EVENSPLAT: COUPLED 2D–3D DECOMPOSITION FOR GAUSSIAN SPLATTING UNDER EXPOSURE AND ILLUMINATION VARIATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A surface photographed under even light presents nearly the same appearance from every angle; the same surface under uneven light does not. Exposure changes between views, illumination varies within a single image, and locally strong light sources leave one region bright and its neighbor in shadow. Multi-view reconstruction methods such as 3D Gaussian Splatting treat these lighting artifacts as if they were properties of the scene, entangling capture-specific illumination with the geometry and color they recover. We present EvenSplat, a framework that separates the two. EvenSplat couples an image-space illumination decomposition with an illumination field carried by the Gaussians, so that the same explanation of the lighting is shared between the two-dimensional and three-dimensional views of the scene; a camera-response network and a local exposure-compensation module absorb the global and residual differences that remain across training images. Through extensive experiments across multiple datasets and diverse forms of uneven illumination (cross-view exposure, spatial illumination variation, and high-contrast lighting) on both real-world captured and simulated benchmarks, EvenSplat generally outperforms state-of-the-art methods, particularly under high-contrast illumination.

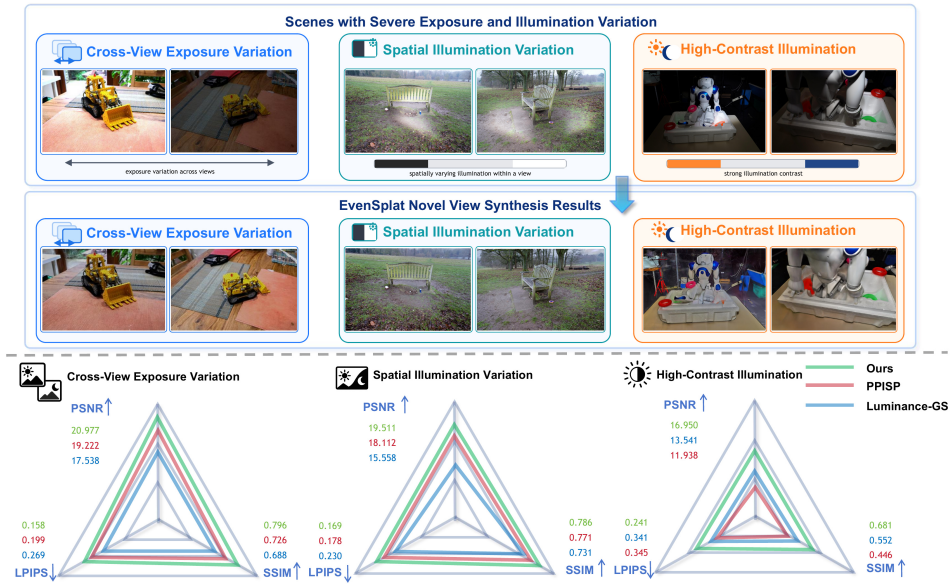

Figure 1: EvenSplat reconstructs base appearance from multi-view images with cross-view exposure variation (CEV), spatial illumination variation (SIV), and high-contrast illumination (HCI). Our method generally outperforms the state-of-the-art methods in PSNR, SSIM, and LPIPS on simulated and real-world 3D Gaussian Splatting novel view synthesis.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

Photographs of the same scene are rarely taken under the same light. A camera's exposure adapts from shot to shot (Shim et al., 2014), a flash or a window lights one side of an object more than the other (Debevec & Malik, 1997), and a bright sky or lamp splits a single frame into washed-out and shadowed regions (Petschnigg et al., 2004). For a human these differences are unremarkable; for a system that reconstructs a scene from many such photographs, they are a problem. Each image reports a different brightness and color for the same physical surface, and the reconstruction is left to reconcile observations that no longer agree.

Novel View Synthesis (NVS) methods such as Neural Radiance Fields (NeRF) (Mildenhall et al., 2020) and 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) reconstruct such scenes with high fidelity, but they assume photometric consistency across views. When the inputs disagree, fitting them faithfully entangles the lighting of each capture with the geometry and color of the scene, so the recovered model carries the artifacts of how it was photographed rather than what was photographed. These inconsistencies take three characteristic forms: *Cross-View Exposure Variation (CEV)*, as when a camera re-meters between shots; *Spatial Illumination Variation (SIV)* within a single view, as under a flash or a side window; and *High-Contrast Illumination (HCI)* with strong bright–dark imbalance, as under a harsh local light source. The first is a property of the camera, the latter two of the scene, but all three break the photometric agreement that reconstruction relies on (Fig. 1).

To evaluate these conditions beyond digital image perturbations, we assembled a nine-scene real-world dataset with exposure-bracketed and flash-illuminated captures. It covers the three settings with three CEV, four SIV, and two HCI scenes, together with evenly illuminated reference captures for appearance evaluation. This dataset forms a central part of our evaluation, complemented by controlled simulated benchmarks and a directional-lighting robustness experiment.

We introduce **EvenSplat**, a coupled 2D–3D appearance decomposition framework inspired by the reflectance–illumination factorization of Retinex theory (Land, 1977). Its central design is to use an *image-space decomposition* optimized together with the Gaussian representation. A multi-scale network predicts per-image illumination and derives a base-appearance estimate, while a *Gaussian-level illumination field* predicts illumination coefficients conditioned on position, surface normal, and viewing direction. Aligning the image-space and rendered illumination maps connects the two branches. Our contributions are: (1) A coupled 2D–3D decomposition framework that transfers online image-space illumination guidance to a Gaussian-level illumination field for reconstruction from exposure- and illumination-inconsistent inputs. (2) An image-specific correction combining global camera response and illumination-guided local exposure compensation for residual variation in training views. (3) A nine-scene real-world captured dataset spanning cross-view exposure variation, spatial illumination variation, and high-contrast illumination, with reference captures for evaluating appearance reconstruction.

2 RELATED WORK

2.1 NVS IN CHALLENGING LIGHTING CONDITIONS

Recent "in-the-wild" NVS methods explicitly model appearance variations that violate the photometric consistency assumed by standard NeRF and 3DGS. NeRF-W (Martin-Brualla et al., 2021) introduced per-image appearance embeddings for unconstrained photo collections. GS-W, SWAG, and WildGaussians (Zhang et al., 2024a; Dahmani et al., 2024; Kulhánek et al., 2024) transfer this idea to Gaussian scenes, while WildSplat and WilLaGS (Zhang et al., 2026; Bai et al., 2026) address unposed or strongly appearance-varying inputs.

A complementary line targets low-light image formation and exposure. RawNeRF (Mildenhall et al., 2022) and HDR-NeRF (Huang et al., 2022b) model HDR image formation, while HDR-GS (Cai et al., 2024) and PhysHDR-GS (Zeng et al., 2026) adapt this objective to Gaussian representations. Lighting-up NeRF (Wang et al., 2023), Aleth-NeRF (Cui et al., 2024), and Gaussian-DK (Ye et al., 2024) instead target dark observations. LITA-GS (Zhou et al., 2025) introduces illumination-invariant physical priors, while LL-Gaussian (Sun et al., 2025b) separates reflectance, illumination, and transient residuals. Luminance-GS (Cui et al., 2025), PPISP (Deutsch et al., 2026), and P2GS (Shimomura et al., 2026) focus on exposure, camera processing, and tone-mapping variation.

Other methods address view harmonization or explicitly changing illumination. CHROMA (Shin et al., 2026) harmonizes multi-view inputs using predicted bilateral grids. DarkGS and SunFaded (Zhang et al., 2024b; Chang et al., 2026) model camera-mounted moving lights, whereas L2DGS and Disco-GS (Kumar & Rajagopalan, 2026b;a) address dynamic or colored lighting.

A related but distinct line performs inverse rendering for relighting. GS-IR, Relightable 3D Gaussians, and GI-GS (Liang et al., 2024; Gao et al., 2024; Chen et al., 2025) recover normals, materials, visibility, and illumination for relighting. Subsequent methods extend this factorization to spatially varying materials, indirect illumination, and unconstrained lights (Sun et al., 2025a; Du et al., 2025; Zhang et al., 2025).

Even these advances remain brittle in extreme spatially uneven lighting, for instance, scenes with severe cross-view exposure discontinuities, mixed HDR regions, and strong, locally varying light sources across large indoor spaces. In such cases, appearance changes overwhelm per-image embeddings or local feature transfers and drive spurious densification or residual tone inconsistency. EvenSplat considers CEV, SIV, and HCI in one formulation without requiring calibrated light geometry. Instead of aiming to recover a full BRDF or a physically calibrated light, EvenSplat learns an illumination-consistent base appearance for novel-view synthesis by coupling image-space decomposition with a shared Gaussian illumination field while retaining the efficient, explicit rendering structure of 3DGS.

2.2 Image-Space Illumination Decomposition

Retinex theory (Land & McCann, 1971) separates observed appearance into reflectance and illumination. RetinexNet, KinD, and URetinex-Net (Wei et al., 2018; Zhang et al., 2019; Wu et al., 2022) replace handcrafted decomposition priors with learned illumination estimation for low-light enhancement. Later work broadens the restoration model: Retinexformer (Cai et al., 2023) uses illumination-guided attention; RetinexMamba and ECMamba (Bai et al., 2024; Dong et al., 2024) employ state-space architectures; and Fourier, video, and diffusion formulations address mixed exposure and general illumination degradation (Huang et al., 2022a; Liu et al., 2024; He et al., 2025).

Beyond Retinex, intrinsic image decomposition methods separate surface appearance from shading rather than directly producing an enhanced photograph. Recent work extends the classical albedo–shading formulation toward more complex illumination and richer intrinsic properties (Careaga & Aksoy, 2024; Dirik et al., 2026). Related multi-view inverse-rendering work further models material, spatial illumination, and camera response from captured images (Lin et al., 2025). These decomposition-based formulations motivate our image-space branch, which separates illumination from base appearance and uses the resulting decomposition to guide the learning of a shared 3D Gaussian illumination representation.

3 EvenSplat

EvenSplat learns a shared Gaussian base-appearance representation from views with inconsistent exposure and spatially uneven illumination. As shown in Fig. 2, an online image-space decomposition guides a Gaussian-level illumination field, while global and local image-specific corrections absorb residual capture variation during training. The resulting scene representation supports base-appearance rendering from unseen viewpoints without fitting a test image.

3.1 Gaussian Appearance and Illumination

Starting from a structure-from-motion point cloud (Schönberger & Frahm, 2016), we use the standard 3DGS geometry, opacity, and spherical-harmonic color parameters (Kerbl et al., 2023). We interpret the SH color as Gaussian base color, while a shared neural field predicts multiplicative illumination from Gaussian position, normal, and viewing direction:

$$\mathbf{C}_j^r = \mathrm{SH}(\mathbf{s}_j, \mathbf{v}_j), \qquad \mathbf{a}_j = \mathcal{F}_{\mathrm{illum}}(\boldsymbol{\mu}_j, \mathbf{n}_j, \mathbf{v}_j). \tag{1}$$

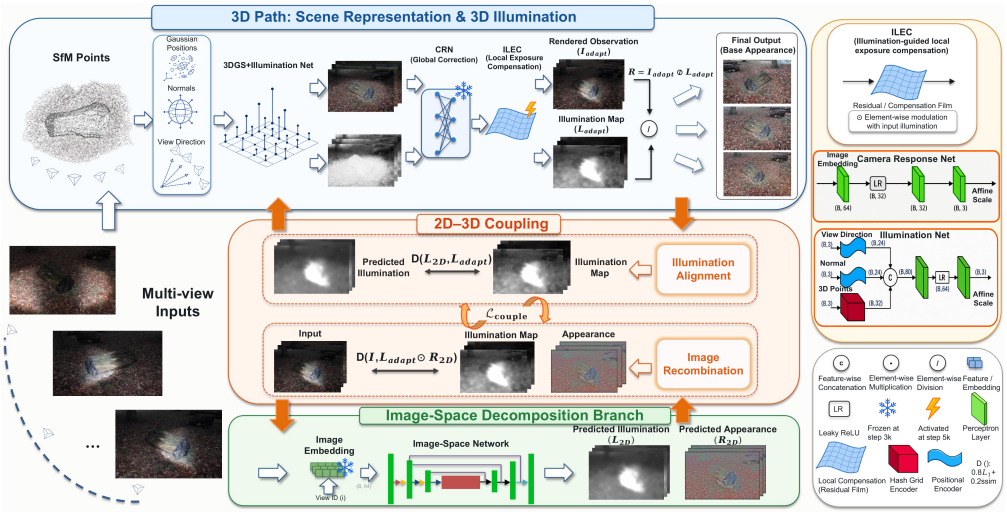

Figure 2: EvenSplat couples image-space decomposition with Gaussian-level illumination. Illumination alignment and image recombination connect the branches. CRN and ILEC absorb global and spatial training-image residuals. Base appearance is distinguished from the observation-fitting path.

Using the same front-to-back compositing weights $w_j^i(\mathbf{p})$, we rasterize the illumination-modulated observation and its illumination map for view i:

$$\hat{\mathbf{I}}^i_{\text{lit}}(\mathbf{p}) = \sum_j w_j^i(\mathbf{p})\big(\mathbf{a}_j \odot \mathbf{C}_j^r\big), \qquad \mathbf{L}^i_{3D}(\mathbf{p}) \;= \sum_j w_j^i(\mathbf{p})\mathbf{a}_j. \tag{2}$$

Illumination therefore acts before compositing and is shared across views. Because the decomposition is learned in sRGB without radiometric calibration and SH base color can remain view-dependent, we describe $\mathbf{C}_j^r$ as *Gaussian base color*, not physical albedo. Image-specific adaptation (Sec. 3.2) and cross-domain supervision (Sec. 3.4) constrain this ambiguity without guaranteeing a unique intrinsic decomposition. Compositing and interpretation details are given in Appendix A.1.1.

3.2 Image-Specific Appearance Corrections

The shared illumination field cannot represent every capture-specific photometric residual. A *camera response network (CRN)* predicts a spatially uniform per-image gain $\mathbf{c}^i$ within the training image-formation path. During training, the CRN compensates for exposure and color-response variations consistently to the rendered observation and illumination map:

$$\hat{\mathbf{I}}^i_{\text{cam}} = \mathbf{c}^i \odot \hat{\mathbf{I}}^i_{\text{lit}}, \qquad \mathbf{L}^i_{\text{cam}} = \mathbf{c}^i \odot \mathbf{L}^i_{3D}. \tag{3}$$

Illumination-guided local exposure compensation (ILEC) models remaining spatial residuals with a low-resolution log-gain map $\Delta^i_{\text{loc}} \in \mathbb{R}^{h\times w\times 1}$. Bilinear upsampling and a gradual activation ramp $r(t)$ yield the adapted observation and illumination map:

$$\hat{\mathbf{I}}^i_{\text{adapt}} = \hat{\mathbf{I}}^i_{\text{cam}} \odot \exp\big(r(t)\,\text{Up}(\Delta^i_{\text{loc}})\big)\,, \qquad \mathbf{L}^i_{\text{adapt}} = \mathbf{L}^i_{\text{cam}} \odot \exp\big(r(t)\,\text{Up}(\Delta^i_{\text{loc}})\big)\,. \tag{4}$$

The low-resolution parameterization limits texture memorization, while illumination-guided magnitude, shrinkage, and promotion terms regularize the map (Appendix A.1.4). CRN and ILEC absorb image-specific photometric variation during training rather than becoming part of the transferable scene representation. They affect training but cancel from the base-appearance ratio because they multiply both paths. At novel views, their image-specific gains are set to the identity.

3.3 Online Image-Space Decomposition

Reconstruction alone leaves the allocation of variation between predicted base appearance and illumination ambiguous. The image-space branch guides this separation through online updates during

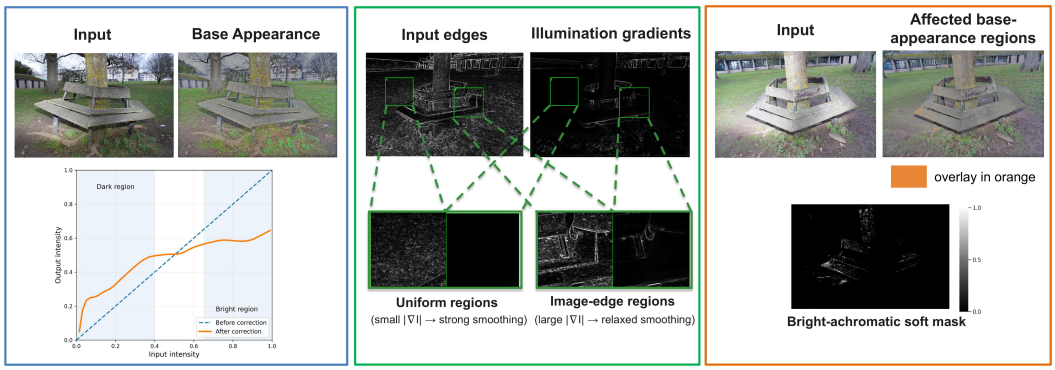

Figure 3: Image-space regularization, from left to right: adaptive curve constraints on base-appearance intensity; edge-aware illumination smoothness; and white preservation using a bright-achromatic soft mask.

scene fitting. For each training image, a multi-scale image-space network conditioned on a learnable image embedding ($\mathbf{e}^i \in \mathbb{R}^{d_{\text{emb}}}$) predicts channel-shared illumination and derives predicted base appearance by division:

$$\mathbf{L}^i_{2D} = 0.05 + 3.95\sigma\big(\mathcal{F}_{2D}(\mathbf{I}^i, \mathbf{e}^i)\big), \qquad \mathbf{R}^i_{2D} = \mathbf{I}^i \oslash (\mathbf{L}^i_{2D} + \epsilon), \tag{5}$$

where $\sigma(x) = 1/(1+e^{-x})$ denotes the sigmoid activation function. The affine scaling constrains $\mathbf{L}^i_{2D}$ to $(0.05, 4.0)$, enabling both attenuation and amplification while preventing unstable near-zero division. The learnable image embedding captures image-specific exposure and appearance variations through FiLM modulation (Perez et al., 2017). The network architecture is in Fig. 9.

As illustrated in Fig. 3, we regularize the decomposition with three priors. *Adaptive curve regularization* selects excessively dark or bright base-appearance pixels and pulls them toward moderate target intensities. *Edge-aware smoothness* penalizes illumination gradients with weights derived from input-image gradients, encouraging smooth lighting without blurring observed boundaries. *White preservation* uses a soft bright, low-chroma CIELAB mask to retain approximately achromatic regions of the input. The complete loss definitions and hyperparameters are given in Appendix A.1.2. This branch is optimized jointly with the scene rather than used as independent preprocessing.

3.4 Coupled Optimization and Rendering

Our *coupled optimization* connects the image-space decomposition to the Gaussian illumination representation. We use a common image discrepancy combining ℓ_1 reconstruction and SSIM:

$$D(\mathbf{A}, \mathbf{B}) = (1 - \lambda_{\text{ssim}})\|\mathbf{A} - \mathbf{B}\|_1 + \lambda_{\text{ssim}}[1 - \text{SSIM}(\mathbf{A}, \mathbf{B})]. \tag{6}$$

The cross-domain *coupled loss* aligns the adapted 3D illumination with the image-space estimate and recombines it with $\mathbf{R}_{2D}$ to reconstruct the input:

$$\mathcal{L}_{\text{couple}} = \mathcal{L}_{\text{align}} + \mathcal{L}_{\text{recomb}} = \lambda_{\text{align}} D(\mathbf{L}_{\text{adapt}}, \mathbf{L}_{2D}) + \lambda_{\text{recomb}} D(\mathbf{L}_{\text{adapt}} \odot \mathbf{R}_{2D}, \mathbf{I}). \tag{7}$$

Direct alignment compares illumination maps, whereas recombination measures compatibility in the observed-image domain. Both branches enter this objective and parameters are updated during training when activated. When activated, no stop-gradient is applied, so gradients from the coupling loss propagate to both the image-space network and the Gaussian illumination branch.

Observation fidelity supervises the corrected Gaussian rendering to agree with captured images, $\mathcal{L}_{\text{obs}} = D(\hat{\mathbf{I}}_{\text{adapt}}, \mathbf{I})$. This route retains the captured image as the reconstruction target throughout scene fitting.

Collecting weighted regularizers into $\mathcal{R}$, we write the objective as:

$$\mathcal{L} = \lambda_{\text{cpl}} \mathcal{L}_{\text{couple}} + \lambda_{\text{obs}} \mathcal{L}_{\text{obs}} + \mathcal{R}. \tag{8}$$

$\mathcal{R}$ contains the image-space regularizers, Gaussian appearance–illumination gradient exclusion, higher-order SH regularization, and ILEC local-map penalties detailed in Appendix A.1.

Staged fitting. The reported training protocol runs for 10,000 iterations. The illumination field, image embeddings, and CRN are frozen after 3,000 iterations; ILEC is activated at iteration 5,000. The image-space decomposition network and $\mathcal{L}_{\text{couple}}$ remain active throughout training. Novel-view rendering uses only the shared Gaussian appearance and illumination field.

4 EXPERIMENTS

We test the central claim behind EvenSplat: coupling image-space priors with a shared 3D illumination model is most useful when photometric variation cannot be explained by a global per-view correction. We evaluate physically captured CEV, SIV, and HCI scenes, repeat the comparison under controlled perturbations, and use cross-lighting and component ablations to probe what is learned.

4.1 DATASETS

Real-world benchmark. Our nine-scene benchmark contains three Cross-View Exposure Variation (CEV), four Spatial Illumination Variation (SIV), and two High-Contrast Illumination (HCI) scenes spanning indoor and outdoor subjects. Unlike a purely synthetic benchmark, these captures retain camera response, highlight saturation, flash falloff, and cast shadows. A second collection under more uniform illumination provides an appearance reference for each scene. Figure 4 demonstrates the three settings; capture parameters, protocols, and structure-from-motion initialization are reported in Appendix A.3.1 and Table 8.

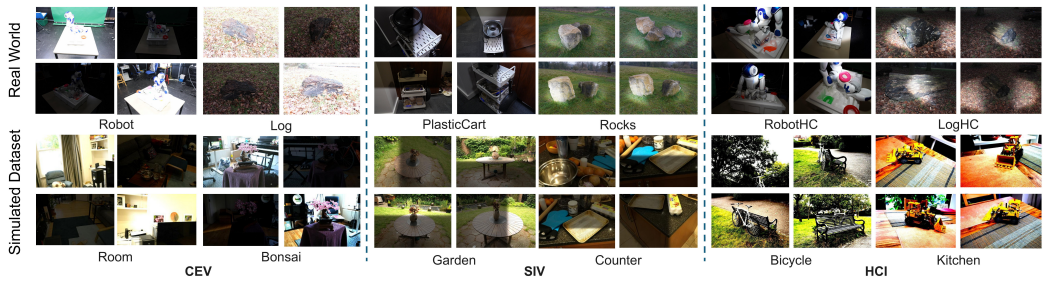

Figure 4: Gallery of the three illumination settings in the real-captured and simulated datasets. CEV changes exposure across views, SIV introduces spatial illumination variation within images, and HCI produces a strong bright–dark imbalance.

Simulated benchmark. We additionally perturb six Mip-NeRF 360 scenes (Barron et al., 2022) while retaining the original images as references. CEV follows the varying-exposure benchmark of Cui et al. (2025); SIV applies smoothly blended exposure factors to two image regions; and HCI increases training-image contrast. Exact operators are in Appendix A.3.2.

4.2 EXPERIMENTAL RESULTS

Training and evaluation protocol. All training and evaluation runs are conducted on two NVIDIA GeForce RTX 3090 GPUs. EvenSplat is implemented with gsplat (Ye et al., 2025) and trained for 10,000 iterations using the staged fitting in Sec. 3.4. We compare against 3DGS (Kerbl et al., 2023), GS-W (Zhang et al., 2024a), Bilateral Grid (Wang et al., 2024), PPISP (Deutsch et al., 2026), and Luminance-GS (Cui et al., 2025). Full settings are in Appendix A.2.

Real-world reconstruction. Tables 1 and 3 reveal a clear condition-dependent pattern. EvenSplat leads every metric for SIV and HCI, including a 3.41 dB PSNR gain over the strongest HCI baseline and a 1.40 dB gain over PPISP for SIV. These are precisely the regimes in which a global exposure or ISP correction cannot explain illumination that changes within an image. Figure 5 shows the corresponding qualitative distinction: several baselines either preserve the capture shadow or flatten bright regions, whereas EvenSplat recovers detail across both sides of the illumination boundary. For the simpler CEV setting, EvenSplat obtains the best PSNR but PPISP retains better SSIM and LPIPS.

Table 1: Scene-averaged real-world results. Entries are PSNR↑/SSIM↑/LPIPS↓. Best values are bold with light-red shading; second-best values are underlined with light-yellow shading.

Method	CEV (3 scenes)	SIV (4 scenes)	HCI (2 scenes)
Bilateral Grid	16.337/0.549/0.524	17.994/0.743/0.227	12.190/0.404/0.349
3DGS	14.197/0.461/0.609	17.902/0.734/0.250	11.702/0.427/0.390
GS-W	12.404/0.523/0.362	17.742/0.702/0.243	11.801/0.425/0.379
Luminance-GS	16.585/0.538/0.485	15.558/0.731/0.230	13.541/0.553/0.341
PPISP	17.794/**0.723**/**0.269**	18.112/0.771/0.178	11.938/0.446/0.345
EvenSplat	**19.526**/0.708/0.305	**19.511**/**0.786**/**0.169**	**16.950**/**0.682**/**0.241**

Table 2: Scene-averaged results on simulated data. Entries are PSNR↑/SSIM↑/LPIPS↓. Best values are bold with light-red shading; second-best values are underlined with yellow shading.

Method	CEV (6 scenes)	SIV (6 scenes)	HCI (6 scenes)
Bilateral Grid	18.777/0.674/0.253	19.102/0.773/0.144	14.056/0.496/0.322
3DGS	17.288/0.605/0.312	18.238/0.773/0.193	13.000/0.452/0.358
GS-W	17.517/0.640/0.215	18.371/0.776/0.173	12.987/0.451/0.360
Luminance-GS	17.538/0.688/0.269	14.485/0.682/0.229	12.963/0.568/0.347
PPISP	19.222/0.726/0.199	20.343/0.814/0.151	15.174/**0.589**/0.302
EvenSplat	**20.977**/**0.796**/**0.158**	**22.000**/**0.862**/**0.097**	**16.736**/0.501/**0.301**

To isolate whether the improvements come merely from the image-space tone prior, Appendix Table 10 additionally reports *Tone-Matched 3DGS*, which applies a single global tone/exposure match preprocessing to vanilla 3DGS using the same mid-tone target of 0.5. This control improves PSNR on two of the three CEV scenes, but it underperforms vanilla 3DGS on every SIV and HCI scene and remains well below EvenSplat. The gains in these more challenging settings therefore cannot be explained by global tone matching alone.

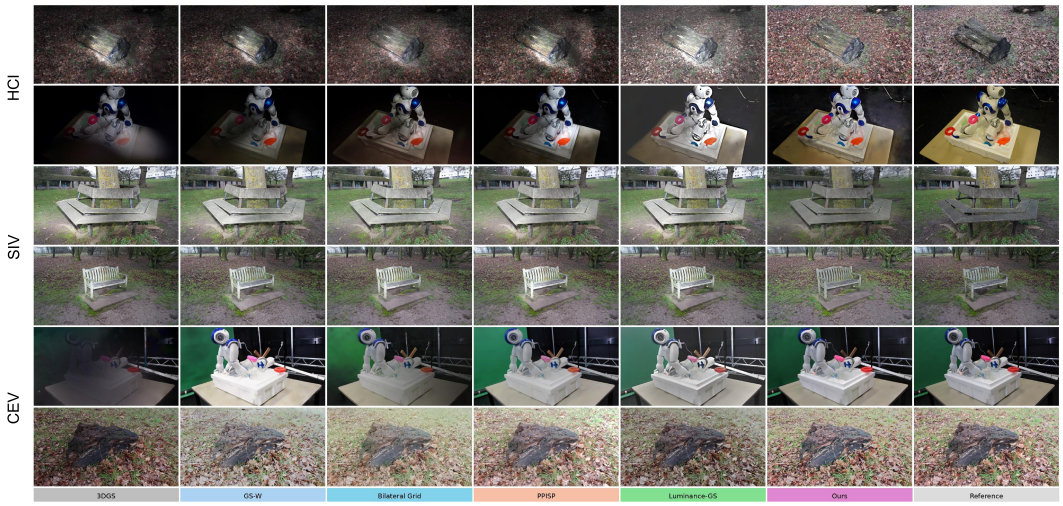

Figure 5: Real-world comparisons. Columns show 3DGS, GS-W, Bilateral Grid, PPISP, Luminance-GS, EvenSplat, and the captured reference. Rows 1–2 show HCI, rows 3–4 SIV, and rows 5–6 CEV. EvenSplat most clearly separates appearance from illumination in the spatially uneven HCI and SIV.

Table 3: Selected real-world scenes under CEV, SIV, and HCI. Each method entry reports PSNR↑/SSIM↑/LPIPS↓. Best values are bold with light-red shading; second-best values are underlined with light-yellow shading. Ties at the reported precision share the same rank. All nine scenes and the complete evaluation results are reported in Appendix Table 10.

Setting	Scene	Bilateral Grid	3DGS	GS-W	Luminance-GS	PPISP	EvenSplat
HCI	*LogHC*	11.312/0.438/0.302	10.533/0.393/0.328	10.686/0.396/0.325	12.614/0.565/0.287	10.568/0.400/0.322	**16.718**/**0.669**/**0.220**
	RobotHC	13.068/0.370/0.395	12.870/0.460/0.452	12.916/0.454/0.432	14.468/0.540/0.394	13.307/0.491/0.367	**17.181**/**0.694**/**0.262**
SIV	*CircularBench*	16.147/0.690/0.233	15.953/0.702/0.244	16.134/0.654/0.257	14.928/0.725/0.214	15.854/**0.741**/0.191	**17.420**/0.739/**0.186**
	ParkBench	19.907/0.803/0.138	20.310/0.810/0.135	20.043/0.761/0.168	18.248/0.803/0.136	19.810/**0.816**/0.142	**21.037**/**0.816**/**0.117**
CEV	*Robot*	16.542/0.677/0.429	14.564/0.601/0.519	14.814/0.798/0.224	16.325/0.656/0.304	21.829/0.861/0.133	**22.781**/**0.884**/**0.097**
	Log	20.130/0.783/0.216	17.072/0.651/0.367	13.432/0.676/0.220	17.745/0.734/0.283	20.547/**0.871**/**0.115**	**20.648**/0.812/0.157

Controlled simulated reconstruction. The controlled results in Table 2 show that the real-world gains are not specific to our capture pipeline. EvenSplat leads all metrics for CEV and SIV, improving PSNR over PPISP by 1.76 and 1.66 dB, respectively. Under HCI it retains the best PSNR

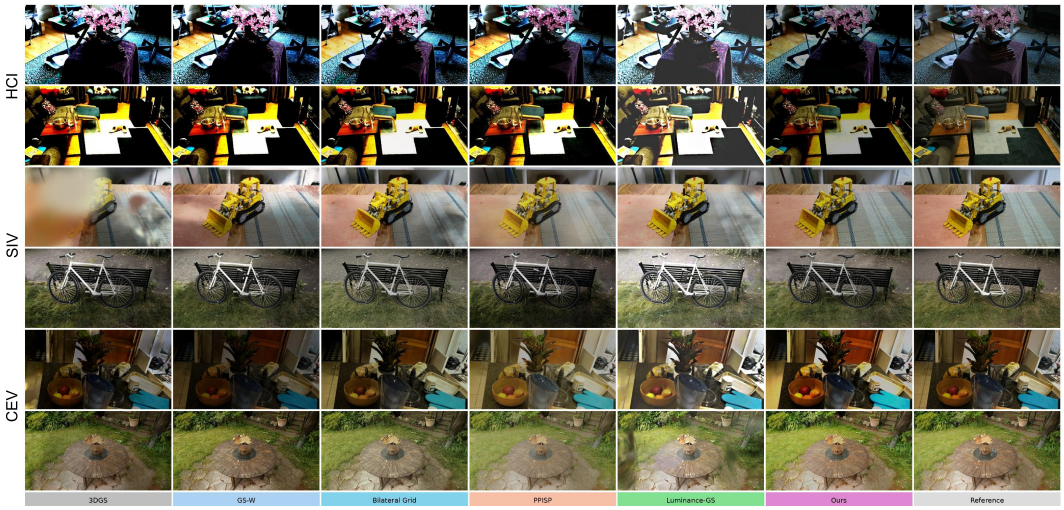

Figure 6: Simulated HCI, SIV, and CEV comparisons. EvenSplat reduces illumination leakage while retaining scene texture; HCI remains the most challenging setting because contrast amplification clips information.

(+1.56 dB) but trails PPISP in SSIM, indicating that correction under severe clipping trades pixel agreement against structural similarity. Per-scene results in Appendix A.4 show that this trend is not driven by a single scene.

Cross-lighting appearance consistency. This test removes the need to choose a lighting-neutral reference: The two capture sets differ only in illumination direction, with light incident predominantly from the left or right, while scene geometry and materials remain unchanged. We therefore assess whether reconstructions from these differently illuminated observations yield similar base appearance.

EvenSplat reaches 15.57 dB, 2.84 dB above Luminance-GS, the strongest baseline. The gain indicates that the recovered representation is less tied to the direction of the capture light. The score alone cannot establish the accuracy of the recovered base appearance, Figure 7 compares the resulting appearance renderings and color-chart crops.

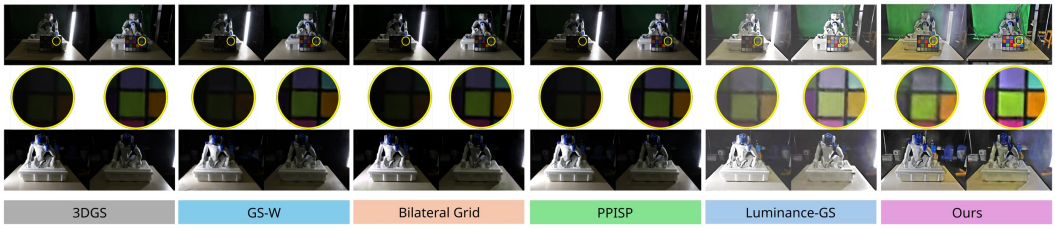

Figure 7: Cross-lighting appearance consistency. The capture sets differ only in left- versus right-dominant illumination. Paired appearance renderings and color-chart crops show the resulting reconstructions. PSNR is computed between corresponding left- and right-dataset appearance renderings, using one as the reference for the other; higher values indicate greater consistency across lighting conditions. PSNR (dB): 3DGS 10.5628; GS-W 10.6427; Bilateral Grid 10.3909; PPISP 10.3641; Luminance-GS 12.7311 (second-best); EvenSplat **15.5723** (best).

Dual outputs evaluation on HDR-NeRF dataset. Our method design makes prediction on both *Rendered Observation* and *Base Appearance* possible. On the simulated HDR-NeRF dataset (Huang et al., 2022b), illumination is severe but consistent across views. EvenSplat retains high-quality full-scene rendering (37.6 dB mean PSNR across scenes) while additionally extracting base appearance, with the highest base appearance PSNR (Appendix Table 12) on every scene against ground truth

under parallel uniform light. If the shared light saturates every view, clipping removes the original color for all methods—an identifiability limit of the observations, not a method-specific failure. Full results are in Appendix A.4.

4.3 Ablation Study

Coupled model and Gaussian-level components. Table 4 supports this design. Removing image-space decomposition costs 3.60 dB on CEV and 4.94 dB on HCI, yet using that branch alone is also substantially worse: the 2D prior needs a shared 3D explanation. CRN removal degrades every setting, confirming the role of global per-image correction. ILEC is sharply specialized: removing it costs 2.88 dB on HCI but is nearly neutral on CEV and SIV. Thus global response, shared illumination, and local compensation address different error scales; the full model offers the most reliable balance across conditions. Appendix A.4 further isolates the ILEC regularizers.

Table 4: Ablation of EvenSplat on the real-world datasets (PSNR/SSIM/LPIPS).

Variant	CEV	SIV	HCI
EvenSplat (Full)	19.526/0.708/0.305	19.511/0.786/0.169	16.950/0.682/0.241
No Image-Space Decomp.	15.924/0.646/0.374	19.216/0.710/0.242	12.012/0.438/0.324
No Normals	16.839/0.661/0.383	19.506/0.782/0.172	16.712/0.674/0.246
No View Directions	16.916/0.665/0.366	19.601/0.786/0.165	16.745/0.670/0.247
No CRN	16.202/0.649/0.411	18.094/0.766/0.185	15.184/0.632/0.280
No ILEC	19.599/0.707/0.308	19.467/0.782/0.168	14.066/0.561/0.266
Image-Space Decomp. Only	15.632/0.587/0.382	12.601/0.483/0.583	15.092/0.531/0.425

Table 5: Ablation only Image-space Decomposition branch's prediction on simulated dataset.

Variant	CEV	SIV	HCI
Image-Space Decomp. Only	17.962/0.706/0.240	15.260/0.676/0.263	16.478/0.547/0.376
No Loss_ (Adaptive Curve)	17.786/0.702/0.249	14.213/0.622/0.284	13.540/0.524/0.396
No Loss_ (Smooth Edge-Aware)	17.568/0.560/0.405	13.922/0.480/0.452	16.508/0.504/0.434
No Loss_ (White Preservation)	17.118/0.688/0.237	14.245/0.626/0.284	15.063/0.545/0.380
Colored Illumination	17.809/0.673/0.239	15.420/0.671/0.229	17.086/0.526/0.308

Image-space decomposition components. Table 5 separates the roles of the three priors in Fig. 3. Adaptive curve regularization is most important under HCI, where its removal costs 2.94 dB. Edge-aware smoothness drives structural and perceptual quality, producing the largest SSIM and LPIPS degradation under CEV and SIV when removed. White preservation provides a smaller but consistent PSNR benefit. Allowing colored illumination improves some PSNR and LPIPS values but lowers SSIM in every setting, reflecting the extra ambiguity of channel-dependent illumination. We therefore retain the more constrained achromatic model for stable separation rather than optimizing one metric through additional decomposition freedom.

4.4 Real Application: Digital Heritage

Uneven illumination limits the legibility of reconstructed cultural assets. We apply EvenSplat to ceiling digitization in a large indoor heritage site (Edwards et al., 2025), where windows and local lighting leave reliefs unevenly exposed. Figure 8 shows that the resulting novel views reveal architectural detail across darker regions, illustrating a practical use beyond the benchmark scenes.



Figure 8: Ceiling digitization under uneven illumination (Edwards et al., 2025). Each panel pairs a held-out photograph (left) with an EvenSplat rendered novel view (right).

5 Conclusion

EvenSplat couples image-space decomposition with shared Gaussian illumination and global–local view-specific correction. Across physical and simulated benchmarks, its largest gains occur under spatially uneven and high-contrast illumination, where global appearance models are insufficient; ablations confirm that the 2D and 3D branches are complementary. The result is a practical route to more illumination-consistent Gaussian scene representations from ordinary sRGB captures. Limitations include sRGB-domain ambiguity and saturation, the training-image-specific ILEC parameters, and unmodeled specularities, motion blur, and sensor noise.

REFERENCES

Jiesong Bai, Yuhao Yin, Qiyuan He, Yuanxian Li, and Xiaofeng Zhang. Retinexmamba: Retinex-based mamba for low-light image enhancement. In *arXiv preprint arXiv:2405.03349*, 2024.

Yuhao Bai, Qianqiu Tan, Lilong Chen, Huanhuan Lv, and Lijun Chen. Willags: Latent-conditional 3d appearance fields for robust gaussian splatting in-the-wild. In *European Conference on Computer Vision*, 2026.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. *CVPR*, 2022.

Yuanhao Cai, Zihao Xiao, Yixun Liang, Minghan Qin, Yulun Zhang, Xiaokang Yang, Yaoyao Liu, and Alan Yuille. Hdr-gs: Efficient high dynamic range novel view synthesis at 1000x speed via gaussian splatting. In *NeurIPS*, 2024.

Yuzhi Cai, Haoran Bian, Jing Lin, Haolin 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*, pp. 12504–12513, 2023.

Chris Careaga and Yağız Aksoy. Colorful diffuse intrinsic image decomposition in the wild. *ACM Transactions on Graphics*, 43(6), 2024. doi: 10.1145/3687984.

Wenjie Chang, Tianle Ding, Wenfei Yang, and Tianzhu Zhang. Sunfaded: Illumination-aware gaussian splatting for dark scenes with camera-mounted active lighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 40876–40885, 2026.

Hongze Chen, Zehong Lin, and Jun Zhang. Gi-gs: Global illumination decomposition on gaussian splatting for inverse rendering. In *International Conference on Learning Representations*, 2025.

Ziteng Cui, Lin Gu, Xiao Sun, Xianzheng Ma, Yu Qiao, and Tatsuya Harada. Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

Ziteng Cui, Xuangeng Chu, and Tatsuya Harada. Luminance-gs: Adapting 3d gaussian splatting to challenging lighting conditions with view-adaptive curve adjustment. In *CVPR*, 2025.

Hiba Dahmani, Moussab Bennehar, Nathan Piasco, Luis Roldao, and Dzmitry Tsishkou. Swag: Splatting in the wild images with appearance-conditioned gaussians. In *European Conference on Computer Vision*, pp. 325–340, 2024.

Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 369–378, 1997.

Isaac Deutsch, Nicolas Moënne-Loccoz, Gavriel State, and Zan Gojcic. Ppisp: Physically-plausible compensation and control of photometric variations in radiance field reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7289–7298, 2026.

Alara Dirik, Tuanfeng Yang Wang, Duygu Ceylan, Stefanos Zafeiriou, and Anna Frühstück. Reasonx: Mllm-guided intrinsic image decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30802–30812, June 2026.

Wei Dong, Han Zhou, Yulun Zhang, Xiaohong Liu, and Jun Chen. Ecmamba: Consolidating selective state space model with retinex guidance for efficient multiple exposure correction. *arXiv preprint arXiv:2410.21535*, 2024. URL https://arxiv.org/abs/2410.21535.

Kang Du, Zhihao Liang, Yulin Shen, and Zeyu Wang. Gs-id: Illumination decomposition on gaussian splatting via adaptive light aggregation and diffusion-guided material priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 26220–26229, 2025.

Jacob Edwards, Stephen D. Laycock, Thomas Roebuck, YingLiang Ma, and Cheng Wang. Level-of-Detail Digitization of High Ceilings in Virtual Reality. In *Digital Heritage*. The Eurographics Association, 2025. ISBN 978-3-03868-277-6. doi: 10.2312/dh.20253301.

Jian Gao, Chun Gu, Youtian Lin, Zhihao Li, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *European Conference on Computer Vision*, pp. 73–89, 2024. doi: 10.1007/978-3-031-72995-9_5.

Chunming He, Chengyu Fang, Yulun Zhang, Tian Ye, Kai Li, Longxiang Tang, Zhenhua Guo, Xiu Li, and Sina Farsiu. Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Jie Huang, Yajing Liu, Fengmei Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *European Conference on Computer Vision*, 2022a.

Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18398–18408, 2022b.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4), July 2023. ISSN 0730-0301. doi: 10.1145/3592433.

Jonáš Kulhánek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. Wildgaussians: 3d gaussian splatting in the wild. *arXiv preprint arXiv:2407.08447*, 2024.

Ashish Kumar and A. N. Rajagopalan. Disco-gs: Gaussian splatting in dynamic color lighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11891–11900, 2026a.

Ashish Kumar and A. N. Rajagopalan. L2dgs: Low-light dynamic gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19096–19106, 2026b.

Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

Edwin Herbert Land and John J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61 1:1–11, 1971.

Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21644–21653, 2024.

Chih-Hao Lin, Jia-Bin Huang, Zhengqin Li, Zhao Dong, Christian Richardt, Tuotuo Li, Michael Zollhöfer, Johannes Kopf, Shenlong Wang, and Changil Kim. Iris: Inverse rendering of indoor scenes from low dynamic range images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 465–474, June 2025.

Hao Liu, Ruofan Zhang, et al. Learning exposure correction in dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *CoRR*, abs/2003.08934, 2020. URL `https://arxiv.org/abs/2003.08934`.

Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16190–16199, 2022.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer, 2017. URL https://arxiv.org/abs/1709.07871.

Georg Petschnigg, Maneesh Agrawala, Hugues Hoppe, Richard Szeliski, Michael Cohen, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, 23(3):664–672, 2004.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Inwook Shim, Joon-Young Lee, and In So Kweon. Auto-adjusting camera exposure for outdoor robotics using gradient information. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.

Kota Shimomura, Hidehisa Arai, Tsubasa Takahashi, Takayoshi Yamashita, and Hironobu Fujiyoshi. P2gs: Physical prior-guided gaussian splatting for photometrically consistent urban reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11736–11745, 2026.

Jisu Shin, Richard Shaw, Seunghyun Shin, Zhensong Zhang, Hae-Gon Jeon, and Eduardo Perez-Pellitero. Chroma: Consistent harmonization of multi-view appearance via bilateral grid prediction. In *International Conference on Learning Representations*, 2026.

Hanxiao Sun, Yupeng Gao, Jin Xie, Jian Yang, and Beibei Wang. Svg-ir: Spatially-varying gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16143–16152, 2025a.

Hao Sun, Fenggen Yu, Huiyao Xu, Tao Zhang, and Changqing Zou. Ll-gaussian: Low-light scene reconstruction and enhancement via gaussian splatting for novel view synthesis. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 4261–4270, 2025b. doi: 10.1145/3746027.3755375.

Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson W. H. Lau. Lighting up nerf via unsupervised decomposition and enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12632–12641, 2023.

Yuehao Wang, Chaoyi Wang, Bingchen Gong, and Tianfan Xue. Bilateral guided radiance field processing. *ACM Transactions on Graphics*, 2024.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Proceedings of the British Machine Vision Conference*, 2018.

Weisheng Wu, Jiazhen Weng, Pei Zhang, Xueyang Wang, Wenhan Yang, and Jiayi Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5901–5910, 2022.

Sheng Ye, Zhen-Hui Dong, Yubin Hu, Yu-Hui Wen, and Yong-Jin Liu. Gaussian in the dark: Real-time view synthesis from inconsistent dark images using gaussian splatting. *Computer Graphics Forum*, 43(7):e15213, 2024. doi: 10.1111/cgf.15213.

Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025.

Huimin Zeng, Yue Bai, Hailing Wang, and Yun Fu. Physically inspired gaussian splatting for hdr novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11808–11817, 2026.

Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. In *European Conference on Computer Vision (ECCV)*, pp. 341–359. Springer, 2024a.

Qi Zhang, Chi Huang, Qian Zhang, Nan Li, and Wei Feng. Su-rgs: Relightable 3d gaussian splatting from sparse views under unconstrained illuminations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 26859–26868, 2025.

Tianyi Zhang, Kaining Huang, Weiming Zhi, and Matthew Johnson-Roberson. Darkgs: Learning neural illumination and 3d gaussians relighting for robotic exploration in the dark. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024b.

Xiyu Zhang, Jingyu Zhuang, Hongjia Zhai, Zizheng Yan, Jinwei Chen, Guofeng Zhang, and Qingnan Fan. Wildsplat: Feedforward gaussian splatting from unposed in-the-wild images. In *European Conference on Computer Vision*, pp. 20–37, 2026.

Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *CVPR*, 2018.

Ying Zhang, Jian Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1632–1640, 2019.

Han Zhou, Wei Dong, and Jun Chen. Lita-gs: Illumination-agnostic novel view synthesis via reference-free 3d gaussian splatting and physical priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21580–21589, 2025.

A Appendix

A.1 Method and Loss Details

This section details Gaussian compositing, image-space decomposition, and the regularizers grouped in $\mathcal{R}$ in Eq. equation 8.

A.1.1 Gaussian Compositing and Appearance Estimates

For the ordered Gaussians contributing to pixel $\mathbf{p}$, the standard front-to-back compositing weight is

$$w_j(\mathbf{p}) = T_j(\mathbf{p})\alpha_j(\mathbf{p}), \qquad T_j(\mathbf{p}) = \prod_{k<j}(1-\alpha_k(\mathbf{p})). \tag{9}$$

Here $\alpha_j(\mathbf{p})$ is Gaussian j's opacity contribution at the pixel and $T_j(\mathbf{p})$ is the transmittance through preceding Gaussians. The same weights render illumination-modulated color and illumination in Eq. equation 2.

The base-appearance quotient documented in the preceding formulation is

$$\hat{\mathbf{R}}_{\mathrm{3D}} = \hat{\mathbf{I}}_{\mathrm{adapt}} \oslash (\mathbf{L}_{adapt} + \epsilon). \tag{10}$$

A quotient of composited maps generally differs from $\sum_j w_j \mathbf{C}_j^r$ during training, it also retains contributions from image-specific corrections.

A.1.2 Image-Space Architecture and Losses

Image-Space Network. Given an input image $\mathbf{I}^i$ and a view-specific embedding $\mathbf{e}^i$, the image-space network uses a lightweight U-Net-style (Ronneberger et al., 2015) architecture with Spatially-FiLM (Perez et al., 2017) conditioning, Retinex blocks, and channel–spatial attention. Multi-scale encoder features are fused with decoder features through skip connections to predict the illumination map $\mathbf{L}_{2D}^i$. The base appearance is then obtained by the element-wise division in Eq. equation 5.

The image-space branch predicts illumination $\mathbf{L}_{2D}$ and derives base appearance $\mathbf{R}_{2D}$. Its intrinsic regularization groups the three perspectives illustrated in Fig. 3:

$$\mathcal{L}_{\mathrm{prior}} = \lambda_{\mathrm{curve}}\mathcal{L}_{\mathrm{curve}} + \lambda_{\mathrm{smooth}}\mathcal{L}_{\mathrm{smooth}}^{2D} + \lambda_{\mathrm{white}}\mathcal{L}_{\mathrm{white}}. \tag{11}$$

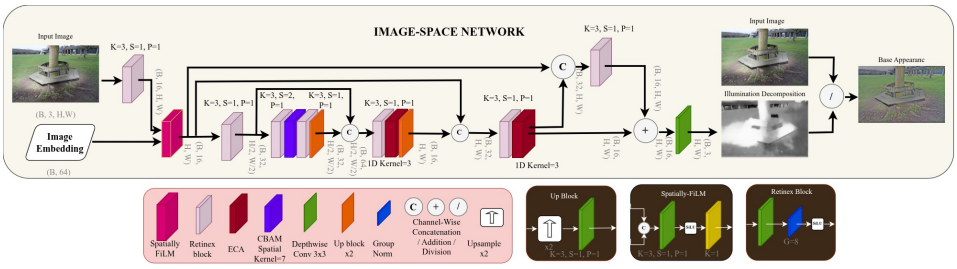

Figure 9: Architecture of the image-space network for base-appearance and illumination decomposition.

Eq. equation 7 connects this branch to the scene through illumination alignment and image recombination. Below, N denotes the number of pixels, and ∇_x and ∇_y denote horizontal and vertical image gradients. Scalar intensity targets are broadcast to the color channels; if a selected set is empty, its corresponding term is taken as zero.

Adaptive Curve Loss. The adaptive curve loss constrains overly dark and bright base-appearance regions:

$$\begin{aligned}\mathcal{L}_{\text{curve}} =& \eta_{\text{dark}} \frac{1}{|\Omega_{\text{dark}}|} \sum_{p \in \Omega_{\text{dark}}} |\mathbf{R}_{2D}(p) - \rho_{\text{dark}}| \\ &+ \eta_{\text{bright}} \frac{1}{|\Omega_{\text{bright}}|} \sum_{p \in \Omega_{\text{bright}}} |\mathbf{R}_{2D}(p) - \rho_{\text{bright}}|\end{aligned} \tag{12}$$

Here Ω_{dark} and Ω_{bright} select base-appearance intensities below T_{dark} and above T_{bright}, respectively. The constants ρ_{dark} and ρ_{bright} are the preferred intensities for the two sets: the thresholds select where the loss applies, while the targets determine the direction of correction. The coefficients η_{dark} and η_{bright} weight the two target penalties within this loss. The single-branch settings in Table 6 use targets 0.45 and 0.55 with selection thresholds 0.25 and 0.65. These target symbols leave α_j reserved for Gaussian opacity.

Edge-Aware Smoothness Loss. To encourage smooth illumination while preserving image edges, we use

$$\mathcal{L}^{2D}_{\text{smooth}} = \frac{1}{N} \sum_{p,d} \exp\left(-\frac{|\nabla_d \mathbf{I}_{\text{gray}}(p)|}{\gamma_{\text{edge}}}\right) |\nabla_d \mathbf{L}_{2D,\text{gray}}(p)|, \tag{13}$$

Here $d \in \{x, y\}$ indexes the gradient direction, and the subscript gray denotes a grayscale map. The parameter $\gamma_{\text{edge}} > 0$ sets the input-gradient scale at which the smoothness weight decays. A smaller value reduces smoothing across weaker image edges, while a larger value retains smoothing across stronger edges.

White Preservation Loss. A soft mask identifies bright and approximately achromatic regions in the input image's CIELAB representation:

$$M_{\text{white}}(p) = \sigma\left(\kappa_{\text{white}}(L^*(p) - T_{\text{white}})\right) \exp\left(-\frac{(a^*(p))^2 + (b^*(p))^2}{2s^2_{\text{chroma}}}\right). \tag{14}$$

Here L^* is CIELAB lightness, and a^* and b^* are its two opponent-color coordinates; values near zero indicate low chroma. T_{white} is the lightness threshold, $\kappa_{\text{white}} > 0$ controls the sigmoid transition steepness, and $s_{\text{chroma}} > 0$ is the chroma tolerance. Increasing s_{chroma} admits more chromatic pixels into the soft mask. These parameters operate in the chosen CIELAB coordinate scale; the brightness targets ρ_{dark} and ρ_{bright} operate in the base-appearance image domain. The corresponding loss is

$$\mathcal{L}_{\text{white}} = \frac{1}{N} \sum_{p} M_{\text{white}}(p) \, |\mathbf{R}_{2D}(p) - \mathbf{I}(p)| \, . \tag{15}$$

The mask expresses a preference for preserving bright, approximately achromatic input regions. Clipped highlights may also receive high mask values, which limits the reliability of this heuristic in saturated regions.

A.1.3 GAUSSIAN-LEVEL LOSSES

The Gaussian-level branch uses reconstruction supervision and regularization of appearance and illumination, together with the cross-domain coupling described in the main text.

Observation Reconstruction. The illumination-aware rendering uses the observation loss from the main text:

$$\mathcal{L}_{\text{obs}} = D(\hat{\mathbf{I}}_{\text{adapt}}, \mathbf{I}). \tag{16}$$

It has already been counted in Eq. equation 8.

Exclusion Loss. To encourage separation of base appearance and illumination, we penalize correlated gradients between their rendered maps:

$$\mathcal{L}_{\text{excl}} = \mathcal{E}\left(\hat{\mathbf{R}}_{\text{3D}}, \mathbf{L}_{\text{adapt}}\right), \tag{17}$$

where the gradient-exclusion operator follows the exclusion loss of Zhang et al. (2018), defined over a multi-scale pyramid as

$$\mathcal{E}(\mathbf{R}, \mathbf{L}) = \sum_{l=0}^{L-1} \sum_{d\in\{x,y\}} \frac{1}{N_l} \sum_{p} \left|\phi\big(\nabla_d \mathbf{R}^{\,l}(p)\big) \cdot \phi\big(\nabla_d \mathbf{L}^{\,l}(p)\big)\right|, \tag{18}$$

where $\mathbf{R}^{\,l}$ and $\mathbf{L}^{\,l}$ denote the reflectance and illumination maps at pyramid level l, respectively, N_l is the number of pixels at that level, ∇_d denotes the spatial gradient along $d \in \{x, y\}$, and $\phi(z) = \tanh(z)$ suppresses small gradient noise and normalizes the gradient responses.

Higher-Order SH Regularization. We regularize higher-order spherical harmonic coefficients to discourage illumination variations from being absorbed into Gaussian appearance:

$$\mathcal{L}_{\text{SH}} = \frac{1}{N_{\text{SH}}} \sum_{i=1}^{N_{\text{SH}}} c_i^2, \tag{19}$$

where c_i denotes a higher-order SH coefficient.

A.1.4 ILEC REGULARIZATION

Relative Illumination Guidance. Let $\hat{\Delta}^i_{\text{loc}} = \text{Up}(\Delta^i_{\text{loc}})$ denote the bilinearly upsampled log-compensation map. For each color channel c, we normalize the rendered illumination over the current image:

$$\mathbf{L}^c_{\text{rel}} = \frac{\mathbf{L}^c_{3D} - \min_{\mathbf{p}} \mathbf{L}^c_{3D}(\mathbf{p})}{\max_{\mathbf{p}} \mathbf{L}^c_{3D}(\mathbf{p}) - \min_{\mathbf{p}} \mathbf{L}^c_{3D}(\mathbf{p}) + \epsilon}. \tag{20}$$

The guidance map is

$$\mathbf{M}_{\text{rel}} = m_{\text{floor}} + (1 - m_{\text{floor}})\sigma\left(\kappa_{\text{guide}}(\mathbf{L}_{\text{rel}} - t_{\text{guide}})\right), \tag{21}$$

Here $m_{\text{floor}} \in [0, 1)$ is the minimum guidance weight, $\kappa_{\text{guide}} > 0$ controls transition steepness, and $t_{\text{guide}} \in [0, 1]$ is the threshold on normalized illumination. Larger values of $\mathbf{M}_{\text{rel}}$ identify regions that are relatively bright within the image; smaller values identify regions receiving less guidance. The map weights the residual regularizers below, and Eq. equation 4 defines its forward exposure adaptation. Image indices are omitted for readability. Products and absolute values in the following expressions are evaluated element-wise, with channel reduction implicit in the image averages.

Residual Magnitude Penalty. An ℓ_1 penalty limits the overall amplitude of the log-exposure residual and favors zero correction where the other objectives provide little evidence for an adjustment:

$$\mathcal{L}_{\text{mag}} = \frac{1}{N} \sum_{p} \left|\hat{\Delta}^i_{\text{loc}}(p)\right|. \tag{22}$$

Guidance-Weighted Shrinkage. This term adds stronger shrinkage where $\mathbf{M}_{\mathrm{rel}}$ is small:

$$\mathcal{L}_{\mathrm{shrink}} = \frac{1}{N} \sum_{p} \left| (1 - \mathbf{M}_{\mathrm{rel}}(p)) \, \hat{\Delta}^{i}_{\mathrm{loc}}(p) \right|. \tag{23}$$

The continuous weight $1 - \mathbf{M}_{\mathrm{rel}}$ concentrates correction in the more strongly guided regions. It provides a soft spatial preference whose strength varies across the image.

Bright-Region Exposure Promotion. This signed term favors positive log-exposure residuals where $\mathbf{M}_{\mathrm{rel}}$ is large:

$$\mathcal{L}_{\mathrm{promote}} = -\frac{1}{N} \sum_{p} \mathbf{M}_{\mathrm{rel}}(p) \hat{\Delta}^{i}_{\mathrm{loc}}(p). \tag{24}$$

For a fixed guidance map, a positive residual decreases this term and increases the corresponding multiplicative factor in Eq. equation 4. It encourages the exposure-adaptation branch to account for additional illumination in relatively bright regions. Its effect is balanced by the magnitude and shrinkage penalties and the remaining training objective.

The combined local regularization objective is

$$\mathcal{L}_{\mathrm{loc}} = \lambda_{\mathrm{mag}} \mathcal{L}_{\mathrm{mag}} + \lambda_{\mathrm{shrink}} \mathcal{L}_{\mathrm{shrink}} + \lambda_{\mathrm{promote}} \mathcal{L}_{\mathrm{promote}}. \tag{25}$$

The local weights are distinguished from the SSIM mixture coefficient λ_{ssim} in Eq. equation 6.

A.2 Experiment and Implementation Settings

This section specifies the training and evaluation settings used throughout our experiments. Unless stated otherwise, all EvenSplat configurations use the same loss weights and optimization schedule across different conditions.

Hyperparameters. Tables 6 and 7 report hyperparameters for the image-space decomposition and coupled Gaussian-level settings used by EvenSplat, respectively.

Table 6: Image-space network settings used across datasets.

Hyperparameter	Value
Loss Weights	
λ_{curve}	0.5
$\lambda_{\mathrm{smooth}}$	1.0
λ_{white}	0.2
Parameters	
η_{dark}	1.0
η_{bright}	1.0
T_{dark}	0.25
T_{bright}	0.65
ρ_{dark}	0.45
ρ_{bright}	0.55
γ_{edge}	0.2
T_{white}	95.0
κ_{white}	9.8
s_{chroma}	2.4
Image embedding dimension d_{emb}	64

Table 7: Coupled Gaussian-level and image correction settings.

Hyperparameter	Value
Loss Weights	
λ_{obs}	1
λ_{excl}	0.3
λ_{SH}	0.8
λ_{mag}	0.1
$\lambda_{\mathrm{shrink}}$	0.1
$\lambda_{\mathrm{promote}}$	0.1
$\lambda_{\mathrm{recomb}}$	0.8
λ_{align}	0.2
λ_{cpl}	1
Parameters	
m_{floor}	0.15
κ_{guide}	8
t_{guide}	0.45
ϵ	10^{-6}

Data splits, resolution, and hardware. All methods use identical camera poses, image resolutions, and train–test splits. Views are ordered with every eighth view held out for evaluation and the remainder used for training. For simulated data, the perturbed images are training observations and the corresponding clean images at the same poses are evaluation references. Simulated scenes are downsampled by a factor of 8 along each spatial dimension, yielding a resolution of approximately 600×400 pixels. Real captures are approximately 3000×2000 pixels and are downsampled by a factor of four, with small scene-dependent variations. RealityScan 2.1 provides fixed camera poses and a sparse point cloud for initializing the real-world scenes. Training uses a batch size of one on two NVIDIA GeForce RTX 3090 GPUs with 24 GB memory.

Optimization schedule. We optimize each scene for 10,000 iterations with Adam. Gaussian learning rates are 1.6×10^{-4} for position, 1.4×10^{-2} for scale, 1.2×10^{-2} for opacity, 1.0×10^{-3} for rotation, 2.5×10^{-3} for the DC spherical-harmonic (SH) coefficients, and 1.25×10^{-4} for higher-order SH coefficients. The image-space network and its appearance embedding use learning rates of 2×10^{-3} and 5×10^{-3}, respectively. Image-space embedding LR, CRN, 3D illumination field, and ILEC map use 5×10^{-3}, 3×10^{-4}, 6×10^{-5}, and 3×10^{-3}, respectively. Position learning rates decay exponentially; the illumination field and CRN use cosine annealing; and the SH degree increases progressively to degree 3.

Gaussian Illumination Field and Camera Response parameterization. The 3D illumination field takes the Gaussian center $\mathbf{x}_j$, Gaussian-to-camera viewing direction $\mathbf{v}_j$, and estimated normal $\mathbf{n}_j$ as input. Positions are scene-normalized and encoded with a 16-level multi-resolution hash grid (2 features per level, base resolution 16, hash size 2^{19}, per-level scale 1.4473), while ℓ_2-normalized viewing directions and normals use four-frequency sinusoidal encodings with frequencies $\{1, 2, 4, 8\}$. The concatenated features are processed by a fully fused MLP with hidden width 128 and LeakyReLU activations, followed by a two-layer prediction head whose final layer has zero bias and weights initialized from $\mathcal{N}(0, 10^{-4})$. The multiplicative illumination gain is parameterized as

$$\mathbf{a}_j = 0.05 + 3.95\,\sigma(\tilde{\mathbf{a}}_j + \mathbf{1}), \tag{26}$$

giving $\mathbf{a}_j \in (0.05, 4.0)^3$. The current renderer uses only $\mathbf{c}_j^{\text{lit}} = \mathbf{a}_j \odot \mathbf{c}_j^{\text{base}}$.

Gaussian normals are taken as the rotation axis corresponding to the smallest Gaussian scale and are ℓ_2-normalized before encoding; no view-dependent normal flipping is applied. The CRN maps each training-image embedding through a width-32 fully fused MLP with LeakyReLU activations and a three-dimensional output head. Its RGB gain is

$$\mathbf{c}^i = \text{softplus}(\tilde{\mathbf{c}}^i), \tag{27}$$

ensuring positivity. The output head uses zero weights, unit gain biases, and zero additive biases, yielding an initial gain of $\text{softplus}(1) \approx 1.313$. Gaussian illumination is alpha-composited with the standard transmittance weights without opacity renormalization, and the illumination used for decomposition is clamped to $[0.05, 4.0]$ to stabilize low-opacity regions.

Gaussian refinement and staged training. Densification and pruning follow the standard 3DGS schedule. Refinement starts at iteration 500, runs every 100 iterations, and stops at iteration 7,000. Gaussian growth uses a projected-gradient threshold of 2×10^{-4} and 2D/3D scale thresholds of 0.05/0.01, while pruning uses an opacity threshold of 0.005 and scale thresholds of 0.15/0.1, with opacity reset every 3,000 iterations. Following Sec. 3.4, the 3D illumination field, Image-space embedding, CRN are frozen after iteration 3,000, and ILEC is enabled at iteration 5,000. All other trainable parameter groups remain active and are updated throughout training. Each training image has a zero-initialized $h \times w$ single-channel local log-gain map, bilinearly upsampled, bounded to $[-1.5, 1.5]$, and exponentiated.

Novel-view inference and comparison protocol. Training-image embeddings and ILEC maps are image specific and are not assigned to held-out views. At inference, the CRN is replaced by the identity response and the ILEC gain is fixed to one; novel-view rendering therefore relies only on the shared Gaussian representation and the learned 3D illumination field. All methods are trained and evaluated on the same input data and paired held-out views. PSNR, SSIM, and LPIPS are computed directly between rendered images and their corresponding references. Unless otherwise stated, no test-time pose refinement, exposure fitting, affine RGB correction, or color matching is applied; corrected metrics are retained only as diagnostics and are excluded from the primary comparisons.

Baseline configuration and test-view inference. We follow the publicly released implementations and their native inference protocols. For 3DGS, we use the default gsplat configuration. For GS-W, we enable its per-image appearance optimization during training and render held-out views from the learned intrinsic/base appearance representation, without optimizing an additional appearance code for the test image. For Bilateral Grid, per-image bilateral grids are used only during training to absorb view-specific photometric discrepancies; the correction branch is bypassed for held-out views, which are rendered directly from the learned 3D representation. Luminance-GS is evaluated

with its released view-adaptive configuration, while PPISP uses its released post-processing pipeline together with the MCMC Gaussian strategy. We do not introduce zero, mean, nearest-view, or other appearance-code heuristics beyond those defined by the corresponding released methods.

Tone-Matched 3DGS baseline. Tone matched preprocessing + 3DGS is a control baseline that applies a fixed global exposure correction to vanilla 3DGS renderings. For each rendered RGB image $\mathbf{I} = (R, G, B)$, we first compute its per-pixel luminance as $Y = 0.2126R + 0.7152G + 0.0722B$, and denote by $\bar{Y}$ the mean luminance averaged over all pixels in the rendered image. We then determine a global scalar gain $\alpha = 0.5/(\bar{Y} + \epsilon)$ and rescale the rendered image as $\mathbf{I}' = \mathrm{clip}(\alpha\mathbf{I}, 0, 1)$, where 0.5 is a fixed prescribed mid-tone target and ϵ is a small constant for numerical stability. This operation does not use ground-truth images and introduces neither additional learnable parameters nor test-time optimization. It serves as a control for quantifying how much improvement can be obtained from simple global tone/exposure normalization alone, thereby separating gains due to overall brightness adjustment from those provided by our image-space decomposition, 3D illumination modeling, and local exposure compensation.

Directional-lighting evaluation. The paired Robot experiment measures appearance consistency under two controlled lighting directions. The room illumination is disabled and the camera follows a fixed trajectory around the robot. At each camera position, we capture one image with a left-side light and one with a right-side light, without changing the camera pose or settings. This produces 74 matched view pairs with shared camera parameters and poses.

The two lighting datasets are fitted independently, with no learned scene parameters shared between them. All 74 views are used for fitting and for the paired-view comparison. This experiment therefore measures consistency at observed, pose-matched views rather than held-out novel-view generalization. For each pose, we compare the two appearance renderings over the full image, without foreground masking or spatial cropping.

PSNR is computed independently for each matched pair and then averaged:

$$\overline{\mathrm{PSNR}} = \frac{1}{74}\sum_{i=1}^{74} \mathrm{PSNR}\Big(\hat{\mathbf{R}}^i_{\mathrm{L}}, \hat{\mathbf{R}}^i_{\mathrm{R}}\Big), \tag{28}$$

where $\hat{\mathbf{R}}^i_{\mathrm{L}}$ and $\hat{\mathbf{R}}^i_{\mathrm{R}}$ are appearances at the same pose from the models fitted to the left- and right-lighting datasets. With RGB values in $[0, 1]$, PSNR uses unit peak intensity. The reported score is the arithmetic mean of the per-view values rather than a score computed from a pooled MSE.

A.3 Dataset Construction and Capture Details

This section gives the construction details omitted from the main paper. The real benchmark contains nine scenes acquired under controlled exposure and illumination variation, while the simulated benchmark applies three perturbation families to six Mip-NeRF 360 (Barron et al., 2022) scenes.

A.3.1 Real-world Capture Datasets

Capture equipment and general protocol. The real benchmark contains three CEV, four SIV, and two HCI scenes, spanning indoor and outdoor subjects. Table 8 records 78–472 reference views per scene and all captured exposure or illumination variants. We use a Canon EOS 700D with an 18–135 mm f/3.5–5.6 lens and a Sony α7 III with a 24–105 mm f/4.0 lens. All scenes are captured handheld to accommodate uneven terrain and viewpoints inaccessible to a tripod. Automatic exposure bracketing and high-rate burst capture reduce motion between consecutive variants, although small residual misalignment can remain.

Real CEV capture. Exposure bracketing changes brightness between views. On the Canon EOS 700D, the Magic Lantern add-on extends the available bracket from the camera's default ± 2 EV to ± 5 EV, creating a wider range of cross-view exposure variation.

Real SIV capture. Localized illumination is produced with an external flash partly obstructed by aluminum foil. For most SIV scenes, the first burst image is deliberately underexposed while

Table 8: Capture log for the nine real-world evaluation scenes. Reference views denote the number of reference photographs; All Images gives the total including captured exposure and illumination variants. CEV, SIV, and HCI denote the three illumination settings defined in Sec. 4.1. AUTO denotes an automatically selected camera setting; shutter speeds are in seconds. Canon and Sony denote the EOS 700D and α7 III, respectively.

Dataset Scene	Reference Views	All Images	ISO	f/stop	Shutter Speed	Category	Camera Model
LogHC	122	244	AUTO	f4	1/200	HCI	Canon
RobotHC	183	366	AUTO	f10	1/200	HCI	Sony
FourLogs	211	633	AUTO	f3.5	AUTO	CEV	Canon
Log	260	780	AUTO	f3.5	AUTO	CEV	Canon
Robot	472	1416	AUTO	f5.6	AUTO	CEV	Canon
ParkBench	124	248	AUTO	f4	1/200	SIV	Sony
CircularBench	78	156	AUTO	f5	1/200	SIV	Sony
PlasticCart	123	492	6400	f5.6	AUTO	SIV	Sony
Rocks	120	240	AUTO	f7.1	1/200	SIV	Sony

the flash fires at maximum power; subsequent reference images are captured without flash while it recharges. For *PlasticCart*, flash–shutter mis-synchronization creates a sharper illumination boundary across the image.

Real HCI capture. Close-range flash, partial obstruction, and low camera exposure create a brightly illuminated region surrounded by deep shadows. The first images in the burst are heavily underexposed, yielding both saturated highlights and severely dark regions. The capture log identifies Canon for *LogHC* and Sony for *RobotHC*.

Reference capture and reconstruction inputs. Reference photographs use nominal 0.0 EV. Outdoor references are captured under overcast skies and indoor references under uniform room lighting to reduce directional shadows and spatial imbalance. Canon references use evaluative metering; per-scene ISO, aperture, and shutter settings are listed in Table 8. These photographs remain sRGB appearance references containing camera response and residual scene illumination, rather than calibrated reflectance. RealityScan 2.1 estimates camera alignment and sparse points for Gaussian initialization.

A.3.2 Simulated Datasets

Simulated scenes and references. The quantitative benchmark uses *Counter*, *Bonsai*, *Bicycle*, *Garden*, *Kitchen*, and *Room* from Mip-NeRF 360 (Barron et al., 2022). Perturbations are applied only to the training images; the original images are retained as evaluation references. Each setting therefore measures recovery toward the same unmodified scene appearance.

Simulated CEV. We use the exposure-variation benchmark introduced by Cui et al. (2025) (their `variance` dataset), which assigns different exposure levels to training views and applies a mild gamma adjustment. This produces view-to-view photometric inconsistency while preserving the scene content and camera trajectory.

Simulated SIV. For each training image, low and high multiplicative exposure factors are sampled from $[0.4, 0.6]$ and $[1.4, 1.6]$. The image is divided horizontally or vertically at a position sampled between 30% and 70% of the corresponding dimension. The two sides receive opposite factors; a Gaussian-smoothed binary mask blends the regions into a gradual transition. The result is clipped to the valid image range.

Simulated HCI. We increase training-image contrast by a factor of three and clip the result to the valid image range, producing stronger bright–dark imbalance than the SIV perturbation.

A.4 Supplementary Results

ILEC regularization ablation. ILEC models image-specific spatial exposure residuals that remain after Gaussian-level illumination modeling. The module-level ablation in Table 4 shows its

largest PSNR benefit under high contrast, where a single global correction cannot absorb the residual bright–dark imbalance. We therefore examine its regularization terms in this setting in Table 9. It shows that all three constraints matter: the magnitude penalty limits arbitrary local edits, exposure promotion recovers persistently dark regions, and guidance-weighted shrinkage provides the strongest contribution (a 0.62 dB PSNR drop when removed). The result supports a constrained, illumination-aware local model rather than an unconstrained per-image correction map.

Table 9: ILEC ablations on real-world High Contrast Illumination (HCI) dataset.

Variant	PSNR/SSIM/LPIPS
Full	16.950/0.682/0.241
No Magnitude Penalty	16.723/0.653/0.243
No Guided Shrinkage	16.330/0.647/0.276
No Exposure Promotion	16.600/0.668/0.260

Additional results on real-world dataset. Table 10 expands the main comparison to all nine scenes. The breakdown confirms that the HCI gain occurs on both high-contrast scenes and that the aggregate SIV improvement is not attributable to one example. It also exposes the expected CEV variability, where methods specialized for global photometric correction can remain competitive. Figure 10 adds difficult cases and separates global color mismatch from spatial illumination error through an affine-corrected visualization for the extremely exposed *FourLogs* scene.

Table 10: Per-scene quantitative comparisons on the real-world dataset under CEV, SIV, and HCI. Each entry reports PSNR↑/SSIM↑/LPIPS↓. **Tone-Matched 3DGS** applies a single global tone/exposure normalization to vanilla 3DGS to match the prescribed mid-tone target of 0.5. This control isolates the contribution of global tone matching from the proposed decomposition. Best results are highlighted in light red and second-best results in light yellow.

Setting	**Scene**	**Bilateral Grid**	**3DGS**	**GS-W**	**Luminance-GS**	**PPISP**	**EvenSplat**	**Tone-Matched 3DGS**
CEV	*FourLogs*	12.339/0.186/0.928	10.954/0.131/0.940	8.965/0.094/0.641	**15.684**/0.223/0.867	11.006/**0.436**/**0.560**	15.148/0.427/0.662	13.035/0.177/0.952
	Log	20.130/0.783/0.216	17.072/0.651/0.367	13.432/0.676/0.220	17.745/0.734/0.283	20.547/**0.871**/**0.115**	**20.648**/0.812/0.157	17.675/0.662/0.321
	Robot	16.542/0.677/0.429	14.564/0.601/0.519	14.814/0.798/0.224	16.325/0.656/0.304	21.829/0.861/0.133	**22.781**/**0.884**/**0.097**	14.117/0.678/0.393
SIV	*PlasticCart*	15.096/0.651/0.415	14.754/0.595/0.498	14.549/0.586/0.402	14.456/0.645/0.400	15.826/0.683/0.258	**18.427**/**0.750**/**0.253**	10.272/0.581/0.513
	Rocks	20.825/0.826/0.122	20.591/0.827/0.121	20.243/0.805/0.144	14.599/0.750/0.171	20.956/**0.842**/0.121	**21.162**/0.838/**0.118**	15.152/0.765/0.156
	ParkBench	19.907/0.803/0.138	20.310/0.810/0.135	20.043/0.761/0.168	18.248/0.803/0.136	19.810/**0.816**/0.142	**21.037**/**0.816**/**0.117**	18.002/0.787/0.146
	CircularBench	16.147/0.690/0.233	15.953/0.702/0.244	16.134/0.654/0.257	14.928/0.725/0.214	15.854/**0.741**/0.191	**17.420**/0.739/**0.186**	14.804/0.671/0.255
HCI	*LogHC*	11.312/0.438/0.302	10.533/0.393/0.328	10.686/0.396/0.325	12.614/0.565/0.287	10.568/0.400/0.322	**16.718**/**0.669**/**0.220**	10.142/0.542/0.315
	RobotHC	13.068/0.370/0.395	12.870/0.460/0.452	12.916/0.454/0.432	14.468/0.540/0.394	13.307/0.491/0.367	**17.181**/**0.694**/**0.262**	11.737/0.603/0.419

Additional results on simulated dataset. Table 11 shows that the aggregate CEV and SIV gains extend across scenes with different geometry and texture. HCI remains more metric-dependent, consistent with the main results: correcting clipped, high-contrast inputs improves intensity agreement but can alter local structure and perceptual similarity. Figures 11– 13 visualize the same pattern, with the clearest advantage near exposure transitions and in shadowed regions.

Additional Experiment on HDR-NeRF dataset. We evaluate the public simulated HDR-NeRF scenes (Huang et al., 2022b) under severe but view-consistent illumination. The *Rendered Observation* target retains this illumination and measures complete-scene reconstruction (Fig. 14). The *Base Appearance* target instead uses the ground-truth appearance rendered under parallel uniform light. EvenSplat directly produces this second output; Luminance-GS produces enhanced renderings; the other comparison methods do not explicitly decompose base appearance, so their standard renderings are compared with the same target (Fig. 15).

EvenSplat maintains high-quality full-scene observation rendering, achieving 37.6-dB PSNR on average and remaining visually comparable to methods dedicated to fitting the illuminated observations. As shown in Table 12, its additional decomposition output has the highest base-appearance PSNR on all four scenes, improving mean PSNR from 13.23 dB for the strongest baseline to 17.91 dB. SSIM and LPIPS remain mixed, particularly on *CatRoom*, so the improvement should be read as better recovery of the uniform-light target rather than a universal perceptual advantage.

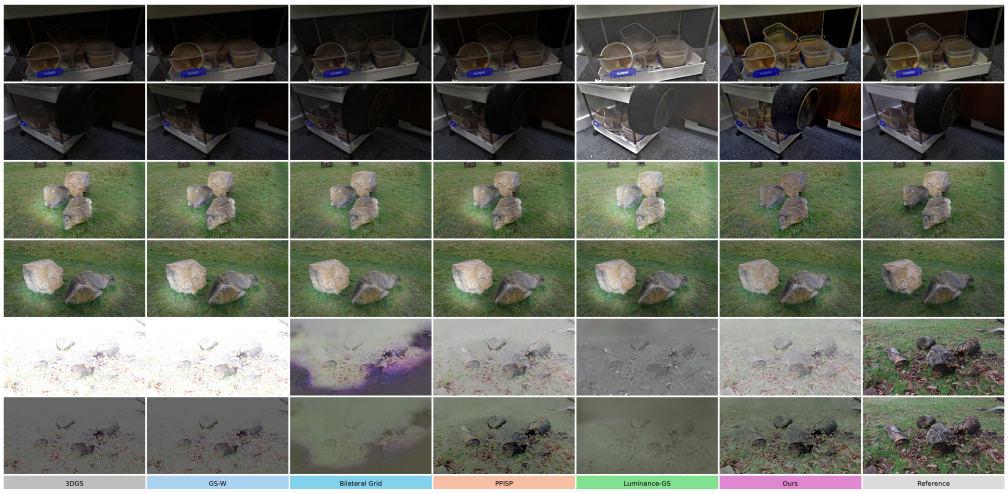

Figure 10: Additional real-world qualitative comparisons. Columns show 3DGS, GS-W, Bilateral Grid, PPISP, Luminance-GS, EvenSplat, and the reference. The first two rows show *PlasticCart* (SIV), the next two show *Rocks* (SIV), and the final two show the extreme *FourLogs* case (CEV). For *FourLogs*, the last row applies a per-image, per-channel affine correction, $y_c = a_c x_c + b_c$, to the preceding outputs. This diagnostic removes global color mismatch and exposes the remaining spatial reconstruction artifacts.

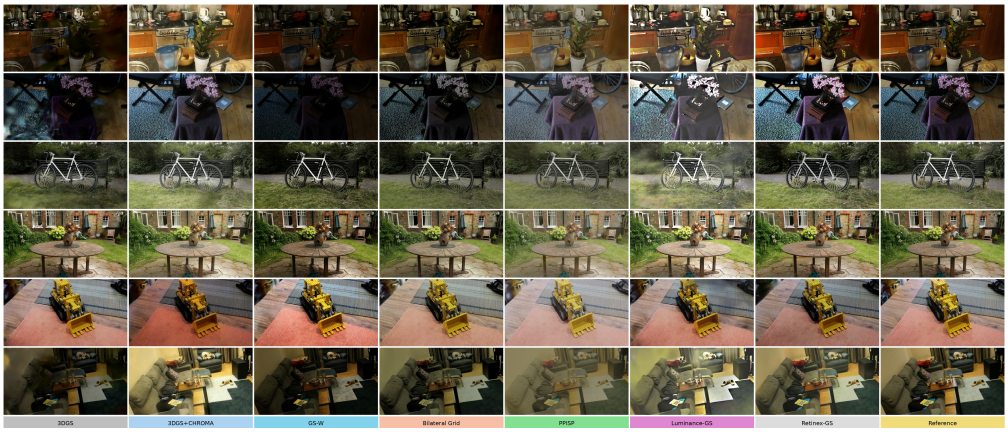

Figure 11: Qualitative comparison on simulated scenes under Cross-View Exposure Variation (CEV). Columns show 3DGS, GS-W, Bilateral Grid, PPISP, Luminance-GS, EvenSplat, and the reference; each row is a held-out view from a different scene. CEV changes the global exposure between views, testing whether each method can recover a consistent scene appearance without retaining view-dependent brightness and color shifts.

This is because in regions saturated by the same strong global light in every view, clamping removes the original color information from the entire observation set. No reconstruction method can identify those colors without additional measurements or priors; this limitation comes from the dataset observations, not from the EvenSplat decomposition.

Table 11: Per-scene quantitative comparisons on simulated data under (a) Cross-View Exposure Variation (CEV), (b) Spatial Illumination Variation (SIV), and (c) High-Contrast Illumination (HCI). Each panel reports the six scenes. Higher PSNR and SSIM and lower LPIPS are better. Following Table 10, best values are bold with light-red shading and second-best values are underlined with light-yellow shading, ranked separately for each scene and metric. Ties at the reported precision share the same rank.

(a) CEV

Method	Counter			Bonsai			Bicycle			Garden			Kitchen			Room		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Bilateral Grid	17.2331	0.5907	0.2759	14.2005	0.4373	0.4298	19.7586	0.6825	0.3256	20.0180	0.7032	0.1812	21.7525	0.7769	0.1228	19.6984	0.8521	0.1842
3DGS	14.1813	0.5011	0.3741	12.7882	0.3272	0.5585	19.0437	0.5934	0.3449	20.1326	0.7720	0.1702	22.9005	0.8686	0.0894	14.6840	0.5686	0.3353
3DGS+CHROMA	15.2677	0.7024	0.1928	**22.0536**	0.7860	0.2160	14.2583	0.6413	0.2882	22.2103	0.8312	0.1588	15.6871	0.7849	0.1373	15.9184	0.7532	0.1974
GS-W	14.7554	0.5824	0.2077	13.1171	0.3968	0.3191	21.2405	0.6334	0.2879	18.9876	0.7605	0.1604	21.1178	0.8408	0.1110	15.8833	0.6255	0.2018
Luminance-GS	15.5841	0.6295	0.2896	15.2206	0.5636	0.3953	16.0880	0.6237	0.3342	19.5610	0.7976	0.1832	23.6768	0.8774	0.1024	15.0968	0.6374	0.3098
PPISP	16.4238	0.6399	0.2238	17.8789	0.6774	0.2163	**22.7584**	**0.7108**	0.2711	21.0851	0.7907	0.1709	18.6551	0.8269	0.1031	18.5316	0.7076	0.2110
EvenSplat	**18.7079**	**0.7644**	**0.1638**	18.2206	**0.7870**	**0.1493**	20.3733	0.6901	**0.2583**	**23.6465**	**0.8310**	**0.1362**	**24.6580**	**0.8806**	**0.0879**	**20.2534**	**0.8242**	**0.1548**

(b) SIV

Method	Counter			Bonsai			Bicycle			Garden			Kitchen			Room		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Bilateral Grid	19.1600	0.7443	0.1515	20.2708	0.8680	0.1200	18.3092	0.7358	0.1392	19.6960	0.7493	0.1637	18.6821	0.7593	0.1550	18.4933	0.7809	0.1325
3DGS	19.0273	0.7834	0.1614	19.5389	0.8409	0.1322	18.4637	0.6237	0.3472	17.2341	0.7695	0.1732	16.3484	0.8043	0.1822	18.8131	0.8171	0.1605
3DGS+CHROMA	19.6773	0.7838	0.1680	19.9524	0.7847	0.1811	16.5234	0.6110	0.3727	18.7528	0.7441	0.2533	17.3001	0.7658	0.2436	18.7915	0.7785	0.2233
GS-W	19.9049	0.7846	0.1450	20.3154	0.8532	0.1072	19.1688	0.6525	0.2876	17.4411	0.7647	0.1746	15.7612	0.8046	0.1484	17.6349	0.7957	0.1775
Luminance-GS	13.6157	0.6462	0.2126	12.7315	0.6407	0.1978	13.1440	0.5931	0.3384	16.4130	0.7530	0.2180	18.2343	0.8309	0.1552	12.7719	0.6265	0.2540
PPISP	20.8596	0.8238	0.1270	22.2038	0.8879	0.0911	18.3908	0.6482	0.3030	**20.3391**	0.8074	0.1508	19.5526	0.8534	0.1122	20.7150	0.8604	0.1236
EvenSplat	**21.7063**	**0.8408**	**0.0976**	**23.8880**	**0.9064**	**0.0669**	**19.7933**	**0.8211**	**0.1146**	19.4974	**0.8090**	**0.1370**	**22.5320**	**0.8956**	**0.0850**	**24.5839**	**0.8965**	**0.0802**

(c) HCI

Method	Counter			Bonsai			Bicycle			Garden			Kitchen			Room		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Bilateral Grid	14.0340	0.4497	0.3523	13.8610	0.4642	0.3070	14.2752	0.4526	0.3141	14.0389	0.5104	0.3583	14.4848	0.6033	0.2833	13.6416	0.4970	0.3195
3DGS	13.0949	0.3921	0.3928	12.5985	0.4295	0.3489	13.2249	0.4210	0.3453	13.0387	0.4530	0.3874	13.5651	0.5538	0.3216	12.4804	0.4633	0.3537
3DGS+CHROMA	13.9290	0.4788	0.3291	13.2153	0.4934	0.3286	14.4233	0.4257	0.3798	13.5515	0.4803	0.3837	16.1563	0.6362	0.2870	13.9091	0.4995	0.3356
GS-W	12.9158	0.3906	0.3846	12.7446	0.4391	0.3444	13.2339	0.4137	0.3528	12.9188	0.4449	0.3909	13.6051	0.5441	0.3263	12.5066	0.4711	0.3602
Luminance-GS	12.3740	0.4502	0.3729	11.3973	0.5395	0.3541	12.2760	0.5627	0.3376	14.7142	**0.6318**	0.3548	15.3304	0.6451	0.3187	11.6883	0.5757	0.3468
PPISP	**17.2665**	**0.5271**	0.3518	14.6068	**0.5727**	0.2930	15.1860	**0.5831**	**0.2864**	**15.3973**	0.5956	0.3254	**15.6006**	**0.6809**	**0.2404**	12.9865	**0.5767**	0.3170
EvenSplat	16.6347	0.4659	**0.3442**	**19.1080**	0.5035	**0.2922**	**16.8896**	0.4645	0.3026	14.7497	0.4699	**0.3215**	14.6440	0.6092	0.2702	**18.3918**	0.4955	**0.2771**

Table 12: Quantitative recovery of base appearance on HDR-NeRF. Outputs are evaluated against ground-truth renderings under parallel uniform illumination. EvenSplat uses its explicit base-appearance output; Luminance-GS uses its enhanced output while other comparison methods use their standard renderings because they do not separately model this output.

Method	CatRoom/desk			BathRoom/bathroom			DogRoom/dog			Sofa/sofa		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS	16.1779	0.8943	0.0989	14.7923	0.5109	0.3258	11.1984	0.7411	0.2668	10.7005	0.5492	0.3190
GS-W	16.1751	0.8945	0.0986	14.7767	0.5151	0.3175	11.1819	0.7386	0.2752	10.7238	0.5507	0.3164
Bilateral Grid	16.1449	0.8735	0.0989	14.7900	0.5110	0.3241	11.1873	0.7383	0.2724	10.7103	0.5497	0.3174
PPISP	16.1873	0.8992	0.0936	14.7960	0.5185	0.3078	11.2013	0.7489	0.2409	10.7288	0.5516	0.3047
Luminance-GS	8.9776	0.5044	0.2877	8.3358	0.4899	0.3275	10.3754	0.7338	0.2808	8.4031	0.4901	0.3164
EvenSplat (Ours)	18.3390	0.8243	0.1550	18.3685	0.6811	0.2863	14.4494	0.7540	0.2483	20.4864	0.7586	0.2346

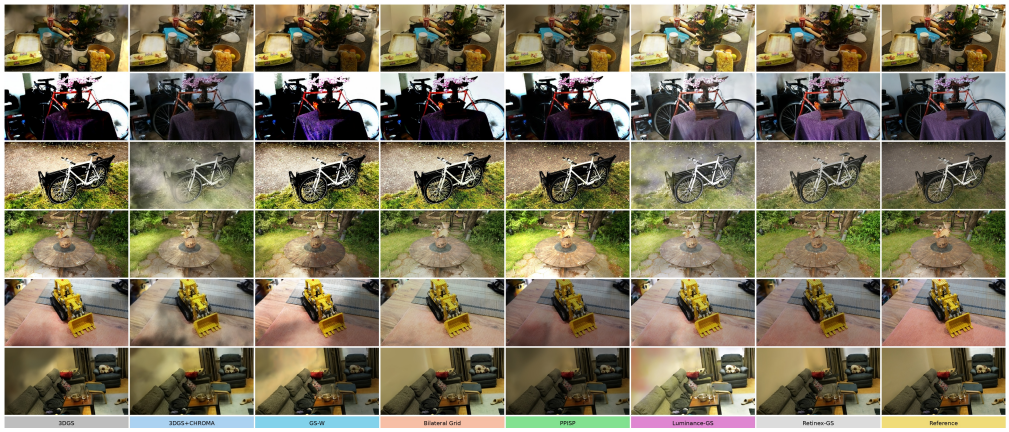

Figure 12: Qualitative comparison on simulated scenes under Spatial Illumination Variation (SIV). Columns show 3DGS, GS-W, Bilateral Grid, PPISP, Luminance-GS, EvenSplat, and the reference; each row is a held-out view from a different scene. The illumination changes within each image, producing adjacent bright and dark regions that cannot be removed by a single global exposure correction.



Figure 13: Qualitative comparison on simulated scenes under High-Contrast Illumination (HCI). Columns show 3DGS, GS-W, Bilateral Grid, PPISP, Luminance-GS, EvenSplat, and the reference; each row is a held-out view from a different scene. Strong highlights and deep shadows create clipping and severe local imbalance, emphasizing whether a method can restore shadow detail without flattening or overexposing bright regions.



Figure 14: Qualitative comparison of *Rendered Observation* on the four HDR-NeRF scenes. Columns show 3DGS, GS-W, Bilateral Grid, PPISP, Luminance-GS, EvenSplat, and the reference. The reference retains the severe but view-consistent illumination used to generate the observations; this comparison therefore measures complete-scene reconstruction fidelity rather than illumination removal.



Figure 15: Qualitative comparison of *Base Appearance* recovery on the four HDR-NeRF scenes. Columns show 3DGS, GS-W, Bilateral Grid, PPISP, Luminance-GS, EvenSplat, and the reference rendered under parallel uniform light. EvenSplat directly renders its decomposed base appearance, Luminance-GS produces enhanced renderings, whereas the other comparison methods contribute their standard scene renderings because they do not expose a separate base-appearance output.

REPRODUCIBILITY STATEMENT

Implementation details, hyperparameters, data construction, and evaluation procedures are provided in the main paper and appendices. An anonymized code repository and an exemplar dataset are provided in the supplementary material. The full code and datasets needed to reproduce the reported results will be publicly released upon acceptance.

AI USE STATEMENT

LLM tools were used solely to improve the clarity, grammar, and style of the manuscript after the authors had independently developed the research idea, designed and conducted the experiments, analyzed the results, and prepared the initial draft. The authors reviewed and verified all AI-assisted revisions and take full responsibility for the final content of the paper.