RL-AWB: Deep Reinforcement Learning for Auto White Balance Correction in Low-Light Nighttime Scenes

Yuan-Kang Lee[1,3*], Kuan-Lin Chen[2*], Chia-Che Chang[1], and Yu-Lun Liu[3]

[1] MediaTek Inc., Taiwan [2] National Taiwan University, Taiwan
[3] National Yang Ming Chiao Tung University, Taiwan
r12942062@ntu.edu.tw, yulunliu@cs.nycu.edu.tw

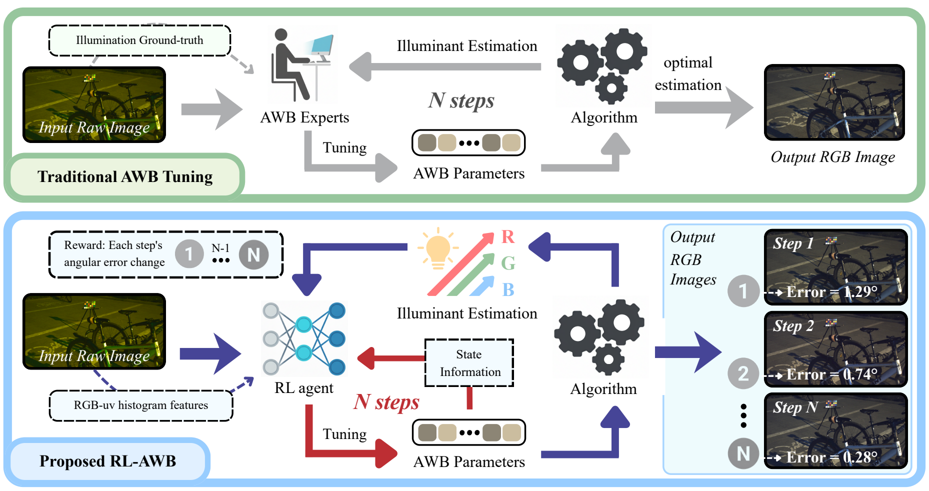


Fig. 1: Our method achieves optimal parameter tuning for automatic white balance (AWB) in nighttime scenes through a hybrid combination of a novel statistical color constancy algorithm and reinforcement learning. Due to the complex lighting conditions in night scenes and the presence of significant noise in the images, traditional AWB tuning faces relatively difficult and time-consuming challenges when adapting parameters for night scene images. Our method can optimize parameters for different night scene images at a faster speed without requiring prior knowledge of illumination ground-truth, and has better cross-sensor generalization advantages.

Abstract. Nighttime color constancy still remains a challenging problem in computational photography due to low-light noise and complex illumination conditions. We present RL-AWB, a novel framework combining statistical methods with deep reinforcement learning for nighttime white balance. Our method begins with a statistical algorithm tailored for nighttime scenes, integrating salient gray pixel detection with novel illuminant estimation. Building on this foundation, we develop the first deep reinforcement learning approach for color constancy

* The first two authors contributed equally to this work.

that leverages the statistical algorithm as its core, mimicking professional AWB tuning experts by dynamically determining image-specific parameters at inference time, without requiring ground-truth illuminants or reference images. To further facilitate cross-sensor evaluation, we introduce the first multi-sensor nighttime dataset. Experiment results demonstrate that our method achieves strong generalization capability across low-light and well-illuminated images. Project page: `https://ntuneillee.github.io/research/rl-awb/`

Keywords: Auto White Balance · Color Constancy · Reinforcement Learning · Nighttime Imaging · Low-light

1 Introduction

Auto White Balance (AWB) is a fundamental component of camera image signal processing (ISP) pipelines that estimates scene illuminant and corrects color casts, ensuring white objects appear neutral across varying lighting conditions [18, 36]. While existing methods achieve robust performance in well-lit daytime scenarios [1, 11, 22], nighttime environments present a fundamentally different challenge. Low illumination, high ISO settings, and severe chromatic noise violate the statistical assumptions underlying conventional AWB algorithms [87, 101], leading to highly unstable illuminant estimates. This instability is further exacerbated in cross-sensor deployment, where the same model often produces significant and unpredictable color shifts across different camera sensors and ISP configurations [1, 28]. Addressing robust cross-sensor nighttime AWB is therefore critical for real-world applications, including mobile photography, surveillance systems, and automotive imaging.

The core difficulty of nighttime AWB stems from the breakdown of reliable color statistics. Statistical methods [18, 36, 101] assume sufficient scene diversity and stable gray pixel detection, which fail under extreme low-light conditions where sensor noise dominates signal [27]. Deep learning approaches [11, 22, 51], while effective in daytime scenarios, require extensive labeled nighttime training data and suffer from severe generalization degradation when deployed on unseen camera sensors [1, 55]. Furthermore, nighttime scenes exhibit heightened sensitivity to algorithmic parameters: small variations in parameter selection can lead to dramatically different illuminant estimates.

To address these challenges, we present RL-AWB, the *first* framework that integrates reinforcement learning into automatic white balance for nighttime color constancy. Our approach fundamentally differs from existing paradigms by formulating AWB as a sequential decision-making problem, where an RL agent learns adaptive parameter selection policies for a novel statistical illuminant estimator. This hybrid design preserves the interpretability and sensor-agnostic nature of statistical methods [36, 87] while gaining the adaptive capability of learning-based approaches [1, 11], all with minimal training data requirements.

In summary, we make the following contributions:

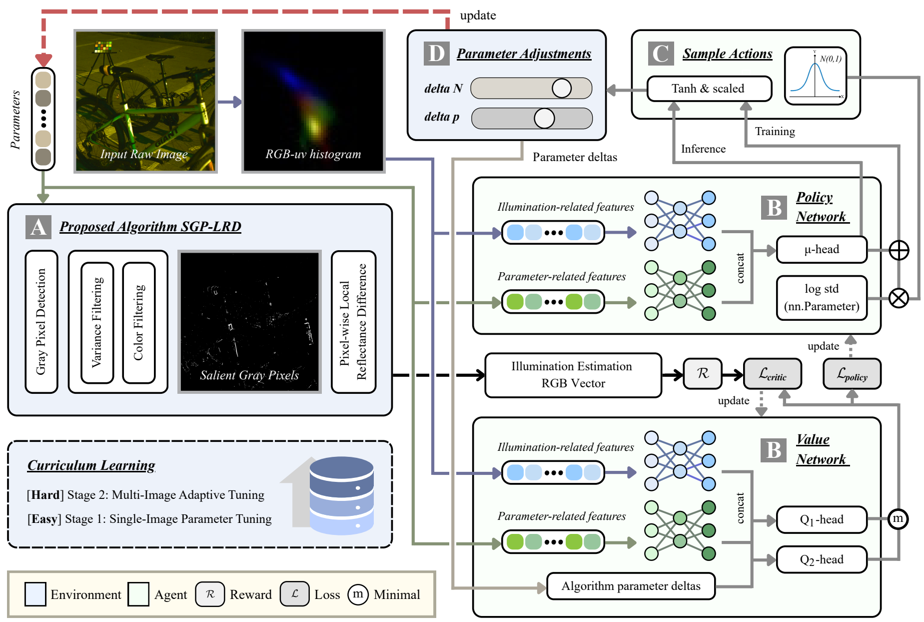


Fig. 2: Overview of the proposed RL-AWB framework. (A) Given an input image, the proposed nighttime color constancy algorithm SGP-LRD estimates the scene illuminant conditioned on two hyperparameters (gray-pixel sampling percentage N and Minkowski order p). (B) A SAC agent selects parameter updates based on image statistics and current AWB settings. (C) The policy outputs one action per parameter; actions are sampled, squashed by tanh to $[-1, 1]$, and rescaled to valid ranges. (D) The rescaled actions update the two hyperparameters and are applied to SGP-LRD to produce the illuminant estimate. Repeat until the termination criterion is met.

- We develop SGP-LRD (Salient Gray Pixels with Local Reflectance Differences), a nighttime-specific color constancy algorithm that achieves state-of-the-art illuminant estimation on public nighttime benchmarks.
- We design the RL-AWB framework with Soft Actor-Critic training and two-stage curriculum learning, enabling adaptive per-image parameter optimization with exceptional data efficiency.
- We contribute LEVI, the first multi-camera nighttime dataset comprising 700 images, enabling rigorous cross-sensor color constancy evaluation.
- With only 5 training images per dataset, RL-AWB generalizes across unseen sensors better than fully-supervised state-of-the-art methods, demonstrating high data efficiency.

2 Related Work

Statistical Color Constancy. Statistical methods exploit scene statistics via achromatic averages (Gray World [18]), maximum responses (Max-RGB) rooted in

Retinex theory [63], unified by Minkowski norms [36]. Gamut mapping methods [34,38,41] constrain feasible illuminants using observed colors, while Bayesian approaches [17, 40] formalize probabilistic priors over illuminant distributions. Edge-based approaches [101] use derivatives with photometric weighting [43] and local statistics [33]. Corrected-moment estimation [35] and spatio-spectral statistics [21] further improve classical estimators. Recent advances include grayness indices [87], FFT acceleration (700 FPS) [12], spectral methods [59], attention mechanisms [62], and multi-scale spatial statistics [100]. For comprehensive evaluations, we refer readers to [10,42]. *Unlike* these methods using *fixed* parameters or heuristics, we employ reinforcement learning to *automatically* learn optimal parameter adjustments for varying nighttime conditions, combining statistical efficiency with adaptive optimization.

Learning-Based Color Constancy. Deep learning approaches feature CNNs with log-chrominance localization and confidence pooling [11, 51], enhanced by cascades [117], contrastive learning [75], and adaptation [1]. Earlier works include CNN-based illuminant classification [15,83], multi-illuminant CNN estimation [16], deep specialized networks with multi-hypothesis selection [91], and metric learning formulations [112]. Multi-hypothesis approaches [49] further enable camera-agnostic probabilistic estimation. GAN-based methods [32, 92] address multi-illuminant scenarios through domain translation. Quasi-unsupervised training [14] and few-shot meta-learning [78] reduce label requirements. Recent diffusion methods [22] achieve 4.32°–5.22° worst-25% error via inpainting, while self-supervised [85] and uncertainty-aware [19] approaches further reduce supervision. Extensions address multi-illuminant [69], post-editing [3], mixed lighting [4], style-based AWB [57], and brightness [26, 111]. Simple features with ensemble learning [29] bridge statistical and deep learning paradigms. *In contrast to* these *data-hungry* approaches requiring extensive labeled nighttime data, our method achieves cross-dataset generalization through *sample-efficient* RL tuning of interpretable statistical parameters, combining both paradigms.

Nighttime and Low-Light Color Constancy. Nighttime scenes present unique challenges that fundamentally violate daytime color constancy assumptions: mixed illumination from multiple light sources, severely low light levels, and elevated sensor chroma noise [23]. Early methods [24, 46, 108] targeted brightness. Recent work employs transformers [20,105,113], adaptive masking [68], synthetic data [86], noise-resistant detection [27], and joint restoration [76, 109, 116, 128]. Diffusion-based approaches [52, 106] and Fourier-domain methods [65, 102] offer new paradigms for low-light enhancement, while novel color spaces [114] address color handling under extreme darkness. Nighttime rendering benchmarks [70, 71] further establish standardized evaluation protocols. For a comprehensive survey, see [66]. However, these methods rely on *pseudo-labels* with error propagation or *fixed* parameters. We formulate nighttime AWB as *sequential decision-making*, where RL learns *dynamic* parameter adjustment strategies.

Reinforcement Learning for ISP. RL enables adaptive ISP policies for exposure [50,120], color enhancement [84], rapid convergence [64] (5 frames, 1ms), and artifact filtering [9]. The seminal RL-Restore [118] introduced toolchain selection for image restoration via DQN. Applications span pixel-wise correction [39], software parameters [60], personalization [115,127], sequential optimization [94], and module selection [107,125]. Our framework adopts the Soft Actor-Critic algorithm [47,48] for its sample-efficient off-policy updates and automatic entropy tuning. These works validate RL for ISP but focus on *daytime* scenarios with *camera settings* as actions. We extend to *nighttime* AWB where actions control *statistical algorithm parameters*, requiring noise-aware rewards and robust optimization.

Cross-Sensor Generalization. Cross-sensor generalization remains a challenge in color constancy, as different camera sensors exhibit varying spectral responses and ISP characteristics [72]. Methods evolved from dataset evaluation [28] through fine-tuning [1], sensor-independent representations [2], embeddings [55], to domain-invariant learning [96,126]. Multi-camera benchmarks [56] establish standardized cross-sensor evaluation. Calibration-light methods use dual-mapping [122] (single D65), self-supervision [31], multi-domain architectures [110], and HDR [5]. Test-time adaptation [54, 73, 95, 103] enables training-free deployment, with recent advances in continual [104] and improved [25] test-time strategies. *Unlike* methods requiring *test-time* multi-image access or *camera-specific* calibration, we achieve generalization through (1) *sensor-agnostic* statistical algorithms and (2) curriculum learning exposing agents to diverse conditions, enabling *few-shot* inference on unseen cameras.

Hybrid Statistical-Learning Approaches. Hybrid methods combine classical algorithms with deep learning. Algorithm unrolling, pioneered by LISTA [45] and surveyed in [79], transforms iterative optimization into trainable networks, as demonstrated for image restoration [81,123,124]. Differentiable ISP pipelines [80, 98,119] enable end-to-end optimization of classical processing modules. Other hybrid strategies include learned parameter dictionaries [30], energy landscapes [8], and hyperparameter prediction [67]. Our work follows this paradigm, combining *statistically-grounded* illuminant estimation with *RL-based* parameter optimization and preserving interpretability.

Curriculum Learning in RL. Curriculum learning, introduced by Bengio *et al.* [13] and extended via self-paced learning [53,61], improves training efficiency through progressive difficulty. In the RL setting, survey works [82, 93] systematize task sequencing strategies, including teacher–student approaches [77], reverse curriculum generation [37, 97], self-paced deep RL [58], automated task selection [44], and strategic sampling [7,89]. Our two-stage approach: (1) single-instance stabilization addresses cold starts, (2) cyclic multi-instance balances exploration and stability.

3 Method

3.1 Nighttime Color Constancy Algorithm

Salient Gray Pixel Detection. Under the narrow-band spectral assumption, the linear image $\mathbf{I}$ is modeled as $\mathbf{I} = \mathbf{W} \cdot \mathbf{L} + \delta$, where $\mathbf{W}$ is the white-balanced image, $\mathbf{L}$ is illuminant, and δ is noise. Neglecting δ and applying log-transform followed by local contrast operator $C\{\cdot\}$ (Laplacian of Gaussian) yields:

$$C\{\log(I_i^{(x,y)})\} \approx C\{\log(W_i^{(x,y)})\}. \tag{1}$$

This shows local contrast depends solely on surface reflectance. Let $\Delta_i(x, y)$ denote the local contrast at (x, y) for channel $i \in \{R, G, B\}$. Following Qian *et al.* [88], achromatic pixels have contrast vectors $\Delta(x, y) = [\Delta_R, \Delta_G, \Delta_B]^{\mathrm{T}}$ aligned with the gray direction $\mathbf{g} = [1, 1, 1]^{\mathrm{T}}$. Grayness is measured as:

$$G(x, y) = \cos^{-1}\left(\frac{\langle \Delta_{(x,y)}, \mathbf{g}\rangle}{\|\Delta_{(x,y)}\|\|\mathbf{g}\|_2}\right). \tag{2}$$

The top $N\%$ pixels ranked by $G(x, y)$ are selected as gray candidates. We then apply a two-layer filtering process to refine this candidate set, mitigating the adverse effects of noise and chromatic outliers prevalent in low-light imagery:

Local Variance Filtering. For each pixel in the initial detected gray pixel mask, we compute the variance across the logarithmic RGB channels. Pixels where the intra-pixel variance is too small often lack reliable color information. By applying a lower bound threshold, VarTh, we filter out these unreliable candidates.

Color Deviation Filtering. The second stage filters out pixels that are too distant from the dominant color cast of the scene's illuminant. We first compute the mean logarithmic intensity for each channel of the image: $\mathbf{M} = [\bar{M}_R, \bar{M}_G, \bar{M}_B]^{\mathrm{T}}$, and then calculate the maximum absolute color deviation, $X(x, y)$, for each pixel from this mean. We define a threshold $T_C = \text{ColorTh} \cdot \min(\mathbf{M})$. Any initial detected gray pixel exceeding this deviation is removed.

Following these two refinement layers to the initial detected gray pixels, we obtain the Salient Gray Pixels (SGPs).

Gray-pixel Confidence Weighting. Let $R(i, j)$, $G(i, j)$, and $B(i, j)$ denote the normalized pixel intensities at location (i, j) in the three channels, respectively. The luminance map for SGPs is then computed as $LM(i, j) = (R(i, j)+G(i, j)+B(i, j))/3$. The skewness value s_{LM} quantifies the asymmetry of the brightness distribution and guides the selection of an adaptive exponent parameter E. We set $E = 1.0$ for highly skewed distributions ($s_{LM} > 1.5$), $E = 2.0$ for moderate skewness ($0.2 < s_{LM} \leq 1.5$), and $E = 4.0$ for uniform illumination ($s_{LM} \leq 0.2$). Let $\overline{LM}$ denote the mean luminance of non-zero SGPs. The gray-pixel confidence weight is then computed as:

$$W_{SGP}(i, j) = 1 - \exp\left[-\left(\frac{LM(i, j)}{\overline{LM}}\right)^E\right]. \tag{3}$$

Pixel-wise Local Reflectance Difference. For each position (i,j) in an image, a local window $\mathcal{W}_{i,j}$ of size 3×3 is centered at that pixel. Let $f_c(x)$ denote the detected SGP intensity for channel c, and $\mathcal{W}^*_{i,j}$ represent the set of non-zero pixels within $\mathcal{W}_{i,j}$. The pixel-wise normalized local reflectance difference at position (i,j) is then defined as:

$$N_c^{i,j} = \frac{\left(\sum_{x \in \mathcal{W}^*_{i,j}} f_c(x) \right) / \, |\mathcal{W}^*_{i,j}|}{\max_{x \in \mathcal{W}_{i,j}} f_c(x)}. \tag{4}$$

For pixels where the maximum value is zero, we set $N_c^{i,j} = 0$ to exclude invalid regions from the computation. Finally, the illuminant is estimated as:

$$\hat{e}_c = \left(\frac{\sum_{\Omega} \left(\mu_c^{i,j} \cdot W_{SGP}(i,j) \right)^p}{\sum_{\Omega} \left(N_c^{i,j} \cdot W_{SGP}(i,j) \right)^p} \right)^{1/p}, \tag{5}$$

where Ω represents all valid pixel positions ($N_c^{i,j} > 0$), μ is the mean intensity of SGPs in $W_{i,j}$, and p is the Minkowski norm parameter. The numerator accumulates weighted SGP intensity, and the denominator accumulates weighted normalized local differences.

Our proposed algorithm, SGP-LRD (Salient Gray Pixels with Local Reflectance Differences), addresses nighttime color constancy through three key design principles:

- **Reliability amplification:** Spatially coherent gray regions are sampled repeatedly across overlapping windows, naturally amplifying high-SNR grayness signals while suppressing isolated spurious pixels.
- **Implicit noise filtering:** While spurious pixels from sensor noise appear in limited windows with minimal contribution, genuine gray regions exhibit consistent responses across neighboring windows, naturally distinguishing signal from noise through spatial redundancy.
- **Spatial prior exploitation:** The overlapping design encodes the natural prior that reliable achromatic surfaces exhibit spatial continuity, ensuring that illuminant estimates are dominated by high-confidence information.

3.2 RL-AWB Framework

Algorithm Parameters. Two parameters critically determine the performance of our SGP-LRD: the gray pixel candidate selection threshold $N\%$ and the Minkowski norm exponent p. Lower $N\%$ for gray-rich scenes (higher purity); raise it when gray cues are sparse (better coverage). Small p yields near-uniform weighting; large p emphasizes high-confidence pixels; it is helpful when detections are reliable but brittle in ambiguous/low-light scenes. In low-light nighttime images, the optimal (N,p) configuration is inherently scene-dependent. We address this challenge through a reinforcement learning framework that learns to adaptively select algorithm parameters based on scene characteristics.

State Design. Our state design differs from RL-AE control [64]: the ground-truth illuminant is unavailable at deployment in RL-AWB and thus cannot appear in the state. We therefore encode rich chromatic statistics without privileged labels and add a compact history descriptor for recent adjustments.

- **Illumination-related Features.** Following prior works [2, 6, 11], we represent an image with a log-chrominance (RGB-uv) histogram $\mathbf{H} \in \mathbb{R}^{m \times m \times 3}$ (granularity $m = 60$). We then apply ℓ_1-normalization to $\mathbf{H}$, followed by element-wise square-root, and flatten it to $\mathbf{s}_{\mathrm{WB}} \in \mathbb{R}^{3m^2}$.
- **Parameter-related Features.** To capture the trajectory of parameter adjustments, analogous to how humans consider past tuning attempts, we append a compact history vector $\mathbf{h}_s \in \mathbb{R}^{11}$ that encodes recent parameter values for both $N\%$ and p, along with a normalized timestep counter.
- **Two-branch Backbone.** The full state is $s_t = (\mathbf{s}_{\mathrm{WB}}, \mathbf{h}_s)$. Both actor and critic employ two-branch MLP encoders that map each input to 64-dimensional embeddings $\mathbf{z}_{\mathrm{WB}}$ and $\mathbf{z}_{\mathrm{hist}}$, which are then fused. The actor outputs $\boldsymbol{\mu}$ and $\log \boldsymbol{\sigma}^2 \in \mathbb{R}^2$ for reparameterized sampling of the two continuous actions, while twin critics compute Q-values with minimum selection to reduce overestimation.

Action Design. We use relative, continuous actions to jointly tune the gray-pixel percentage N and Minkowski order p: $\mathrm{param}_{t+1} = \mathrm{param}_t + a_t$, with $\mathrm{param} \in \{N, p\}$. Actions are sampled from the policy, squashed by tanh to $[-1, 1]$, and rescaled to valid ranges (e.g., $a_t^{(N)} \in [-0.6, 0.6]$, $a_t^{(p)} \in [-4, 4]$). This yields smooth updates and coordinated adaptation.

Reward Design. We measure the quality of illuminant estimation by the angular error. To stabilize the training across images with different initial errors E_0, the main reward signal is the relative error improvement:

$$E = \arccos\left(\frac{\langle \hat{\mathbf{e}}, \mathbf{e} \rangle}{\|\hat{\mathbf{e}}\| \, \|\mathbf{e}\|}\right), \tag{6}$$

$$R_{\mathrm{err}} = \frac{E_0 - E_t}{E_0 + \left(\dfrac{E_0}{c_1}\right)^{\alpha}}, \tag{7}$$

where c_1 is the average initial error and $\alpha = 0.6$. To discourage overly large moves, we add an action cost $R_{\mathrm{act}} = -\lambda \sqrt{(a_1/0.6)^2 + (a_2/4)^2}$, where $\lambda = 0.1$ and a_1, a_2 control the gray-pixel selection percentage $N\%$ and Minkowski order p. A difficulty-aware relaxation scales the penalty: $R_{\mathrm{step}} = R_{\mathrm{err}} + (1 - E_0/c_2) \times R_{\mathrm{act}}$, where c_2 is the maximum initial error. Episodes stop after three steps of estimation stability. At termination, we add bonus $R_\rho \in \{+50, +30, +20, +10, -10\}$ for improvement ratio $\rho = E_t / \max(E_0, 10^{-12})$ in ranges $[0, 0.8)$, $[0.8, 0.9)$, $[0.9, 0.95)$, $[0.95, 1.0)$, ≥ 1.0, yielding $R_{\mathrm{final}} = R_{\mathrm{step}} + R_\rho$.

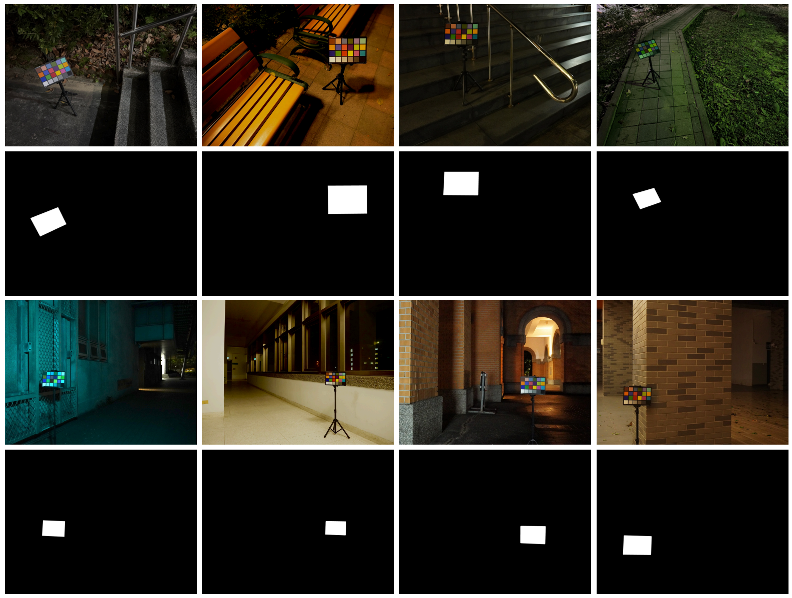


Fig. 3: Sample images from the proposed LEVI dataset with their corresponding Color Checker mask annotations. The dataset captures diverse nighttime scenes with complex mixed lighting, low illumination, and high ISO conditions.

Optimization. We adopt the off-policy Soft Actor-Critic (SAC) algorithm [47] with a stochastic policy and critics implemented with twin Q-value heads.

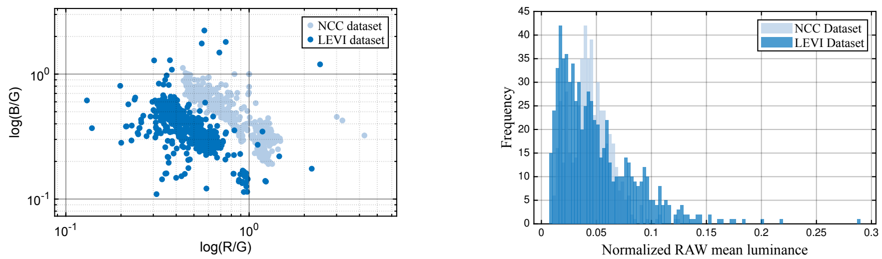
3.3 Curriculum Learning

Stage 1: Single-Image Parameter Tuning. Stage 1 uses a fixed training image to train the agent on error reduction through sequential parameter adjustments and termination detection. Once behavior stabilizes and the agent reliably stops at convergence, we proceed to Stage 2.

Stage 2: Multi-Image Tuning. We use a curriculum pool $\mathcal{D}_c = \{x_1, \ldots, x_M\}$ (M=5). For each training data x_i, the agent runs 5 consecutive episodes, then cycles to x_{i+1} (wrapping after x_M). This cyclic schedule reduces environment resets, captures short-horizon patterns in the replay buffer, and exploits SAC's off-policy experience reuse for stable updates.

4 LEVI Dataset

Prior to our work, NCC dataset [27] was the only public nighttime color constancy benchmark, containing 513 images from a single camera. To enable cross-sensor evaluation, we introduce the **L**ow-light **E**vening **V**ision **I**llumination (**LEVI**)



(a) Illuminant distribution. **(b)** Normalized mean luminance histogram.

Fig. 4: Dataset statistics of the LEVI and NCC datasets. (a) Illuminant distribution over all collected nighttime images. (b) Normalized mean luminance histogram showing that LEVI contains more low-luminance images.

dataset. The first multi-camera nighttime benchmark comprising 700 linear RAW images from two systems: iPhone 16 Pro (images #1–370, 4320×2160, 12-bit) and Sony ILCE-6400 (images #371–700, 6000×4000, 14-bit), with ISO ranging from 500 to 16,000. Each scene contains a Macbeth Color Checker with manual annotations. Ground-truth illuminants are computed as median RGB values of non-saturated achromatic patches. All images are black-level corrected and converted to linear RGB. Fig. 3 shows sample images with Color Checker masks; Fig. 4a and Fig. 4b compare the illuminant and luminance distributions of NCC and LEVI datasets. LEVI complements NCC by covering broader lighting conditions and containing more low-luminance nighttime images, offering a new benchmark for low-light color constancy evaluation.

5 Experiments

5.1 Implementation Details

We normalize image resolutions across datasets by downsampling: iPhone 16 Pro captures in LEVI are resized by 0.25×, while Sony ILCE-6400 images in LEVI and all NCC images are resized by 0.125×. The performance of color constancy methods is evaluated by the standard angular error metric (measured in degrees). RL-AWB (SAC) is trained on an Intel Core i5-13600K CPU. Training batch size is 256, $\gamma = 0.1$, $\tau = 0.005$, learning rate 3×10^{-4}, and 16 parallel environments over 15,000 timesteps, with updates starting after 100 initial steps. At inference, the agent iteratively updates parameters until convergence (average 3 iterations), with a runtime of approximately 1.1 s per image (∼360 ms per call) on an NVIDIA RTX 3080 GPU.

5.2 Results and Comparisons

Experimental results in Tab. 1 show that the proposed SGP-LRD consistently outperforms other statistical baselines, yielding the minimum angular errors on

Table 1: Evaluation of statistics-based methods on both the NCC and LEVI datasets. Angular error in degrees.

	NCC					LEVI				
Method	Med.	Mean	Tri.	B-25	W-25	Med.	Mean	Tri.	B-25	W-25
GE-1st [101]	4.14	5.17	4.35	1.25	10.87	3.94	4.31	3.97	1.82	7.45
GE-2nd [101]	3.58	4.64	3.78	1.11	9.93	4.17	4.49	4.19	1.80	7.76
MSGP [88]	2.48	3.52	2.70	0.80	8.02	3.12	3.34	3.14	1.54	5.52
GI [87]	3.13	4.52	3.40	0.91	10.60	3.10	3.42	3.15	1.49	5.91
BCC [99]	3.06	3.81	3.23	1.05	7.78	4.23	4.53	4.28	2.52	7.06
RGP [27]	2.22	3.33	2.44	0.68	7.81	3.21	3.56	3.29	1.63	6.12
SGP-LRD (Ours)	2.12	3.11	2.29	0.68	7.22	3.08	3.25	3.07	1.40	5.46

both NCC and LEVI datasets. Following our ablation study (Sec. 5.3), we train RL-AWB using only 5 images per dataset. In contrast, all learning-based baselines (FFCC [12], C^4 [117], C^5 [1], FC^4 [51], PCC [121], GCC [22], ePCC [74]) are trained on the full training sets according to their official three-fold cross-validation protocols. As shown in Tab. 2, our proposed RL-AWB framework further enhances illuminant estimation performance. Despite the minimal supervision (5 training images), RL-AWB remains competitive with fully-trained models, showcasing a superior accuracy-data trade-off in the few-shot regime.

Existing learning-based baselines suffer from a substantial performance degradation under cross-dataset evaluation. Specifically, for both NCC$\rightarrow$ LEVI and LEVI$\rightarrow$ NCC, the median and worst-25% errors increase markedly compared to their in-dataset counterparts. This degradation highlights the severe impact of domain and sensor shifts, as the two datasets differ significantly in both scene content and camera characteristics, hindering the generalization of models trained on a single domain. Rather than directly regressing illuminant RGB as in deep learning-based methods, RL-AWB focuses on adaptively tuning the control parameters of SGP-LRD on a per-image basis. Illuminant estimation is then performed by the underlying statistical model, whose inherent robustness to distribution shifts supports reliable cross-dataset generalization. These results show that combining statistical estimation with reinforcement learning effectively mitigates the severe generalization degradation typical of purely learning-based methods. Fig. 5 shows qualitative results of cross-sensor performance on several nighttime scenes from NCC and LEVI datasets.

Beyond nighttime color constancy, we further examine whether the proposed method generalizes to daytime scenarios. To adapt SGP-LRD to well-lit datasets, we remove the local variance and the color deviation filtering modules, and subsequently perform our RL-based parameter tuning. The evaluation results on Gehler-Shi dataset [40, 90] are shown in Tab. 3. Despite being tailored for low-light nighttime scenes, RL-AWB achieves state-of-the-art generalization capability compared with other baselines.

Table 2: Cross-dataset evaluation of the learning-based methods between NCC and LEVI datasets. All learning-based baselines are implemented using three-fold cross-validation protocols and trained on the complete dataset.

Train → Test	NCC → LEVI					LEVI → NCC				
Method	Med.	Mean	Tri.	B-25	W-25	Med.	Mean	Tri.	B-25	W-25
FC^4 [51]	11.8	12.3	11.9	6.36	19.2	13.2	14.4	13.8	6.17	24.1
FFCC [12]	4.44	7.12	5.17	1.96	16.7	7.54	8.51	7.60	4.36	14.8
C^4 [117]	3.09	3.47	3.18	1.18	6.37	5.85	6.73	6.04	2.86	12.2
C^5 [1]	9.12	11.7	9.85	3.76	23.4	4.47	5.46	4.68	1.70	10.9
PCC [121]	11.1	12.5	11.7	7.36	19.7	8.96	10.4	9.35	6.24	16.7
GCC [22]	28.1	43.0	40.5	12.2	90.0	9.77	41.1	28.1	2.26	90.0
ePCC [74]	9.59	10.2	9.67	7.05	14.3	7.92	8.77	8.04	5.01	13.9
RL-AWB (Ours)	3.03	3.24	3.04	1.45	5.36	1.99	3.12	2.25	0.67	7.39

Table 3: Evaluation results on Gehler-Shi dataset trained on NCC dataset. FC^4, C^4, C^5, PCC, GCC, ePCC, and the proposed RL-AWB are trained on the NCC dataset and evaluated on the Gehler–Shi dataset. Compared with our SGP-LRD, the proposed RL-AWB framework achieves a reduction of 5.9% in the median angular error and 9.8% in the best-25% angular error, showing that RL-AWB generalizes well across low-light and well-illuminated images.

Method	Med.	Mean	Tri.	B-25	W-25
FC^4 [51]	13.8	15.8	14.6	6.38	18.8
C^4 [117]	5.62	6.52	5.84	2.43	12.0
C^5 [1]	3.34	3.97	3.43	1.32	7.80
PCC [121]	4.97	8.50	6.07	1.64	20.9
GCC [22]	8.44	20.2	9.68	3.06	58.9
ePCC [74]	4.10	5.08	4.11	1.49	10.7
SGP-LRD (Ours)	2.38	3.64	2.64	0.51	8.89
RL-AWB (Ours)	2.24	3.50	2.51	0.46	8.67

5.3 Ablation Studies

Necessity of the RL framework. We validate the necessity of the proposed RL-AWB framework by comparing it against one-step parameter regressors. In principle, a dense per-image grid search over the algorithm parameters can provide parameters ground truth. Nonetheless, such a procedure requires illuminant ground truth at test time to select the optimal configuration via an argmin operation, making it incompatible with real-world deployment. For example, evaluating 440 parameter configurations per image (11 values for N and 40 values for p) requires approximately 94 hours of computation. We utilize this grid-search procedure only offline to generate supervisory targets for one-step regressors. Specifically, we train both MLP and CNN models to predict (N, p) directly from the same RGB-uv histogram used by RL-AWB, where the MLP backbone is identical to the SAC actor network. The regressors are trained on NCC (N), LEVI (L), and Gehler-Shi (G) datasets following the same evalu-

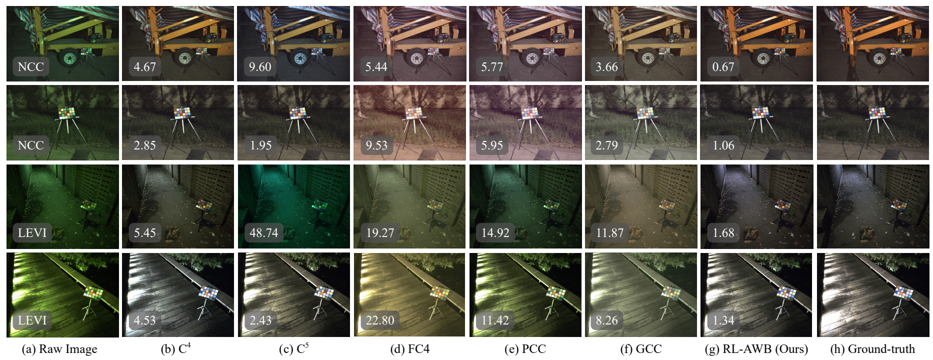


Fig. 5: Visual comparison of cross-dataset performance under domain shift. Top: LEVI $\rightarrow$ NCC; Bottom: NCC $\rightarrow$ LEVI. Unlike prior learning-based approaches that struggle to generalize to unseen sensor domains, RL-AWB shows superior robustness across diverse sensor characteristics. Images are gamma-corrected for visualization.

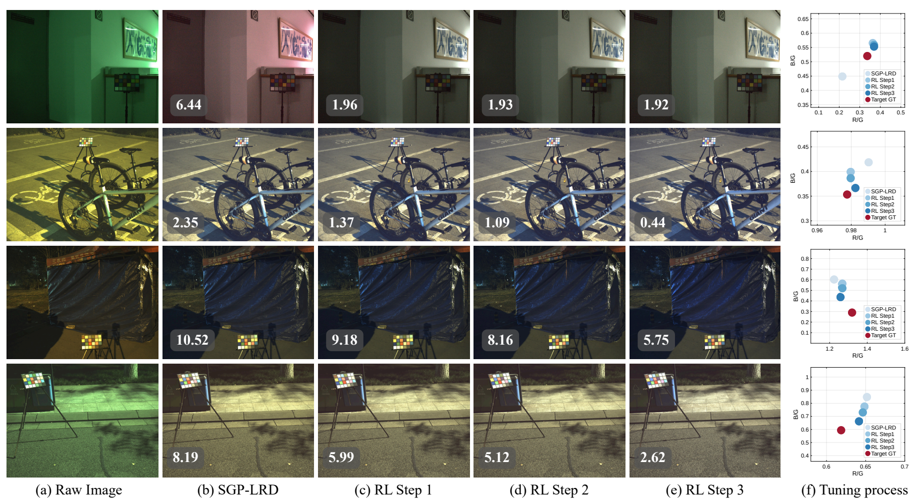


Fig. 6: Illustration of the RL-AWB auto-tuning process for representative nighttime scenes. For each image, we visualize the initial input, several intermediate correction results along the trajectory of the RL policy, and the final output, together with the corresponding angular error at each step. As the agent iteratively updates the SGP-LRD parameters, the corrected images gradually approach the ground-truth white-balanced results, and the angular error decreases. Note that the images shown are gamma-corrected for visualization.

ation protocol as RL-AWB, including three-fold in-domain evaluation and all cross-domain transfer settings. The median angular error results are shown in Tab. 4. This performance advantage arises from RL-AWB's iterative refinement mechanism. Unlike one-step regressors, which must commit to a single parame-

Table 4: Evaluation results on RL-AWB compared to one-step regressors. RL-AWB consistently achieves lower final angular error than both one-step MLP and CNN regressors, despite being trained with only five images.

Method	Train: NCC N→N	Train: NCC N→L	Train: NCC N→G	Train: LEVI L→N	Train: LEVI L→L	Train: LEVI L→G	Train: Gehler G→N	Train: Gehler G→L	Train: Gehler G→G
MLP	2.01	3.07	2.29	2.25	3.02	2.39	2.20	3.13	2.34
CNN	2.06	3.05	2.34	2.17	3.04	2.35	2.07	3.17	2.30
RL-AWB	**1.98**	**3.03**	**2.24**	**1.99**	**3.01**	**2.27**	**1.99**	**3.12**	**2.16**

Table 5: Ablation on curriculum pool size (SAC). The results exhibit a U-shaped performance trend. Small pools (<5 images) lack sufficient diversity to learn robust policies, while large pools (>5 images) reduce per-sample visitation under a fixed replay budget, leading to excessive exploration noise.

	NCC Med.	NCC Mean	NCC W-25	LEVI Med.	LEVI Mean	LEVI W-25
3	2.16	3.29	7.69	3.05	3.28	5.55
5	**1.98**	**3.07**	**7.22**	**3.01**	**3.22**	**5.32**
7	2.09	3.19	7.54	3.04	3.28	5.53
9	2.13	3.23	7.63	3.03	3.23	5.39
15	2.24	3.21	7.47	3.06	3.24	5.41

ter estimate and cannot revise it afterward, the RL agent progressively adjusts (N, p) over multiple steps and can recover from suboptimal intermediate decisions. Such error-correction behavior is fundamentally unavailable to one-step regression models by construction. Fig. 6 shows the stepwise corrections.

Training data number. We vary the Stage-2 curriculum pool size $M \in \{3, 5, 7, 9, 15\}$ (Tab. 5). We adopt $M = 5$ as the best trade-off.

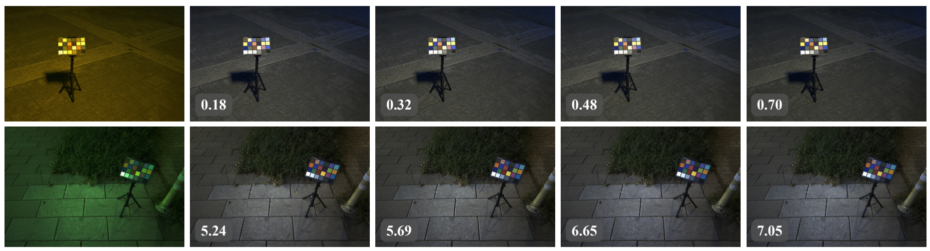
SGP-LRD. We ablate the two key filtering stages of SGP-LRD to assess their individual contributions. As shown in Tab. 6, optimal performance is attained only when both the noise and color filtering modules are incorporated. It is worth noting that the noise filter is designed as a robustness safeguard to mitigate the impact of shot noise inherent in nighttime imagery, not an accuracy driver.

6 Conclusion

This study is the first to apply reinforcement learning to color constancy, showing that DRL can be effectively used for white balance tuning. Our work makes three contributions: (1) SGP-LRD, a novel statistical algorithm for nighttime color constancy, (2) RL-AWB, a reinforcement learning framework for adaptive parameter optimization, and (3) LEVI, the first multi-camera nighttime color

Table 6: Ablation study on SGP-LRD components. Angular error in degrees.

	NCC			LEVI		
	Med.	Mean	W-25	Med.	Mean	W-25
w/o Noise Filtering	2.12	3.12	7.32	3.09	3.26	5.48
w/o Color Filtering	2.51	3.90	9.54	3.25	3.68	6.61
Full SGP-LRD	**2.12**	**3.11**	**7.22**	**3.08**	**3.25**	**5.46**



(a) Raw Image (b) SGP-LRD (c) RL Step 1 (d) RL Step 2 (e) RL Step 3

Fig. 7: Failure case analysis. Examples of failure cases in challenging nighttime scenes, where RL-AWB over-corrects the illumination and produces visually degraded outputs. Images are gamma-corrected for visualization.

constancy dataset. Experiments on both nighttime and daytime datasets demonstrate that the proposed method achieves competitive performance compared to existing baselines, while exhibiting strong cross-sensor robustness. Compared to one-step regressors, RL-AWB provides a practical and deployment-compatible solution that remains ground-truth free at inference time while generalizing effectively to unseen datasets and camera sensors.

Limitations and future work. While RL-AWB consistently reduces overall angular error, we observe one failure mode: for images with already low initial estimation error, the RL agent may over-correct, leading to slightly degraded outputs. Fig. 7 illustrates such cases where the agent's parameter adjustments increase the angular error from an initial estimate. Future work will explore safety-aware reward formulations and constrained optimization strategies to mitigate over-correction, as well as hierarchical policies to efficiently coordinate additional tunable ISP parameters beyond the current two-parameter action space.

Acknowledgements

This research was funded by the National Science and Technology Council, Taiwan, under Grants NSTC 112-2222-E-A49-004-MY2 and 113-2628-E-A49-023-. The authors are grateful to Google, NVIDIA, and MediaTek Inc. for their generous donations. Yu-Lun Liu acknowledges the Yushan Young Fellow Program by the MOE in Taiwan.

References

1. Afifi, M., Barron, J.T., LeGendre, C., Tsai, Y.T., Bleibel, F.: Cross-camera convolutional color constancy. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1981–1990 (2021)
2. Afifi, M., Brown, M.S.: Sensor-independent illumination estimation for dnn models. arXiv preprint arXiv:1912.06888 (2019)
3. Afifi, M., Brown, M.S.: Deep white-balance editing. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. pp. 1397–1406 (2020)
4. Afifi, M., Brubaker, M.A., Brown, M.S.: Auto white-balance correction for mixed-illuminant scenes. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1210–1219 (2022)
5. Afifi, M., Hu, Z., Liang, L.: Optimizing illuminant estimation in dual-exposure hdr imaging. In: European Conference on Computer Vision. pp. 202–219. Springer (2024)
6. Afifi, M., Price, B., Cohen, S., Brown, M.S.: When color constancy goes wrong: Correcting improperly white-balanced images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1535–1544 (2019)
7. Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O., Zaremba, W.: Hindsight experience replay. Advances in neural information processing systems **30** (2017)
8. Bai, M., Urtasun, R.: Deep watershed transform for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5221–5229 (2017)
9. Bajaj, C., Yang, Y., Wang, Y.: Reinforcement learning of self-enhancing camera image and signal processing. In: Advances in Data-driven Computing and Intelligent Systems: Selected Papers from ADCIS 2022, Volume 2, pp. 281–303. Springer (2023)
10. Barnard, K., Cardei, V., Funt, B.: A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. IEEE transactions on Image Processing **11**(9), 972–984 (2002)
11. Barron, J.T.: Convolutional color constancy. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 379–387 (2015)
12. Barron, J.T., Tsai, Y.T.: Fast fourier color constancy. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 886–894 (2017)
13. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th annual international conference on machine learning. pp. 41–48 (2009)
14. Bianco, S., Cusano, C.: Quasi-unsupervised color constancy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12212–12221 (2019)
15. Bianco, S., Cusano, C., Schettini, R.: Color constancy using cnns. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 81–89 (2015)
16. Bianco, S., Cusano, C., Schettini, R.: Single and multiple illuminant estimation using convolutional neural networks. IEEE Transactions on Image Processing **26**(9), 4347–4362 (2017)
17. Brainard, D.H., Freeman, W.T.: Bayesian color constancy. Journal of the optical Society of America A **14**(7), 1393–1411 (1997)

18. Buchsbaum, G.: A spatial processor model for object colour perception. Journal of the Franklin institute **310**(1), 1–26 (1980)
19. Buzzelli, M., Bianco, S.: Uncertainty estimation in color constancy. Pattern Recognition **160**, 111175 (2025)
20. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 12504–12513 (2023)
21. Chakrabarti, A., Hirakawa, K., Zickler, T.: Color constancy with spatio-spectral statistics. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(8), 1509–1519 (2011)
22. Chang, C.W., Fan, C.D., Chang, C.C., Lo, Y.C., Tseng, Y.C., Huang, J.L., Liu, Y.L.: Gcc: Generative color constancy via diffusing a color checker. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 10868–10878 (2025)
23. Chang, K.C., Wang, R., Lin, H.J., Liu, Y.L., Chen, C.P., Chang, Y.L., Chen, H.T.: Learning camera-aware noise models. In: European Conference on Computer Vision. pp. 343–358. Springer (2020)
24. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3291–3300 (2018)
25. Chen, L., Zhang, Y., Song, Y., Shan, Y., Liu, L.: Improved test-time adaptation for domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24172–24182 (2023)
26. Chen, S.K., Yen, H.L., Liu, Y.L., Chen, M.H., Hu, H.N., Peng, W.H., Lin, Y.Y.: Learning continuous exposure value representations for single-image hdr reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12990–13000 (2023)
27. Cheng, C., Yang, K.F., Wan, X.M., Chan, L.L.H., Li, Y.J.: Nighttime color constancy using robust gray pixels. Journal of the Optical Society of America A **41**(3), 476–488 (2024)
28. Cheng, D., Prasad, D.K., Brown, M.S.: 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)
29. Cheng, D., Price, B., Cohen, S., Brown, M.S.: Effective learning-based illuminant estimation using simple features. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1000–1008 (2015)
30. Conde, M.V., McDonagh, S., Maggioni, M., Leonardis, A., Pérez-Pellitero, E.: Model-based image signal processors via learnable dictionaries. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 481–489 (2022)
31. Cun, X., Wang, Z., Pun, C.M., Liu, J., Zhou, W., Jia, X., Li, H.: Learning enriched illuminants for cross and single sensor color constancy. arXiv preprint arXiv:2203.11068 (2022)
32. Das, P., Liu, Y., Karaoglu, S., Gevers, T.: Generative models for multi-illumination color constancy. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1194–1203 (2021)
33. Ebner, M.: Color constancy based on local space average color. Machine Vision and Applications **20**(5), 283–301 (2009)
34. Finlayson, G., Hordley, S.: Improving gamut mapping color constancy. IEEE Transactions on Image Processing **9**(10), 1774–1783 (2000)

35. Finlayson, G.D.: Corrected-moment illuminant estimation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1904–1911 (2013)
36. Finlayson, G.D., Trezzi, E.: Shades of gray and colour constancy. In: Color and imaging conference. vol. 12, pp. 37–41. Society of Imaging Science and Technology (2004)
37. Florensa, C., Held, D., Wulfmeier, M., Zhang, M., Abbeel, P.: Reverse curriculum generation for reinforcement learning. In: Conference on robot learning. pp. 482–495. PMLR (2017)
38. Forsyth, D.A.: A novel algorithm for color constancy. International Journal of Computer Vision **5**(1), 5–35 (1990)
39. Furuta, R., Inoue, N., Yamasaki, T.: Pixelrl: Fully convolutional network with reinforcement learning for image processing. IEEE Transactions on Multimedia **22**(7), 1704–1719 (2019)
40. Gehler, P.V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2008)
41. Gijsenij, A., Gevers, T., Van De Weijer, J.: Generalized gamut mapping using image derivative structures for color constancy. International Journal of Computer Vision **86**(2), 127–139 (2010)
42. Gijsenij, A., Gevers, T., Van De Weijer, J.: Computational color constancy: Survey and experiments. IEEE transactions on image processing **20**(9), 2475–2489 (2011)
43. Gijsenij, A., Gevers, T., Van De Weijer, J.: Improving color constancy by photometric edge weighting. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(5), 918–929 (2011)
44. Graves, A., Bellemare, M.G., Menick, J., Munos, R., Kavukcuoglu, K.: Automated curriculum learning for neural networks. In: international conference on machine learning. pp. 1311–1320. Pmlr (2017)
45. Gregor, K., LeCun, Y.: Learning fast approximations of sparse coding. In: Proceedings of the 27th international conference on international conference on machine learning. pp. 399–406 (2010)
46. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1780–1789 (2020)
47. Haarnoja, T., Zhou, A., Abbeel, P., Levine, S.: Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. International Conference on Machine Learning (ICML) (2018)
48. Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., et al.: Soft actor-critic algorithms and applications. arXiv preprint arXiv:1812.05905 (2018)
49. Hernandez-Juarez, D., Parisot, S., Busam, B., Leonardis, A., Slabaugh, G., McDonagh, S.: A multi-hypothesis approach to color constancy. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2270–2280 (2020)
50. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: A white-box photo post-processing framework. ACM Transactions on Graphics (TOG) **37**(2), 1–17 (2018)
51. Hu, Y., Wang, B., Lin, S.: Fc4: Fully convolutional color constancy with confidence-weighted pooling. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4085–4094 (2017)

52. Jiang, H., Luo, A., Fan, H., Han, S., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. ACM Transactions on Graphics (ToG) **42**(6), 1–14 (2023)
53. Jiang, L., Meng, D., Zhao, Q., Shan, S., Hauptmann, A.: Self-paced curriculum learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 29 (2015)
54. Karmanov, A., Guan, D., Lu, S., El Saddik, A., Xing, E.: Efficient test-time adaptation of vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14162–14171 (2024)
55. Kim, D., Afifi, M., Kim, D., Brown, M.S., Kim, S.J.: Ccmnet: Leveraging calibrated color correction matrices for cross-camera color constancy. arXiv preprint arXiv:2504.07959 (2025)
56. Kim, D., Kim, J., Nam, S., Lee, D., Lee, Y., Kang, N., Lee, H.E., Yoo, B., Han, J.J., Kim, S.J.: Large scale multi-illuminant (lsmi) dataset for developing white balance algorithm under mixed illumination. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2410–2419 (2021)
57. Kınlı, F., Yılmaz, D., Özcan, B., Kıraç, F.: Modeling the lighting in scenes as style for auto white-balance correction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4903–4913 (2023)
58. Klink, P., D'Eramo, C., Peters, J.R., Pajarinen, J.: Self-paced deep reinforcement learning. Advances in Neural Information Processing Systems **33**, 9216–9227 (2020)
59. Koskinen, S., Acar, E., Kämäräinen, J.K.: Single pixel spectral color constancy. International Journal of Computer Vision **132**(2), 287–299 (2024)
60. Kosugi, S., Yamasaki, T.: Unpaired image enhancement featuring reinforcement-learning-controlled image editing software. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 11296–11303 (2020)
61. Kumar, M., Packer, B., Koller, D.: Self-paced learning for latent variable models. Advances in neural information processing systems **23** (2010)
62. Laakom, F., Passalis, N., Raitoharju, J., Nikkanen, J., Tefas, A., Iosifidis, A., Gabbouj, M.: Bag of color features for color constancy. IEEE Transactions on Image Processing **29**, 7722–7734 (2020)
63. Land, E.H., McCann, J.J.: Lightness and retinex theory. Journal of the Optical society of America **61**(1), 1–11 (1971)
64. Lee, K., Shin, U., Lee, B.U.: Learning to control camera exposure via reinforcement learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2975–2983 (2024)
65. Li, C., Guo, C.L., Zhou, M., Liang, Z., Zhou, S., Feng, R., Loy, C.C.: Embedding fourier for ultra-high-definition low-light image enhancement. arXiv preprint arXiv:2302.11831 (2023)
66. Li, C., Guo, C., Han, L., Jiang, J., Cheng, M.M., Gu, J., Loy, C.C.: Low-light image and video enhancement using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(12), 9396–9416 (2021)
67. Li, J., Chen, C., Huang, W., Lang, Z., Song, F., Yan, Y., Xiong, Z.: Learning steerable function for efficient image resampling. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5866–5875 (2023)
68. Li, S., Tan, R.T.: Nightcc: Nighttime color constancy via adaptive channel masking. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25522–25531 (2024)
69. Li, S., Wang, J., Brown, M.S., Tan, R.T.: Transcc: Transformer-based multiple illuminant color constancy using multitask learning. CoRR (2022)

70. Li, Z., Yi, S., Ma, Z.: Rendering nighttime image via cascaded color and brightness compensation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 897–905 (2022)
71. Liu, X., Wu, Z., Li, A., Vasluianu, F.A., Zhang, Y., Gu, S., Zhang, L., Zhu, C., Timofte, R., Jin, Z., et al.: Ntire 2024 challenge on low light image enhancement: Methods and results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6571–6594 (2024)
72. Liu, Y.L., Lai, W.S., Chen, Y.S., Kao, Y.L., Yang, M.H., Chuang, Y.Y., Huang, J.B.: Single-image hdr reconstruction by learning to reverse the camera pipeline. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1651–1660 (2020)
73. Liu, Y., Kothari, P., Van Delft, B., Bellot-Gurlet, B., Mordan, T., Alahi, A.: Ttt++: When does self-supervised test-time training fail or thrive? Advances in Neural Information Processing Systems **34**, 21808–21820 (2021)
74. Liu, Y., Wei, M.: Color constancy from a pure color view: An edge-aware algorithm for a wider application. Color Research & Application **51**(1), e70036 (2026)
75. Lo, Y.C., Chang, C.C., Chiu, H.C., Huang, Y.H., Chen, C.P., Chang, Y.L., Jou, K.: Clcc: Contrastive learning for color constancy. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8053–8063 (2021)
76. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5637–5646 (2022)
77. Matiisen, T., Oliver, A., Cohen, T., Schulman, J.: Teacher–student curriculum learning. IEEE transactions on neural networks and learning systems **31**(9), 3732–3740 (2019)
78. McDonagh, S., Parisot, S., Zhou, F., Zhang, X., Leonardis, A., Li, Z., Slabaugh, G.: Formulating camera-adaptive color constancy as a few-shot meta-learning problem. arXiv preprint arXiv:1811.11788 (2018)
79. Monga, V., Li, Y., Eldar, Y.C.: Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Processing Magazine **38**(2), 18–44 (2021)
80. Mosleh, A., Sharma, A., Onzon, E., Mannan, F., Robidoux, N., Heide, F.: Hardware-in-the-loop end-to-end optimization of camera image processing pipelines. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7529–7538 (2020)
81. Mou, C., Wang, Q., Zhang, J.: Deep generalized unfolding networks for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17399–17410 (2022)
82. Narvekar, S., Peng, B., Leonetti, M., Sinapov, J., Taylor, M.E., Stone, P.: Curriculum learning for reinforcement learning domains: A framework and survey. Journal of Machine Learning Research **21**(181), 1–50 (2020)
83. Oh, S.W., Kim, S.J.: Approaching the computational color constancy as a classification problem through deep learning. Pattern Recognition **61**, 405–416 (2017)
84. Park, J., Lee, J.Y., Yoo, D., Kweon, I.S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5928–5936 (2018)
85. Peng, R., Wu, C.: Cccg: Self-supervised color constancy with collaborative generative network. In: Proceedings of the 30th Annual International Conference on Mobile Computing and Networking. pp. 2055–2060 (2024)

86. Punnappurath, A., Abuolaim, A., Abdelhamed, A., Levinshtein, A., Brown, M.S.: Day-to-night image synthesis for training nighttime neural isps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10769–10778 (2022)
87. Qian, Y., Kamarainen, J.K., Nikkanen, J., Matas, J.: On finding gray pixels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8062–8070 (2019)
88. Qian, Y., Pertuz, S., Nikkanen, J., Kämäräinen, J.K., Matas, J.: Revisiting gray pixel for statistical illumination estimation. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP). VISAPP, vol. 4, pp. 36–46 (2019). https://doi.org/10.5220/0007406900360046, https://dblp.org/rec/conf/visapp/QianPNKM19
89. Schaul, T., Quan, J., Antonoglou, I., Silver, D.: Prioritized experience replay. arXiv preprint arXiv:1511.05952 (2015)
90. Shi, L., Funt, B.: Re-processed version of the gehler color constancy dataset of 568 images. Simon Fraser University (2010), technical Report
91. Shi, W., Loy, C.C., Tang, X.: Deep specialized network for illuminant estimation. In: European conference on computer vision. pp. 371–387. Springer (2016)
92. Sidorov, O.: Conditional gans for multi-illuminant color constancy: Revolution or yet another approach? In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0 (2019)
93. Soviany, P., Ionescu, R.T., Rota, P., Sebe, N.: Curriculum learning: A survey. International Journal of Computer Vision **130**(6), 1526–1565 (2022)
94. Sun, X., Zhao, Z., Wei, L., Lang, C., Cai, M., Han, L., Wang, J., Li, B., Guo, Y.: Rl-seqisp: Reinforcement learning-based sequential optimization for image signal processing. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 5025–5033 (2024)
95. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: International conference on machine learning. pp. 9229–9248. PMLR (2020)
96. Tang, Y., Kang, X., Li, C., Lin, Z., Ming, A.: Transfer learning for color constancy via statistic perspective. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 2361–2369 (2022)
97. Tao, S., Shukla, A., Chan, T.k., Su, H.: Reverse forward curriculum learning for extreme sample and demonstration efficiency in reinforcement learning. arXiv preprint arXiv:2405.03379 (2024)
98. Tseng, E., Yu, F., Yang, Y., Mannan, F., Arnaud, K.S., Nowrouzezahrai, D., Lalonde, J.F., Heide, F.: Hyperparameter optimization in black-box image processing using differentiable proxies. ACM Trans. Graph. **38**(4), 27–1 (2019)
99. Ulucan, O., Ulucan, D., Ebner, M.: Block-based color constancy: The deviation of salient pixels. In: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
100. Ulucan, O., Ulucan, D., Ebner, M.: Multi-scale color constancy based on salient varying local spatial statistics. The Visual Computer **40**(9), 5979–5995 (2024)
101. Van De Weijer, J., Gevers, T., Gijsenij, A.: Edge-based color constancy. IEEE Transactions on image processing **16**(9), 2207–2214 (2007)
102. Wang, C., Wu, H., Jin, Z.: Fourllie: Boosting low-light image enhancement by fourier frequency information. In: Proceedings of the 31st ACM international conference on multimedia. pp. 7459–7469 (2023)

103. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726 (2020)
104. Wang, Q., Fink, O., Van Gool, L., Dai, D.: Continual test-time domain adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7201–7211 (2022)
105. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 2654–2662 (2023)
106. Wang, Y., Yu, Y., Yang, W., Guo, L., Chau, L.P., Kot, A.C., Wen, B.: Exposurediffusion: Learning to expose for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12438–12448 (2023)
107. Wang, Y., Xu, T., Fan, Z., Xue, T., Gu, J.: Adaptiveisp: Learning an adaptive image signal processor for object detection. Advances in Neural Information Processing Systems **37**, 112598–112623 (2024)
108. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018)
109. Wu, Y., Pan, C., Wang, G., Yang, Y., Wei, J., Li, C., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1662–1671 (2023)
110. Xiao, J., Gu, S., Zhang, L.: Multi-domain learning for accurate and few-shot color constancy. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3258–3267 (2020)
111. Xie, M., Zhong, C., He, Y., Qin, Z., Fang, M.: Boosting illuminant estimation in deep color constancy through brightness robustness enhancement. Pattern Recognition p. 112153 (2025)
112. Xu, B., Liu, J., Hou, X., Liu, B., Qiu, G.: End-to-end illuminant estimation based on deep metric learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3616–3625 (2020)
113. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17714–17724 (2022)
114. Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., Zhang, Y.: Hvi: A new color space for low-light image enhancement. In: Proceedings of the computer vision and pattern recognition conference. pp. 5678–5687 (2025)
115. Yang, H., Wang, B., Vesdapunt, N., Guo, M., Kang, S.B.: Personalized exposure control using adaptive metering and reinforcement learning. IEEE transactions on visualization and computer graphics **25**(10), 2953–2968 (2018)
116. Yi, X., Xu, H., Zhang, H., Tang, L., Ma, J.: Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 12302–12311 (2023)
117. Yu, H., Chen, K., Wang, K., Qian, Y., Zhang, Z., Jia, K.: Cascading convolutional color constancy. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12725–12732 (2020)
118. Yu, K., Dong, C., Lin, L., Loy, C.C.: Crafting a toolchain for image restoration by deep reinforcement learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2443–2452 (2018)

119. Yu, K., Li, Z., Peng, Y., Loy, C.C., Gu, J.: Reconfigisp: Reconfigurable camera image processing pipeline. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4248–4257 (2021)
120. Yu, R., Liu, W., Zhang, Y., Qu, Z., Zhao, D., Zhang, B.: Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. Advances in neural information processing systems **31** (2018)
121. Yue, S., Wei, M.: Color constancy from a pure color view. Journal of the Optical Society of America A **40**(3), 602–610 (2023)
122. Yue, S., Wei, M.: Effective cross-sensor color constancy using a dual-mapping strategy. Journal of the Optical Society of America A **41**(2), 329–337 (2024)
123. Zhang, J., Ghanem, B.: Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1828–1837 (2018)
124. Zhang, K., Gool, L.V., Timofte, R.: Deep unfolding network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3217–3226 (2020)
125. Zhang, S., He, J., Zhu, Y., Wu, J., Yuan, J.: Efficient camera exposure control for visual odometry via deep reinforcement learning. IEEE Robotics and Automation Letters (2024)
126. Zhang, Z., Kang, X., Ming, A.: Domain adversarial learning for color constancy. In: IJCAI. pp. 1693–1699 (2022)
127. Zhao, M., Liu, P., Zhang, T., Zhang, Z.: Ref-lle: Personalized low-light enhancement via reference-guided deep reinforcement learning. arXiv preprint arXiv:2506.22216 (2025)
128. Zhou, S., Li, C., Change Loy, C.: Lednet: Joint low-light enhancement and deblurring in the dark. In: European conference on computer vision. pp. 573–589. Springer (2022)