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

Supplementary Material

In the supplementary material, we first provide more implementation details of our work in Sec. 1. We also provide visual comparisons on different sequences with different direct lighting system variants in Sec. 2. In addition, we provide and discuss failure cases and limitations of the proposed work in Sec. 3. Sec. 4 provides general remarks regarding the study expansion in Ambient Lighting Normalization (ALN), the posed challenges, and how they impact the proposed solution.

## 1. Benchmarking setup

The methods evaluated for the benchmark summarized in Tab. 3 (main manuscript) were trained using the publicly available implementations, adopting the default parameters, and procedures for supervision, as described in the published works. We limited the set of compared methods to works for which the codes (including the training scripts or parameter configurations) and the evaluated pretrained models are publicly available. In the case of the AMBIENT6k benchmark [6], the settings described by the authors were adopted for the evaluation of the different **RLN**$^2$ variants as part of their benchmark.

**Training data:** The models were trained with data sampled from the training split of the introduced **CL3AN** dataset. The model uses the RGB representation of the data, with images resized to a resolution of 1440×1080px to fit in the requirements of the remote storage, especially in terms of data loading latency. All compared methods are then trained on image patches randomly cropped from the input image, with a resolution of 960×960px selected as the size of the crop. The resolution is gradually increased to 1080×1080px as training progresses. Regarding data augmentation, a mixture of random rotations and random flips represent the core of this component. The procedure is completed by the addition of the *mixup* augmentation [2], in which the input images are linearly blended to the ground-truth images through a parameter $\lambda \sim Beta(\alpha = 0.2)$. Given the nature of the training data, in which the input images are acquired under RGB direct lighting, and the reference images represent white ambient-lit equivalents of the same scene, the *mixup* will add variance in terms of color hue, but especially in terms of color saturation.

**RLN**$^2$ **Optimization parameters:** Consistent with the previous benchmarks [6], the standard $L1$ loss was set as the optimization objective for the *Adam* algorithm [4]. The learning rate used was set to $0.0002$, with a *cyclic cosine scheduler* updating the learning rate, under two periods split between 10000 and 20000 data batches seen in training. The length of the training procedure is set to 30000 data batches, with a batch of 15 image crops. The batch size value was chosen as the maximum possible value given the 48GB VRAM of the used GPU, and the requirements of RLN$^2$-Lf.

**Evaluation:** As the evaluation was performed on a local NVIDIA 4090Ti GPU, with 24 GB available VRAM, the original 24 MP images were resized to 1440×1080, due to computational restrictions. The reported evaluations are performed consistently at this image resolution. The values for PSNR and SSIM are calculated after the restored images are mapped to byte-sized RGB values, before saving them to local storage. In the case of LPIPS [7], the default deployment with image tensors in the $[-1, 1]$ range is used to compare the restored variants of the input images to the reference equivalents. The shift in representation domain, from the $[0, 1]$ interval to $[-1, 1]$ is standard for the AlexNet feature extractor [5].

## 2. Additional Qualitative Results

Fig. A provides for visual evaluation additional equivalent more scenes from the test split of the CL3AN dataset. Note that since the data is acquired under controlled laboratory conditions, there are **no ethics concerns**. Each image also emphasizes a representative region through a 2× upscaled patch. The upscaling method uses the *nearest interpolation* algorithm, to avoid introducing new colors at the restored image level. Here, the behavior of the top three performing models can be followed, under changing lighting conditions, including white-aligned direct lighting (setup similar to AMBIENT6K [6]) (rows 1 and 5), and various effects under multiple RGB directional lights. On rows 2 and 6, we can observe restored images, with inputs under the effect of consistent color shift, while rows 3, 4 7, and 8 show equivalent inputs under the effect of multiple color direct lighting. When HINet [3] shows a stronger influence of the incoming light color at the input level in its renderings, the outputs of NAFNet [1] are affected by a significant level of local color artifacts and remaining self-shadows. IFBlend [6] and the proposed RLN$^2$-Lf produce improved quality renderings, with IFBlend showing inconsistent colors in the case of some surfaces. Note that the inconsistencies seem systematic, regardless of the incoming lighting setup that describes the color shift at the input image level. On rows 7-8, IFBlend [6] and RLN$^2$-Lf show better-aligned rendered colors, and significantly lower self-shadows. Moreover, RLN$^2$-Lf has an edge over IFBlend [6] in terms of rendered highlights.

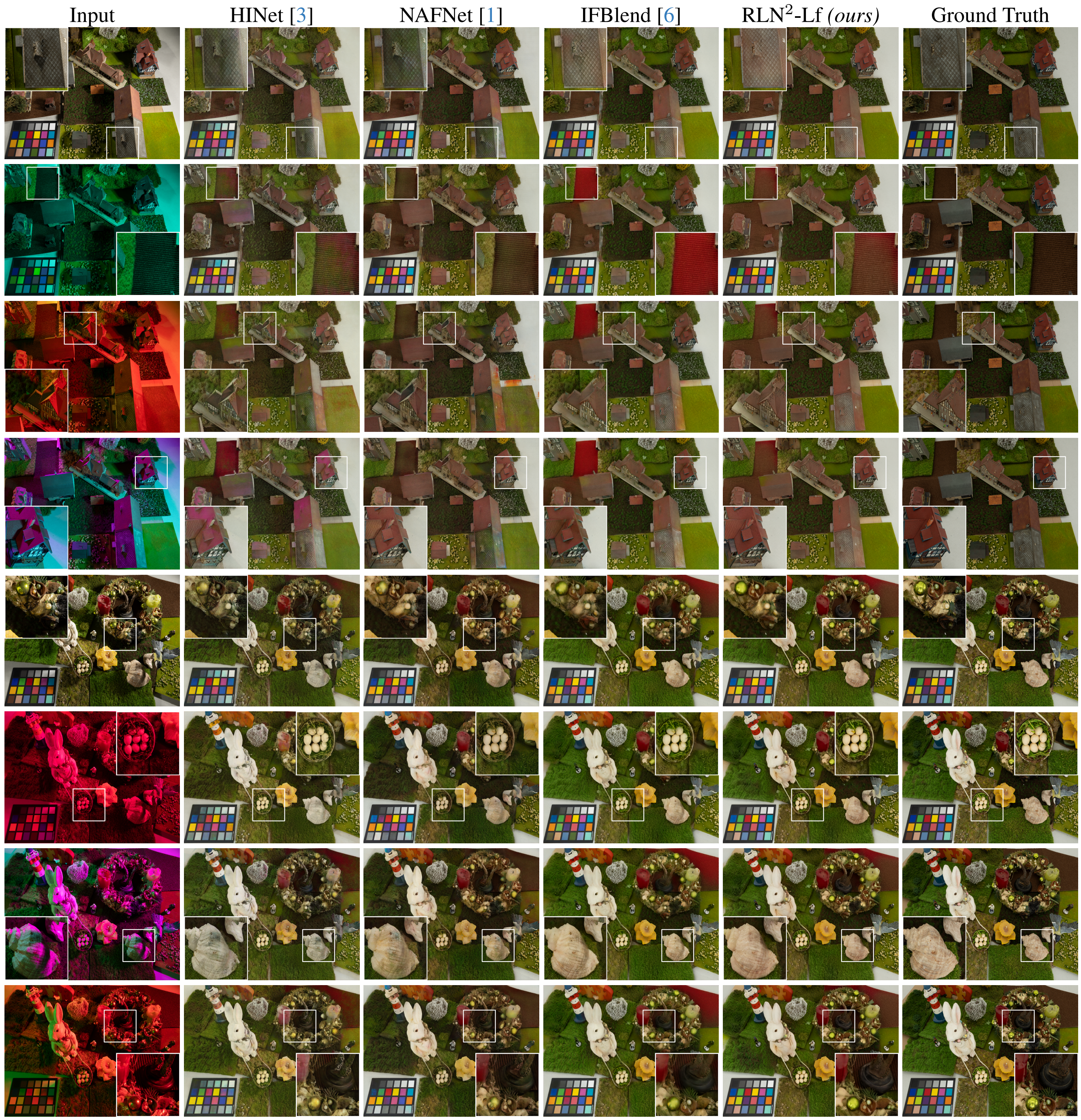

Figure A. **Additional Qualitative Comparisons** on different scenes from test split of the **CL3AN** dataset. Best viewed in the electronic version.

To showcase the generalization ability of the proposed RLN$^2$-Lf model, in Fig. B we compare its renderings against HINet [3] and NAFNet [1] on a set of *DeepAI text2image* generated images[1] representing different contents under single-color or multi-color direct lighting. Here, we can observe RLN$^2$-Lf having a clear advantage over HINet [3] and NAFNet [1] in terms of color shifts still present in their outputs. Under non-natural shadows (row 1), RLN$^2$-Lf can reconstruct the ambient-lit images in sharp details and improve lighting uniformity. On the second raw, RLN$^2$-Lf shows an improved ability in terms of color rendering, with the restored ambient-lit image being character-

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

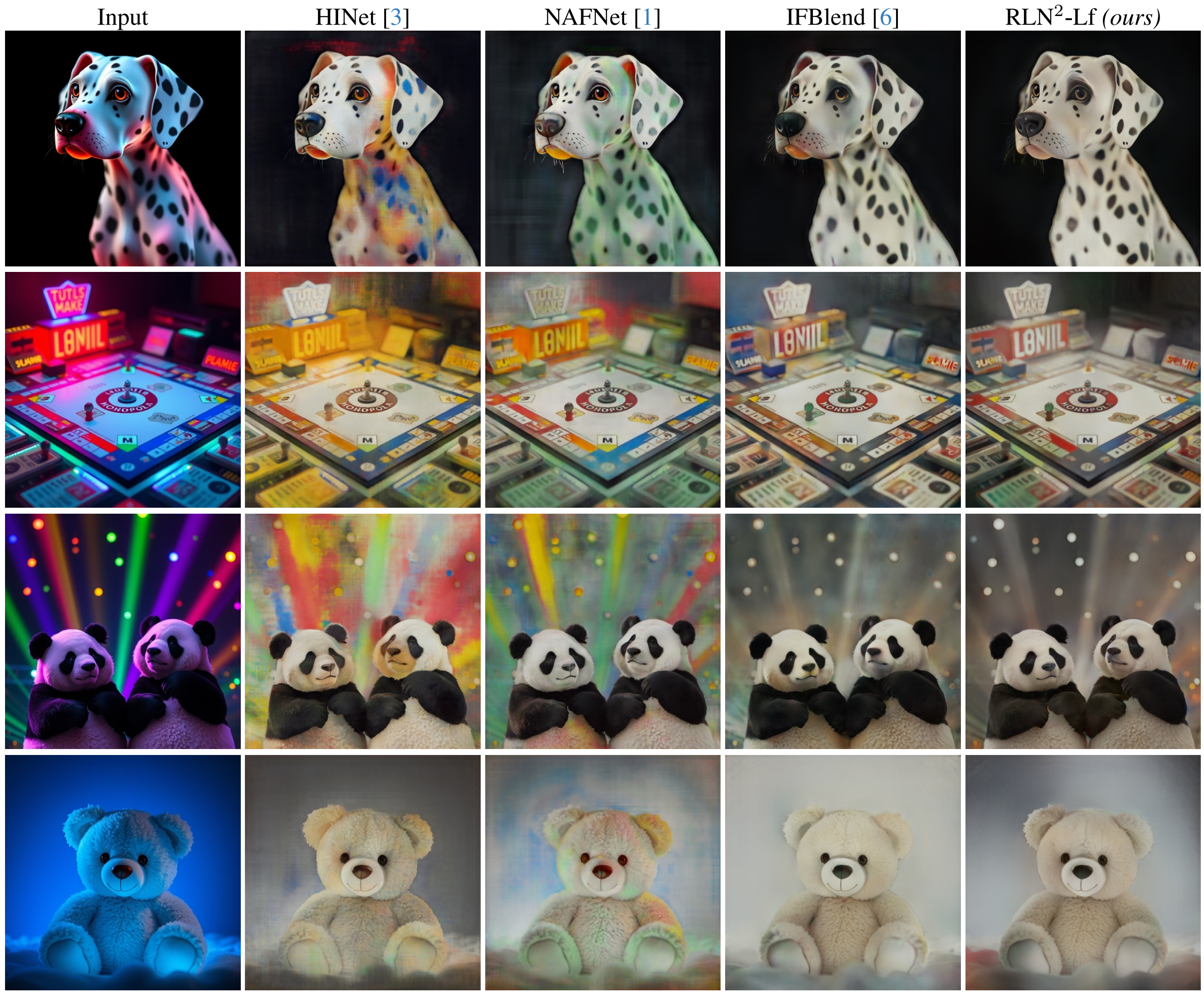


Figure B. **Additional Qualitative Comparisons** on various AI-generated[1] images, representing objects subjected to color direct lighting.

ized by a naturally rendered color palette and limited influence in terms of artifacts visible in the output image. On rows 3 and 4, RLN$^2$-Lf shows an advantage over IFBlend [6] with better handling of incoming light intensity, lower self-shadows influence, and better-appearing backgrounds. Note that the input images are AI-generated, without control in terms of content appearance. Even if the images are not natural in terms of light occlusion and observed intensity, the ambient lit restored images show uniform lighting distribution, limited self-shadow influence, and homogeneous colors in semantically connected image segments.

## 3. Failure Cases

**CL3AN** is a difficult benchmark, shown by the statistics characterizing the input data provided in Tab. 3 (main manuscript). Direct RGB lighting produces color shifts under the RGB arithmetic described by the properties of the represented materials. Various highly reflective surfaces show complicated highlights. Moreover, observed self-shadows color and strength are driven by the scene geometry under the color, intensity, and orientation of the directional lights illuminating the scene.

Fig. C presents a scene in which smooth surfaces characterizing caustics are encased in a black opaque case. Black objects have the property of appearing dark regardless the color of the incoming light, given the low reflectivity at the surface level. Therefore, the input color variance for this class of objects is limited. However, one can observe that when dark objects dominate the scene, a tendency to mistakenly render other contents black exists. This can be explained by the RLN$^2$-Lf over-relying on global image properties in its renderings. However, note that HINet [3], NAFNet [1], and IFBlend [6] also struggle under these conditions. HINet [3] produces visible color artifacts in areas

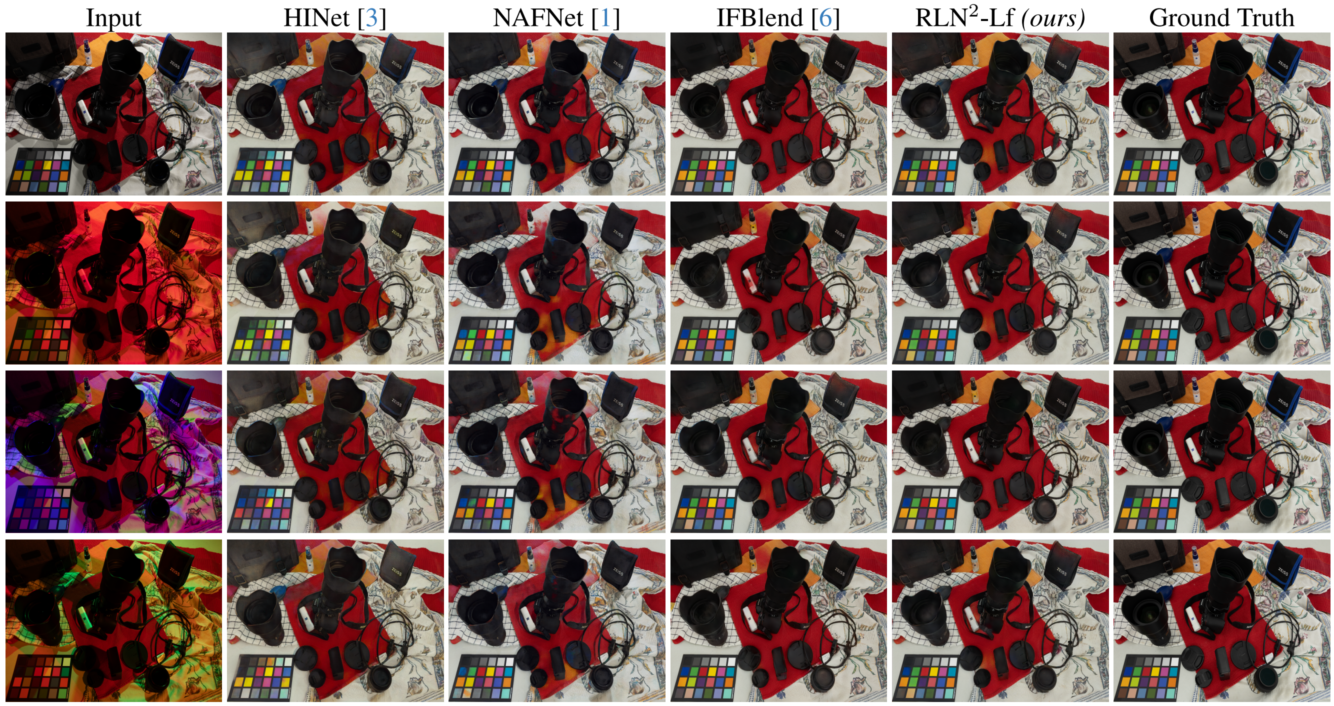


Figure C. **Failcases** on challenging materials represented in the test split of the **CL3AN** dataset.

affected by shadows, while NAFNet [1] borrows from the incoming light color in rendering some of the backgrounds. All models are struggling to correctly render very detailed low-intensity reflections, like those of the internal elements of the depicted lens systems.

## 4. Discussion

**CL3AN** extends the study in Ambient Lighting Normalization (ALN) to the influence of variable multi-color direct lighting in the normalized images. Given that the high complexity of the ALN is already emphasized in public literature [6], the extension of the current study to non-uniform color shifts leads to complex adverse conditions in which a proposed solution has to operate.

This comparison is easy to make, given the statistics provided in Tab. 3 (main manuscript), in which the images acquired under direct multi-RGB lighting are characterized by a loss of 2.56 dB in PSNR, 0.207 lower SSIM, and a degradation of 0.265 in terms of LPIPS. Naturally, ALN remains a particularly challenging task, especially given the current extension, which shows great potential for future research.

RLN$^2$ represents a strong baseline for Color-to-Ambient Lighting Normalization, as a robust model which is able to account for the challenges posed by the introduction of the multi-color direct lighting, while setting the current state-of-the-art on the public AMBIENT6K benchmark, in the white-aligned direct lighting setting. On novel contents, RLN$^2$ shows improved handling of novel contents and represented lighting scenarios, producing high-quality renderings with improved color and lighting distribution, supporting the identified advantages, and becoming a new benchmark ALN solution.

## References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022. 1, 2, 3, 4

[2] Yann N. Dauphin David Lopez-Paz Hongyi Zhang, Moustapha Cisse. mixup: Beyond empirical risk minimization. *International Conference on Learning Representations*, 2018. 1

[3] 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. 1, 2, 3, 4

[4] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[5] 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. 1

[6] 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

[7] 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. 1