Learning Camera-Agnostic White-Balance Preferences

Luxi Zhao Mahmoud Afifi Michael S. Brown
AI Center-Toronto, Samsung Electronics
{lucy.zhao, m.afifi1, michael.b1}@samsung.com

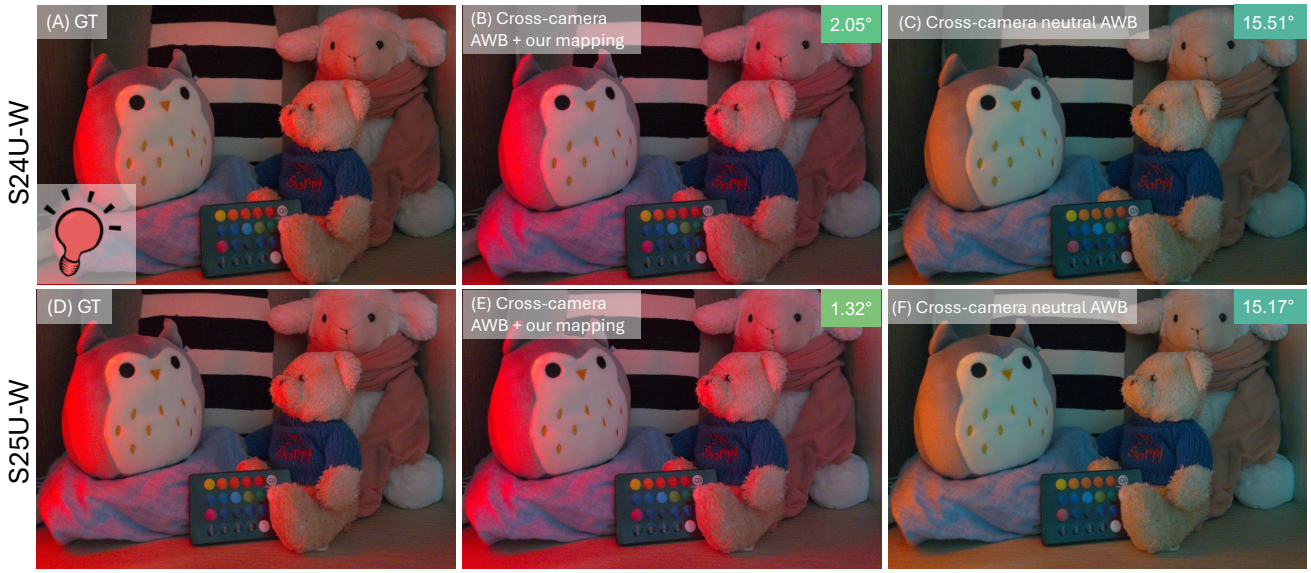

Figure 1. Cross-camera AWB methods typically aim for neutral color correction (subfigures C and F). In practice, however, camera manufacturers often design AWB algorithms to reflect white-balance preferences that achieve a desired aesthetic (subfigures A and D). We introduce a camera-agnostic learnable mapping function that transforms the neutral illuminant predicted by cross-camera AWB algorithms into a target white-balance preference, enabling consistent rendering across different cameras (B and E). Shown are: (A) a scene captured by the Galaxy S24 Ultra wide (S24U-W) camera, rendered using the "ground-truth" white-balance preference; (B) the same scene corrected using our mapping trained on S24U-W data; (C) the same image corrected using C5 [11]'s neutral white balance; (D) the same scene captured by the Galaxy S25 Ultra wide (S25U-W) camera, rendered using the "ground-truth" preference; (E) the same S25U-W image corrected using our mapping trained on S24U-W; and (F) the same S25U-W image corrected using C5's neutral white balance.

Abstract

The image signal processor (ISP) pipeline in modern cameras consists of several modules that transform raw sensor data into visually pleasing images in a display color space. Among these, the auto white balance (AWB) module is essential for compensating for scene illumination. However, commercial AWB systems often strive to compute aesthetic white-balance preferences rather than accurate neutral color correction. While learning-based methods have improved AWB accuracy, they typically struggle to generalize across different camera sensors—an issue for smartphones with multiple cameras. Recent work has explored cross-camera AWB, but most methods remain focused on achieving neutral white balance. In contrast, this paper is the first to address aesthetic consistency by learning a post-illuminant-estimation mapping that transforms neutral illuminant corrections into aesthetically preferred corrections in a camera-agnostic space. Once trained, our mapping can be applied after any neutral AWB module to enable consistent and stylized color rendering across unseen cameras. Our proposed model is lightweight—containing only ∼500 parameters—and runs in just 0.024 milliseconds on a typical flagship mobile CPU. Evaluated on a dataset of 771 smartphone images from three different cameras, our method achieves state-of-the-art performance while remaining fully compatible with existing cross-camera AWB techniques, introducing minimal computational and memory overhead.

1. Introduction

Auto white balance (AWB) is a fundamental module in the camera image signal processor (ISP) pipeline [18, 23], which converts raw sensor data into aesthetically pleasing images in a display color space. Ideally, the AWB module approximates color constancy, aiming to render object colors consistently regardless of the scene's illumination [14]. Specifically, AWB attempts to map the scene's illuminant color to the achromatic line in the camera's raw space—that is, to align raw RGB values corresponding to illumination with the $R = G = B$ "white" line [7].

However, in practice, commercial camera ISPs often incorporate aesthetic considerations that go beyond purely neutral white balancing. These preferences are influenced by photographic trends and do not always adhere to the white-light assumption [5, 8, 21, 24, 35, 46]. In fact, neutral white balancing—which attempts to fully neutralize the lighting color in the scene—can sometimes produce images that appear unnatural. This is partly due to the human visual system's incomplete chromatic adaptation [42, 49], which limits our ability to fully discount strong environmental lighting. Furthermore, camera manufacturers frequently bias the white balance away from neutral to reflect aesthetic preferences or maintain a brand-specific tonal style. These aesthetic adjustments cannot be captured by a neutral illuminant alone (see Fig. 1).

AWB typically involves two main steps: (1) illuminant estimation and (2) color correction [30]. In the first step, the goal is to estimate a single color vector representing the color bias introduced by the scene's illumination in the camera-captured raw image. This estimated vector—commonly referred to as the illuminant color—is then used in the second step, which applies global scaling to the R, G, and B channels of the raw image to compensate for the color bias [7]. Most prior work focuses on the illuminant estimation step and frames the problem within the context of color constancy, aiming to neutralize the effect of the scene's illumination [15, 16, 19, 20, 26, 31, 38, 41, 43, 44]. These methods seek to accurately predict the global illuminant color in the camera's raw color space to achieve a neutral white balance.

Recent state-of-the-art illuminant estimation methods are learning-based and require training on paired datasets consisting of raw images captured by a single camera along with ground-truth illuminants [12, 16, 32, 34, 39]. These ground truths are typically obtained using a calibration object (e.g., a color chart) placed in the scene to accurately capture the scene illumination. Such models are inherently camera-specific and struggle to generalize to new cameras with different sensor spectral sensitivities. This is because each camera interprets raw RGB values differently (see Fig. 2), making a model trained on one sensor unreliable when applied to another [4, 36].

To deploy AWB models across devices, manufacturers would need to retrain or fine-tune the model for each new sensor, which is impractical—especially for mobile manufacturers who release multiple devices per year, each with potentially different cameras. To address this, recent work has attempted to retain the power of learning-based methods while improving their generalization without requiring retraining [4, 11, 36, 50]. These models, termed *cross-camera AWB*, are typically trained on data from multiple cameras and incorporate architectural or training strategies that help the model adapt to new sensors. However, to ensure consistent supervision across sensors—each with its own raw RGB color space—these models are often restricted to using neutral illuminants as training targets. While practical, this constraint limits their ability to model aesthetic variations tailored to visual preferences or brand-specific styles, highlighting the need for a cross-camera approach that goes beyond color constancy to account for both perceptual and stylistic consistency.

This motivated us to develop a method that enables accurate, preference-aware white balance across cameras. Specifically, we propose a technique that augments existing cross-camera AWB methods by introducing an additional learnable mapping function to model target white-balance preferences. Our method is designed to be camera-agnostic, meaning that once the mapping is trained, it can be applied to unseen cameras without re-training, making it highly practical for camera manufacturers.

To evaluate our method's ability to produce consistent illuminant predictions with brand-specific or aesthetically-preferred biases across different sensors, we collected a test set from devices made by the same manufacturer, assuming that their onboard AWB systems reflect a consistent aesthetic preference. We used these in-camera AWB illuminant outputs as ground truth to train and evaluate our method. Additionally, we tested our model on the S24 dataset [13], which includes illuminants manually annotated by a professional photographer to reflect aesthetic preferences. Our results show consistent improvements—both qualitatively and quantitatively—across testing sets captured with different cameras, demonstrating the method's potential for real-world deployment.

Contribution In this paper, we propose a method for incorporating white-balance preferences into cross-camera AWB frameworks. Our approach introduces a post-illuminant-estimation mapping function trained to convert neutral illuminants predicted by existing cross-camera AWB models into manufacturer-preferred illuminants within a camera-agnostic space. This allows the method to generalize effectively to unseen cameras with varying characteristics. We evaluate our approach on a dataset of 771 smartphone images captured by three cam-

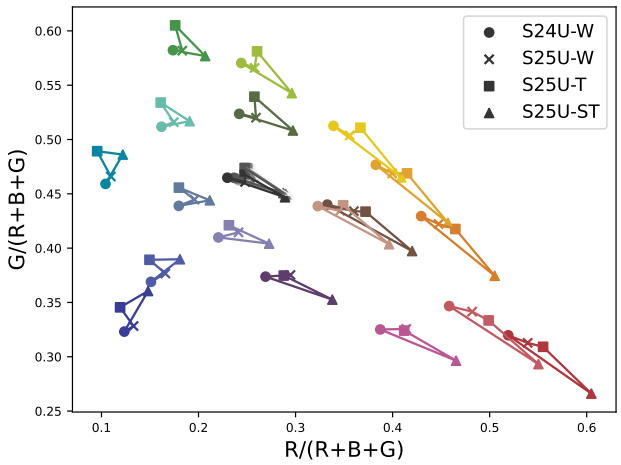

Figure 2. Each point in the figure represents a raw color from a Macbeth color checker. For the same object under identical illumination, different sensors produce different raw values due to variations in spectral response and other camera characteristics. Without special handling, models must be trained per sensor to accurately predict raw illuminants. Shown are examples from the Samsung S24 Ultra's wide camera (S24U-W), Samsung S25 Ultra's wide camera (S25U-W), Samsung S25 Ultra's telephoto camera (S25U-T), and Samsung S25 Ultra's super-telephoto camera (S25U-ST).

eras and show consistent improvements over standard cross-camera AWB methods in achieving the desired white balance style. Our model is lightweight, containing only ∼500 tunable parameters, and runs in under 0.05 milliseconds on a typical flagship mobile CPU.

2. Related Work

As stated earlier, most prior research has focused on accurately predicting the scene's illuminant color from a given image or auxiliary inputs for neutral AWB [2, 11, 12, 16, 34, 43]. Going beyond this, our method targets cross-camera illuminant estimation while explicitly modeling target white-balance preferences through a post-illuminant estimation mapping. To provide context, we briefly review related work in three categories: (1) camera-specific illuminant estimators, (2) cross-camera illuminant estimators, and (3) color mapping techniques related to white balance and in-camera color correction.

2.1. Camera-Specific Illuminant Estimators

Camera-specific illuminant estimators are learning-based models trained on paired datasets captured by a single camera (e.g., [2, 9, 12, 13, 15, 16, 34, 38, 41, 48, 51]). These models learn a mapping from input images (or derived color features), optionally combined with additional cues (e.g., [2, 12, 13]), to ground-truth illuminant colors obtained from a calibration object placed in the scene. These ground-truth illuminants are used to supervise models for neutral AWB, which aims to normalize both sensor sensitivity bias and the color cast introduced by scene illumination. As a result, most of these methods are inherently designed for neutral white balancing, due to the nature of how their ground-truth data is collected. A recent exception is the S24 dataset [13], which introduces ground-truth labels based on aesthetic preference, manually curated by a professional photographer.

Regardless of the supervision type, these models typically struggle to generalize to new cameras with different sensor characteristics [4]. This is because each sensor has its own raw RGB color space, and the same RGB value can correspond to different physical colors across sensors [3]. Consequently, re-training or fine-tuning is often required to adapt such models to new sensors—an impractical requirement in real-world scenarios, particularly in the smartphone industry, where manufacturers release multiple devices with varying camera modules each year.

2.2. Cross-Camera Illuminant Estimators

Cross-camera illuminant estimators are designed to deliver consistent performance regardless of the testing camera's characteristics. These methods fall into two categories: (1) inherently camera-agnostic approaches, such as statistical-based methods (e.g., [17, 19, 20, 31, 43, 44]), which estimate the illuminant directly from the image content without relying on learned priors; and (2) learning-based models that incorporate design elements to enhance generalization to unseen cameras at test time (e.g., [4, 11, 36, 50]).

Recent state-of-the-art cross-camera learning-based methods often rely on additional inputs beyond the image to help the model adapt to new sensors. For instance, C5 [11] requires auxiliary images captured by the test-time camera to better adapt to the color space of the testing camera. CCMNet [36] leverages the camera's color correction matrices (CCMs)—also known as color space transformation (CST) matrices—typically calibrated within the ISP, to aid adaptation to the test camera's color space.

While these methods achieve promising results across cameras, they are generally trained on data from multiple sensors with diverse characteristics and are limited to neutral AWB. This restriction arises because training on preferred white balance would require ground-truth preference data to be consistent across all cameras—a condition that is difficult to guarantee. Collecting such data would involve a highly controlled and labor-intensive process to ensure uniform aesthetic preferences across different sensors, making it impractical for the large-scale datasets required to train these models.

2.3. Color Mapping

Camera ISPs apply various color mapping transformations to render raw sensor data into display-ready images [23]. One such transformation is performed immediately after white balancing, using pre-calibrated CST matrices [18]. These matrices are typically calibrated in the lab during manufacturing using color charts, and are designed to map white-balanced raw data into a canonical, device-independent color space (e.g., CIE XYZ) [10, 28].

Color mapping has also been utilized within illuminant estimation and white balance methods. For example, sensor sharpening [27] applies a transformation to raw images to improve illuminant estimation accuracy. SIIE [4] learns a 3×3 invertible matrix to map raw images from various camera color spaces into a unified "working" space to enhance cross-camera generalization. APAP [9] introduces a post-illumination-estimation camera-specific mapping to refine initial illuminant estimates. The work in [6] proposes a rectification mapping function that projects initial illuminant estimates into a higher-order polynomial space to address non-linear color shifts caused by in-camera white balancing errors. While our method also learns a post-illuminant-estimation mapping, to the best of our knowledge, no prior work has explored learning a camera-agnostic post-illuminant-estimation mapping in a canonical space specifically aimed at cross-camera AWB with white-balance preferences.

3. Method

Given an RGB illuminant color, $l_{\texttt{raw}} \in \mathbb{R}^3$, in the testing camera's raw space—estimated by a cross-camera *neutral* AWB method—our goal is to estimate a mapped RGB illuminant, $\hat{l}_{\texttt{raw}} \in \mathbb{R}^3$, in the same raw space, aligned with a target white-balance preference that may incorporate aesthetic intent or a deliberate bias away from neutral white balance. We aim to learn a unified mapping regardless of the testing camera's raw space. In other words, the mapping should be camera-agnostic to ensure practicality and align with the goals of cross-camera AWB—eliminating the need for tuning or retraining for new cameras.

An overview of our method is shown in Fig. 3. As illustrated, our approach is learning-based and requires access to a paired training dataset from a source camera, consisting of raw images and their corresponding ground-truth white-balance preference illuminants in that camera's raw space. These target illuminants can be obtained by an expert photographer or derived through systematic criteria with manual review. Importantly, this dataset is needed only once—we do not require retraining our model for new, unseen cameras. To achieve this and ensure practicality, our method learns a mapping from neutral illuminants—produced by any cross-camera AWB algorithm—to the corresponding white-balance preference bias, in a camera-agnostic space. Specifically, we first apply a cross-camera AWB method, f_{cc}, to estimate neutral illuminants for the source raw images. Both the estimated neutral illuminants and the ground-truth preferred illuminants are then transformed from the source camera's raw space into the CIE XYZ space (Sec. 3.1). We then learn a mapping function, g, that translates neutral illuminants to their white-balance-preferred counterparts in this common space (Sec. 3.2).

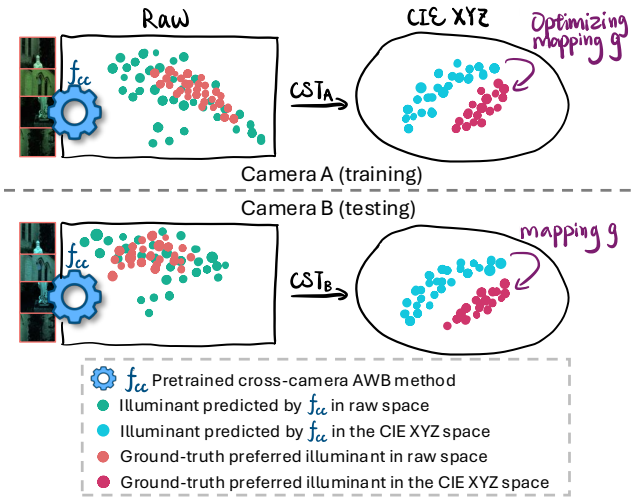

Figure 3. Our method learns a camera-agnostic mapping from illuminants predicted by a cross-camera neutral AWB algorithm to a target white-balance preference in the CIE XYZ space. At test time, it acts as a plug-and-play module that maps neutral illuminants to the desired preference for unseen cameras—without retraining or tuning. The mapped illuminant is then transformed back to the unseen camera's raw RGB space.

At inference time, this learned mapping can be directly applied to raw images from new, unseen cameras. That is, we estimate the neutral illuminant in the testing camera's raw space using a cross-camera AWB method, transform it to the CIE XYZ space, apply the learned mapping, and finally convert the result back to the testing camera's raw space. Our overall framework can be summarized as:

$$l_{\texttt{raw}} = f_{cc}(\mathbf{I}_{\texttt{raw}}), \tag{1}$$

$$\hat{l}_{\texttt{raw}} = \texttt{CST}^{-1}_{l_{\texttt{raw}}} g\left(\texttt{CST}_{l_{\texttt{raw}}} l_{\texttt{raw}}\right), \tag{2}$$

where $\mathbf{I}_{\texttt{raw}}$ is the downsampled raw image (which may optionally include additional images or metadata required by the cross-camera AWB method, f_{cc}), and $l_{\texttt{raw}} \in \mathbb{R}^3$ is the illuminant in raw space predicted by a cross-camera AWB method—either a pretrained learning-based model or a statistical method that is inherently cross-camera. $\texttt{CST}_{l_{\texttt{raw}}}$ denotes the 3×3 color space transformation matrix computed

based on $l_{\texttt{raw}}$. The function g represents the learned mapping applied in the CIE XYZ space.

3.1. Raw-to-XYZ Conversion

The CIE XYZ color space is a standardized, device-independent color space defined by the International Commission on Illumination (CIE). It was designed to model human color perception and serves as a foundational reference for many other color spaces (e.g., sRGB, LAB, etc.). In our work, we map estimated neutral illuminants to the CIE XYZ space before applying our mapping. The motivation for choosing the CIE XYZ space is that camera manufacturers typically pre-calibrate CST matrices for each new camera as part of the manufacturing process. These matrices are designed to map the camera's raw RGB space—under a specific correlated color temperature (CCT)—to the CIE XYZ space. They are accessible to camera ISPs and are often stored in the metadata of raw files. In the case of DNG files—a standardized raw format—these matrices are stored in the 'ColorMatrix1' and 'ColorMatrix2' tags, each calibrated under a standard illuminant specified by the 'CalibrationIlluminant1' and 'CalibrationIlluminant2' tags, respectively.

To compute a CST matrix for an arbitrary illuminant (different from the two calibration illuminants), we follow the DNG specification [1] and perform linear interpolation based on the inverse CCT of the target illuminant [47]:

$$\texttt{CST}_{l_{\texttt{raw}}} = \alpha\left(l_{\texttt{raw}}\right)\texttt{CST}_{l_a} + \left(1-\alpha\left(l_{\texttt{raw}}\right)\right)\texttt{CST}_{l_b}, \quad (3)$$

$$\alpha\left(l_{\texttt{raw}}\right) = \frac{1/c_{l_{\texttt{raw}}} - 1/c_{l_b}}{1/c_{l_a} - 1/c_{l_b}}, \quad (4)$$

where $\texttt{CST}_{l_a}$ and $\texttt{CST}_{l_b}$ are the CST matrices pre-calibrated for standard illuminants l_a and l_b, and c_l denotes the CCT of illuminant l, computed using Robertson's method [45].

We use this transformation to convert all training illuminants (both neutral and preferred) from the training camera's raw space to the CIE XYZ space. At inference time, we apply the same transformation to the neutral illuminant estimated by the cross-camera AWB method, f_{cc}, converting it from the testing camera's raw space to the CIE XYZ space. After our learned mapping (Sec. 3.2), we apply the inverse of the computed CST matrix (Eq. 3) to transform the mapped illuminant back from the CIE XYZ space to the testing camera's raw space. This design enables our mapping to be camera-agnostic, allowing it to serve as a plug-and-play module for any unseen camera without requiring retraining or tuning.

3.2. Learnable White-Balance Preference Mapping

We design g as a learnable function implemented by a lightweight neural network, trained to map neutral illuminant colors in the CIE XYZ space to their corresponding target white-balance preferences. Our model consists of 539 parameters and comprises four linear layers with ELU activations [22], with batch normalization applied to the first layer. To enrich the input feature, we first project the input neutral illuminant color, $l_{\texttt{xyz}} = \texttt{CST}_{l_{\texttt{raw}}} l_{\texttt{raw}}$, $l_{\texttt{xyz}} \in \mathbb{R}^3$, to a higher-dimensional space and process it using our network, g, as follows:

$$\hat{l}_{\texttt{xyz}} = g(\phi(l_{\texttt{xyz}})), \quad (5)$$

where ϕ is a polynomial kernel [33], which transforms the $[x, y, z]^T$ illuminant vector into $[x, y, z, xy, xz, yz, x^2, y^2, z^2, xyz]^T$. Each layer of our network has the following (input, output) dimensions: (10, 16), (16, 8), (8, 16), and (16, 3), respectively. The last layer outputs the predicted illuminant, $\hat{l}_{\texttt{xyz}}$, and has no activation. $\hat{l}_{\texttt{xyz}}$ is then normalized via division by its L2 norm before transforming back to the target sensor's raw space.

The weights of our mapping function, g, are optimized using the Adam optimizer [37] for 2000 epochs, with β values set to (0.9, 0.999), a weight decay of 10^{-9}, and a cosine annealing learning rate schedule [40]. The objective is to minimize the angular error between the mapped illuminants (in CIE XYZ) and the ground-truth illuminants with the target white-balance preference in the training dataset.

4. Experiments

We evaluate our method in a cross-camera setting, where the model is trained on data captured by a source camera and tested on data from three different target cameras to assess generalization. To validate performance in a camera-specific scenario, we also report results on the same camera used for training. In addition, we compare our approach against two baseline mappings and analyze its performance when the mapping is learned in an alternative color space. We follow standard evaluation practices in color constancy and white balance literature, reporting angular error statistics including the mean, median, and tri-mean, as well as the arithmetic mean of the best 25%, worst 25%, and worst 5% angular errors between the predicted and ground-truth illuminants, in the target camera's raw space.

4.1. Data

To quantitatively evaluate our method, we require paired data captured by different cameras, each with distinct characteristics, where a ground-truth aesthetically-preferred illuminant accompanies each image. To ensure consistency in the aesthetic adjustments of the ground-truth illuminants across all cameras, we use the in-camera AWB estimated illuminant values saved in the raw image metadata. We use these values as our target white-balance preference because AWB algorithms from the same manufacturer, particularly across successive smartphone generations, are typically tuned to maintain a consistent aesthetic style.

Table 1. **Cross-camera evaluation:** We report the angular errors on the *S25U-W* test camera for different cross-camera methods combined with mapping baselines and our proposed mapping.

Cross camera method	Mapping	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max
C5	No mapping	2.77	1.34	0.43	7.61	19.96	1.52	41.22
	3×3	2.53	1.91	0.77	5.29	11.91	2.01	26.78
	Polynomial	2.21	1.10	0.36	6.03	17.34	1.24	45.75
	Ours	**1.34**	**0.92**	**0.30**	**3.07**	**5.66**	**1.02**	**9.07**
C4	No mapping	3.33	2.06	0.61	8.24	23.38	2.24	48.11
	3×3,	3.49	2.78	1.43	6.49	9.44	3.06	17.59
	Polynomial	2.51	1.88	**0.49**	5.88	14.82	1.88	44.57
	Ours	**1.95**	**1.71**	0.68	**3.63**	**5.51**	**1.75**	**8.81**
GW	No mapping	3.60	2.43	0.63	9.02	24.25	2.40	49.89
	3×3	2.74	1.86	0.72	6.11	12.74	2.16	33.70
	Polynomial	2.27	1.14	0.44	6.32	21.47	1.21	72.45
	Ours	**1.14**	**0.93**	**0.35**	**2.34**	**4.18**	**0.94**	**8.10**
wGE	No mapping	1.83	**0.89**	**0.25**	5.13	11.24	**1.04**	18.74
	3×3	2.80	2.23	0.99	5.51	7.63	2.45	**10.33**
	Polynomial	1.79	1.21	0.43	4.13	8.36	1.29	17.01
	Ours	**1.59**	1.09	0.41	**3.66**	**7.34**	1.15	11.72

Table 2. **Cross-camera evaluation:** We report the angular errors on the *S25U-T* test camera for different cross-camera methods combined with mapping baselines and our proposed mapping.

Cross camera method	Mapping	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max
C5	No mapping	3.90	3.06	0.97	8.17	17.41	3.18	34.77
	3×3	2.47	1.79	0.57	5.61	12.26	1.90	23.90
	Polynomial	2.26	1.70	**0.53**	5.02	9.83	1.78	15.90
	Ours	**2.10**	**1.61**	0.58	**4.44**	**7.60**	**1.71**	**10.14**
C4	No mapping	4.05	2.81	0.75	9.74	24.89	2.84	49.18
	3×3	4.00	3.16	1.41	7.97	14.15	3.39	30.72
	Polynomial	2.60	2.03	0.66	5.64	11.01	2.10	20.03
	Ours	**2.17**	**1.78**	**0.56**	**4.51**	**7.43**	**1.84**	**10.84**
GW	No mapping	4.53	3.18	0.97	10.57	25.11	3.38	47.37
	3×3	3.23	2.32	1.12	6.84	15.57	2.47	39.54
	Polynomial	3.15	1.56	0.54	8.62	26.97	1.70	82.51
	Ours	**1.77**	**1.31**	**0.52**	**3.84**	**6.64**	**1.40**	**14.48**
wGE	No mapping	3.60	1.77	**0.41**	10.27	24.74	2.08	44.09
	3×3	3.63	2.69	1.12	7.71	15.75	2.87	34.68
	Polynomial	3.44	1.71	0.67	9.49	25.02	1.92	64.48
	Ours	**2.20**	**1.29**	0.48	**5.50**	**9.73**	**1.48**	**13.03**

Table 3. **Cross-camera evaluation:** We report the angular errors on the *S25U-ST* test camera for different cross-camera methods combined with mapping baselines and our proposed mapping.

Cross camera method	Mapping	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max
C5	No mapping	4.18	3.41	1.04	8.89	17.58	3.47	33.92
	3×3	2.53	**1.82**	**0.58**	5.68	11.93	**1.94**	22.07
	Polynomial	2.49	1.96	0.61	5.28	9.15	2.04	11.82
	Ours	**2.45**	1.98	0.69	**5.14**	**8.50**	2.03	**11.00**
C4	No mapping	4.20	2.75	0.82	10.05	22.11	3.01	43.68
	3×3	3.68	2.81	1.23	7.38	12.12	3.11	20.18
	Polynomial	**2.67**	**1.98**	**0.56**	6.05	11.48	**2.11**	24.50
	Ours	2.69	2.20	0.85	**5.45**	**9.67**	2.27	**13.90**
GW	No mapping	4.87	3.37	0.94	11.53	24.34	3.60	43.99
	3×3	2.86	2.07	0.72	6.38	13.89	2.20	31.41
	Polynomial	2.16	1.50	**0.54**	5.00	10.71	1.58	15.98
	Ours	**1.89**	**1.42**	0.55	**4.06**	**7.60**	**1.48**	**10.40**
wGE	No mapping	4.32	2.00	**0.56**	12.23	27.86	2.33	51.39
	3×3	3.54	2.47	0.79	8.17	17.56	2.64	43.53
	Polynomial	2.98	1.77	0.58	7.69	15.10	1.96	26.30
	Ours	**2.55**	**1.67**	0.64	**6.06**	**10.46**	**1.82**	**15.30**

In our experiments, we use the Samsung S24 Ultra's wide camera (S24U-W) for training, utilizing the S24 dataset introduced in [13], which includes 2,619 training images, 205 validation images, and 400 testing raw images captured under diverse lighting conditions. This dataset also provides aesthetically-preferred white-balance annotations, where an expert photographer has manually adjusted the white balance of each image.

In the cross-camera setup, we use the in-camera AWB estimated illuminants as the target white-balance preference to evaluate generalization. Additionally, we report results using the aesthetically-preferred white-balance annotations in the S24 dataset [13] in the camera-specific evaluation. For cross-camera testing, we collected 257 additional scenes using three cameras from the same manufacturer as the training device. Specifically, the testing images were captured using the Samsung S25 Ultra's wide, telephoto, and super-telephoto cameras—referred to as S25U-W, S25U-T, and S25U-ST, respectively—resulting in a total of 771 raw images.

4.2. Cross-Camera Experiments

We report the results of our model, trained on S24U-W, on the S25U-W test set in Table 1, the S25U-T test set in Table 2, and the S25U-ST test set in Table 3. We show results basing our method on four cross-camera AWB methods, including two learning-based methods, C5 [11] and C4 [50], and two statistical-based methods, gray-world (GW) [19] and weighted gray-edge (wGE) [31]. Qualitative results are shown in Figs. 4–5, with additional examples in the supplemental materials.

We compare our method against three mapping baselines: (1) no mapping, (2) polynomial mapping [29], and (3) 3×3 correction [25]. For both the polynomial mapping and the 3×3 correction, we use the S24 dataset [13]—consistent with our method—to fit the coefficients of the respective mapping functions. In the 'no mapping' baseline, we compute the angular error between the cross-camera method's output and the white-balance preferred ground-truth illuminant. As expected, the angular error is relatively high, since cross-camera methods are designed to predict neutral rather than stylistically biased illuminants. The 3×3 correction baseline also yields high errors, sometimes even higher than the 'no mapping' baseline, likely due to its limited expressive capacity, thus underfitting the training dataset. The polynomial mapping performs better, offering higher flexibility than the 3×3 correction. Our model combines high expressiveness with lightweight computation, consistently outperforming all baselines in accuracy.

Notably, our mapping performs particularly well when combined with the GW method [19], often surpassing combinations with learning-based approaches. This may be attributed to GW's more systematic and predictable failures [52], which make it easier for our learned mapping to not only bias the initial neutral illuminant estimates towards the

Table 4. **Camera-specific evaluation:** Angular errors on the test set of the *S24U-W* camera [13], which is also the same camera used for training our model and fitting the baseline mappings. Results are reported for both the in-camera AWB estimated illuminant ground truth and the aesthetically-preferred ground truth provided in the S24 dataset, shown in the order: camera-estimated / aesthetically-preferred.

Cross-camera method	Mapping	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max
C5	No Mapping	2.92 / 2.96	1.71 / 1.93	0.50 / 0.57	7.34 / 7.22	14.71 / 14.18	2.03 / 2.14	35.44 / 28.35
	3×3	2.57 / 2.33	2.26 / 2.01	0.90 / 0.79	4.83 / 4.46	8.77 / 7.94	2.29 / 2.04	24.83 / 16.72
	Ours	**1.44 / 1.56**	**0.98 / 1.10**	**0.30 / 0.32**	**3.31 / 3.61**	**5.81 / 6.53**	**1.10 / 1.19**	**13.06 / 12.90**
C4	No Mapping	3.67 / 3.01	2.63 / 1.88	0.81 / 0.61	8.46 / 7.21	15.77 / 14.03	2.91 / 2.21	37.53 / 34.68
	3×3	4.24 / 3.89	3.84 / 3.54	1.89 / 1.98	7.28 / 6.42	10.14 / 8.74	3.91 / 3.63	14.90 / 12.69
	Ours	**1.96 / 1.73**	**1.47 / 1.41**	**0.57 / 0.39**	**4.17 / 3.64**	**6.62 / 5.93**	**1.62 / 1.48**	**11.66 / 10.38**
GW	No Mapping	4.72 / 5.63	3.38 / 4.65	0.98 / 1.18	10.65 / 11.89	19.40 / 19.69	3.70 / 4.79	36.01 / 31.63
	3×3	2.96 / 2.87	2.60 / 2.52	1.19 / 0.98	5.41 / 5.35	9.11 / 8.55	2.63 / 2.58	24.55 / 18.14
	Ours	**1.73 / 1.97**	**1.24 / 1.43**	**0.37 / 0.36**	**3.98 / 4.52**	**6.36 / 7.10**	**1.36 / 1.54**	**8.47 / 9.64**
wGE	No Mapping	3.15 / 3.33	1.82 / 2.01	0.47 / 0.53	8.19 / 8.45	16.94 / 16.01	2.03 / 2.33	37.62 / 31.94
	3×3	3.28 / 2.95	2.66 / 2.47	1.23 / 1.09	6.42 / 5.65	10.85 / 9.02	2.80 / 2.58	29.95 / 22.73
	Ours	**1.70 / 1.79**	**1.28 / 1.17**	**0.37 / 0.30**	**3.82 / 4.35**	**6.83 / 7.91**	**1.37 / 1.32**	**10.80 / 17.49**

Table 5. **Ablation study:** We report results for different cross-camera methods using our learnable mapping, optimized in (1) the training camera's raw space and (2) the CIE XYZ space, applied across three unseen test cameras: *S25U-W*, *S25U-T*, and *S25U-ST*.

Camera	Cross-camera method	Mapping space	Mean	Med.	Best 25%	Worst 25%	Worst 5%	Tri.	Max
S25U-W	C5	Camera's raw	1.67	1.26	0.48	3.52	6.00	1.34	9.25
		CIE XYZ (ours)	**1.34**	**0.92**	**0.30**	**3.07**	**5.66**	**1.02**	**9.07**
	C4	Camera's raw	2.89	2.90	1.12	4.79	6.93	2.78	10.58
		CIE XYZ (ours)	**1.95**	**1.71**	**0.68**	**3.63**	**5.51**	**1.75**	**8.81**
	GW	Camera's raw	2.10	2.03	0.85	3.52	5.05	2.00	**7.66**
		CIE XYZ (ours)	**1.14**	**0.93**	**0.35**	**2.34**	**4.18**	**0.94**	8.10
	wGE	Camera's raw	1.95	1.75	0.65	3.75	**6.71**	1.72	**11.10**
		CIE XYZ (ours)	**1.59**	**1.09**	**0.41**	**3.66**	7.34	**1.15**	11.72
S25U-T	C5	Camera's raw	2.13	1.65	0.59	4.57	8.60	1.71	11.41
		CIE XYZ (ours)	**2.10**	**1.61**	**0.58**	**4.44**	**7.60**	**1.71**	**10.14**
	C4	Camera's raw	3.49	3.22	1.38	6.11	9.16	3.24	12.56
		CIE XYZ (ours)	**2.17**	**1.78**	**0.56**	**4.51**	**7.43**	**1.84**	**10.84**
	GW	Camera's raw	2.18	1.95	0.85	4.01	6.78	1.94	**13.93**
		CIE XYZ (ours)	**1.77**	**1.31**	**0.52**	**3.84**	**6.64**	**1.40**	14.48
	wGE	Camera's raw	2.41	1.87	0.68	**5.19**	**8.72**	1.95	**12.80**
		CIE XYZ (ours)	**2.20**	**1.29**	**0.48**	5.50	9.73	**1.48**	13.03
S25U-ST	C5	Camera's raw	4.06	3.79	1.55	7.20	11.27	3.78	15.04
		CIE XYZ (ours)	**2.45**	**1.98**	**0.69**	**5.14**	**8.50**	**2.03**	**11.00**
	C4	Camera's raw	5.34	5.17	2.78	8.24	12.04	5.14	17.77
		CIE XYZ (ours)	**2.69**	**2.20**	**0.85**	**5.45**	**9.67**	**2.27**	**13.90**
	GW	Camera's raw	4.33	4.14	2.17	6.90	9.79	4.11	12.40
		CIE XYZ (ours)	**1.89**	**1.42**	**0.55**	**4.06**	**7.60**	**1.48**	**10.40**
	wGE	Camera's raw	4.29	4.03	1.88	7.34	11.67	3.96	19.86
		CIE XYZ (ours)	**2.55**	**1.67**	**0.64**	**6.06**	**10.46**	**1.82**	**15.30**

preferred white-balance target but also correct those systematic errors. A similar trend was observed in APAP [9].

Ablation. We perform an ablation study on the choice of color space by learning the mapping directly in raw space instead of the camera-agnostic CIE XYZ space. As shown in Table 5, learning the mapping in the CIE XYZ consistently reduces angular error across all cameras and cross-camera base methods used for initial prediction. This aligns with our expectations: mappings learned in raw space tend to capture camera-specific spectral responses and characteristics (see Fig. 2), which limits their generalization to other cameras.

4.3. Camera-Specific Experiments

For completeness, we report results on the test split of the S24 dataset [13] in Table 4, to evaluate the effectiveness of our method for camera-specific white balance. In addition to using the camera-estimated illuminant values as ground truth, we also present results based on photographer-annotated aesthetically-preferred illuminants in the S24 dataset. Our proposed mapping consistently outperforms all baselines under both ground-truth settings.

5. Conclusion

We propose a cross-camera auto white balance method that explicitly accounts for target white-balance preferences. Specifically, we present a lightweight model (with only $\sim$500 parameters) that learns in a camera-agnostic space to map neutral illuminants—predicted by a pretrained cross-camera method—to illuminant colors that reflect a target white-balance aesthetic preference. Our approach is efficient, running in just 0.024 milliseconds on the Samsung S24 Ultra CPU, and is capable of supporting high frame rates on modern smartphone devices.

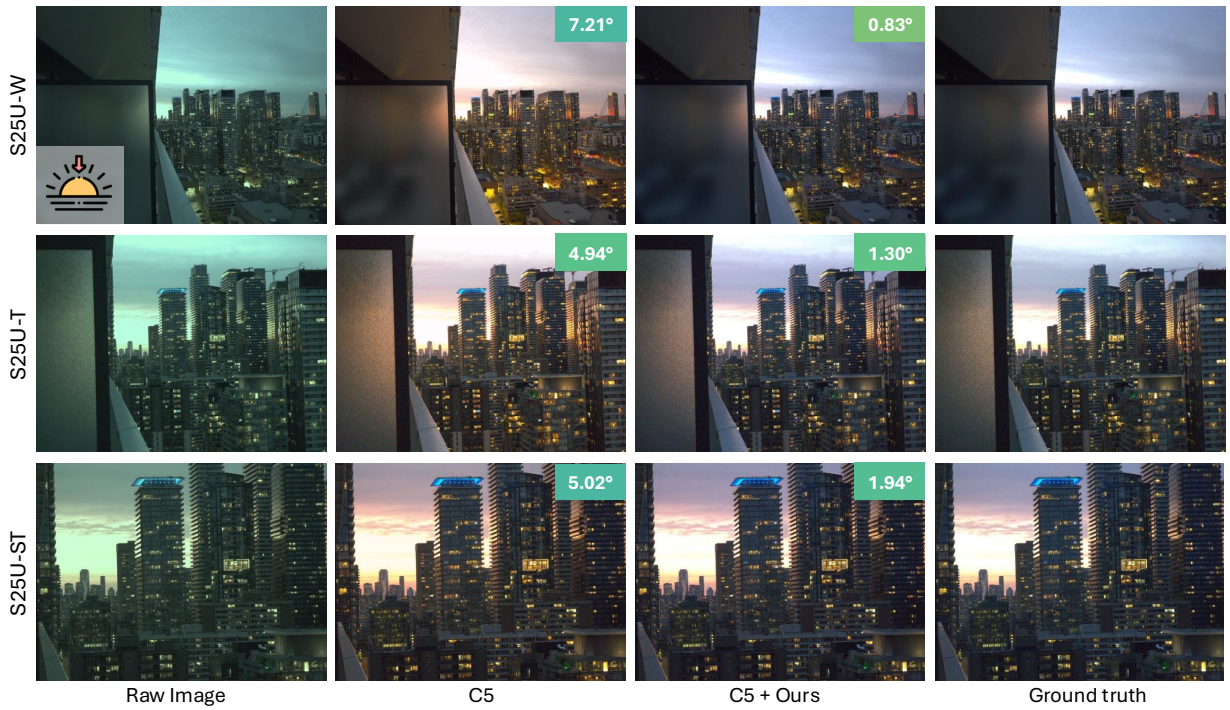

Figure 4. Qualitative results. Shown is a sunset scene captured by the S25U wide, telephoto, and super-telephoto sensors, white-balanced using C5 [11] predicted illuminant, C5 corrected with our mapping function, and the ground truth aesthetically-preferred illuminant.

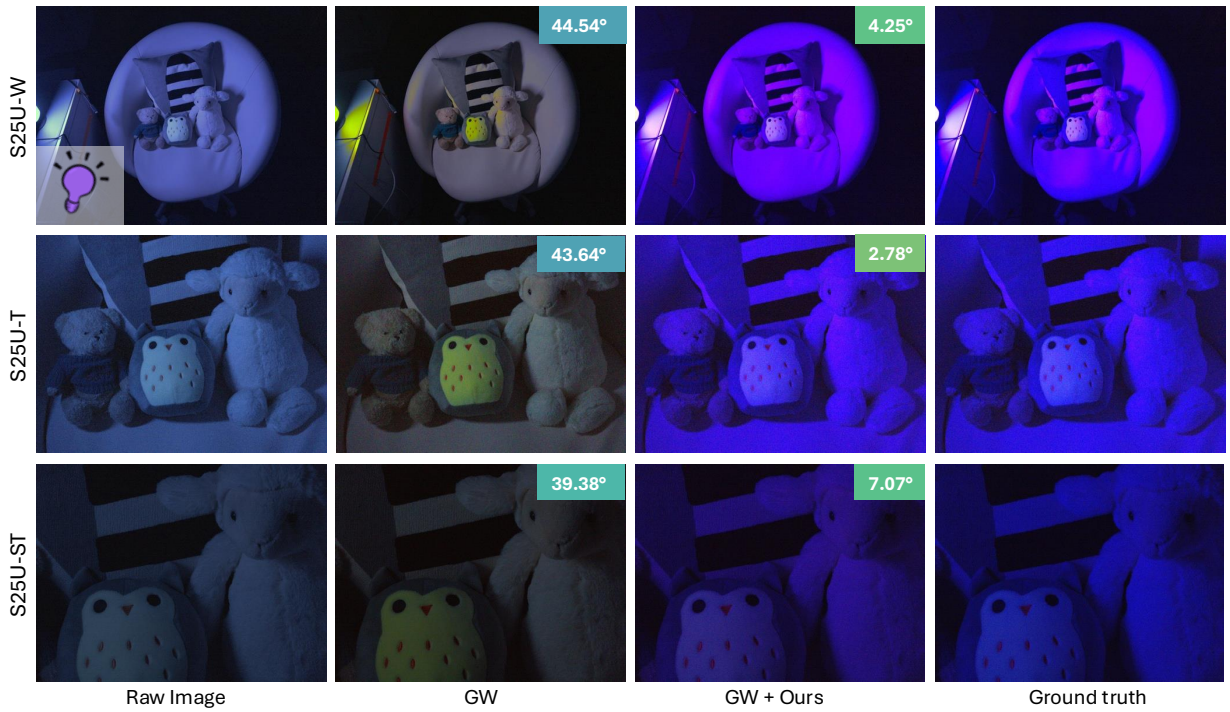

Figure 5. Qualitative results. Shown is an indoor scene illuminated by a purple LED light, captured by the S25U wide, telephoto, and super-telephoto sensors, white-balanced using GW [19] predicted illuminant, GW corrected with our mapping function, and the ground truth aesthetically-preferred illuminant.

References

[1] Adobe Digital Negative (DNG). https://helpx.adobe.com/camera-raw/digital-negative.html. Accessed: 2022-11-06. 5

[2] Abdelrahman Abdelhamed, Abhijith Punnappurath, and Michael S Brown. Leveraging the availability of two cameras for illuminant estimation. In *CVPR*, 2021. 3

[3] Mahmoud Afifi and Abdullah Abuolaim. Semi-supervised raw-to-raw mapping. In *BMVC*, 2021. 3

[4] Mahmoud Afifi and Michael S Brown. Sensor-independent illumination estimation for DNN models. In *BMVC*, 2019. 2, 3, 4

[5] Mahmoud Afifi and Michael S Brown. Deep white-balance editing. In *CVPR*, 2020. 2

[6] Mahmoud Afifi and Michael S Brown. Interactive white balancing for camera-rendered images. In *CIC*, 2020. 4

[7] Mahmoud Afifi, Brian Price, Scott Cohen, and Michael S Brown. When color constancy goes wrong: Correcting improperly white-balanced images. In *CVPR*, 2019. 2

[8] Mahmoud Afifi, Abhijith Punnappurath, Abdelrahman Abdelhamed, Hakki Can Karaimer, Abdullah Abuolaim, and Michael S Brown. Color temperature tuning: Allowing accurate post-capture white-balance editing. In *CIC*, 2019. 2

[9] Mahmoud Afifi, Abhijith Punnappurath, Graham Finlayson, and Michael S Brown. As-projective-as-possible bias correction for illumination estimation algorithms. *Journal of the Optical Society of America A*, 36(1):71–78, 2019. 3, 4, 7

[10] 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. 4

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

[12] Mahmoud Afifi, Zhenhua Hu, and Liang Liang. Optimizing illuminant estimation in dual-exposure HDR imaging. In *ECCV*, 2024. 2, 3

[13] Mahmoud Afifi, Luxi Zhao, Abhijith Punnappurath, Mohammed A Abdelsalam, Ran Zhang, and Michael S Brown. Time-aware auto white balance in mobile photography. *arXiv preprint arXiv:2504.05623*, 2025. 2, 3, 6, 7

[14] Vivek Agarwal, Besma R Abidi, Andreas Koschan, and Mongi A Abidi. An overview of color constancy algorithms. *Journal of Pattern Recognition Research*, 1(1):42–54, 2006. 2

[15] Jonathan T Barron. Convolutional color constancy. In *ICCV*, 2015. 2, 3

[16] Jonathan T Barron and Yun-Ta Tsai. Fast Fourier color constancy. In *CVPR*, 2017. 2, 3

[17] 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. 3

[18] Michael S. Brown. Color processing for digital cameras. *Fundamentals and applications of colour engineering*, pages 81–98, 2023. 2, 4

[19] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1): 1–26, 1980. 2, 3, 6, 8

[20] 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. *Journal of the Optical Society of America A*, 31(5):1049–1058, 2014. 2, 3

[21] Dongliang Cheng, Abdelrahman Abdelhamed, Brian Price, Scott Cohen, and Michael S Brown. Two illuminant estimation and user correction preference. In *CVPR*, 2016. 2

[22] Djork-Arné Clevert. Fast and accurate deep network learning by exponential linear units (ELUs). *arXiv preprint arXiv:1511.07289*, 2015. 5

[23] Mauricio Delbracio, Damien Kelly, Michael S Brown, and Peyman Milanfar. Mobile computational photography: A tour. *Annual Review of Vision Science*, 7(1):571–604, 2021. 2, 4

[24] Egor Ershov, Vasily Tesalin, Ivan Ermakov, and Michael S Brown. Physically-plausible illumination distribution estimation. In *ICCV*, 2023. 2

[25] Graham D Finlayson. Corrected-moment illuminant estimation. In *ICCV*, 2013. 6

[26] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *CIC*, 2004. 2

[27] Graham D Finlayson, Mark S Drew, and Brian V Funt. Spectral sharpening: sensor transformations for improved color constancy. *Journal of the optical society of america A*, 11 (5):1553–1563, 1994. 4

[28] Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. Color correction using root-polynomial regression. *IEEE Transactions on Image Processing*, 24(5):1460–1470, 2015. 4

[29] Graham D. Finlayson, Michal Mackiewicz, and Anya Hurlbert. Color correction using root-polynomial regression. *IEEE Transactions on Image Processing*, 24(5):1460–1470, 2015. 6

[30] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, 20(9):2475–2489, 2011. 2

[31] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Improving color constancy by photometric edge weighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):918–929, 2011. 2, 3, 6

[32] Daniel Hernandez-Juarez, Sarah Parisot, Benjamin Busam, Ales Leonardis, Gregory Slabaugh, and Steven McDonagh. A multi-hypothesis approach to color constancy. In *CVPR*, 2020. 2

[33] Guowei Hong, M Ronnier Luo, and Peter A Rhodes. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Research & Application*, 26 (1):76–84, 2001. 5

[34] Yuanming Hu, Baoyuan Wang, and Stephen Lin. FC4: Fully convolutional color constancy with confidence-weighted pooling. In *CVPR*, 2017. 2, 3

[35] Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics (TOG)*, 37(2): 1–17, 2018. 2

[36] Dongyoung Kim, Mahmoud Afifi, Dongyun Kim, Michael S Brown, and Seon Joo Kim. CCMNet: Leveraging calibrated color correction matrices for cross-camera color constancy. *arXiv preprint arXiv:2504.07959*, 2025. 2, 3

[37] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[38] 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. 2, 3

[39] Yi-Chen Lo, Chia-Che Chang, Hsuan-Chao Chiu, Yu-Hao Huang, Chia-Ping Chen, Yu-Lin Chang, and Kevin Jou. CLCC: Contrastive learning for color constancy. In *CVPR*, 2021. 2

[40] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 5

[41] Zhongyu Lou, Theo Gevers, Ninghang Hu, Marcel P Lucassen, et al. Color constancy by deep learning. In *BMVC*, 2015. 2, 3

[42] Ming Luo. A review of chromatic adaptation transforms. *Review of Progress in Coloration and Related Topics*, 30:77 – 92, 2008. 2

[43] Yanlin Qian, Joni-Kristian Kamarainen, Jarno Nikkanen, and Jiri Matas. On finding gray pixels. In *CVPR*, 2019. 2, 3

[44] Yanlin Qian, Said Pertuz, Jarno Nikkanen, Joni-Kristian Kämäräinen, and Jiri Matas. Revisiting gray pixel for statistical illumination estimation. In *VISSAP*. 2019. 2, 3

[45] A. R. Robertson. Computation of correlated color temperature and distribution temperature. *Journal of the Optical Society of America*, (11):1528–1535, 1968. 5

[46] Michael Scuello, Israel Abramov, James Gordon, and Steven Weintraub. Museum lighting: Why are some illuminants preferred? *Journal of the Optical Society of America A*, 21(2): 306–311, 2004. 2

[47] Donghwan Seo, Abhijith Punnappurath, Luxi Zhao, Abdelrahman Abdelhamed, Sai Kiran Tedla, Sanguk Park, Jihwan Choe, and Michael S. Brown. Graphics2RAW: Mapping computer graphics images to sensor raw images. In *ICCV*, 2023. 5

[48] Wu Shi, Chen Change Loy, and Xiaoou Tang. Deep specialized network for illuminant estimation. In *ECCV*, 2016. 3

[49] Shoji Tominaga, Takahiko Horiuchi, Shiori Nakajima, and Mariko Yano. Prediction of incomplete chromatic adaptation under illuminant A from images. In *CIC*, 2014. 2

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

[51] 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. 3

[52] Roshanak Zakizadeh, Michael S Brown, and Graham D Finlayson. A hybrid strategy for illuminant estimation targeting hard images. In *ICCV Workshops*, 2015. 6