This ICCV 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.

After the Party: Navigating the Mapping From Color to Ambient Lighting

Florin-Alexandru Vasluianu Tim Seizinger Zongwei Wu* Radu Timofte
Computer Vision Lab, CAIDAS & IFI, University of Würzburg
{firstname.lastname}@uni-wuerzburg.de

Abstract

Illumination in practical scenarios is inherently complex, involving colored light sources, occlusions, and diverse material interactions that produce intricate reflectance and shading effects. However, existing methods often oversimplify this challenge by assuming a single light source or uniform, white-balanced lighting, leaving many of these complexities unaddressed. In this paper, we introduce CL3AN, the first large-scale, high-resolution dataset of its kind designed to facilitate the restoration of images captured under multiple Colored Light sources to their Ambient-Normalized counterparts. Through benchmarking, we find that leading approaches often produce artifacts—such as illumination inconsistencies, texture leakage, and color distortion—primarily due to their limited ability to precisely disentangle illumination from reflectance. Motivated by this insight, we achieve such a desired decomposition through a novel learning framework that leverages explicit chromaticity-luminance components guidance, drawing inspiration from the principles of the Retinex model. Extensive evaluations on existing benchmarks and our dataset demonstrate the effectiveness of our approach, showcasing enhanced robustness under non-homogeneous color lighting and material-specific reflectance variations, all while maintaining a highly competitive computational cost. The benchmark, codes, and models are available at www.github.com/fvasluianu97/RLN2.

1. Introduction

The goal of Ambient Lighting Normalization is recovering an ambient-lit image, under uniform lighting distribution, regardless of the lighting scenario characteristic to the depicted contents [65]. The ability to recover ambient-normalized images—free from spatially varying illumination artifacts—is critical for robust visual understanding [33, 51, 53]. In both practical imaging scenarios [27, 32] and the rapidly evolving domain of image generation and synthesis [57, 81], illumination always involves complex interactions of multiple colored light sources, occlusions, and material properties, resulting in intricate shadows, glare, color shifts, and texture distortions. Lighting and material characteristics are the driving factors for explaining the manifestation of such artifacts, thus explaining the need for data-level representativity prior to proposing image restoration solutions aiming for improved degradation robustness or restoration performance.

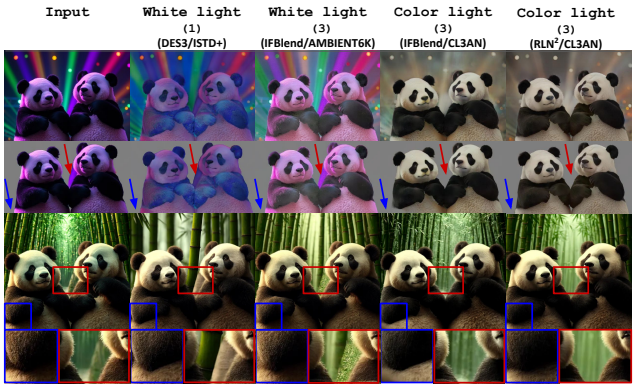


Figure 1. Effective deployment of Ambient Lighting Normalization (ALN) as a pre-processing step in Neural Image Editing [19]. We compare the effectiveness of our proposed RLN^2 against state-of-the-art models in fields like Shadow Removal (DES3 [36]), or ALN (IFBlend [65]), trained on settings including single (1) [68] or multiple (3) white-domain-tuned [65] or color light sources.

While a wide range of applications require illumination-invariant representations [7, 31, 59, 67], existing normalization methods struggle to disentangle these intertwined factors [19, 35, 62, 81], as illustrated in Fig. 1. Even if data contributions further away from the natural light in ISTD+ [44, 68], or multiple white-light-tuned light sources from AMBIENT6K [65] can improve general effectiveness, model design can further drive robustness.

This limitation stems primarily from oversimplified assumptions, such as single white light sources [64, 68] or homogeneous color fields [56, 68, 87], which fail to capture the richness of complex illumination. A key technical challenge lies in the lack of comprehensive datasets that

*Corresponding

adequately represent the complexity of real-world lighting, especially outside the realm of natural white-color-tuned lighting. Prior efforts in shadow removal [64, 68], color constancy [6, 15, 16], or color correction [1, 22, 34, 52, 73] address isolated aspects of illumination but fall short of modeling multi-light interactions. Recent attempts [38, 65] introduce variable-intensity lighting datasets but lack chromatic diversity and material-specific reflectance variations. Furthermore, wide access to ambient-lit images outside the scope of scene-rendering software remains a challenge. As a result, learning-based methods often produce artifacts.

In this paper, we address these limitations through two core contributions. First, we introduce CL3AN, the first large-scale dataset capturing high-resolution scenes under multiple colored directional lights paired with ambient-lit references. Each scene triplet includes variations beyond white-balanced lighting, colored lighting, and a diffuse-lit ground truth, enabling systematic study of illumination effects like self-shadows, highlights, glare areas, color spill or saturation, and other material-dependent artifacts. By benchmarking leading image restoration methods [9, 23, 65, 80], we observe that these models frequently suffer from critical shortcomings, such as the conflation of chromatic variations caused by illumination, texture leakage, inconsistent partial normalization, etc.

We identify the fundamental limitation of existing methods as their inability to systematically disentangle reflectance and illumination—a critical flaw exacerbated by the lack of explicit guidance for separating chromaticity and luminance under multi-colored lighting. Most current approaches [23, 80, 84] rely on single-stream, end-to-end feature modeling, which implicitly maps input images to ambient-normalized counterparts without incorporating physical awareness of light-material interactions. While some recent works attempt to address this by introducing pixel- and frequency-domain supervision [12, 65], more flexible towards reflectance- and chromaticity-information extraction, they remain limited in learning appropriate frequency filters — a challenging task without explicit guidance — resulting in suboptimal performance under the challenging multi-colored lighting conditions.

In this paper, from a novel perspective, we propose **RLN**^{2}, a Retinex-inspired framework that introduces parallel, physics-grounded streams to independently model reflectance (material properties) and illumination (lighting geometry/color). We observe that these intrinsic specificities are naturally encoded in the cylindrical representation of the input image, where the Hue (H) and Saturation (S) maps align with the observed chromaticity under the effect of the incoming light, while the Value (V) map provides strong priors for materials-specific reflectance. By embedding these domain-specific clues — such as reflectance information entropy (V), and smoothness constraints (H, S) — into feature refinement, our framework achieves precise decomposition, effectively eliminating undesirable artifacts while preserving fine-grained details.

Our contributions can be summarized as follows:

- The introduction of the **CL3AN**, a first-of-its-kind dataset enabling the study of ambient lighting normalization under multiple RGB lights;
- The extensive study of a wide range of models, seminal works in various image restoration tasks, when trained and tested on the **CL3AN** dataset;
- We propose **RLN**^{2}, a robust algorithm built on the principle of simultaneous image reflection and lightness components compensation to recover an ambient lit image, characterized by uniform lighting distribution. **RLN**^{2} is proposed as a strong baseline for the ambient lighting normalization task, achieving state-of-the-art results on both the public AMBIENT6K dataset and the proposed **CL3AN** benchmark.

2. Related Work

This work expands Ambient Lighting Normalization to multicolor direct lighting, addressing previously unexplored effects appearing in cluttered scenes, and analyzing shadow interactions and highlights appearing under multicolor direct lighting.

Lighting and Color: Color analysis is central to image processing and understanding, with various downstream tasks such as facial recognition [61, 76, 77], or tracking and matching applications [5, 14, 41, 59, 63] relying on the color similarity principle, thus benefiting from increased appearance robustness, especially under challenging lighting.

Image Quality Assessment (IQA) methods [3, 79] show a strong correlation between details-abundance and image quality, motivating the focus on professional scene lighting rigs in all expert image photography applications. Given the scarcity of such equipment, alternative strategies [1, 2, 6, 45] are developed around image sensor sensor properties and cross-device constancy [8, 20].

Naturally, registered image stacks either for tasks such as multiple view geometry estimation [17, 90] or different image stitching applications require consistent color and content appearance [4, 11, 42, 60]. In such applications, the effort is focused on consistent lighting in the image acquisition, rather than post-processing correction. In-the-wild applications focusing on widely available amateur images [17, 24, 60] document extensive data curation strategies prior to model estimations.

Later applications, given the newly-found popularity of neural 3D modeling, approach scene relighting as part of their rendering pipeline [58, 62]. Alternatively, approaches based on complex generative image models [35, 55] successfully tackle lighting manipulation, but at a high computational cost. Unfortunately, the limited availability of

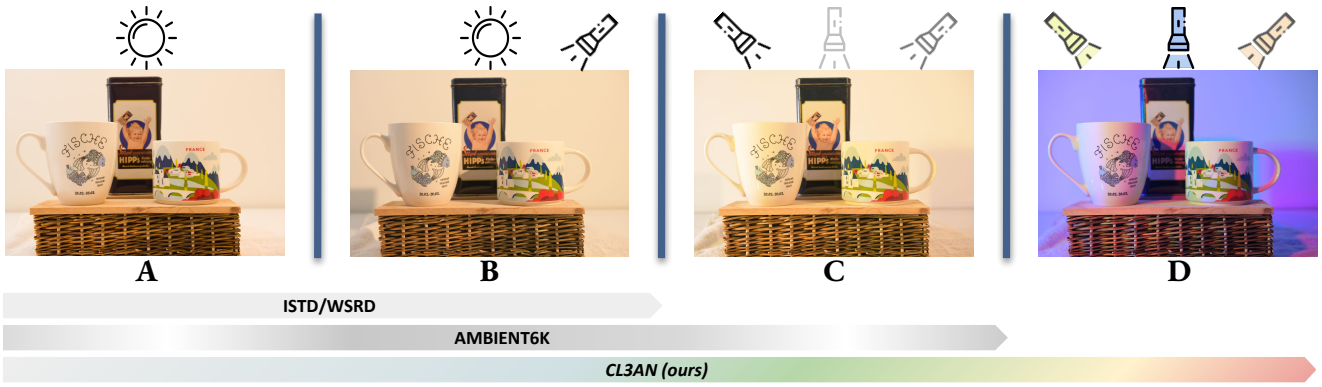


Figure 2. A comparison between the lighting scenarios represented in datasets targeting different applications of Ambient Lighting Normalization (ISTD [68], WSRD[64], or AMBIENT6K [65]). In **(A)**, ambient lighting is used to acquire a uniformly-lit reference image. In **(B)**, direct lighting is used for shadow casting, while constraining consistency to the reference image in shadow-free segments. In **(C)**, more directional lights are added to the direct lighting system. In our proposed setup **(D)**, the direct lighting setup is based on RGB lights, dropping the color consistency constraint, and showcasing complex material-light interactions under various conditions.

real-world ambient lit images, primarily due to the complexity of an ambient lighting setup, explains the difficulty in addressing real-world lighting normalization, even with the obvious availability of large-complexity models.

Reliance on natural light for ambient-lit images proved a viable option, but hindered the variety of possible observed degradations, especially in terms of chromaticity shifts. Unfortunately, the large number of factors involved in light occlusion interaction greatly constrains a representative simulation in general settings. While ISTD/ISTD+ [44, 68], SRD [56] or USR [26] show exclusively flat surfaces under the influence of natural light, the influence of object self-shadows was just recently represented in WSRD [64], only possible under a laboratory setup. Recently, multiple source lighting gained in data-level representation. For example, in LSMI [38], the authors target White-Balance correction robustness under natural and artificial lighting. Alternatively, effects characteristic to multiple illuminants were captured in works such as [30, 54], or in the context of flash photography [29]. However, a uniformly-lit ambient normalized image can not be acquired in these settings.

Alternatively, the introduction of AMBIENT6K [65] coupled multiple light sources of varying intensity and orientation for complex interactions, while also providing an ambient-lit image based on an alternative lighting setup. Unfortunately, the strong connection between the operating parameters set for the direct and ambient lighting systems limits the study to exposure correction for white-aligned scene lighting (see Fig. 2 **(A)**, **(B)**, and **(C)**).

Image Restoration (IR): Joint color and exposure correction, while restoring details lost to occlusion or saturation, defines Ambient Lighting Normalization as an Image Restoration task. Various Convolution Neural Networks (CNNs) [10, 37, 83, 85] became strong baselines for the family of IR tasks. The later introduction of the attention mechanism [66] in image-based applications [13] enabled consistent performance advantages in the wide IR field. Focusing on different feature fusion techniques through specifically tailored attention mechanisms, based on the IR task particularities, defines the success of multiple models [10, 47, 48, 71, 80, 84] easily becoming benchmark solutions for various IR problems. Separate refinement for image reflectance and lightness under the Retinex model [43] proved particularly effective in low-light image enhancement (LLIE) [9, 72, 78, 88], or image shadow removal [28].

Global range dependencies benefits drive the development of State Space Models (SSM) [21], with promising results in general IR [23]. Alternatively, global knowledge can be leveraged through the integration of alternative frequency image representations in IR solutions. Solutions such as FFTFormer [39], DCFormer [46], SFNet [12], DW-GAN [18], and IFBlend [65] successfully extract priors based on Fourier Transform, Discrete Cosine Transform, or other Wavelet Transform variants.

Finally, diffusion-based iterative recurrent solutions [25] started to mature as benchmark solutions in Image Restoration [50, 69, 75, 89], especially in terms of perceived image quality. However, their characteristic increased computational effort hinders their deployment.

3. Method

The core of our work is extending the study of Ambient Lighting Normalization to direct color lighting.

Fig. 3 compares the properties of a scene depicted under multi-source color direct lighting, and an equivalent image under white-aligned direct lighting. The goal to recover the ambient-lit reference image. Provided statistics, such as the RGB image histogram, and the Hue/Saturation footprint showcase the complexity of the ALN transformation, espe-

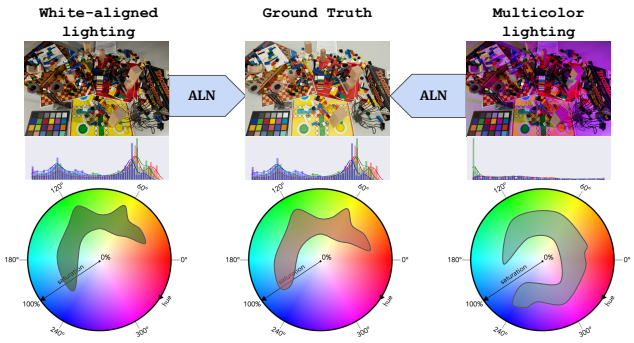


Figure 3. A graphical representation of the Ambient Lighting Normalization transformation in the context of the white-aligned lighting, compared to the challenges posed by the introduction of a multicolor direct lighting system. The RGB image histogram, coupled with the Hue-Saturation rough projection is provided as a reference. Note the complexity of the change in image statistics.

cially increasing with the introduction of variable intensity multicolor direct light sources. Naturally, the most notable difference comes from the color hue deviation. Materials adopt the color of incoming light, with the reflected light wavelength depending on the reflectance defining the represented materials, and the incoming light properties.

Furthermore, the complex geometry of the scene drives light occlusions, with numerous self-shadows at the image level. In this case, given the non-uniform color distribution, the color of the observed shadows depends on the local lighting availability and their light source-defining properties. A successful restoration of the ambient-lit image can not be performed without a full understanding of both scene geometry (for local exposure compensation) and lighting system configuration (for chromaticity uniformization, given the arithmetic of the locally available light radiation given the material properties).

3.1. CL3AN

CL3AN represents an image collection representing **Tri**plets consisting of images acquired under both **C**olor and White **L**ighting, and **A**mbient **N**ormalized images, under uniformly distributed diffuse light.

As introduced in Fig. 2, **CL3AN** is the only dataset providing equivalent representations of the same scene under white-aligned direct lighting (setups **(B)**, **(C)**), varying intensity multicolor lighting (setup **(D)**), while also providing an ambient-lit reference image (setup **(A)**). This makes **CL3AN** a *first-of-its-kind* dataset.

The capturing system relies on a professional photography studio setup, in which a dark room is used to ensure that all image-capturing parameters become trackable and controllable. Two separate scene lighting systems are concurrently developed, optimizing for uniform diffuse light distribution, respectively direct lighting, being triggered only for the acquisition of the corresponding samples. Note that directional lights are fully controllable, so they can be tuned to the white natural light domain, producing images similar to those available in datasets such as AMBIENT6K [65].

The images are captured with a tripod-mounted and remotely operated Canon R6 MKII camera, at constant focal range. This leads to a perfect pixel alignment between the images characteristic to each triplet. For consistency and traceability, the camera-related parameters are kept fixed during the image-capturing process. All hardware parameters relevant to data acquisition are documented in Tab. 1.

For the direct lighting system, three programmable mobile RGB lights are used to capture scene variants under variable light color, intensity, position, and orientation. The direct consequence of this set of parameters is a non-homogeneous distribution of the incoming light color across the scene, implying that global color correction can not be used for restoration in this scenario (see Fig. 2 and Fig. 3).

The diffuse ambient lighting system is composed of five white softbox lights, covering all sides of the cubic acquisition environment, except the base, which is the surface supporting the captured contents. Thus, the possibility for self-shadows in the ambient-lit images is almost fully mitigated, being furthermore enforced through geometry-based object filtering. Calibration over a 24-sample color checker under full white light intensity enables the alignment of both lighting setups, by determining the set of operating parameters producing a similar colors under optimal exposure.

CL3AN is a scene-based dataset, covering 105 cluttered scenes representing an extensive set of materials, both conductors and dielectrics, with varying transparency, color, and surface roughness. The data is available as both RAW and RGB images, at 24 MP, the camera sensor resolution. Data splitting is performed based on the represented contents, with 10 scenes used exclusively for testing, and 10 scenes for validation. Both testing and validation splits represent objects unseen during training, but under lighting conditions consistent with the training split. Additional metadata listing the objects within a scene is provided. The training split consists of a number of 3667 samples, with 437 images used for validation and 431 for testing.

3.2. RLN^2

Fig. 2 and Fig. 3 show the challenging environment posed by the introduction of multicolor direct lighting in the ALN field. Dramatic changes can be observed when mapping from the input image to the ambient-lit reference, especially in terms of observed chromaticity. Moreover, light occlusion by the many rough surfaces corresponding to the represented objects drives a wide variety of intricate effects, with a strong information loss. Learning a robust ALN mapping requires a robust design, able to disentangle significant information compensation in confined image segments from

Camera parameters					Diffuse Lighting				Direct Lighting			
Exposure time (s)	Aperture	ISO	White balance	Focal dist. (mm)	# Active lights	White light temp. (K)	White light temp. range	Intensity	# Active lights	White light temp. (K)	Intensity range	Color Hue range
1/60	f11	100	6400K	50	5	6400	±5%	100%	≥ 1	6400	30-100%	0°-360°

Table 1. **Dataset capturing parameters** regarding the setup of the camera, the diffuse lighting system, and the direct lighting hardware.

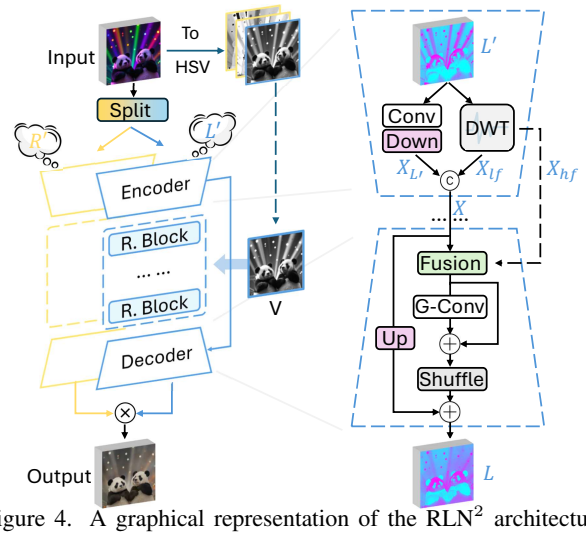


Figure 4. A graphical representation of the RLN^2 architecture, including detailed representations for an Encoder (E) - Decoder (D) blocks pair on the Luminance L-branch. The Reflectance R-branch design is similar, being omitted for clarity.

the large magnitude chromaticity shifts.

Given these observations, problem sub-division is the core of our proposed design RLN^2, aiming for joint independent compensation in both Image Reflectance and Luminance components. Under the Retinex theory [43], a given image I can be decomposed as a product of two multiplicative terms, representing image illumination or lightness L, and the material specific image reflectance R.

Thus, an image pair (I, I') in which I' is the degraded version of the reference image I, can be represented by the component set $\{(L, R), (L', R')\}$, in line with Eq. (1).

$$\begin{aligned} I(x,y) &= L(x,y) \cdot R(x,y) \\ I'(x,y) &= L'(x,y) \cdot R'(x,y) \end{aligned} \tag{1}$$

Therefore, our goal is to determine the pair $(\overline{L}, \overline{R})$ in which $\overline{L}$ restores the details lost due to inadequate lighting, while $\overline{R}$ aligns the input chromaticity with the white-aligned ambient lit setup.

$$\begin{aligned} L(x,y) &= L'(x,y) + \overline{L}(x,y) \\ R(x,y) &= R'(x,y) + \overline{R}(x,y) \end{aligned} \tag{2}$$

Fig. 4 provides a graphical representation of the proposed RLN^2, in which the (R', L') components of the input I' are subjected to joint independent compensation, while the alternative HSV image representation information is used for feature refinement. Aligned with the properties of the HSV image representation, which codes color in a cylindrical coordinate system, in which the Value (V) channel is similar to explicit illumination intensity, and the Hue-Saturation (HS) represents color tones, the V channel is used as guidance on the L-branch of our model, while the HS pair drives feature refinement on the R-branch of RLN^2.

At the encoder level, the information goes through successive decomposition, in which Haar-DWT is used for spectral dissemination, alongside RGB feature downsampling. The dual-domain information will be altered in the feature refinement stage, at the *R.block* level, through the Cross-Domain Feature Fusion Attention module (CDFFA).

Here, the low-frequency information X_{lf}, and the RGB extracted features $X_{L'}$, projected to an n-dimensional space, are then subjected to relevance enhancement, through a set of parameters based on dot product vector similarity (see Fig. 5) to features extracted from the guidance component. The V channel controls the estimation of the $\overline{L}$ residual, while H and S drive the reflectance estimation $\overline{R}$.

$$\begin{aligned} X_{L'} &= proj_n(X_{L'}) \\ X_{lf} &= proj_n(X_{lf}) \\ \alpha_i &= sim(X_{L',i}, X_{V,i}) \\ \beta_i &= sim(X_{lf,i}, X_{V,i}) \end{aligned} \tag{3}$$

The set of similarity coefficients is used to define a relevance probability distribution, then used to enhance the relevant information by filtering outlier features. The enhanced information is further refined before output.

$$\begin{aligned} \alpha &= softmax(\alpha) \\ \beta &= softmax(\beta) \\ X_{out} &= \{\alpha_i \cdot X_{L',i}\}_{i=0}^{n} + \{\beta_i \cdot X_{lf,i}\}_{i=0}^{n} \end{aligned} \tag{4}$$

At the decoder stage, the refined information is concatenated to the high-frequency spectral information from the corresponding encoder block. Moreover, an ImageNet [40] pretrained ConvNeXt-based feature extractor module [49], provides wide contextual information in feature fusion. A cross-attention module handles information exchange between the wide context information and the dual-domain features refined in the previous stage.

For a fair comparison, at this level, RLN^2 benefits from priors identical to [65], even if the implemented strategies are different. After linear projections, irrelevant features are suppressed through Channel Attention [74], being further refined for Reflectance/Luminance compensation [82].

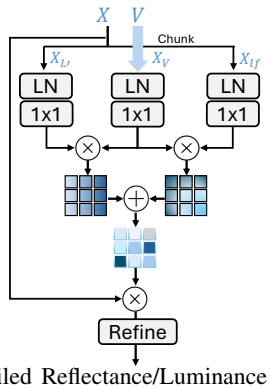


Figure 5. A detailed Reflectance/Luminance Refinement Block (R.Block), including the CDFFA attention module.

Model	Fusion	Guidance	MACs (G.)	PSNR↑	SSIM↑	LPIPS↓
RLN^2	*None*	*None*	20.95	19.892	0.689	0.271
RLN^2	*concat.*	*RGB*	23.67	20.018	0.703	0.233
RLN^2	*concat.*	*L*a*b**	23.68	20.149	0.724	0.226
RLN^2	*concat.*	*HSV*	23.68	20.216	0.731	0.220
RLN^2	*(CDFFA)*	*HSV*	22.72	**20.523**	**0.746**	**0.208**

Table 2. Inner-stage guidance feature fusion ablation study conducted on the testing split of the CL3AN dataset.

Ablative study: Tab. 2 compares the standard RLN^2 configuration, as represented in Fig. 4, to three similar variants in which the control mechanism implemented in the middle stage is modified. Reducing the *R.block* to simple feature refinement by removing any control mechanism shows an important performance deficit, consistent in image restoration fidelity and perceptual quality. Alternatively, using the RGB domain images, passed through a dynamical filter and then split based on the filtered output and its alternative residual shows a slight improvement, at a relatively low induced computational cost.

Task-aligned image representations, with explicit coding of image luminance and chromaticity, prove rich in relevant information, adding only the image conversion cost to the general computation. Focusing on relevant feature enhancement through the CDFFA introduction further increases model robustness and restoration performance.

Implementation: RLN^2 was developed in PyTorch, and the documented experiments were performed on up to three NVIDIA L40s 48 GB VRAM GPUs running a CUDA 12.6 environment. The optimization is split into multiple stages, with the resolution of the training patches gradually increasing as the training procedure progresses. The optimizer used was Adam, with the learning rate $lr = 0.0002$, and the loss optimized was the L1 distance. A cosine learning rate scheduler was used, with two oscillation periods. Optimization deploys gradient clipping, with a standard value of 0.01 under the L1 norm. Further details regarding the training procedure description can be found in the attached supplementary material.

4. Results

Experimental Setup: Additionally to the introduced **CL3AN** dataset, we perform experiments on the AMBIENT6K image database [65], which consists of 6000 images representing scenes under the effect of up to three white-aligned LED directional lights (see Fig. 3).

As compared metrics, we report restoration fidelity through Peak Signal to Noise Ratio (PSNR), and Structure Similarity Index Measure (SSIM) [70]. To quantify the perceptual quality of the restored images, we deploy the AlexNet-based Learned Perceptual Image Patch Similarity score (LPIPS) [86].

In terms of compared methods, we chose various types of solutions proposed in the image restoration field, which became seminal work after their introduction. For an extensive benchmark, we combine refined CNN solutions such as MPRNet [83] and HINet [37], with transformer-based architectures like SwinIR [47], NAFNet [10], Uformer [71], Restormer [84], and GRL [80]. Additionally, we consider solutions integrating state-space-models (*e.g.* MAMBAIR [21]), or solutions optimizing joint Spatial-Frequency entropy such as SFNet [12] and IFBlend [65]. Ultimately, we include Retinexformer [9], representing the class of solutions based on Retinex image decomposition. We train all the aforementioned methods using their official public implementations and settings.

Quantitative Results: In Tab. 3, we compare the aforementioned solutions with different variants of the proposed RLN^2. The different variants represent our proposed models at different complexity levels, such that when comparing to similar existing solutions, advantages become easily identifiable. The computational complexity, as the Multiplications-Additions count (MAC), is reported for a patch of 128×128 pixels, using *ptflops* for consistency.

The naming convention can be easily followed. Firstly, the difference between the RLN^2-S and the RLN^2-L is given by the complexity of the Wide Context Feature Extractor found in the feature fusion module, at the decoder stage, with RLN^2-S using a ConvNeXt-T configuration, while RLN^2-L benefits from a large ConvNeXt-XL model [49]. Secondly, we distinguish between RLN^2 variants that have/have not access to the dual image frequency domain representation. Thus RLN^2-S and RLN^2-L drop the frequency domain feature at the encoder, refinement stage, and decoder stage, using cross-attention for fusion, while RLN^2-Sf and RLN^2-Lf use all the available information.

All the compared RLN^2 variants are trained under the same conditions, including the optimizer, batch size, data augmentation strategies, optimized objective, and the number of samples seen during training. For AMBIENT6K [65], we adopt their publicly available benchmark.

Reading Tab. 3, it becomes obvious that RLN^2 is a high-performance solution, improving over similar state-of-the-

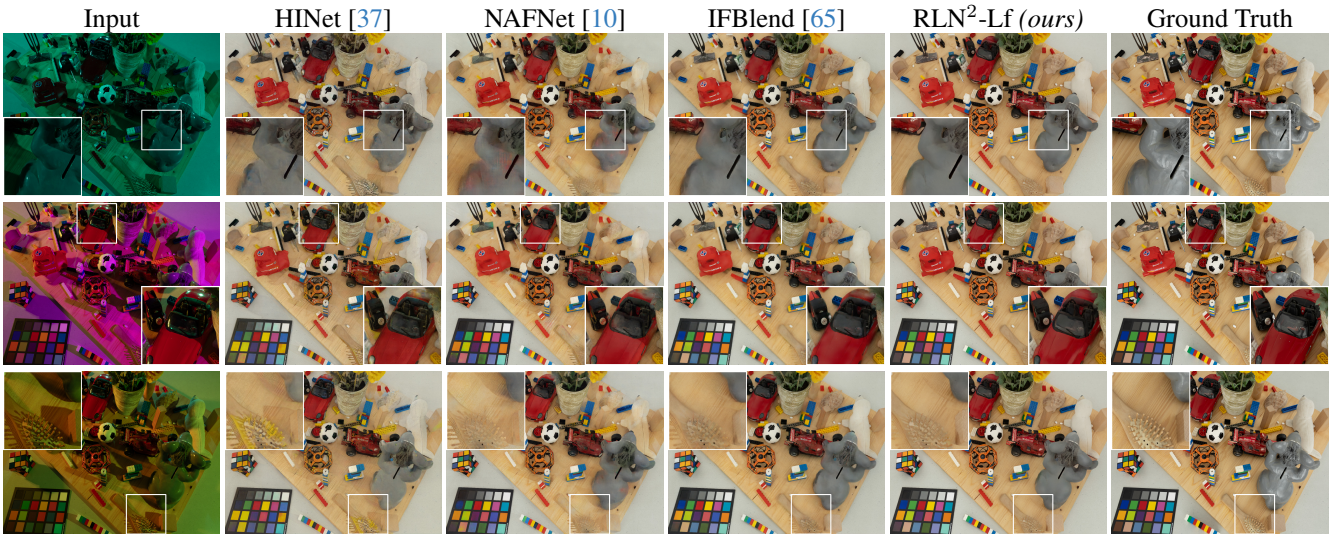


Figure 6. **Qualitative comparison** on scene #1 from test split of the **CL3AN** dataset. Best viewed in the electronic version.

Method name	General information Type	Restoration Task	Prior	MACs (G.)	CL3AN *(ours)* PSNR↑	SSIM↑	LPIPS↓	AMBIENT6K [65] PSNR↑	SSIM↑	LPIPS↓
unprocessed	-	-	-	-	10.837	0.447	0.518	13.403	0.652	0.250
NAFNet[10]	Conv.	Multi-task IR	RGB	15.92	19.476	0.709	0.249	20.580	0.808	0.142
MPRNet [83]	Conv.	Multi-task IR	RGB	37.21	18.453	0.688	0.291	20.947	0.820	0.129
SFNet [12]	Transf.	Multi-task IR	RGB + Freq.	31.27	18.382	0.686	0.291	20.519	0.812	0.141
SwinIR [47]	Transf.	Multi-task IR	RGB	37.81	16.386	0.643	0.372	20.528	0.817	0.131
Uformer [71]	Transf.	Multi-task IR	RGB	19.33	17.508	0.655	0.313	20.776	0.818	0.131
Restormer [84]	Transf.	Multi-task IR	RGB	35.31	18.560	0.691	0.278	21.141	0.817	0.132
HINet [37]	Conv.	Multi-task IR	RGB	42.68	19.388	0.707	0.248	20.856	<u>0.821</u>	0.129
IFBlend [65]	Conv	ALN	RGB + Freq.	26.01	20.370	0.720	0.228	21.443	0.819	0.128
MAMBAIR [23]	Transf. + SSM	Multi-task IR	RGB	34.32	18.970	0.704	0.254	-	-	-
GRL [80]	Transf.	Multi-task IR	RGB	2.16	18.089	0.672	0.308	-	-	-
Retinexformer [9]	Transf.	LLIE	RGB	4.86	18.649	0.683	0.281	-	-	-
RLN^2-S *(ours)*	Conv	ALN	RGB	3.95	19.992	0.718	0.236	21.181	0.815	0.131
RLN^2-Sf *(ours)*	Conv	ALN	RGB + Freq.	4.24	20.128	<u>0.730</u>	0.223	21.333	0.819	0.128
RLN^2-L *(ours)*	Conv	ALN	RGB	22.45	<u>20.383</u>	0.723	<u>0.222</u>	<u>21.553</u>	<u>0.821</u>	<u>0.123</u>
RLN^2-Lf *(ours)*	Conv	ALN	RGB + Freq.	22.72	**20.523**	**0.746**	**0.208**	**21.712**	**0.825**	**0.120**

Table 3. Various RLN^2 variants versus current SOTA image restoration models, on the proposed **CL3AN** dataset test split (1920 × 1440 px.), and AMBIENT6K benchmark [65] (1280 × 960 px. default size). The **best** and the <u>second best</u> performing models are emphasized.

art solutions at all complexity levels, on both CL3AN and AMBIENT6K [65]. While RLN^2-S largely outperforms RetinexFormer [9], at a lower computational cost, RLN^2-Sf improves over the perceptual performance of IFBlend [65], with a fraction of its required floating point operations.

Against IFBlend [65], a strong recently proposed solution in the ALN field, RLN^2-L shows an advantage, especially in joint RGB-Frequency domains processing, with an ≈15% complexity reduction, while outperforming it.

Another important observation is regarding the architectural core of the compared methods. Under both white-aligned and RGB direct lighting, ALN tends to favor solutions showing an extensive receptive field [37, 83], rather than transformers optimized for local dependencies [47, 71, 84], even when balancing to wider context blocks [80]. However, MPRNet [83] and HINet [37] are larger models, with an additional cost of approx. 63%-87%, compared to the highest complexity RLN^2-Lf variant. While state-state models prove computationally ineffective for the level of performance achieved, frequency-domain information access shows a consistent advantage, especially in terms of observed perceptual quality.

Qualitative Results: In Fig. 6, we provide for qualitative evaluation a set of samples representing a cluttered scene under the effect of non-homogeneous RGB direct lighting. Note the complex geometry and the wide RGB spectrum represented at the ambient lit image level. RLN^2-Lf shows improved behavior in terms of color normalization (row

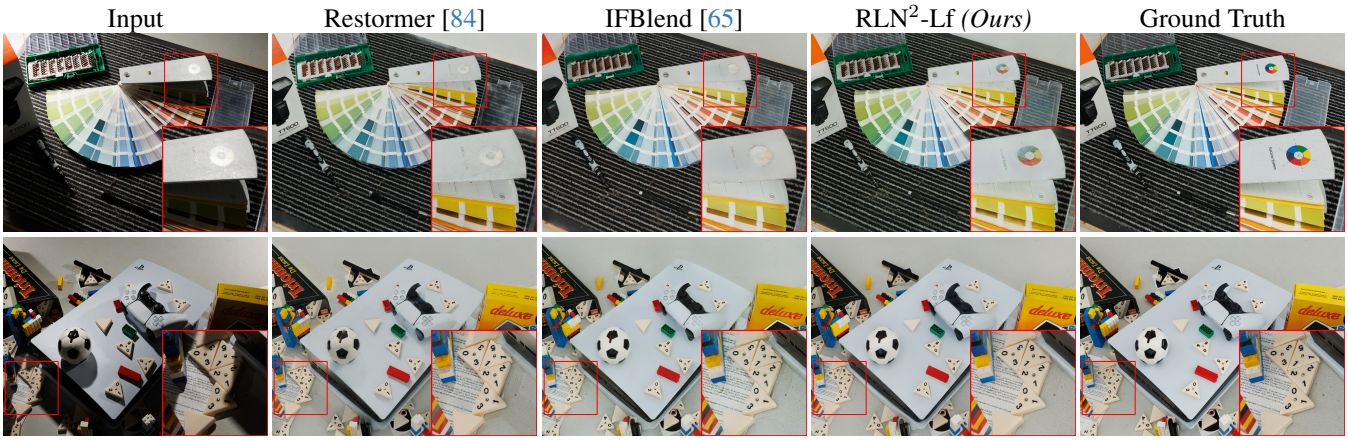


Figure 7. Qualitative comparison challenging samples from the AMBIENT6K testing set [65]. Our RLN^2 produces improved quality ambient normalized images, with sharper images, and a lower amount of visible artifacts.

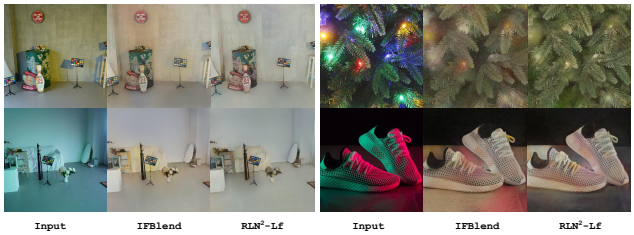


Figure 8. IFBlend[65] and RLN^2-Lf on real-world images from LSMI [38] (*left*), or real product photography samples (*right*).

1), superior restorations in highlights-rich areas, and better high-frequency shadow areas compensation (row 2).

Fig. 7 provides for visual comparison equivalent renderings against the top performing state-of-the-art methods the AMBIENT6K [65] benchmark. Consistently, the restored ambient normalized images are characterized by more natural colors, an increased amount of restored details, and geometrically correct renderings even in areas severely affected by underexposure (second row). Overcoming the challenges specific to non-homogeneous direct color lighting, RLN^2 shows consistent performance, confirming the advantages quantified through evaluations provided in Tab. 3.

To test RLN^2 on out-of-domain data, in Fig. 8 we compare IFBlend [65] and our proposed RLN^2-Lf on inputs from LSMI [38], or complex lighting scenarios from product photography. Our RLN^2-Lf shows increased robustness, with plausible rendered colors under uniform white ambient lighting conditions. Moreover, in Fig. 9 we deploy it on AI-generated images[1] representing various contents in challenging lighting environments. Applying our ALN solution as a pre-processing step in Neural Image Editing proves beneficial for IC-Light [19], with improved layer separation leading to less semantic discontinuities (rows 1, 2, and 4), and improves the rendered results quality in term of shadow casting (row 3) and foreground clarity (row 2).

[1] deepai.org/machine-learning-model/text2img

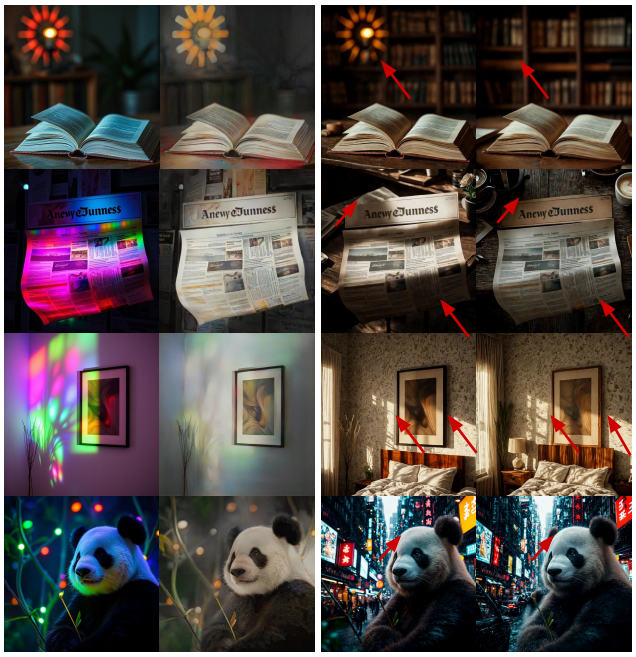


Figure 9. ALN as a preprocessing step before image relighting with foreground condition. Results rendered with ICLight [19] for the input image (col. 3) vs. the RLN^2 output (col. 4).

5. Conclusion

In this work, we propose extending the current ALN study to the challenges added by multicolor and multi-intensity RGB direct lighting. We propose **CL3AN**, a high-resolution image database showcasing a multitude of materials under complex RGB lighting scenarios while providing white-light ambient-lit reference images using a professional photography setup. We introduce RLN^2, a high-performance ALN algorithm setting a new standard for the ALN task, showing how alternative image representations can provide valuable control signals in feature refinement, or offer natural bypassing options for detail preservation.

Acknowledgments. We thank the reviewers and ACs for their sustained effort and helpful feedback. This research was partially supported by the Alexander von Humboldt Foundation.

References

[1] Mahmoud Afifi and Michael S Brown. Deep white-balance editing. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 1397–1406, 2020. 2

[2] Mahmoud Afifi, Jonathan T Barron, Chloe LeGendre, Yun-Ta Tsai, and Francois Bleibel. Cross-camera convolutional color constancy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1981–1990, 2021. 2

[3] Lorenzo Agnolucci, Leonardo Galteri, Marco Bertini, and Alberto Del Bimbo. Arniqa: Learning distortion manifold for image quality assessment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 189–198, 2024. 2

[4] Alexandre Araujo, Jean Ponce, and Julien Mairal. Towards real-world focus stacking with deep learning. *arXiv preprint arXiv:2311.17846*, 2023. 2

[5] Monika Bansal, Munish Kumar, and Manish Kumar. 2d object recognition: a comparative analysis of sift, surf and orb feature descriptors. *Multimedia Tools and Applications*, 80 (12):18839–18857, 2021. 2

[6] Jonathan T Barron. Convolutional color constancy. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2015. 2

[7] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014. 1

[8] Ivaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Frédo Durand. User-guided white balance for mixed lighting conditions. *ACM Trans. Graph.*, 31(6):200–1, 2012. 2

[9] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 2, 3, 6, 7

[10] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022. 3, 6, 7

[11] Yen-Chi Cheng, Chieh Hubert Lin, Hsin-Ying Lee, Jian Ren, Sergey Tulyakov, and Ming-Hsuan Yang. Inout: Diverse image outpainting via gan inversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11431–11440, 2022. 2

[12] Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3, 6, 7

[13] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3

[14] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7157–7173, 2022. 2

[15] Graham D. Finlayson, Steven D. Hordley, and Paul M. Hubel. Color by correlation: A simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001. 2

[16] David H Foster. Color constancy. *Vision research*, 51(7): 674–700, 2011. 2

[17] Jan-Michael Frahm, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, and Marc Pollefeys. Building rome on a cloudless day. In *Computer Vision – ECCV 2010*, pages 368–381, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. 2

[18] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 203–212, 2021. 3

[19] Tsu-Jui Fu, Wenze Hu, Xianzhi Du, William Yang Wang, Yinfei Yang, and Zhe Gan. Guiding instruction-based image editing via multimodal large language models. *arXiv preprint arXiv:2309.17102*, 2023. 1, 8

[20] Arjan Gijsenij, Rui Lu, and Theo Gevers. Color constancy for multiple light sources. *IEEE Transactions on image processing*, 21(2):697–707, 2011. 2

[21] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 3, 6

[22] Julian Jorge Andrade Guerreiro, Mitsuru Nakazawa, and Björn Stenger. Pct-net: Full resolution image harmonization using pixel-wise color transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5917–5926, 2023. 2

[23] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision*, pages 222–241. Springer, 2025. 2, 3, 7

[24] Kamal Gupta, Varun Jampani, Carlos Esteves, Abhinav Shrivastava, Ameesh Makadia, Noah Snavely, and Abhishek Kar. Asic: Aligning sparse in-the-wild image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4134–4145, 2023. 2

[25] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3

[26] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE/CVF interna-*

tional conference on computer vision, pages 2472–2481, 2019. 3

[27] Jie Huang, Pengfei Zhu, Mingrui Geng, Jiewen Ran, Xingguang Zhou, Chen Xing, Pengfei Wan, and Xiangyang Ji. Range scaling global u-net for perceptual image enhancement on mobile devices. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 1

[28] Yan Huang, Xinchang Lu, Yuhui Quan, Yong Xu, and Hui Ji. Image shadow removal via multi-scale deep retinex decomposition. *Pattern Recognition*, page 111126, 2024. 3

[29] Zhuo Hui, Kalyan Sunkavalli, Sunil Hadap, and Aswin C Sankaranarayanan. Illuminant spectra-based source separation using flash photography. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6209–6218, 2018. 3

[30] Zhuo Hui, Ayan Chakrabarti, Kalyan Sunkavalli, and Aswin C Sankaranarayanan. Learning to separate multiple illuminants in a single image. In *Computer Vision and Pattern Recognition (CVPR 2019)*, 2019. 3

[31] Anne Humeau-Heurtier. Texture feature extraction methods: A survey. *IEEE access*, 7:8975–9000, 2019. 1

[32] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 536–537, 2020. 1

[33] Jitesh Jain, Jiachen Li, Mang Tik Chiu, Ali Hassani, Nikita Orlov, and Humphrey Shi. Oneformer: One transformer to rule universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2989–2998, 2023. 1

[34] Jong Ju Jeon, Jun Young Park, and Il Kyu Eom. Low-light image enhancement using gamma correction prior in mixed color spaces. *Pattern Recognition*, 146:110001, 2024. 2

[35] Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. Neural gaffer: Relighting any object via diffusion. *Advances in Neural Information Processing Systems*, 37:141129–141152, 2025. 1, 2

[36] Yeying Jin, Wei Ye, Wenhan Yang, Yuan Yuan, and Robby T Tan. Des3: Adaptive attention-driven self and soft shadow removal using vit similarity. *arXiv preprint arXiv:2211.08089*, 2022. 1

[37] Junpeng Jing, Xin Deng, Mai Xu, Jianyi Wang, and Zhenyu Guan. Hinet: Deep image hiding by invertible network. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4733–4742, 2021. 3, 6, 7

[38] Dongyoung Kim, Jinwoo Kim, Seonghyeon Nam, Dongwoo Lee, Yeonkyung Lee, Nahyup Kang, Hyong-Euk Lee, ByungIn Yoo, Jae-Joon Han, and Seon Joo Kim. 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*, pages 2410–2419, 2021. 2, 3, 8

[39] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023. 3

[40] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc. 5

[41] Avisek Lahiri, Vivek Kwatra, Christian Frueh, John Lewis, and Chris Bregler. Lipsync3d: Data-efficient learning of personalized 3d talking faces from video using pose and lighting normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2755–2764, 2021. 2

[42] Wei-Sheng Lai, Orazio Gallo, Jinwei Gu, Deqing Sun, Ming-Hsuan Yang, and Jan Kautz. Video stitching for linear camera arrays. *arXiv preprint arXiv:1907.13622*, 2019. 2

[43] EDWIN H. LAND. The retinex. *American Scientist*, 52(2): 247–264, 1964. 3, 5

[44] Hieu Le and Dimitris Samaras. Physics-based shadow image decomposition for shadow removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9088–9101, 2021. 1, 3

[45] Shuwei Li, Jikai Wang, Michael S Brown, and Robby T Tan. Mimt: Multi-illuminant color constancy via multi-task local surface and light color learning. *arXiv preprint arXiv:2211.08772*, 2022. 2

[46] Xinyu Li, Yanyi Zhang, Jianbo Yuan, Hanlin Lu, and Yibo Zhu. Discrete cosin transformer: Image modeling from frequency domain. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5468–5478, 2023. 3

[47] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 3, 6, 7

[48] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *IEEE Transactions on Image Processing*, 33:2171–2182, 2024. 3

[49] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022. 5, 6

[50] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1680–1691, 2023. 3

[51] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024. 1

[52] Radoslaw Mantiuk, Rafal Mantiuk, Anna Tomaszewska, and Wolfgang Heidrich. Color correction for tone mapping. In *Computer graphics forum*, pages 193–202. Wiley Online Library, 2009. 2

[53] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3523–3542, 2021. 1

[54] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A dataset of multi-illumination images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4080–4089, 2019. 3

[55] Puntawat Ponglertnapakorn, Nontawat Tritrong, and Supasorn Suwajanakorn. Difareli: Diffusion face relighting. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 22646–22657, 2023. 2

[56] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4067–4075, 2017. 1, 3

[57] Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun BR, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, et al. Lite2relight: 3d-aware single image portrait relighting. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. 1

[58] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Nerf for outdoor scene relighting. In *European Conference on Computer Vision*, pages 615–631. Springer, 2022. 2

[59] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020. 1, 2

[60] Richard Szeliski and Richard Szeliski. Image alignment and stitching. *Computer Vision: Algorithms and Applications*, pages 401–441, 2022. 2

[61] Hironobu Takano, Hiroki Kobayashi, and Kiyomi Nakamura. Rotation invariant iris recognition method adaptive to ambient lighting variation. *IEICE transactions on information and systems*, 90(6):955–962, 2007. 2

[62] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20762–20772, 2023. 1, 2

[63] Prune Truong, Martin Danelljan, Radu Timofte, and Luc Van Gool. Pdc-net+: Enhanced probabilistic dense correspondence network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10247–10266, 2023. 2

[64] Florin-Alexandru Vasluianu, Tim Seizinger, and Radu Timofte. Wsrd: A novel benchmark for high resolution image shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1826–1835, 2023. 1, 2, 3

[65] Florin-Alexandru Vasluianu, Tim Seizinger, Zongwei Wu, Rakesh Ranjan, and Radu Timofte. Towards image ambient lighting normalization. *arXiv preprint arXiv:2403.18730*, 2024. 1, 2, 3, 4, 5, 6, 7, 8

[66] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 3

[67] Qiyang Wan, Ruiping Wang, and Xilin Chen. Interpretable object recognition by semantic prototype analysis. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 800–809, 2024. 1

[68] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1788–1797, 2018. 1, 2, 3

[69] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C.K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. 2024. 3

[70] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[71] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 3, 6, 7

[72] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 3

[73] Ching-Chih Weng, Homer Chen, and Chiou-Shann Fuh. A novel automatic white balance method for digital still cameras. In *2005 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 3801–3804. IEEE, 2005. 2

[74] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 5

[75] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13095–13105, 2023. 3

[76] Xudong Xie and Kin-Man Lam. An efficient illumination normalization method for face recognition. *Pattern Recognition Letters*, 27(6):609–617, 2006. 2

[77] Xiaohua Xie, Wei-Shi Zheng, Jianhuang Lai, Pong C. Yuen, and Ching Y. Suen. Normalization of face illumination based on large-and small-scale features. *IEEE Transactions on Image Processing*, 20(7):1807–1821, 2011. 2

[78] Han Xu, Hao Zhang, Xunpeng Yi, and Jiayi Ma. Cretinex: A progressive color-shift aware retinex model for low-light image enhancement. *International Journal of Computer Vision*, 132(9):3610–3632, 2024. 3

[79] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200, 2022. 2

[80] Xiaoyu Xiang Denis Demandolx Rakesh Ranjan Radu Timofte Yawei Li, Yuchen Fan and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 2, 3, 6, 7

[81] Jae Shin Yoon, Zhixin Shu, Mengwei Ren, Cecilia Zhang, Yannick Hold-Geoffroy, Krishna Kumar Singh, and He Zhang. Generative portrait shadow removal. *ACM Transactions on Graphics (TOG)*, 43(6):1–13, 2024. 1

[82] Weihao Yu and Xinchao Wang. Mambaout: Do we really need mamba for vision? *arXiv preprint arXiv:2405.07992*, 2024. 5

[83] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 3, 6, 7

[84] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 2, 3, 6, 7, 8

[85] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 3

[86] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 6

[87] Xuaner Zhang, Jonathan T Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E Jacobs. Portrait shadow manipulation. *ACM Transactions on Graphics (TOG)*, 39(4):78–1, 2020. 1

[88] Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. Zero-shot restoration of underexposed images via robust retinex decomposition. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 3

[89] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1219–1229, 2023. 3

[90] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R. Oswald, and Marc Pollefeys. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12786–12796, 2022. 2