45 / 0.76	11.65 / 12.07	19.87 / 21.06	3.52 / 3.46	36.34 / 36.02
TECC [9]	3.92 / 3.39	2.95 / 2.11	0.67 / 0.58	8.85 / 8.41	13.68 / 15.43	3.21 / 2.46	21.85 / 31.83
Gamut (pixels) [23]	3.53 / 2.53	2.53 / 1.46	0.69 / 0.59	7.95 / 6.39	12.35 / 12.86	2.88 / 1.65	21.53 / 26.62
Gamut (edges) [23]	4.28 / 4.06	3.30 / 3.03	1.07 / 0.96	9.23 / 9.07	14.47 / 16.86	3.48 / 3.18	28.50 / 30.72
Gamut (1st) [23]	3.87 / 3.81	2.80 / 2.51	0.70 / 0.85	8.88 / 8.83	14.01 / 16.17	3.03 / 2.86	21.38 / 32.62
NIS [22]	4.53 / 4.03	3.69 / 2.74	0.67 / 0.78	9.80 / 9.40	14.83 / 16.12	3.77 / 3.10	20.76 / 32.75
Classification-CC [35]	2.71 / 1.68	2.07 / 1.19	0.60 / 0.35	5.94 / 3.78	9.83 / 6.28	2.21 / 1.32	19.31 / 10.26
FFCC [8]	2.61 / 1.54	1.43 / 0.85	0.37 / 0.26	6.83 / 4.07	16.24 / 8.46	1.66 / 0.98	48.98 / 18.60
FFCC (capture info) [8]	2.19 / 1.37	1.37 / 0.82	0.30 / 0.24	5.49 / 3.53	11.86 / 6.90	1.53 / 0.92	48.53 / 16.40
FC4 [26]	3.92 / 2.67	2.77 / 2.23	0.84 / 0.89	9.17 / 5.18	17.30 / 7.77	2.88 / 2.36	45.83 / 11.83
APAP (GW) [4]	3.77 / 2.13	3.20 / 1.70	0.98 / 0.48	7.63 / 4.49	10.96 / 6.69	3.34 / 1.80	14.66 / 8.91
SIIE [2]	4.16 / -	3.37 / -	0.91 / -	9.06 / -	16.13 / -	3.49 / -	43.76 / -
SIIE (tuned) [2]	3.16 / -	2.25 / -	0.49 / -	7.27 / -	12.25 / -	2.50 / -	34.53 / -
SIIE (tuned-CS) [2]	3.17 / 1.79	2.20 / 1.21	0.49 / 0.32	7.50 / 4.23	13.76 / 6.94	2.41 / 1.34	39.01 / 9.54
KNN (raw) [3]	2.41 / 1.44	1.50 / 0.85	0.34 / 0.21	6.14 / 3.70	12.05 / 7.12	1.65 / 1.00	28.95 / 13.49
Quasi-U-CC [10]	3.84 / 3.38	2.94 / 1.89	0.55 / 0.60	8.55 / 8.56	13.62 / 16.02	3.20 / 2.36	24.74 / 32.79
Quasi-U-CC (tuned) [10]	3.02 / 2.70	2.18 / 1.51	0.48 / 0.48	6.95 / 6.94	11.38 / 14.05	2.37 / 1.74	22.67 / 33.64
BoCF [30]	3.44 / 2.14	2.67 / 1.60	0.89 / 0.49	7.16 / 4.72	11.07 / 7.84	2.89 / 1.71	21.38 / 19.68
C4 [47]	1.73 / 1.45	1.18 / 0.90	0.35 / 0.24	4.09 / 3.67	7.60 / 7.05	1.29 / 1.00	21.55 / 18.09
CWCC [31]	3.46 / 2.31	2.48 / 1.71	0.76 / 0.70	7.62 / 4.98	11.84 / 9.54	2.82 / 1.85	20.73 / 20.81
C5 [6]	3.18 / -	2.49 / -	0.78 / -	6.82 / -	10.55 / -	2.67 / -	16.70 / -
C5 (tuned-CS) [6]	1.81 / 1.27	1.22 / 0.86	0.36 / 0.22	4.31 / 2.98	7.45 / 5.08	1.36 / 0.95	16.78 / 9.23
TLCC [40]	2.64 / 2.87	2.06 / 2.01	0.65 / 0.69	5.69 / 6.58	9.68 / 13.13	2.16 / 2.18	21.44 / 33.20
PCC [48]	2.87 / 1.68	2.03 / 1.16	0.51 / 0.38	6.83 / 3.89	10.62 / 7.05	2.25 / 1.25	14.60 / 10.34
RGP [16]	4.57 / 4.50	3.21 / 2.88	0.43 / 0.67	10.95 / 11.16	18.06 / 20.05	3.55 / 3.33	32.11 / 33.98
CFCC [13]	2.82 / 1.50	2.01 / 1.02	0.70 / 0.38	6.24 / 3.46	10.62 / 6.39	2.19 / 1.10	17.11 / 10.19
Ours (w/o $\mathbf{n}$, w/o $\mathbf{r}$)	1.86 / 1.26	1.31 / 0.78	0.37 / 0.23	4.33 / 3.15	8.71 / 6.00	1.41 / 0.89	22.63 / 16.07
Ours (w/o $\mathbf{n}$, w/ $\mathbf{r}$)	1.84 / 1.25	1.18 / 0.80	0.34 / 0.24	4.55 / 3.04	9.36 / 5.15	1.31 / 0.90	24.99 / 13.02
Ours (w/ $\mathbf{n}$, w/o $\mathbf{r}$)	1.84 / 1.22	1.18 / 0.73	0.35 / 0.24	4.56 / 3.06	9.85 / 5.76	1.25 / 0.82	29.46 / 15.24
Ours (w/ $\mathbf{n}$, w/ $\mathbf{r}$)	1.82 / 1.21	1.19 / 0.78	0.36 / 0.20	4.42 / 2.97	9.42 / 5.18	1.25 / 0.87	35.42 / 12.93

without capture information, and C5 has differently tuned models.

As shown in Fig. 2, the worst 5% example is a scene with limited colors, a typical challenge in illuminant estimation, where color information can mislead any model from achieving accurate estimates. Although our method has a relatively high error, other methods, such as TLCC [40] and FFCC [8], exhibit even higher errors. However, our method results in more perceptually acceptable differences compared to these methods when compared to the ground truth.

To further examine our method on scenes with limited colors, we present additional qualitative examples in Fig. 3. As shown, our method performs reasonably well in these

Table 4. **Results on the validation set without masking**. We report the mean, median, best 25%, worst 25%, tri-mean, and maximum angular errors for each method on neutral and user-preference white-balance ground-truth illuminants, presented in the format (neutral / user-preference). Symbols $\mathbf{n}$ and $\mathbf{r}$ refer to noise and SNR stats, respectively. The best and second-best results are highlighted.

Method	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max
GW [12]	5.76 / 5.30	4.83 / 3.94	0.78 / 0.89	12.45 / 11.68	19.83 / 18.98	4.91 / 4.29	30.03 / 30.51
SoG [20]	4.42 / 3.74	3.27 / 2.35	0.59 / 0.71	10.01 / 8.76	15.91 / 14.41	3.65 / 2.77	30.43 / 30.82
GE-1st [41]	4.15 / 3.40	2.99 / 2.30	0.63 / 0.61	9.42 / 8.13	16.51 / 14.63	3.28 / 2.61	32.02 / 32.37
GE-2nd [41]	3.88 / 3.10	2.74 / 2.01	0.69 / 0.58	8.68 / 7.27	15.31 / 13.45	3.06 / 2.29	28.72 / 29.15
Max-RGB [11]	3.76 / 2.57	2.98 / 1.79	0.96 / 0.92	7.94 / 5.56	13.21 / 10.13	3.17 / 1.91	21.06 / 17.57
wGE [24]	4.11 / 3.28	2.65 / 2.06	0.59 / 0.53	9.66 / 8.30	17.98 / 15.37	3.00 / 2.33	33.88 / 34.29
PCA [17]	4.36 / 3.77	3.04 / 2.15	0.56 / 0.54	10.27 / 9.32	17.99 / 16.41	3.37 / 2.69	32.27 / 32.71
MSGP [37]	6.39 / 5.81	5.48 / 3.87	0.80 / 0.99	14.08 / 13.17	23.97 / 23.36	5.35 / 4.51	34.23 / 35.95
GI [36]	4.21 / 4.53	2.75 / 2.79	0.42 / 0.76	10.82 / 11.08	21.20 / 20.40	2.85 / 3.17	32.52 / 32.96
TECC [9]	3.89 / 3.08	2.80 / 2.00	0.64 / 0.56	8.80 / 7.35	15.47 / 13.64	3.03 / 2.26	28.28 / 28.69
Gamut (pixels) [23]	3.72 / 2.54	2.83 / 1.61	0.73 / 0.61	8.10 / 5.99	13.66 / 10.52	3.01 / 1.83	21.81 / 17.82
Gamut (edges) [23]	4.42 / 3.92	3.37 / 3.09	1.05 / 1.13	9.48 / 8.12	15.14 / 13.49	3.65 / 3.25	19.24 / 15.91
Gamut (1st) [23]	4.33 / 3.85	3.34 / 2.58	0.68 / 0.98	9.87 / 8.77	15.61 / 13.86	3.52 / 2.83	22.16 / 19.03
NIS [22]	4.36 / 3.80	3.44 / 2.73	0.80 / 0.85	9.35 / 8.21	14.88 / 13.33	3.70 / 3.09	31.43 / 31.87
Classification-CC [35]	2.58 / 1.61	2.25 / 1.23	0.58 / 0.35	5.32 / 3.53	9.23 / 5.33	2.24 / 1.29	18.53 / 6.95
FFCC [8]	2.19 / 1.51	1.29 / 0.91	0.42 / 0.25	5.42 / 3.76	10.20 / 7.70	1.53 / 1.01	17.18 / 13.79
FFCC (capture info) [8]	1.97 / 1.35	1.29 / 0.91	0.35 / 0.25	4.70 / 3.22	8.23 / 5.59	1.47 / 1.02	16.01 / 8.74
FC4 [26]	4.02 / 2.87	2.92 / 2.72	0.90 / 0.83	9.26 / 5.24	19.13 / 7.23	2.98 / 2.72	39.64 / 11.83
APAP (GW) [4]	3.38 / 1.93	2.54 / 1.52	0.86 / 0.51	7.15 / 4.02	11.04 / 5.96	2.80 / 1.58	16.12 / 7.16
SIIE [2]	3.67 / -	3.16 / -	0.88 / -	7.44 / -	10.75 / -	3.24 / -	16.91 / -
SIIE (tuned) [2]	2.90 / -	2.23 / -	0.45 / -	6.44 / -	10.27 / -	2.38 / -	13.97 / -
SIIE (tuned-CS) [2]	2.65 / 1.61	1.91 / 1.27	0.45 / 0.32	6.13 / 3.58	10.76 / 5.34	2.05 / 1.35	20.53 / 8.12
KNN (raw) [3]	2.49 / 1.39	1.61 / 0.99	0.35 / 0.26	6.13 / 3.18	11.35 / 5.70	1.72 / 1.05	20.63 / 7.91
Quasi-U-CC [10]	3.66 / 3.19	2.61 / 1.95	0.55 / 0.53	8.58 / 7.79	14.24 / 12.77	2.84 / 2.29	22.69 / 23.25
Quasi-U-CC (tuned) [10]	2.82 / 2.36	1.99 / 1.51	0.55 / 0.47	6.46 / 5.60	9.93 / 8.77	2.21 / 1.70	11.32 / 11.99
BoCF [30]	3.18 / 1.97	2.49 / 1.42	0.82 / 0.50	6.53 / 4.20	9.29 / 6.70	2.68 / 1.61	12.17 / 10.60
C4 [47]	1.72 / 1.42	1.04 / 0.86	0.30 / 0.26	4.22 / 3.49	7.19 / 5.44	1.21 / 1.03	14.36 / 6.49
CWCC [31]	3.42 / 2.27	2.71 / 1.75	0.89 / 0.73	7.28 / 4.69	11.93 / 8.36	2.90 / 1.82	17.99 / 13.95
C5 [6]	2.97 / -	2.27 / -	0.81 / -	6.21 / -	10.13 / -	2.42 / -	18.38 / -
C5 (tuned) [6]	2.00 / -	1.21 / -	0.31 / -	5.09 / -	9.71 / -	1.31 / -	17.56 / -
C5 (tuned-CS) [6]	2.01 / 1.46	1.36 / 0.96	0.36 / 0.26	4.82 / 3.53	8.82 / 5.68	1.49 / 1.06	15.99 / 8.17
TLCC [40]	2.70 / 2.36	2.24 / 1.69	0.69 / 0.56	5.54 / 5.19	8.87 / 8.24	2.30 / 1.91	14.35 / 11.87
PCC [48]	3.32 / 1.90	2.19 / 1.34	0.48 / 0.39	7.89 / 4.33	13.12 / 7.49	2.52 / 1.45	24.28 / 11.30
RGP [16]	4.28 / 4.38	2.82 / 2.76	0.37 / 0.72	10.79 / 10.62	20.02 / 18.62	3.04 / 3.24	33.76 / 34.19
CFCC [13]	2.98 / 1.70	2.13 / 1.18	0.62 / 0.43	6.83 / 3.90	11.92 / 7.85	2.33 / 1.28	19.49 / 14.06
Ours (w/o $\mathbf{n}$, w/o $\mathbf{r}$)	1.85 / 1.27	1.32 / 0.90	0.35 / 0.25	4.26 / 3.01	7.53 / 5.64	1.44 / 0.95	17.44 / 10.14
Ours (w/o $\mathbf{n}$, w/ $\mathbf{r}$)	1.72 / 1.11	1.16 / 0.83	0.30 / 0.24	4.13 / 2.58	8.16 / 4.66	1.23 / 0.86	18.50 / 7.54
Ours (w/ $\mathbf{n}$, w/o $\mathbf{r}$)	1.67 / 1.11	1.07 / 0.70	0.29 / 0.24	4.04 / 2.68	7.90 / 4.97	1.21 / 0.76	24.40 / 8.28
Ours (w/ $\mathbf{n}$, w/ $\mathbf{r}$)	1.66 / 1.14	1.20 / 0.73	0.33 / 0.26	3.77 / 2.69	6.95 / 4.70	1.29 / 0.86	19.80 / 6.98

challenging cases when compared to other methods (e.g., TLCC [40]).

2.6. Cross-camera generalization

In the main paper, we explained that our method is inherently camera-specific by design. However, this limitation can be mitigated through calibration. Here, we present an experimental evaluation of a calibration-based solution to address the camera-specific nature of our approach. Specifically, we calibrate a polynomial mapping for metadata (ISO, shutter speed) and a 3×3 color mapping matrix between our primary camera (Samsung S24 Ultra main camera) and the Samsung S25 Ultra telephoto camera. We evaluate two strategies on 257 test scenes captured by the S25 Ultra telephoto camera:

1. Mapping the S24 Ultra main camera's training data to the S25 Ultra telephoto camera's space *offline*, followed by training on the mapped data.
2. Mapping S25 Ultra telephoto images and metadata *online* to the S24 Ultra main camera's space, applying the model trained on the S24 main camera, and then mapping the predicted illuminant back to the S25 Ultra telephoto camera's space.

Results are shown in Table 5, alongside C4 [47] and C5 [6], both of which are cross-camera methods. None of the methods, including ours, had access to training examples from the target camera (S25 Ultra telephoto).

Table 5. **Results on cross-camera generalization.** We report the mean angular error for each method on 257 test scenes captured using the S25 Ultra telephoto camera. None of the listed methods were trained on any data from the test camera. For our method, we present results for the model trained on data from the S24 Ultra main camera under three settings: without calibration, with *offline* calibration, and with *online* calibration. The best result is highlighted.

Method	C4 [47]	C5 [6]	Ours		
			w/o calibration	w/ offline calibration	w/ online calibration
#params (K)	5,116	172	4.8	4.8	4.8
Mean AE (S25-T)	1.61	1.63	2.06	1.70	1.53

3. Additional details of dataset

In the main paper, we presented our dataset of 3,224 images captured by the Samsung S24 Ultra's main camera. Example scenes from our dataset are shown in Fig. 4. As shown, our dataset includes diverse scenes captured under various weather and lighting conditions.

A distinctive feature of our dataset is the inclusion of a "user-preference" white-balance ground truth that focuses on matching real-world scene observations and enhancing image aesthetics. Figure 5-A shows the chromaticity distribution of both the neutral ground truth (obtained from the color chart) and the user-preference ground truth in our dataset. As shown, the neutral ground truth spans a larger area in the rg chromaticity space, which is intuitive, as it represents the true color of the illuminant lighting the scene. In contrast, the user-preference ground truth has a narrower distribution near the Planckian locus. This explains the lower angular errors observed in most methods when compared to the neutral illuminant estimation results.

3.1. Statistics

Figure 6 shows the statistics of lighting classes (i.e., artificial lights such as incandescent and fluorescent, and natural lights such as outdoor daylight) and scene classes (daylight, sunset/sunrise, night, and indoor) in our dataset. The training, validation, and testing splits are evenly distributed across the different lighting and scene classes.

3.2. Data labeling

To facilitate the annotation process, we developed a Matlab graphical user interface (GUI) tool; see Fig. 7-A. An expert photographer was instructed to select a reference white point of the scene from the raw image of the color chart for each scene, copy it, and paste it to assign as the ground-truth neutral illuminant color for the sequential scene(s) sharing the same lighting condition. In addition, the annotator was asked to assign a "user-preference" ground truth, which may not align with the neutral white-balance appearance or the in-camera white-balance result of the Samsung S24 Ultra; see Fig. 5-B. Notably, the user-preference ground truth is intended to reflect real-world observations and enhance the scene's aesthetics, and therefore, may differ from both the neutral and camera-based ground truths. The mean angular error between the annotated user-preference ground truth and the neutral white-balance ground truth is 2.67°, and the error between the user-preference and the illuminant colors from the in-camera AWB module is 1.34°.

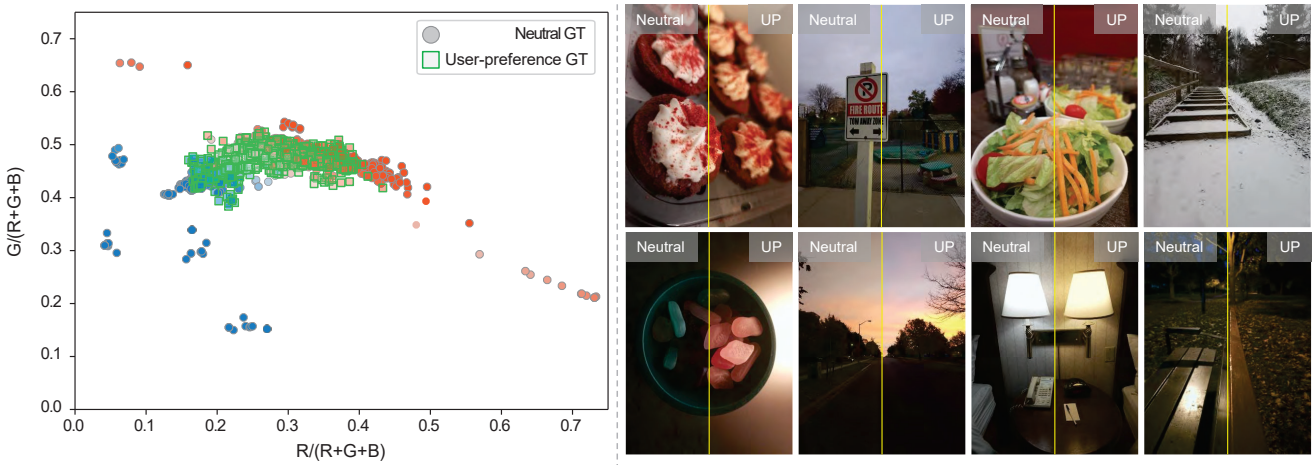
The user-preference tools allow the annotator to interpolate between the camera white balance setting (produced by the in-camera illuminant estimation method) and the annotated neutral white balance. Additionally, the annotator can adjust the user-preference white point to make the scene appear cooler or warmer by modifying the correlated color temperature (CCT).

To map between illuminant RGB colors in the camera raw space and CCTs, we captured a color chart under various CCTs ranging from 1,325K to 10,000K using a controllable light booth, see Fig. 7-B. We then measured the raw RGB color corresponding to each CCT by manually selecting gray patches from the color chart and averaging them for each raw image. We then fit a linear regression model to map the R/G and B/G chromaticity values of raw white points to the corresponding CCT. To convert the CCT value back to the normalized RGB illuminant color, we locate the nearest CCT value within the calibrated CCTs. Subsequently, we linearly interpolate between the corresponding measured chroma values of the nearest lower and higher CCTs.

While this is a simplified method for converting between chromaticity values and CCTs, it was sufficient for our goal to enable the annotator to adjust the white balance in an interpretable manner. These adjustments could be achieved either by modifying the CCT or interpolating between the camera's white balance and the neutral white balance settings, rather than directly adjusting the RGB values of the illuminant, which can be more challenging to fine-tune for the desired results.

Our dataset includes a diverse range of scenes captured under various lighting conditions, including night scenes, making it challenging to ensure the presence of a single light source in each scene. To address this, we complemented white-balance labeling with binary masks for scenes containing multiple light sources. These masks identify regions illuminated by light sources different from the dominant light used to label the ground truth. See Fig. 8 for example masks.

Additionally, to ensure privacy, we applied blurring to personal information (e.g., faces, license plates, phone num-



(A) *rg* chromaticity distribution of neutral and user-preference ground-truth labels in our dataset.

(B) White-balanced images using neutral, and user-preference (UP) ground-truth illuminants.

Figure 5. Our dataset includes the ground-truth illuminant color for each scene under neutral white balance (obtained from gray patches of a color chart) and user-preference white balance, where an expert photographer adjusts the white-balance illuminant color of each image to match real scene observations and enhance image aesthetics. In (A), we show the rg chromaticity distribution of both neutral and user-preference ground-truth illuminants, and in (B), example linear images from our dataset corrected using these ground-truth illuminants. Color correction matrix (CCM) and gamma correction are applied to enhance visualization.

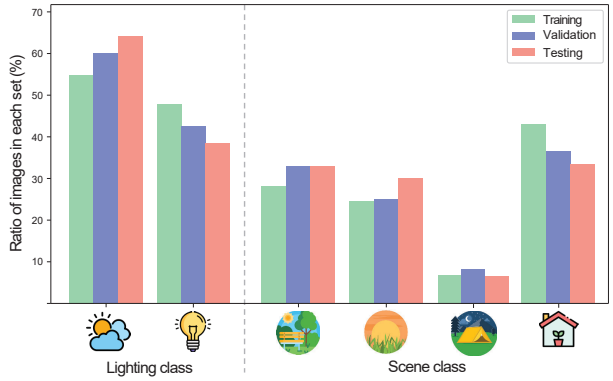


Figure 6. Statistics of our dataset categorized by lighting classes ('natural' and 'artificial' light sources) and semantic scene classes ('outdoor [daylight]', 'outdoor [sunset/sunrise]', 'outdoor [night]', and 'indoor').

bers, etc.) across both the raw and camera sRGB images in our dataset.

Each scene is further annotated with its scene class (daylight, sunset/sunrise, night, and indoor) and lighting condition class (artificial or natural light). Although these labels were primarily used for dataset statistics, we believe they hold significant potential for future research. For example, the scene class could be an additional input feature to improve model accuracy.

The GUI tool also facilitates the assignment of images to one of three sets: training, testing, or validation. The primary criterion was to ensure that testing and validation sets were distinct, containing no overlapping scenes with the training set. We further evaluated the testing and validation sets by applying the gray-world algorithm [12] to selected images, generating real-time statistics that provided insights into their complexity. Since the gray-world algorithm is a simple baseline, its angular error served as a useful indicator of the difficulty of these sets. Finally, we visually reviewed the testing and validation sets to ensure they comprised unique and diverse scenes.

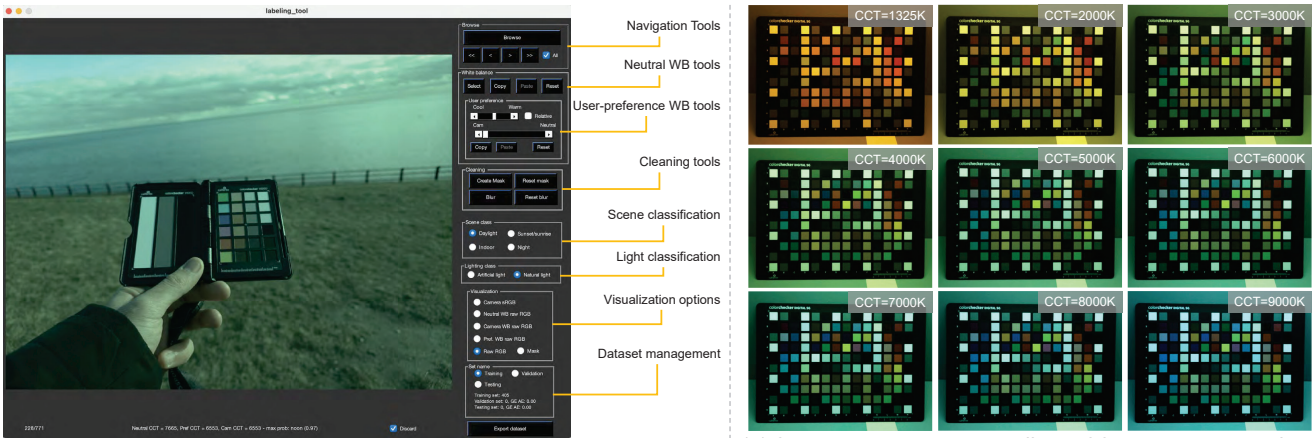
3.3. User study

To validate the annotation of user-preference ground truth, we conducted a user study with 20 participants who had normal vision. We first sorted the images by the angular error between the user-preference ground truth and the neutral ground truth.

From the images with the highest angular errors between the user-preference and neutral ground truths, we randomly selected 100 images.

For each participant, we performed 100 trials, showing these 100 images one by one. In each trial, we present the participant with two versions of white-balanced images corresponding to each ground truth, after applying the color correction matrices and gamma correction for better visualization on a calibrated monitor.

Participants were asked to select the image that appeared most natural. To provide context, we also showed the time of day at which the image was captured (e.g., sunset, sun-



(A) Graphical user interface tool developed to facilitate data annotation

(B) Captured color-chart under different CCTs, used in the tool for user preference annotation.

Figure 7. (A) The graphical user interface (GUI) tool developed to facilitate the annotation process. The tool provides features such as navigation tools, neutral white balance (WB) tools, user preference WB tools, cleaning tools, scene and light classification options, visualization options, and dataset management functionalities. In the neutral WB tools, the annotator can select a reference white point from a raw color chart image, copy it, and paste it into the sequential scene(s) sharing the same lighting condition. In the user preference WB tools, the annotator can interpolate between the camera WB and the neutral WB. Additionally, the annotator can adjust the corresponding correlated color temperature (CCT) of the selected WB setting to create a cooler or warmer appearance. (B) Color charts captured under different CCTs, used within the GUI tool, to calculate the corresponding CCT of illuminant colors in the camera raw space. The images shown in (B) are in raw space with a gamma correction applied to enhance visualization.

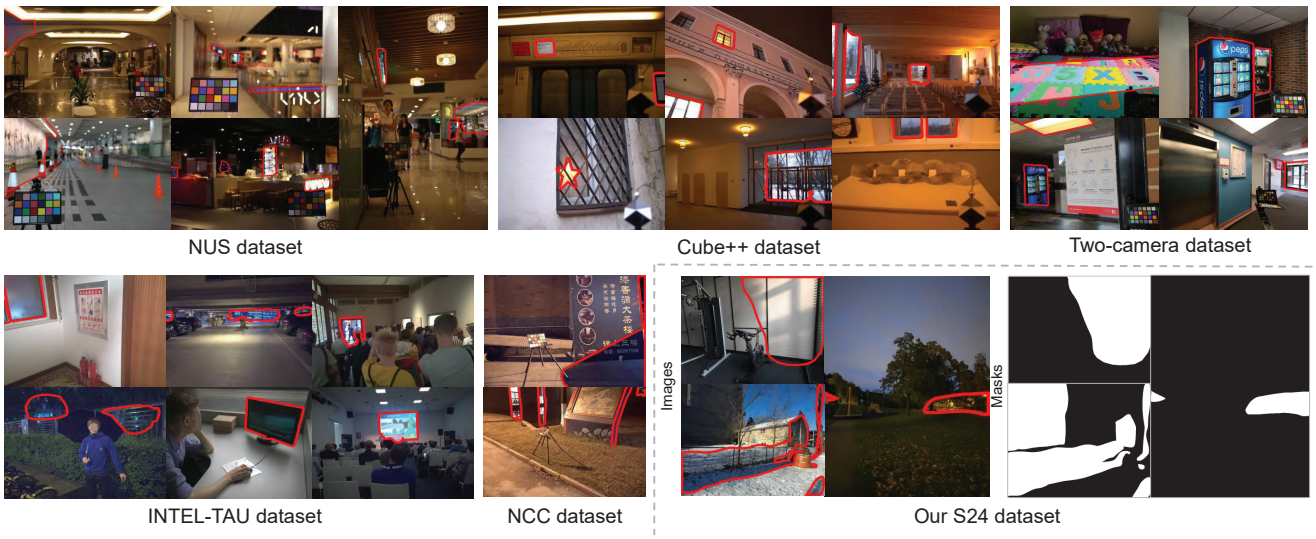


Figure 8. Unlike other color constancy datasets (e.g., NUS [17], Cube++ [19], Two-camera [1], INTEL-TAU [32], NCC [15]), our dataset provides masks for scenes lit by illuminants different from the dominant illuminant used as the ground-truth illuminant color. In this figure, we show examples from each dataset, including ours, where each scene contains regions (highlighted with red borders) that are either lit by or have source lighting different from the dominant illuminant color of the scene. For our dataset, we also show the corresponding manually created masks for the shown images. All shown images are in the sRGB color space.

rise, daylight, etc.). Overall, participants selected the user-preference ground truth in 71.95% of the trials, and selected the neutral ground truth in 28.05% of the trials, confirming that the user-preference ground truth is preferred by users most of the time.

4. Additional ground truth and applications

Denoising: We used Adobe Lightroom AI denoiser to generate denoised raw images that simulate in-camera denoising. These denoised images were used to compute noise

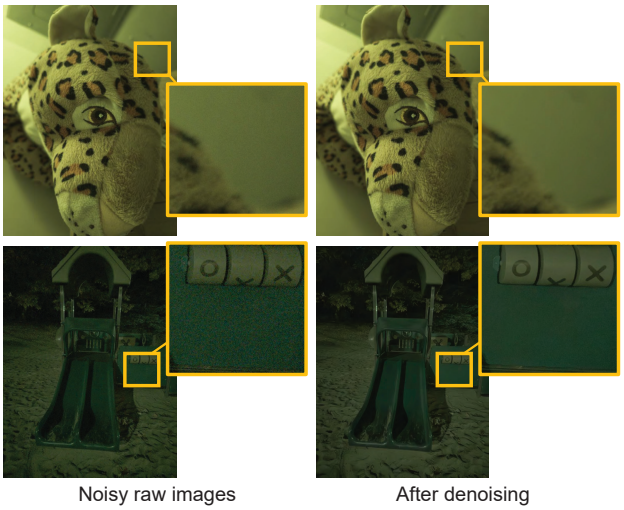


Figure 9. Our dataset includes denoised images that can serve as proxy ground truth for learnable denoisers. To illustrate the effect of denoising, we present raw images before and after denoising, with gamma correction applied for better visualization.

stats, $\mathbf{n}$, which served as one of the input features for our model. The Adobe Lightroom AI denoiser may leverage camera-specific information to achieve effective denoising in the linear raw space, requiring minimal manual adjustment to produce satisfactory results.

These denoised raw images, included as additional "ground-truth" data in our dataset, can serve as a proxy for evaluating denoising algorithms (e.g., [33, 44]) in the linear raw space across diverse lighting conditions, including dark scenes; see Fig. 9.

Expert sRGB rendering: Our dataset was captured in Samsung Pro mode using the Samsung S24 Ultra, providing pre-processed raw images after early-stage operations such as demosaicing. The device also produces an sRGB image by rendering the raw image through a simplified version of the camera's native ISP, which lacks accurate local tone mapping and denoising.

To improve sRGB rendering, we manually processed the raw images in Adobe Lightroom, including local tone mapping adjustments. Specifically, an expert photographer rendered all 3,224 denoised raw images to sRGB, enhancing their aesthetic appeal through global and local tone mapping adjustments using manually created spatial masks in Adobe Lightroom. Although our rendering approach may seem similar to that of the MIT-Adobe FiveK dataset [14], which also involves expert manual rendering using Adobe Lightroom, our dataset, captured with the more recent Samsung S24 Ultra, provides a more up-to-date representation than the older DSLR cameras used in Adobe FiveK. Moreover, Adobe FiveK lacks local tonal adjustments in expert rendering. Our rendered sRGB images leverage the latest denoising techniques in Adobe Lightroom and its advanced functionality to achieve high-quality tone mapping, including local tone adjustments (see Fig. 10).

These high-quality rendered sRGB images make our dataset a valuable resource for the raw-to-sRGB rendering task. In contrast to existing raw-to-sRGB datasets (e.g., the Zurich raw-to-RGB dataset [27] and the Samsung S7 dataset [38]), which suffer from input–ground truth misalignment [27] or contain limited numbers of images (e.g., fewer than 250 full-resolution images [27, 38]) with restricted scene diversity and lighting conditions (e.g., primarily daylight [27]), our dataset offers well-aligned, high-resolution (4000×3000) raw, denoised raw, and sRGB ground-truth images across diverse scenes and lighting conditions. This makes it a reliable resource for training neural ISP methods aimed at rendering raw images into high-quality sRGB images (e.g., [25, 28, 29]).

Table 6. Quantitative results for rendering raw images to expert-rendered sRGB on our test set. Each method was trained on our training set to map raw images to expert-rendered sRGB.

Method	PSNR	SSIM	LPIPS	ΔE_{2000}	#params (K)
CIE XYZ Net [5]	23.32	0.8596	0.1242	7.0239	1,348.8
Invertible ISP [46]	22.87	0.8197	0.1468	7.3739	1,413.8
Param ISP [29]	24.32	0.8411	0.1145	6.1353	1,420.0
Lite ISP [50]	25.49	0.8967	0.0744	5.5213	9,094.0
Fourier ISP [25]	24.50	0.9125	0.0962	5.9276	7,589.8

sRGB picture styles: In addition to the sRGB images from the Samsung S24 Ultra (Pro mode) and our expert-rendered sRGB images, we provide five additional sRGB versions for each raw image using Adobe Lightroom presets, similar to [18]. These can serve as ground truth for picture style transfer or raw-to-multiple-style sRGB rendering. See Fig. 11.

Results of raw-to-sRGB rendering: We evaluate different neural ISP methods that aim to render raw images into corresponding sRGB images using our dataset, which includes expert-rendered sRGB ground truth and five additional picture styles. Specifically, we trained the methods in [5, 25, 29, 46, 50] on our training set to map noisy raw images to expert-rendered sRGB. Additionally, we trained each method to map raw images to each of our five picture styles. Table 6 shows PSNR, SSIM [45], LPIPS [49], and ΔE_{2000} [39] results on our test set for the evaluated neural ISP methods. We also report results for the five different styles, where each model was trained to map raw images to a specific target style in the sRGB space, as shown in Table 7. Figures 12 and 13 provide qualitative examples comparing these methods' outputs with the ground-truth

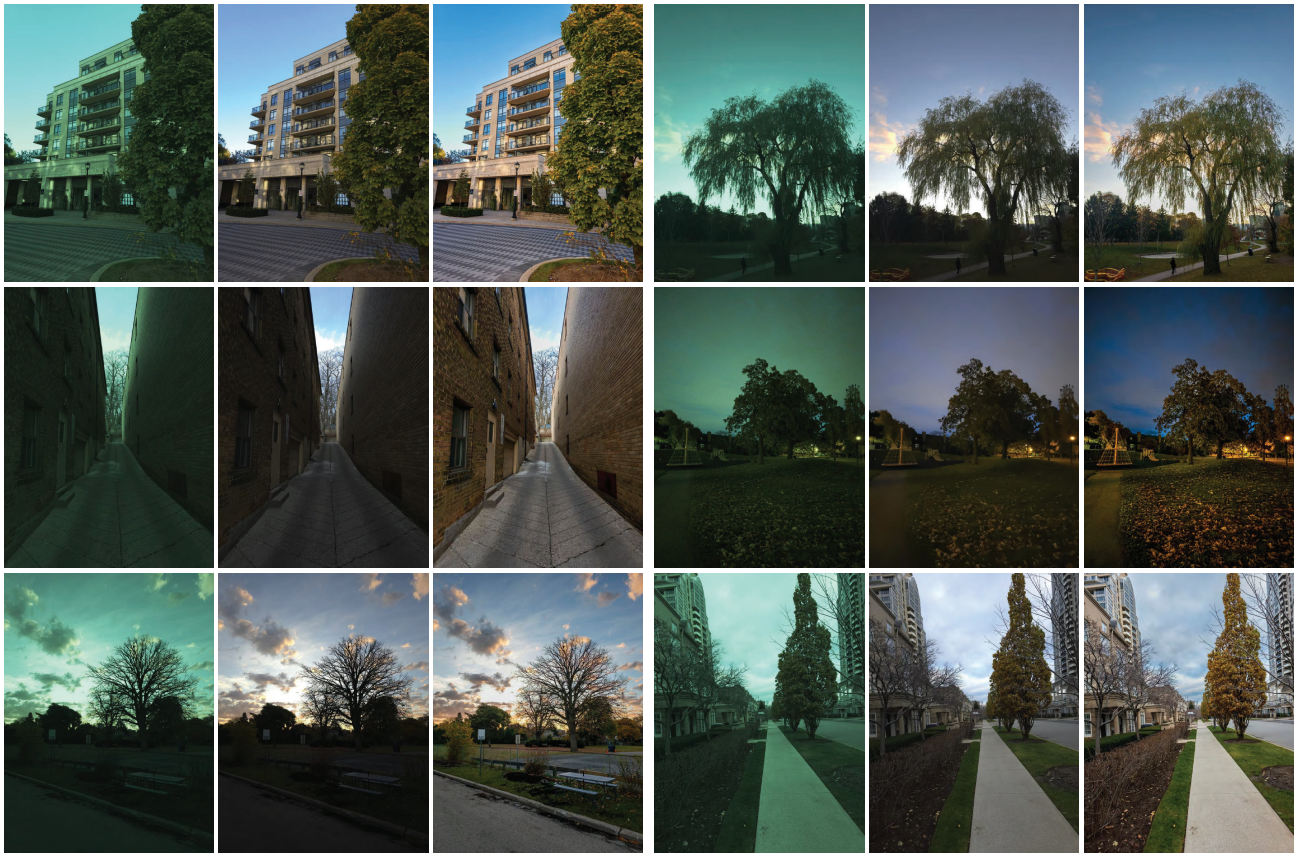


Figure 10. Our dataset includes sRGB images produced by the in-camera lightweight ISP and expert-rendered sRGB images from Adobe Lightroom, which incorporate local tone mapping adjustments to enhance aesthetic appeal.

Table 7. Quantitative results for rendering raw images to five different styles in sRGB. Each method was trained on our training set to map raw images to sRGB images rendered in a specific style.

Method	Style 1				Style 2				Style 3				Style 4				Style 5			
	PSNR	SSIM	LPIPS	ΔE_{2000}	PSNR	SSIM	LPIPS	ΔE_{2000}	PSNR	SSIM	LPIPS	ΔE_{2000}	PSNR	SSIM	LPIPS	ΔE_{2000}	PSNR	SSIM	LPIPS	ΔE_{2000}
CIE XYZ Net [5]	22.40	0.8586	0.1762	8.2912	24.05	0.8720	0.1199	6.6336	22.00	0.8536	0.1615	8.9058	22.26	0.8457	0.1468	7.6241	24.67	0.8967	0.1326	5.2699
Invertible ISP [46]	23.48	0.8322	0.1571	6.6418	26.35	0.8606	0.115	5.2289	23.84	0.8518	0.1437	7.3474	23.33	0.8422	0.1445	7.6935	24.9	0.8748	0.1626	5.9271
Param ISP [29]	24.97	0.8559	0.123	5.7238	27.81	0.8750	0.0952	4.9218	27.11	0.869	0.1005	4.9472	24.18	0.8533	0.1184	6.1376	25.43	0.8667	0.1344	4.4993
Lite ISP [50]	26.66	0.9145	0.0668	4.7019	28.33	0.9224	0.0636	4.2839	26.31	0.9126	0.0729	5.2200	25.04	0.8942	0.082	5.5165	28.07	0.9353	0.0707	3.4339
Fourier ISP [25]	25.19	0.925	0.0985	5.432	28.03	0.9276	0.0819	4.4879	25.38	0.9186	0.0997	5.7031	24.74	0.9063	0.0996	5.5908	27.41	0.9468	0.0889	3.5606

images.

References

[1] Abdelrahman Abdelhamed, Abhijith Punnappurath, and Michael S Brown. Leveraging the availability of two cameras for illuminant estimation. In *CVPR*, 2021. 10

[2] Mahmoud Afifi and Michael S Brown. Sensor-independent illumination estimation for DNN models. In *BMVC*, 2019. 1, 5, 6, 7

[3] Mahmoud Afifi, Brian Price, Scott Cohen, and Michael S Brown. When color constancy goes wrong: Correcting improperly white-balanced images. In *CVPR*, 2019. 1, 5, 6, 7

[4] Mahmoud Afifi, Abhijith Punnappurath, Graham Finlayson, and Michael S Brown. As-projective-as-possible bias correction for illumination estimation algorithms. *JOSA A*, 36 (1):71–78, 2019. 1, 5, 6, 7

[5] Mahmoud Afifi, Abdelrahman Abdelhamed, Abdullah Abuolaim, Abhijith Punnappurath, and Michael S Brown. CIE XYZ Net: Unprocessing images for low-level computer vision tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4688–4700, 2021. 11, 12, 14, 15

[6] Mahmoud Afifi, Jonathan T Barron, Chloe LeGendre, Yun-Ta Tsai, and Francois Bleibel. Cross-camera convolutional color constancy. In *ICCV*, 2021. 1, 2, 3, 4, 5, 6, 7, 8

[7] Jonathan T Barron. Convolutional color constancy. In *ICCV*, 2015. 5

Figure 11. In addition to the expert rendering (style #0), our dataset includes sRGB images with multiple styles, which can serve as ground truth for picture style transfer or raw-to-multi-style sRGB rendering.

[8] Jonathan T Barron and Yun-Ta Tsai. Fast Fourier color constancy. In *CVPR*, 2017. 1, 2, 3, 4, 5, 6, 7

[9] Simone Bianco and Marco Buzzelli. Truncated edge-based color constancy. In *ICCE*, 2022. 2, 5, 6, 7

[10] Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy. In *CVPR*, 2019. 1, 5, 6, 7

[11] David H Brainard and Brian A Wandell. Analysis of the Retinex theory of color vision. *Journal of the Optical Society of America A*, 3(10):1651–1661, 1986. 5, 6, 7

[12] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1): 1–26, 1980. 1, 5, 6, 7, 9

[13] Marco Buzzelli and Simone Bianco. A convolutional framework for color constancy. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2024. 5, 6, 7

[14] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *CVPR*, 2011. 11

[15] Cheng Cheng, Kai-Fu Yang, Xue-Mei Wan, Leanne Lai Hang Chan, and Yong-Jie Li. Nighttime color constancy using robust gray pixels. *JOSA A*, 41(3):476–488, 2024. 10

[16] Cheng Cheng, Kai Fu Yang, Xue Mei Wan, Leanne Lai Hang Chan, and Yonge Jie Li. Nighttime color constancy using robust gray pixels. *JOSA A*, 41(3):476–488, 2024. 5, 6, 7

[17] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: Why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 1, 5, 6, 7, 10

[18] Omar Elezabi, Marcos V Conde, Zongwei Wu, and Radu Timofte. INRetouch: Context aware implicit neural representation for photography retouching. *arXiv preprint arXiv:2412.03848*, 2024. 11

[19] Egor Ershov, Alexey Savchik, Illya Semenkov, Nikola Banić, Alexander Belokopytov, Daria Senshina, Karlo Koščević, Marko Subašić, and Sven Lončarić. The cube++ illumination estimation dataset. *IEEE Access*, 8:227511–227527, 2020. 1, 10

[20] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *CIC*, 2004. 5, 6, 7

[21] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *CVPR*, 2008. 1

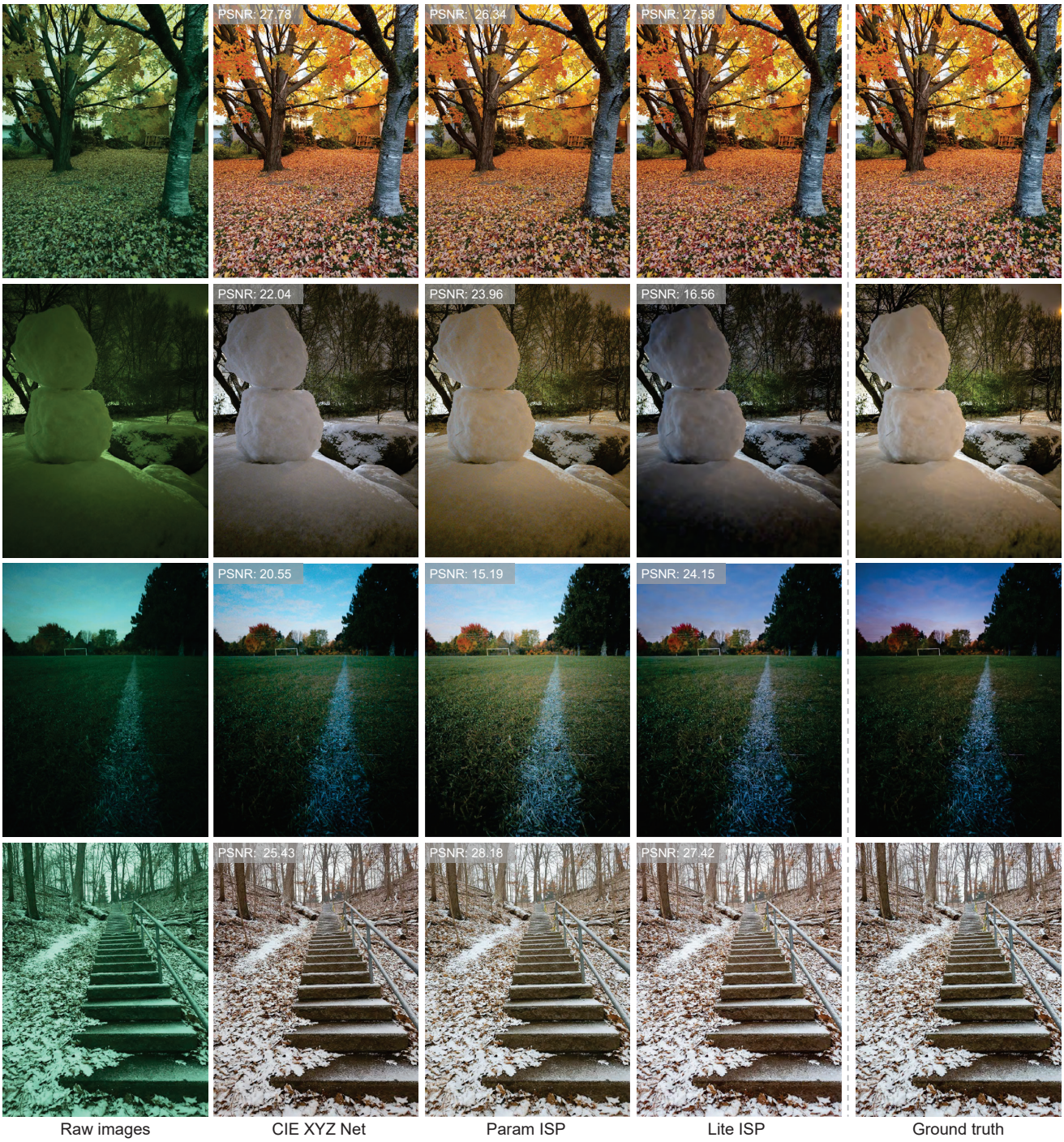


Figure 12. Qualitative results from our testing set for trained models rendering raw images to our expert-rendered sRGB. Results are shown for CIE XYZ Net [5], Param ISP [29], and Lite ISP [50]. Raw images have undergone gamma correction for better visualization.

[22] Arjan Gijsenij and Theo Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4): 687–698, 2011. 5, 6, 7

[23] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2):127–139, 2010. 2, 5, 6, 7

[24] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Improving color constancy by photometric edge weighting.

Figure 13. Qualitative results from our testing set for trained models rendering raw images to the sRGB ground truth of our five additional picture styles. The top row shows the raw image (gamma corrected for visualization) and the corresponding ground-truth image with different picture styles. The second row presents results from CIE XYZ Net [5], Param ISP [29], and Lite ISP [50].

IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(5):918–929, 2011. 5, 6, 7

[25] Xuanhua He, Tao Hu, Guoli Wang, Zejin Wang, Run Wang, Qian Zhang, Keyu Yan, Ziyi Chen, Rui Li, Chengjun Xie, et al. Enhancing RAW-to-sRGB with decoupled style structure in Fourier domain. In *AAAI*, 2024. 11, 12

[26] Yuanming Hu, Baoyuan Wang, and Stephen Lin. FC4: Fully convolutional color constancy with confidence-weighted pooling. In *CVPR*, 2017. 5, 6, 7

[27] Andrey Ignatov, Radu Timofte, Sung-Jea Ko, Seung-Wook Kim, Kwang-Hyun Uhm, Seo-Won Ji, Sung-Jin Cho, Jun-Pyo Hong, Kangfu Mei, Juncheng Li, et al. AIM 2019 challenge on raw to RGB mapping: Methods and results. In *ICCVW*, 2019. 11

[28] Andrey Ignatov, Anastasia Sycheva, Radu Timofte, Yu Tseng, Yu-Syuan Xu, Po-Hsiang Yu, Cheng-Ming Chiang, Hsien-Kai Kuo, Min-Hung Chen, Chia-Ming Cheng, et al. MicroISP: processing 32MP photos on mobile devices with deep learning. In *ECCV*, 2022. 11

[29] Woohyeok Kim, Geonu Kim, Junyong Lee, Seungyong Lee, Seung-Hwan Baek, and Sunghyun Cho. ParamISP: Learned forward and inverse ISPs using camera parameters. In *CVPR*,

2024. 11, 12, 14, 15

[30] Firas Laakom, Nikolaos Passalis, Jenni Raitoharju, Jarno Nikkanen, Anastasios Tefas, Alexandros Iosifidis, and Moncef Gabbouj. Bag of color features for color constancy. *IEEE Transactions on Image Processing*, 29:7722–7734, 2020. 5, 6, 7

[31] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj. Robust channel-wise illumination estimation. In *BMVC*, 2021. 5, 6, 7

[32] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj. Intel-TAU: A color constancy dataset. *IEEE Access*, 9:39560–39567, 2021. 10

[33] Yang Liu, Zhenyue Qin, Saeed Anwar, Pan Ji, Dongwoo Kim, Sabrina Caldwell, and Tom Gedeon. Invertible denoising network: A light solution for real noise removal. In *CVPR*, 2021. 11

[34] Jean H Meeus. *Astronomical algorithms*. Willmann-Bell, Incorporated, 1991. 4

[35] Seoung Wug Oh and Seon Joo Kim. Approaching the computational color constancy as a classification problem through deep learning. *Pattern Recognition*, 61:405–416, 2017. 1, 5, 6, 7

[36] Yanlin Qian, Joni-Kristian Kamarainen, Jarno Nikkanen, and Jiri Matas. On finding gray pixels. In *CVPR*, 2019. 5, 6, 7

[37] Yanlin Qian, Said Pertuz, Jarno Nikkanen, Joni-Kristian Kämäräinen, and Jiri Matas. Revisiting gray pixel for statistical illumination estimation. In *VISSAP*. 2019. 5, 6, 7

[38] Eli Schwartz, Raja Giryes, and Alex M Bronstein. DeepISP: Toward learning an end-to-end image processing pipeline. *IEEE Transactions on Image Processing*, 28(2):912–923, 2018. 11

[39] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005. 11

[40] Yuxiang Tang, Xuejing Kang, Chunxiao Li, Zhaowen Lin, and Anlong Ming. Transfer learning for color constancy via statistic perspective. In *AAAI*, 2022. 1, 2, 3, 4, 5, 6, 7

[41] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007. 2, 5, 6, 7

[42] TC Van Flandern and KF Pulkkinen. Low-precision formulae for planetary positions. *Astrophysical Journal Supplement Series*, 41:391–411, 1979. 4

[43] Robert Walraven. Calculating the position of the sun. *Solar energy*, 20(5):393–397, 1978. 4

[44] Yuzhi Wang, Haibin Huang, Qin Xu, Jiaming Liu, Yiqun Liu, and Jue Wang. Practical deep raw image denoising on mobile devices. In *ECCV*, 2020. 11

[45] 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. 11

[46] Yazhou Xing, Zian Qian, and Qifeng Chen. Invertible image signal processing. In *CVPR*, 2021. 11, 12

[47] Huanglin Yu, Ke Chen, Kaiqi Wang, Yanlin Qian, Zhaoxiang Zhang, and Kui Jia. Cascading convolutional color constancy. In *AAAI*, 2020. 2, 3, 4, 5, 6, 7, 8

[48] Shuwei Yue and Minchen Wei. Color constancy from a pure color view. *Journal of the Optical Society of America A*, 40 (3):602–610, 2023. 5, 6, 7

[49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 11

[50] Zhilu Zhang, Haolin Wang, Ming Liu, Ruohao Wang, Wangmeng Zuo, and Jiawei Zhang. Learning RAW-to-sRGB mappings with inaccurately aligned supervision. In *ICCV*, 2021. 11, 12, 14, 15

[51] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. 1