This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

# Leveraging Multispectral Sensors for Color Correction in Mobile Cameras

Luca Cogo[1] Marco Buzzelli[1] Simone Bianco[1] Javier Vazquez-Corral[2,3] Raimondo Schettini[1]
[1] University of Milano-Bicocca [2] Computer Vision Center [3] Universitat Autònoma de Barcelona
{luca.cogo, marco.buzzelli, simone.bianco, raimondo.schettini}@unimib.it
javier.vazquez@cvc.uab.cat

## Abstract

*Recent advances in snapshot multispectral (MS) imaging have enabled compact, low-cost spectral sensors for consumer and mobile devices. By capturing richer spectral information than conventional RGB sensors, these systems can enhance key imaging tasks, including color correction. However, most existing methods treat the color correction pipeline in separate stages, often discarding MS data early in the process. We propose a unified, learning-based framework that performs end-to-end color correction and jointly leverages data from a high-resolution RGB sensor and an auxiliary low-resolution MS sensor. Our approach integrates the full pipeline within a single model, producing coherent and color-accurate outputs. We demonstrate the flexibility and generality of our framework by refactoring two different state-of-the-art image-to-image architectures. To support training and evaluation, we construct a dedicated dataset by aggregating and repurposing publicly available spectral datasets, rendering under multiple RGB camera sensitivities. Extensive experiments show that our approach improves color accuracy and stability, reducing error by up to 50% compared to RGB-only and MS-driven baselines. Code, models and dataset available at: https://lucacogo.github.io/Mobile-Spectral-CC/.*

## 1. Introduction

Color correction is a fundamental component of camera imaging pipelines, responsible for transforming raw sensor measurements into perceptually accurate, device-independent color representations within standard color spaces such as CIE XYZ [40] or ProPhoto RGB [28], typically under a canonical illuminant (e.g., D65). This task is particularly challenging in consumer digital photography applications, where a wide variety of lighting conditions must be handled robustly.

Traditional color correction pipelines rely on high-resolution RGB sensors and implement the process as a sequence of modular stages, namely automatic white balancing (AWB) and color space transformation (CST). AWB consists of illuminant estimation and illuminant discounting, which compensate for the color of the scene illumination. The subsequent transformation maps the white-balanced image from the camera's raw color space to a device-independent representation. Although this modular approach simplifies the pipeline development, it also often propagates errors between stages, since the steps are treated independently. Furthermore, RGB sensors capture information in only three broad spectral channels, providing only limited spectral information about the observed scene. As a result, ambiguities arise when disentangling surface reflectance from illumination.

Recent progress in nanophotonics, semiconductor fab-

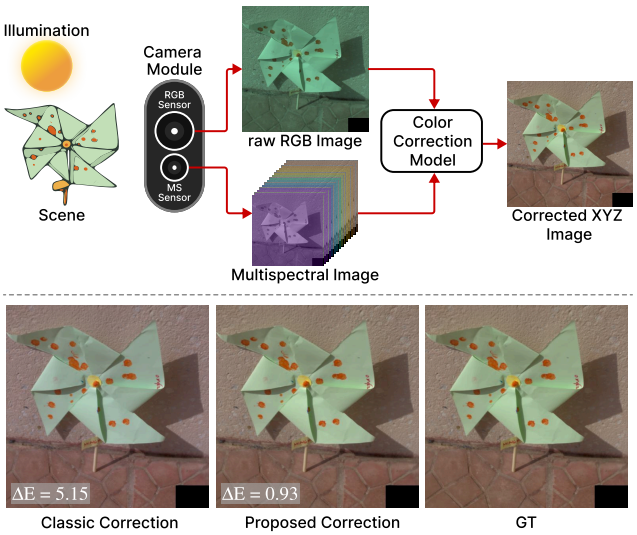



Figure 1. **Top**: Overview of the proposed color correction framework. A high-resolution RGB sensor is paired with a low-resolution multispectral sensor within a camera module. The two inputs are fused by a unified model that jointly performs illuminant estimation, illuminant discounting, and color correction to produce a color-accurate output. **Bottom**: Visual comparison between our proposed approach and a classic correction pipeline (using FC$^4$ [26] for illuminant estimation). We report the average $\Delta E_{00}$ color distance and convert the images to sRGB for visualization purposes.

rication, and on-chip optics has enabled compact snapshot multispectral sensors [46]. In contrast to high-end laboratory-grade systems that use line-scan mechanisms or tunable filters, snapshot MS sensors utilize spectral filter arrays (SFAs), where each pixel (or small cluster of pixels) is coated with a distinct spectral filter. Although the former offer high spectral resolution, they come with drawbacks such as mechanical complexity, bulky design, and lengthy acquisition times [32, 48]. In comparison, SFAs enable the simultaneous capture of multiple spectral bands within a single exposure. As a result, snapshot MS sensors are compact, cost-effective, and capable of real-time operation, prompting the community to explore their use in computational photography applications [45, 47, 51]. Despite their lower spatial resolution and coarser spectral sampling compared to high-end devices, they provide richer spectral cues than RGB sensors, which can be leveraged to improve the color correction process. The benefit of incorporating multispectral information into the color correction pipeline can be understood through the image formation model:

$$I_c = \int_{\lambda_{min}}^{\lambda_{max}} R(\lambda)\, E(\lambda)\, S_c(\lambda)\, d\lambda, \tag{1}$$

where $I_c$ denotes the sensor response in channel $c$, $R(\lambda)$ is the surface reflectance, $E(\lambda)$ is the spectral power distribution of the illuminant, and $S_c(\lambda)$ is the spectral sensitivity of the channel $c$ of the camera. Accurate color correction requires disentangling $R(\lambda)$ and $E(\lambda)$ to achieve illuminant-invariant and device-independent color representations. Since RGB sensors provide only three measurements of this integral, the problem is underdetermined, resulting in the ambiguities described earlier. By sampling the spectrum more densely, multispectral sensors provide additional constraints that enable more accurate estimation of both $R(\lambda)$ and $E(\lambda)$, supporting more reliable illuminant estimation, illuminant discounting, and color space transformation. However, most existing research, both in RGB-only and MS-driven settings, still treats these stages as independent [6–9, 26]. In many multispectral-based pipelines, the spectral data is used only for illuminant estimation and then discarded, thereby ignoring valuable information that could enhance subsequent processing steps [22, 31]. Recent works that attempt to jointly model the entire color correction process [33] perform the computation directly on multispectral images, which limits their applicability to mobile platforms where the MS sensor typically has low spatial resolution and serves as an auxiliary modality.

In this work, we introduce a unified, learning-based framework that jointly performs illuminant estimation, illuminant discounting, and color space transformation by fusing high-resolution RGB imagery with auxiliary low-resolution MS measurements. As illustrated in Figure 1, rather than using spectral cues only for illuminant estimation, our method integrates them throughout the entire pipeline, producing more stable and color-accurate results. The framework is designed for mobile configurations where a compact MS sensor complements a primary high-resolution RGB camera. Our contributions are threefold:

- We propose a dual-input, end-to-end color correction framework that fuses high-resolution RGB and low-resolution MS inputs, jointly modeling all stages of the color correction process.
- We construct a physically grounded dataset comprising 116,688 RGB–MS image pairs with ground-truth color references, covering a wide range of illuminants and camera sensitivities.
- We refactor two lightweight image-to-image architectures [12, 39] into our framework, demonstrating the flexibility of our approach.

## 2. Related Work

**Illuminant Estimation** Illuminant estimation is the first step of automatic white balancing and aims to infer the color of the scene illumination from the captured image. Classical statistical methods rely on global color assumptions such as the Gray-World, White-Patch, and Gray-Edge hypotheses [49], which assume that the illuminant color can be inferred from simple image statistics (e.g., mean or max). These simple approaches are computationally inexpensive but often fail when their assumptions do not hold. Convolutional neural network based methods improve robustness by predicting the illuminant from learned image features; representative examples include the works of Barron et al. [6, 7], Hu et al. [26], and Bianco et al. [8, 9]. Recent research has tackled this task using different strategies, such as learning post-capture adjustments from rendered sRGB images [1–3, 43] or exploiting additional scene information [4]. In this context, the use of multispectral data has been increasingly investigated for improving illuminant estimation [14, 15, 22, 29, 31]. However, most existing methods rely solely on MS data rather than fusing it with RGB information, and typically discard the MS cues after illuminant estimation instead of leveraging them in subsequent processing stages. In this work, we address this limitation by introducing a unified RGB–MS framework that exploits multispectral information beyond illuminant estimation to achieve more accurate and consistent color correction.

**Illuminant Discounting** After the illuminant is estimated, its influence must be removed to recover the scene reflectance. This illuminant discounting stage is commonly implemented using the von Kries adaptation model [17], which applies a channel wise scaling that assumes the illuminant contribution can be neutralized independently in each channel. This yields a diagonal correction matrix that enforces a neutral response under a canonical white light.

Although the von Kries model is efficient and widely used, it is known to be suboptimal, as it generally achieves a neutral appearance for gray regions but introduces noticeable chromatic errors for non-neutral surfaces [11]. Several works have proposed more advanced correction strategies that integrate spatial or contextual information [11, 18], yet these remain constrained by the limited spectral information available in RGB data. Spectral measurements from multispectral sensors enable a more physically grounded separation of reflectance and illumination and directly address this limitation. In this work, given the end-to-end nature of our framework, we incorporate multispectral cues throughout the discounting process so that the model can perform more accurate illumination discount.

**Color Space Transformation** Color space transformation converts white-balanced camera RGB values into a device-independent color representation such as CIE XYZ [40] or ProPhoto RGB [28]. Conventional pipelines adopt pre-calibrated $3 \times 3$ CST matrices under multiple reference illuminants using standard color targets such as the Macbeth ColorChecker [38]. At runtime the estimated correlated color temperature determines interpolation weights between them, yielding the final transformation. This approach effectively normalizes the camera response but does not explicitly model interactions between illumination and reflectance, which limits color accuracy under varying lighting conditions. Other studies replace the linear transform with non-linear mappings learned end-to-end from data [19, 24]. These methods can improve perceptual fidelity but they still rely solely on RGB inputs, which inherently constrains their accuracy. In contrast, the end-to-end formulation of our framework leverages multispectral information to inform the color space transformation (CST) directly, resulting in more perceptually accurate colors.

**Traditional Camera Color Correction Pipelines** Conventional camera pipelines perform color correction in two stages: automatic white balancing (AWB) and color space transformation (CST), both derived from the image formation model in Equation (1). The AWB stage discounts for the estimated illuminant using the von Kries adaptation model [17], which applies channel-wise scaling to enforce neutrality under a canonical white light. The CST stage then maps the white-balanced RGB values to a device-independent color space such as CIE XYZ [40]. Two $3 \times 3$ CST matrices are typically calibrated under reference illuminants at 2500°K and 6500°K using Macbeth ColorChecker targets [38], and the final transformation is obtained by interpolating between them based on the estimated correlated color temperature. This modular pipeline, though efficient, treats illuminant estimation, discounting, and transformation as independent steps, which limits color accuracy under complex lighting. Nevertheless, it represents the standard color correction structure adopted in commercial cameras. For a fair and realistic comparison, we employ this pipeline (hereafter referred to as the traditional correction pipeline) when evaluating state-of-the-art illuminant estimation methods, as they operate solely on the first stage of the color correction pipeline. In contrast, our framework learns all stages jointly, guided by both RGB and multispectral cues.

## 3. Method

We aim at performing the full color correction pipeline by leveraging both a high-resolution RGB image and an auxiliary low-resolution multispectral capture. The two modalities are fused within an image-to-image framework that implicitly estimates the scene illuminant, compensates for both its color cast and the spectral sensitivity of the camera, and produces color-corrected outputs directly in the CIE XYZ color space. We adapt to this setting two recent lightweight architectures, LPIENet [12] and cmKAN [39], modifying their designs to process the additional MS input. Figure 2 shows a schematic of both architectures.

**LPIENet** was originally designed for general-purpose image restoration and optimized for lightweight deployment on mobile hardware. Its structure follows a U-Net-like topology [42], which makes it a suitable foundation for image-to-image problems requiring spatially coherent transformations. The core component of the architecture is the IRA block (Inverted Residual Attention), which integrates convolutional blocks inspired by the MobileNet architecture [25] with parallel channel and spatial attention mechanisms [50]. The network is composed of six main blocks: three encoder blocks and two decoder blocks, connected through skip connections, followed by a final refinement block. To integrate the multispectral modality, we introduce an additional spectral encoder that mirrors the original encoder architecture and consists of three IRA blocks without downsampling. The features extracted from each encoder of the multispectral branch are fused with those from the RGB encoder through the skip connections, where element-wise addition is applied before each decoding stage. The decoder structure is maintained identical to the original LPIENet. We conduct experiments using two model configurations. The first follows the original feature sizes $[16, 32, 64, 32, 16]$ across the encoder-decoder blocks, while the second one uses a "small" configuration $[8, 16, 32, 16, 8]$. Both models are highly lightweight, containing approximately 220K and 60K trainable parameters, respectively.

**cmKAN** is an image-to-image architecture that differs from conventional encoder–decoder designs. Originally designed for color matching, it employs a hypernetwork gen-

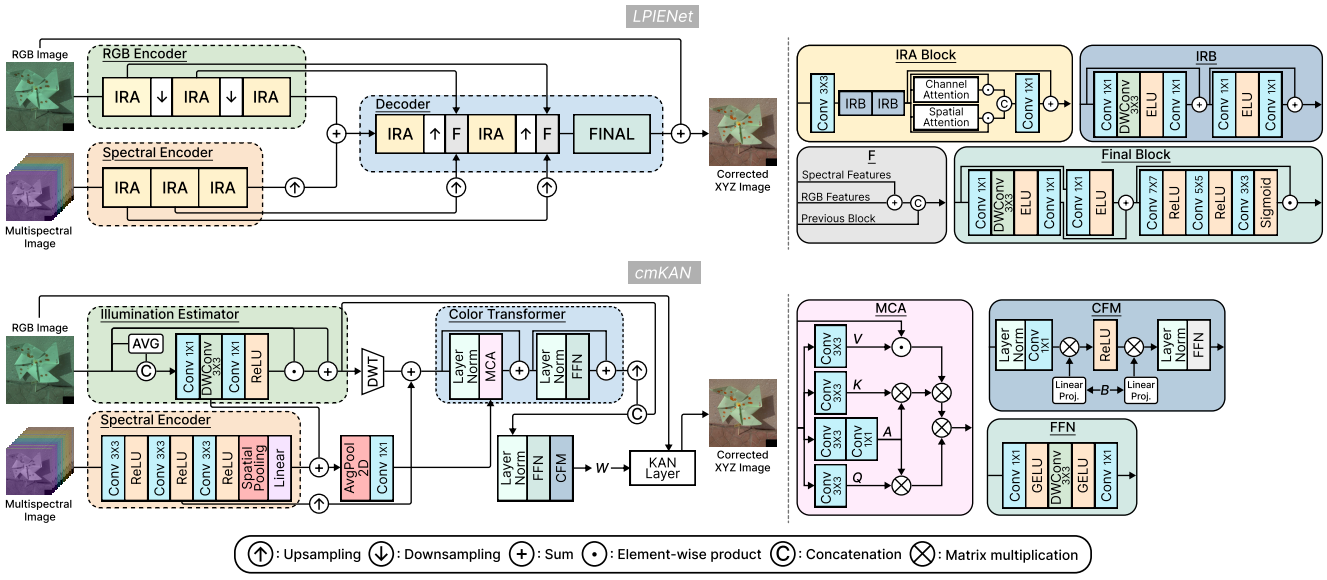



Figure 2. Overview of the proposed architectures adapted from (**top**) LPIENet [12] and (**bottom**) cmKAN [39]. For both architectures, we add a spectral encoder module and fuse the features with the ones extracted from the RGB image. For visualization purposes the output and the input RGB images are respectively converted to sRGB and gamma corrected.

erator that predicts spatially varying parameters controlling the splines of a Kolmogorov-Arnold Network (KAN) layer [35], enabling smooth and non-linear color transformations across the image. In this work, we refactor the lightweight generator variant proposed by the original authors, which comprises three principal modules: the Illumination Estimator (IE), which captures illumination information; the Color Transformer (CT), which incorporates a Multi-Scale Color Attention (MCA) mechanism to encode inter-channel dependencies; and the Color Feature Modulator (CFM), which fuses the learned features and projects them into the KAN parameter space. Consistent with the original implementation, a single KAN layer is employed, using B-splines of order $k = 3$ and a grid size $G = 5$. To integrate multispectral information, a compact spectral encoder composed of three convolutional layers with $3 \times 3$ kernels and ReLU activations is introduced. Spectral features are incorporated within the generator at two distinct feature levels through element-wise addition with the outputs produced by the Illumination Estimator block. The entire architecture is highly compact, with only 18K trainable parameters in total.

## 4. Dataset

To validate the proposed framework, we require a dataset comprising high-resolution RGB and low-resolution MS acquisitions, along with ground-truth images in the CIE XYZ color space under a canonical illuminant (D65). Such data allow quantitative evaluation of the color correction process within a physically consistent environment. Existing datasets for color constancy and color correction [10, 16, 21] or for RGB+MS imaging [22] lack the spectral and colorimetric information necessary to derive ground-truth reflectance and illumination. In practice, accurate color ground-truths cannot be obtained from RGB or MS captures alone, since this requires hyperspectral reflectance measurements that permit re-illumination under a known illuminant. Consequently, no real dataset currently satisfies the requirements of our task. To overcome this limitation, we construct a physically grounded simulated dataset derived from two publicly available hyperspectral datasets with densely sampled spectral reflectance images [13, 34]. Because these scenes were captured with real hyperspectral cameras, they inherently include realistic sensor noise, optical blur, and acquisition artifacts. Using these reflectance spectra, we synthesize corresponding RGB and MS measurements according to the image formation model in Equation (1) and render ground-truth color references under the canonical D65 illuminant. Figure 3 illustrates the dataset generation procedure and some representative scenes. Each scene is re-illuminated under 102 illuminants from the dataset of Barnard et al. [5], providing a wide variety of lighting conditions. RGB images are simulated using several representative camera spectral sensitivities from mobile devices[1] and mirrorless cameras[2], while the MS images are rendered using the sensitivities of a Spectricity S1 sensor [44], comprising 15 narrowband channels across the visible range. RGB images are rendered at $512 \times 512$ pix-

[1]Google Pixel 3, iPhone Xs Max, Huawei Mate 20 Pro, Samsung Galaxy Note 9

[2]Canon R5, Sony $\alpha$9 III, Nikon Zf

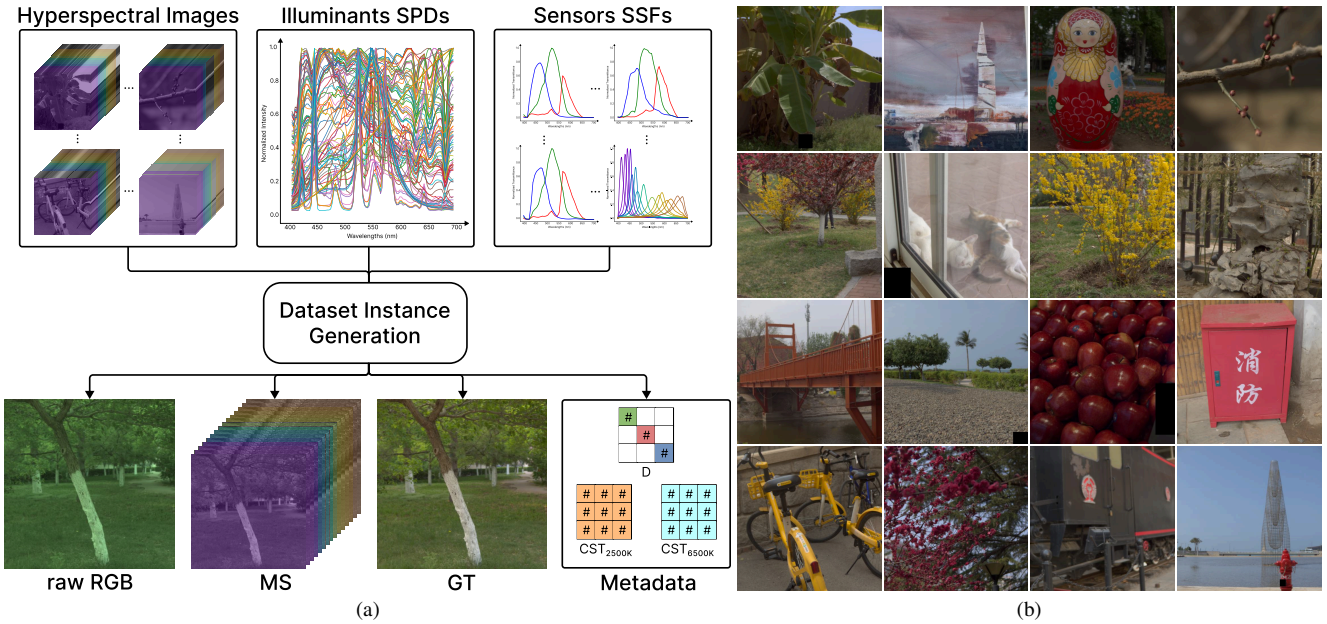



Figure 3. (a) Overview of the dataset generation pipeline. Hyperspectral reflectance images are rendered using multiple camera sensitivities under different illuminants with known spectral power distributions (SPDs) to produce RGB, MS and GT image triplets. (b) Representative samples of the proposed dataset. For visualization purposes, images are converted to sRGB.

els, matching the native resolution of the hyperspectral data, and MS images at $64 \times 64$ pixels to emulate a compact snapshot spectral sensor. Prior to rendering, we mask the white reference patches used for calibration and perform manual quality control to remove corrupted or incomplete data. After filtering, 1,144 valid scenes remain. Combined with the 102 illuminants, this results in 116,688 image triplets of high-resolution RGB, low-resolution MS, and CIE XYZ ground-truth under D65 illuminant. In addition, we compute metadata for each image, including the ground-truth illuminant RGB triplet and the color space transformation (CST) presets required for the traditional color correction pipeline. To prevent data leakage, we apply a scene-wise split so that test and validation scenes remain entirely unseen during training. Specifically, 80% of scenes are used for training and 20% for testing, with 20% of the training subset reserved for validation.

**Simulating Spatial Misalignments** To emulate geometric inconsistencies in dual-sensor systems, we generate a misaligned version of the aforementioned dataset by introducing spatial offsets to the RGB–MS pairs. Realistic warping transformations are derived from the Zurich dataset [27], which provides 168 real misaligned image pairs. For each simulated scene, a homography randomly sampled from this distribution is applied to the MS image, introducing plausible misalignment while preserving physical consistency. The misaligned dataset mirrors the structure and partitioning of the aligned version.

## 5. Experiments

**Experimental Setting** All models are trained using the proposed dataset, conducting separate experiments for each simulated RGB sensor. Training is performed with the Adam optimizer [30], using a learning rate of $1 \times 10^{-4}$. We employ early stopping with a maximum of 300 epochs and a patience value of 5, retaining the model that achieves the best performance on the validation set. Cosine annealing scheduling [36] is used to adapt the learning rate during training. Illuminant estimation networks are optimized using the Angular Error as the loss function, while image-to-image models are trained using the $\Delta E_{76}$ color difference metric [41]. For quantitative evaluation, we compute the $\Delta \mathrm{E}_{00}$ color distance [37] and the Reproduction Error [20].

**Experiments on Aligned Data** We compare the proposed framework against several state-of-the-art illuminant estimation methods, including both statistical and learning-based approaches. Statistical baselines are drawn from the edge-based color constancy framework [49], comprising Gray-World (GW), White-Patch (WP), Shades-of-Gray (SoG), General Gray-World (GGW), and the first and second-order Gray-Edge variants (GE1 and GE2). Learning-based baselines include FC$^4$ [26], ConvMean [23], and QU [8]. Methods that apply post-capture adjustments on already rendered sRGB images [1–3, 43] are excluded from the comparison, as they do not operate within the camera color correction pipeline. To enable a

Table 1. Results on the aligned version of the proposed dataset, aggregated by camera type: mirrorless (Canon R5, Nikon Zf, Sony $\alpha$9 III) and mobile (Google Pixel 3, Huawei Mate 20 Pro, iPhone Xs Max, Samsung Galaxy Note 9). Aggregation is performed by averaging the results of each camera. We highlight best, second best and third best results for each metric. Per-camera results are provided in the Supplementary Material.

| | | $\Delta E_{00} \downarrow$ | | | | | | | | Reproduction Error $\downarrow$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Camera | Method | Mean | Med. | Tri. | B-25 | W-25 | 95-P | 99-P | Max | Mean | Med. | Tri. | B-25 | W-25 | 95-P | 99-P | Max |
| Mirrorless Sensors | GW [49] | 7.80 | 7.15 | 7.28 | 2.81 | 13.63 | 15.92 | 21.89 | 26.15 | 9.42 | 8.07 | 8.24 | 2.75 | 18.91 | 23.89 | 33.06 | 42.69 |
| | WP [49] | 5.08 | 4.23 | 4.57 | 1.89 | 9.62 | 11.09 | 14.60 | 22.60 | 5.93 | 4.47 | 4.90 | 1.43 | 12.73 | 15.64 | 20.78 | 38.99 |
| | GGW [49] | 6.13 | 5.47 | 5.61 | 2.05 | 11.53 | 13.65 | 19.49 | 24.70 | 7.31 | 5.88 | 5.98 | 1.79 | 15.51 | 19.82 | 29.29 | 40.85 |
| | SoG [49] | 5.94 | 5.32 | 5.43 | 2.07 | 11.08 | 13.35 | 17.75 | 30.62 | 7.06 | 5.70 | 5.77 | 1.70 | 15.02 | 19.01 | 26.19 | 45.11 |
| | GE1 [49] | 5.53 | 4.91 | 4.92 | 1.84 | 10.54 | 12.72 | 18.79 | 24.90 | 6.40 | 4.79 | 5.08 | 1.47 | 14.03 | 17.65 | 22.60 | 40.34 |
| | GE2 [49] | 5.49 | 5.05 | 5.04 | 1.99 | 9.99 | 12.17 | 14.82 | 25.03 | 6.35 | 5.00 | 5.29 | 1.49 | 13.66 | 16.58 | 22.37 | 56.92 |
| | Squeezenet-FC$^4$ [26] | 3.28 | 2.86 | 2.97 | 1.45 | 5.82 | 6.72 | 9.90 | 18.64 | 4.38 | 2.78 | 3.17 | 0.92 | 10.57 | 13.68 | 18.37 | 37.72 |
| | ConvMean [23] | 3.55 | 2.98 | 3.09 | 1.43 | 6.67 | 7.89 | 12.34 | 20.58 | 4.61 | 2.91 | 3.35 | 0.95 | 11.13 | 13.90 | 19.58 | 41.24 |
| | QU [8] | 6.37 | 5.73 | 5.79 | 2.31 | 11.70 | 13.90 | 18.84 | 26.52 | 7.88 | 5.88 | 6.36 | 2.03 | 17.04 | 21.42 | 29.43 | 41.75 |
| | QU+ft [8] | 5.30 | 4.65 | 4.76 | 1.95 | 9.89 | 11.90 | 16.62 | 24.42 | 6.60 | 5.12 | 5.35 | 1.72 | 13.99 | 17.51 | 27.04 | 39.77 |
| | SpectralFC$^4$ [26] | 3.25 | 2.79 | 2.88 | 1.35 | 5.98 | 6.90 | 10.66 | 20.42 | 4.31 | 2.73 | 3.14 | 0.88 | 10.45 | 13.01 | 18.74 | 36.91 |
| | SpectralConvMean [23] | 3.17 | 2.73 | 2.83 | 1.33 | 5.77 | 6.56 | 10.50 | 23.40 | 4.24 | 2.70 | 3.05 | 0.87 | 10.27 | 13.03 | 18.47 | 40.77 |
| | LPIENet (Ours) | 1.74 | 1.55 | 1.59 | 0.99 | 2.80 | 3.18 | 4.54 | 14.10 | 3.23 | 1.53 | 2.01 | 0.67 | 8.70 | 10.78 | 17.29 | 38.49 |
| | LPIENet-small (Ours) | 2.09 | 1.88 | 1.92 | 1.26 | 3.29 | 3.73 | 5.41 | 21.10 | 3.57 | 1.91 | 2.35 | 0.88 | 9.09 | 11.24 | 17.56 | 38.05 |
| | cmKAN-light (Ours) | 1.60 | 1.30 | 1.36 | 0.71 | 3.02 | 3.55 | 5.74 | 21.19 | 2.91 | 1.21 | 1.71 | 0.42 | 8.26 | 10.33 | 16.66 | 38.21 |
| Mobile Sensors | GW [49] | 7.71 | 7.06 | 7.19 | 2.79 | 13.49 | 15.92 | 21.45 | 25.87 | 9.48 | 8.16 | 8.35 | 2.86 | 18.76 | 23.31 | 33.47 | 42.85 |
| | WP [49] | 5.04 | 4.20 | 4.50 | 1.95 | 9.46 | 10.96 | 14.84 | 18.75 | 6.16 | 4.76 | 5.10 | 1.59 | 13.02 | 15.68 | 20.83 | 40.11 |
| | GGW [49] | 6.15 | 5.46 | 5.61 | 2.09 | 11.47 | 13.43 | 19.03 | 24.22 | 7.57 | 6.09 | 6.21 | 1.94 | 15.83 | 20.25 | 29.29 | 41.54 |
| | SoG [49] | 5.96 | 5.32 | 5.46 | 2.12 | 11.06 | 13.28 | 17.41 | 22.50 | 7.33 | 5.94 | 6.01 | 1.86 | 15.39 | 19.45 | 27.20 | 41.53 |
| | GE1 [49] | 5.50 | 4.87 | 4.90 | 1.88 | 10.46 | 12.71 | 17.95 | 21.99 | 6.59 | 4.98 | 5.28 | 1.58 | 14.36 | 17.81 | 23.59 | 41.85 |
| | GE2 [49] | 5.44 | 4.91 | 4.97 | 2.01 | 9.90 | 12.35 | 14.91 | 22.28 | 6.50 | 5.04 | 5.36 | 1.55 | 13.83 | 16.62 | 22.29 | 42.06 |
| | Squeezenet-FC$^4$ [26] | 3.16 | 2.75 | 2.83 | 1.40 | 5.63 | 6.65 | 9.47 | 19.76 | 4.51 | 2.87 | 3.29 | 0.98 | 10.80 | 13.63 | 18.71 | 38.39 |
| | ConvMean [23] | 3.54 | 2.95 | 3.07 | 1.44 | 6.68 | 7.98 | 12.29 | 21.16 | 4.76 | 3.05 | 3.49 | 1.03 | 11.40 | 14.16 | 19.53 | 40.20 |
| | QU [8] | 6.37 | 5.80 | 5.87 | 2.31 | 11.65 | 14.29 | 19.28 | 21.94 | 8.10 | 6.34 | 6.71 | 2.09 | 17.17 | 21.39 | 30.17 | 42.07 |
| | QU+ft [8] | 5.41 | 4.79 | 4.87 | 1.97 | 10.04 | 11.87 | 17.33 | 20.87 | 7.03 | 5.55 | 5.80 | 1.85 | 14.70 | 18.53 | 28.58 | 40.69 |
| | SpectralFC$^4$ [26] | 3.19 | 2.77 | 2.85 | 1.39 | 5.76 | 6.68 | 10.29 | 19.80 | 4.38 | 2.76 | 3.17 | 0.96 | 10.53 | 13.38 | 19.05 | 37.70 |
| | SpectralConvMean [23] | 3.16 | 2.73 | 2.82 | 1.39 | 5.68 | 6.43 | 10.45 | 22.38 | 4.35 | 2.86 | 3.21 | 0.98 | 10.28 | 12.77 | 18.76 | 40.70 |
| | LPIENet (Ours) | 1.66 | 1.49 | 1.53 | 0.94 | 2.69 | 3.06 | 4.38 | 12.73 | 3.13 | 1.52 | 1.96 | 0.66 | 8.38 | 10.24 | 16.53 | 37.70 |
| | LPIENet-small (Ours) | 1.84 | 1.64 | 1.68 | 1.07 | 2.96 | 3.37 | 5.00 | 21.20 | 3.28 | 1.66 | 2.11 | 0.75 | 8.59 | 10.60 | 16.68 | 37.43 |
| | cmKAN-light (Ours) | 1.47 | 1.19 | 1.24 | 0.65 | 2.79 | 3.30 | 5.61 | 20.78 | 2.83 | 1.13 | 1.62 | 0.40 | 8.15 | 10.25 | 16.35 | 37.56 |

Table 2. Results on the misaligned version of the proposed dataset. Aggregation by camera type, metric reporting, and highlighting follow the same format as in Table 1. Our proposed models are adapted to the new input by fine-tuning their spectral encoder, while MS-driven baselines are completely retrained. RGB-only methods are omitted as they are unaffected by spatial misregistration. Per-camera results are provided in the Supplementary Material.

| | | $\Delta E_{00} \downarrow$ | | | | | | | | Reproduction Error $\downarrow$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Camera | Method | Mean | Med. | Tri. | B-25 | W-25 | 95-P | 99-P | Max | Mean | Med. | Tri. | B-25 | W-25 | 95-P | 99-P | Max |
| Mirrorless Sensors | SpectralFC$^4$ [26] | 3.25 | 2.79 | 2.88 | 1.35 | 5.93 | 6.91 | 10.47 | 23.17 | 4.34 | 2.73 | 3.13 | 0.90 | 10.47 | 13.53 | 18.61 | 39.57 |
| | SpectralConvMean [23] | 3.18 | 2.78 | 2.86 | 1.34 | 5.72 | 6.61 | 10.00 | 20.13 | 4.26 | 2.71 | 3.05 | 0.89 | 10.30 | 13.21 | 18.24 | 40.76 |
| | LPIENet (Ours) | 1.84 | 1.65 | 1.69 | 1.02 | 2.99 | 3.46 | 4.66 | 13.41 | 3.39 | 1.74 | 2.22 | 0.74 | 8.92 | 10.81 | 17.04 | 38.94 |
| | LPIENet-small (Ours) | 2.34 | 2.14 | 2.17 | 1.40 | 3.66 | 4.17 | 5.78 | 19.95 | 3.86 | 2.21 | 2.67 | 1.02 | 9.56 | 11.73 | 17.44 | 39.42 |
| | cmKAN-light (Ours) | 1.84 | 1.51 | 1.58 | 0.82 | 3.44 | 4.07 | 6.26 | 20.93 | 3.20 | 1.50 | 1.96 | 0.50 | 8.75 | 11.10 | 16.99 | 39.55 |
| Mobile Sensors | SpectralFC$^4$ [26] | 3.23 | 2.74 | 2.83 | 1.43 | 5.88 | 6.92 | 10.50 | 22.05 | 4.54 | 2.97 | 3.31 | 1.01 | 10.77 | 13.79 | 19.71 | 39.29 |
| | SpectralConvMean [23] | 3.12 | 2.69 | 2.78 | 1.37 | 5.59 | 6.45 | 9.95 | 20.87 | 4.31 | 2.82 | 3.15 | 0.97 | 10.21 | 12.94 | 18.33 | 40.69 |
| | LPIENet (Ours) | 1.83 | 1.62 | 1.67 | 1.00 | 3.01 | 3.48 | 4.85 | 13.06 | 3.36 | 1.70 | 2.18 | 0.75 | 8.85 | 10.74 | 17.01 | 38.41 |
| | LPIENet-small (Ours) | 2.25 | 2.03 | 2.07 | 1.31 | 3.58 | 4.10 | 5.81 | 20.45 | 3.77 | 2.12 | 2.58 | 0.99 | 9.39 | 11.60 | 17.39 | 38.41 |
| | cmKAN-light (Ours) | 1.75 | 1.43 | 1.50 | 0.78 | 3.29 | 3.88 | 6.31 | 21.45 | 3.14 | 1.41 | 1.89 | 0.50 | 8.67 | 11.14 | 16.88 | 37.89 |

fair comparison in multispectral settings, we adapt the FC$^4$ and ConvMean architectures to process MS data instead of RGB inputs. Since all these baselines address only the first stage of the color correction pipeline (i.e., illuminant estimation), we complete the remaining steps following the traditional correction procedure described in Section 2. Quantitative results are summarized in Table 1, where the $\Delta E_{00}$ color distance and the Reproduction Error are presented with their statistical distribution over the dataset, including the mean, median, trimean, best-quartile (B-25), worst-quartile (W-25), 95th and 99th percentiles, and maximum values. Results are aggregated by camera type (mirrorless and mobile) by averaging the statistics of individual cameras, while the full per-camera results are reported in the

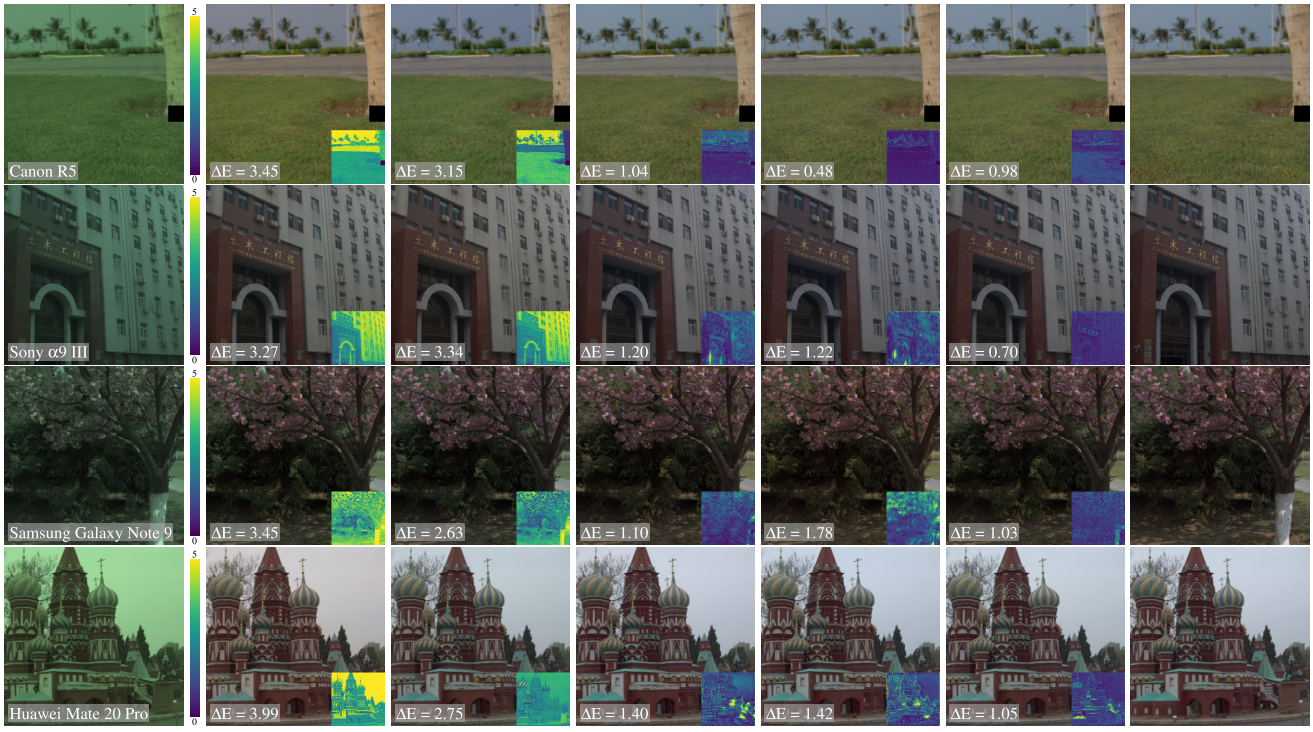



Figure 4. Qualitative results of the best performing methods: FC$^4$ [26], SpectralConvMean [23], and our three proposed models. We show results from two mirrorless and two mobile cameras, with $\Delta E_{00}$ map reported in the bottom-right corner and average $\Delta E_{00}$ reported in the bottom-left one. For visualization purposes, we gamma correct the raw visualizations (first column) and convert the results (columns two to six) to sRGB. More qualitative results are reported in the Supplementary Material.

Supplementary Material. Our framework consistently outperforms both RGB-only and MS-driven baselines across camera categories. It is noteworthy that the proposed models achieve an average $\Delta E_{00}$ that is ~50% lower than that of competing methods, whose average performance is comparable to the worst-quartile (W-25) results of our approaches. Qualitative comparisons, reported in Figure 4, further corroborate these findings, demonstrating that our framework produces visually more accurate and perceptually consistent results across diverse scenes, illumination conditions, and cameras.

**Experiments on Misaligned Data** We further assess the adaptability of our proposed framework under realistic acquisition conditions. To this end, we also conduct an experiment using the misaligned dataset described in Section 4. This experiment simulates the spatial inconsistencies that typically arise in RGB–MS imaging systems due to imperfect optical alignment. For our proposed architectures, the models are fine-tuned on the misaligned data by keeping all weights frozen except for the spectral encoder branch, which is responsible for extracting features from the MS modality. This strategy allows the network to adjust its spectral representation to the new, geometrically perturbed inputs without altering the overall color correction mapping. For the MS variants of ConvMean and FC$^4$, which do not include an explicit spectral encoder, we retrain the models from scratch following the same procedure described for the previous experiment. Methods relying solely on RGB inputs are unaffected and thus excluded from this experiment. Table 2 summarizes the results obtained on the misaligned dataset, aggregated by camera type (mirrorless and mobile) following the same procedure described for Table 1. Full per-camera results are reported in the Supplementary Material. As shown, our proposed models exhibit only a minor decrease in accuracy relative to the aligned case, remaining the top-performing methods by a significant margin and confirming their adaptability to realistic cross-sensor misalignments. As for the multispectral illuminant estimation baselines, their performance remains comparable to the results reported in Table 1, since their architectures strongly downsample the input and thus mitigate the effect of spatial misregistration.

**Ablation 1: Impact of Multispectral Information** To assess the contribution of the multispectral information in our framework, we conduct an ablation study by training all models without their spectral encoder branch. This con-

figuration effectively reduces the input to RGB-only data, allowing us to isolate the impact of the multispectral modality on overall performance. The results, summarized in Table 3, highlight that our proposed modifications for integrating multispectral cues are the key factor driving the superior performance of these architectures in the color correction task, yielding improvements of up to 50% over the RGB-only setup. This confirms that leveraging complementary spectral information substantially enhances color correction accuracy.

**Ablation 2: Robustness to Exposure Changes** One of the key advantages of traditional color correction pipelines lies in their inherent exposure invariance, a consequence of their linear formulation. Our approach, in contrast, relies on non-linear transformations and therefore does not possess this property by design. To evaluate the extent to which exposure variations affect performance, we simulate different exposure levels by multiplying both the input and ground-truth images by a factor $\alpha$. We consider two reduced exposure levels, $\alpha = 0.75$ and $\alpha = 0.5$, and report the corresponding results in Table 4. For this experiment, we focus on Reproduction Error, as $\Delta E_{00}$ decreases at lower brightness levels due to the way it is mathematically formulated. As expected, the Reproduction Error slightly increases at lower exposure levels, indicating a modest degradation in performance. Nonetheless, the models maintain higher accuracy than all baselines built on the traditional color correction pipeline, including the spectral variant of ConvMean [23], which represents the strongest baseline.

**Limitations** Our method aims at directly providing the best final color-corrected estimate given an input non-corrected image. However, the end-to-end nature of our approach does not rely on explicit intermediate estimates such as illuminant or reflectance. Consequently, it offers less interpretability than modular pipelines, which can provide diagnostic or reusable intermediate outputs. Nonetheless, in applications where the primary goal is achieving the highest color accuracy, our end-to-end approach remains advantageous.

## 6. Conclusion

We presented a unified framework for end-to-end color correction that jointly leverages high-resolution RGB data and low-resolution multispectral information. Unlike prior works that treat the stages of the color correction pipeline independently, our approach integrates them within a single learning-based model. To enable training and evaluation under physically consistent conditions, we generated a new dataset by rendering hyperspectral reflectance data into realistic RGB and MS images, including a second version with spatial misalignments that replicate real-world acquisition imperfections. We demonstrated the flexibility of our framework by refactoring two state-of-the-art image-to-image architectures to operate within our dual-input setting. Extensive experiments demonstrate that the proposed framework consistently outperforms existing RGB-only and MS-driven approaches across multiple camera models. The results confirm that jointly learning illuminant estimation, discounting, and color transformation, while exploiting multispectral cues, leads to substantial gains in both color accuracy and stability. By considering two distinct architectures and performing comprehensive ablations, we show that these improvements stem primarily from the proposed framework rather than the specific network design. This suggests that the framework is not tied to a specific architecture: any future backbone, if more powerful, can be seamlessly integrated within the same formulation to further enhance performance. Furthermore, the proposed architectures maintain strong performance under challenging conditions such as exposure variations and spatial misalignments, validating their robustness for practical deployment. We anticipate that our framework and dataset will support further advancement in color correction research.

Table 3. Impact of adding multispectral information to the proposed framework. We report mean $\Delta E_{00}$ and Reproduction Error obtained with (w/) and without (w/o) MS data, showing the increase of performance when spectral data is used.

| | | Mean $\Delta E_{00}$ ↓ | | Mean Repr. Error ↓ | |
|---|---|---|---|---|---|
| Camera | Method | w/o MS data | w/ MS data | w/o MS data | w/ MS data |
| Mirrorless Sensors | LPIENet | 3.23 | 1.74 | 4.71 | 3.23 |
| | LPIENet-small | 3.33 | 2.09 | 4.85 | 3.56 |
| | cmKAN-light | 3.06 | 1.60 | 4.40 | 2.92 |
| Mobile Sensors | LPIENet | 3.29 | 1.66 | 4.71 | 3.11 |
| | LPIENet-small | 3.48 | 1.80 | 4.84 | 3.27 |
| | cmKAN-light | 3.03 | 1.49 | 4.38 | 2.84 |

Table 4. Mean Reproduction Error for different exposure levels, simulated by scaling both input and ground-truth images by a factor $\alpha$. Although Reproduction Error slightly increases for lower exposures, all models remain robust and outperform the other baselines. We report SpectralConvMean [23] as the representative baseline, since it is the best-performing one.

| | | Mean Reproduction Error ↓ | | |
|---|---|---|---|---|
| Camera | Method | $\alpha = 1$ | $\alpha = 0.75$ | $\alpha = 0.5$ |
| Mirrorless Sensors | SpectralConvMean [23] | 4.24 | 4.25 | 4.30 |
| | LPIENet (Ours) | 3.23 | 3.41 | 3.84 |
| | LPIENet-small (Ours) | 3.57 | 3.70 | 4.04 |
| | cmKAN-light (Ours) | 2.91 | 2.98 | 3.18 |
| Mobile Sensors | SpectralConvMean [23] | 4.35 | 4.35 | 4.36 |
| | LPIENet (Ours) | 3.11 | 3.28 | 3.73 |
| | LPIENet-small (Ours) | 3.27 | 3.38 | 3.67 |
| | cmKAN-light (Ours) | 2.84 | 2.90 | 3.08 |

## Acknowledgments

JVC acknowledges Grants PID2021-128178OB-I00 and PID2024-162555OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF "A way of making Europe", the Generalitat de Catalunya CERCA Program, and the 2025 Leonardo Grant for Scientific Research and Cultural Creation from the BBVA Foundation. The BBVA Foundation accepts no responsibility for the opinions, statements and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors. This work was partially supported by the MUR under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy.

## References

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

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

[3] Mahmoud Afifi, Marcus A Brubaker, and Michael S Brown. Auto white-balance correction for mixed-illuminant scenes. In *WACV*, 2022. 2, 5

[4] Mahmoud Afifi, Luxi Zhao, Abhijith Punnappurath, Mohamed A Abdelsalam, Ran Zhang, and Michael S Brown. Time-aware auto white balance in mobile photography. In *ICCV*, 2025. 2

[5] Kobus Barnard, Lindsay Martin, Brian V. Funt, and Adam Coath. A data set for color research. *Color Research & Application*, 27:147–151, 2002. 4

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

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

[8] Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy. In *CVPR*, 2019. 2, 5, 6

[9] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Color constancy using cnns. In *CVPR*, 2015. 2

[10] 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. 4

[11] Dongliang Cheng, Brian Price, Scott Cohen, and Michael S Brown. Beyond white: Ground truth colors for color constancy correction. In *ICCV*, 2015. 3

[12] Marcos V Conde, Florin Vasluianu, Javier Vazquez-Corral, and Radu Timofte. Perceptual image enhancement for smartphone real-time applications. In *WACV*, 2023. 2, 3, 4

[13] Zhuoran Du, Shaodi You, Cheng Cheng, and Shikui Wei. Automatic spectral calibration of hyperspectral images: Method, dataset and benchmark. In *CVPR*, 2025. 4

[14] Ilaria Erba, Marco Buzzelli, and Raimondo Schettini. Rgb color constancy using multispectral pixel information. *JOSA A*, 41(2):185–194, 2024. 2

[15] Ilaria Erba, Marco Buzzelli, Jean-Baptiste Thomas, Jon Y. Hardeberg, and Raimondo Schettini. Improving rgb illuminant estimation exploiting spectral average radiance. *JOSA A*, 41 3:516–526, 2024. 2

[16] 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. 4

[17] Mark D Fairchild. *Color appearance models*. John Wiley & Sons, 2013. 2, 3

[18] Graham D Finlayson, Mark S Drew, and Brian V Funt. Diagonal transforms suffice for color constancy. In *ICCV*, 1993. 3

[19] Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. Color correction using root-polynomial regression. *IEEE TIP*, 24(5):1460–1470, 2015. 3

[20] Graham D Finlayson, Roshanak Zakizadeh, and Arjan Gijsenij. The reproduction angular error for evaluating the performance of illuminant estimation algorithms. *IEEE TPAMI*, 39(7):1482–1488, 2016. 5

[21] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *CVPR*, 2008. 4

[22] Ortal Glatt, Yotam Ater, Woo-Shik Kim, Shira Werman, Oded Berby, Yael Zini, Shay Zelinger, Sangyoon Lee, Heejin Choi, and Evgeny Soloveichik. Beyond rgb: a real world dataset for multispectral imaging in mobile devices. In *WACV*, 2024. 2, 4

[23] Han Gong. Convolutional mean: A simple convolutional neural network for illuminant estimation. In *BMVC*, 2019. 5, 6, 7, 8

[24] 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. 3

[25] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 3

[26] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *CVPR*, 2017. 1, 2, 5, 6, 7

[27] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *CVPR Workshops*, 2020. 5

[28] ISO. 22028-2:2013 photography and graphic technology — extended colour encodings for digital image storage, manipulation and interchange-part 2: Reference output medium metric rgb colour image encoding (romm rgb). *International Organization for Standardization*, 2013. 1, 3

[29] Haris Ahmad Khan, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Olivier Laligant. Illuminant estimation in multispectral imaging. *JOSA A*, 34(7):1085–1098, 2017. 2

[30] Diederik P Kingma. Adam: A method for stochastic optimization. In *ICLR*, 2015. 5

[31] Samu Koskinen, Erman Acar, and Joni-Kristian Kämäräinen. Single pixel spectral color constancy. *IJCV*, 132(2):287–299, 2024. 2

[32] Pierre-Jean Lapray, Xingbo Wang, Jean-Baptiste Thomas, and Pierre Gouton. Multispectral filter arrays: Recent advances and practical implementation. *Sensors*, 14(11): 21626–21659, 2014. 2

[33] Jiacheng Li, Chang Chen, Xue Hu, Fenglong Song, Youliang Yan, and Zhiwei Xiong. Multi-spectral image color reproduction. In *WACV*, 2025. 2

[34] Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *ICCV*, 2021. 4

[35] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks. In *ICLR*, 2025. 4

[36] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. In *ICLR*, 2017. 5

[37] M Ronnier Luo, Guihua Cui, and Bryan Rigg. The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research & Application*, 26(5):340–350, 2001. 5

[38] Calvin S McCamy, Harold Marcus, James G Davidson, et al. A color-rendition chart. *Journal of Applied Photographic Engineering*, 2(3):95–99, 1976. 3

[39] Artem Nikonorov, Georgy Perevozchikov, Andrei Korepanov, Nancy Mehta, Mahmoud Afifi, Egor Ershov, and Radu Timofte. Color matching using hypernetwork-based kolmogorov-arnold networks. In *ICCV*, 2025. 2, 3, 4

[40] International Commission on Illumination (CIE). Colour-matching functions of cie 1931 standard colorimetric observer. https://cie.co.at/datatable/cie-1931-colour-matching-functions-2-degree-observer, 2019. 1, 3

[41] Alan R Robertson. The cie 1976 color-difference formulae. *Color Research & Application*, 2(1):7–11, 1977. 5

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 3

[43] David Serrano-Lozano, Aditya Arora, Luis Herranz, Konstantinos G Derpanis, Michael S Brown, and Javier Vazquez-Corral. Revisiting image fusion for multi-illuminant white-balance correction. In *ICCV*, 2025. 2, 5

[44] Spectricity — Multispectral Imaging Solutions. https://spectricity.com/. [Accessed 17-11-2025]. 4

[45] Haonan Su, Cheolkon Jung, and Long Yu. Multi-spectral fusion and denoising of color and near-infrared images using multi-scale wavelet analysis. *Sensors*, 21(11):3610, 2021. 2

[46] Jean-Baptiste Thomas, Pierre-Jean Lapray, and Steven Le Moan. Trends in snapshot spectral imaging: systems, processing, and quality. *Sensors*, 25(3):675, 2025. 2

[47] Shulin Tian, Yufei Wang, Renjie Wan, Wenhan Yang, Alex C Kot, and Bihan Wen. Enhancing low-light images using infrared encoded images. In *ICIP*, 2023. 2

[48] Fabrizio Vagni. Survey of hyperspectral and multispectral imaging technologies (etude sur les technologies d'imagerie hyperspectrale et multispectrale). Technical report, 2007. 2

[49] Joost Van De Weijer and Theo Gevers. Color constancy based on the grey-edge hypothesis. In *ICIP*, 2005. 2, 5, 6

[50] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV*, 2018. 3

[51] Kailai Zhou, Lijing Cai, Yibo Wang, Mengya Zhang, Bihan Wen, Qiu Shen, and Xun Cao. Joint rgb-spectral decomposition model guided image enhancement in mobile photography. In *ECCV*, 2024. 2