GS-ID: Illumination Decomposition on Gaussian Splatting via Adaptive Light Aggregation and Diffusion-Guided Material Priors

Supplementary Material

This supplementary material provides a detailed description of our method's implementation, followed by additional results from various datasets. We then introduce a new experiment to compare the illumination decomposition results under simple and complex light sources, reinforcing the arguments presented in the main paper. Finally, we present further application results. The supplementary material video files include additional videos demonstrating our work.

1. Implementation Details

We implement GS-ID using CUDA extensions and modify 3DGS to output G-buffer properties, including albedo, roughness, metallic, normal, and depth, visibility vector.

1.1. Representation

In the vanilla 3DGS, each 2D Gaussian utilizes learnable parameters $\mathcal{T} = \{\boldsymbol{p}, \boldsymbol{s}, \boldsymbol{q}\}$ and $\mathcal{C} = \{\alpha, \boldsymbol{f}_c, \boldsymbol{N}, \boldsymbol{D}\}$ to describe its geometric properties and volumetric appearance, respectively. Here, $\boldsymbol{p}$ denotes the position vector, $\boldsymbol{s}$ denotes the scaling vector, $\boldsymbol{q}$ denotes the unit quaternion for rotation, α denotes the opacity, $\boldsymbol{N}$ denotes the normal, $\boldsymbol{D}$ denotes the depth, and $\boldsymbol{f}_c$ denotes the spherical harmonics (SH) coefficients for view-dependent color. In GS-ID, we extend $\mathcal{C}$ to $\{\alpha, \boldsymbol{f}_c, \boldsymbol{N}, \boldsymbol{D}, \boldsymbol{A}, R, M\}$ to describe the material properties of the 2D Gaussian, and $\boldsymbol{A}, R, M$ denote Albedo, Roughness, and Metallic values. Additionally, on each pixel in screen space, we can calculate the world space positions $\boldsymbol{P}$ by rasterizing $\boldsymbol{D}$ with the camera transformation matrix.

1.2. Training Details

Our training pipeline consists of two stages: pre-20k and post-10k iterations.

Pre-20k Stage. The primary goal of this stage is to establish a stable geometric structure, particularly accurate surface normals and point positions for rasterization. We run 20k iterations using a modified 3D Gaussian Splatting (3DGS) framework, where surface normals predicted by a pretrained diffusion model serve as supervision. This stage ensures geometric consistency before introducing illumination modeling.

Post-10k Stage. In this stage, we freeze the Gaussian point positions and use the modified 3DGS to predict G-Buffer components, including albedo, roughness, metallic, and a unit shadow direction vector. These outputs are supervised by corresponding maps generated from the diffusion model. Subsequently, we optimize our illumination model, which comprises a set of learnable spherical Gaussian mixtures (SGMs) and a trainable environment light. We introduce a lighting regularization term and train for an additional 10k iterations. The final supervision is provided by comparing rendered RGB images against ground truth.

$$
\begin{aligned}
f_r(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) &= \underbrace{(1-M)\frac{\boldsymbol{A}}{\pi}}_{\text{diffuse component}} + \underbrace{\frac{DFG}{4(\boldsymbol{n}\cdot\boldsymbol{\omega}_i)(\boldsymbol{n}\cdot\boldsymbol{\omega}_o)}}_{\text{specular component}}, \\
\boldsymbol{h} &= \text{normalize}(\boldsymbol{\omega}_o + \boldsymbol{\omega}_i), \\
F_0 &= (1-M)\cdot 0.04 + M\cdot \boldsymbol{A}, \\
D(\boldsymbol{n}, \boldsymbol{h}) &= \frac{R^4}{\pi\left(\boldsymbol{n}\cdot\boldsymbol{h}\,(R^4-1)+1\right)^2}, \\
F(\boldsymbol{\omega}_i, \boldsymbol{n}) &= F_0 + (1-F_0)\left(1-\boldsymbol{n}\cdot\boldsymbol{\omega}_i\right)^5, \\
G(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i, \boldsymbol{h}) &= G_1(\boldsymbol{\omega}_o, \boldsymbol{h})\cdot G_1(\boldsymbol{\omega}_i, \boldsymbol{h}), \\
G_1(\boldsymbol{n}, \boldsymbol{h}) &= \frac{1}{1+\boldsymbol{n}\cdot\boldsymbol{h}\sqrt{R^4+\boldsymbol{n}\cdot\boldsymbol{h}-R^4\cdot\boldsymbol{n}\cdot\boldsymbol{h}}},
\end{aligned} \tag{1}
$$

where $\boldsymbol{A}$, R, and M denote the albedo, roughness, and metallicity, respectively. The Normal Distribution Function (NDF) D, Fresnel function F, and Geometry function G are derived from physical materials. We use the Adam optimizer for training, and the training process is divided into three stages.

We employ a mixture model of a set of spherical Gaussians (SGMs) to represent the localized highlight illumination as $L_o^{\text{SGMs}}(\boldsymbol{x}, \boldsymbol{\omega}_o)$:

$$
L_o^{\text{SGMs}}(\boldsymbol{x}, \boldsymbol{\omega}_o) = \sum_i^{N_{\text{light}}} \frac{{f_r}_i^{\text{SGM}}(\boldsymbol{n}\cdot\boldsymbol{\omega}_i) * \mathbf{V}_i}{d_i^2} \sum_j^{N_{\text{sg}}} W_{i,j} SG(j), \tag{2}
$$

where d_i denotes the distance from the i-th spherical Gaussian mixture to the surface point $\boldsymbol{x}$. ${f_r}_i^{\text{SGM}}$ denotes the BRDF function, and SG denotes the spherical Gaussian Function, respectively, as defined in the Methodology section. Each spherical Gaussian mixture contains NSG spherical Gaussians. Utilizing pretrained 3DGS $\mathcal{T}$ and $\mathcal{C}$, along with supervision of material properties generated by a pretrained diffusion model, we optimize M and L over 15,000 iterations. The optimization is guided by the following loss function:

$$
\mathcal{L}_{\text{reg}} = \lambda_R \text{L1}(R, \hat{R}) + \lambda_M \text{L1}(M, \hat{M}) + \lambda_{\boldsymbol{A}} \text{L1}(\boldsymbol{A}, \hat{\boldsymbol{A}}), \tag{3}
$$

where the weights for the loss function are set as: $\lambda_R = 0.1$, $\lambda_M = 0.1$, and $\lambda_{\boldsymbol{A}} = 0.5$. The entire illumination optimization process can be visualized as shown in Figure 1.

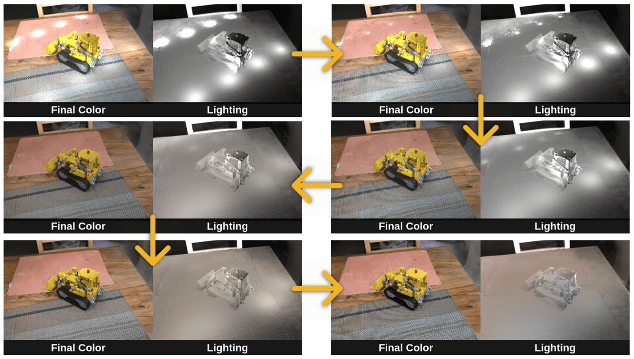

Figure 1. We optimize for N_{light} light sources with SGMs, in conjunction with a learnable environment map. Our representation is sufficiently expressive to capture detailed emissions while remaining controllable for light editing purposes.

2. Results on the MipNeRF 360, DB, and T&T Datasets

Tabs. 1, 2, and 3 present the results for novel view synthesis using the Mip-NeRF 360 Dataset. Tabs. 4, 5, and 6 display the results for Deepblend and the T&T Dataset. Additionally, Figure 3 illustrates the ID results for these scenes.

3. Results on the TensoIR Synthetic Dataset

Table 8 presents the outcomes for normal estimation, novel view synthesis, albedo reconstruction, and relighting across all four scenes.

4. Results on the ADT Dataset

Table 8 presents the outcomes for albedo estimation and novel view synthesis across all four scenes.

Method	bicycle	flowers	garden	stump	treehill	room	counter	kitchen	bonsai
NeRF++	22.64	20.31	24.32	24.34	22.20	28.87	26.38	27.80	29.15
Plenoxels	21.91	20.10	23.49	20.66	22.25	27.59	23.62	23.42	24.67
INGP-Base	22.19	20.35	24.60	23.63	22.36	29.27	26.44	28.55	30.34
INGP-Big	22.17	20.65	25.07	23.47	22.37	29.69	26.69	29.48	30.69
Mip-NeRF 360	24.40	21.64	26.94	26.36	22.81	29.69	26.69	29.48	30.69
3DGS	25.25	21.52	27.41	26.55	22.49	30.63	28.70	30.32	31.98
2DGS	24.87	21.15	26.95	26.47	22.27	31.06	28.55	30.50	31.52
GaussianShader	23.12	20.34	26.44	23.92	20.17	24.27	26.35	27.72	28.09
GSIR	23.80	20.57	25.72	25.37	21.79	28.79	26.22	27.99	28.18
Ours (3^3 SGMs)	24.41	20.90	26.76	24.35	22.01	30.47	27.42	29.95	30.18

Table 1. PSNR scores for Mip-NeRF360 scenes.

5. Applications

Light Editing The use of a spherical Gaussian (SG) mixture for lighting provides a flexible and intuitive representation of scene illumination. After training, the emission weights,

Method	bicycle	flowers	garden	stump	treehill	room	counter	kitchen	bonsai
NeRF++	0.526	0.453	0.635	0.594	0.530	0.530	0.802	0.816	0.876
Plenoxels	0.496	0.431	0.606	0.523	0.509	0.842	0.759	0.648	0.814
INGP-Base	0.491	0.450	0.649	0.574	0.518	0.855	0.798	0.818	0.890
INGP-Big	0.512	0.486	0.701	0.594	0.542	0.871	0.817	0.858	0.906
Mip-NeRF 360	0.693	0.583	0.816	0.746	0.632	0.913	0.895	0.920	0.939
3DGS	0.771	0.605	0.868	0.775	0.638	0.914	0.905	0.922	0.938
2DGS	0.752	0.588	0.852	0.765	0.627	0.912	0.900	0.919	0.933
GaussianShader	0.700	0.542	0.842	0.667	0.572	0.847	0.874	0.887	0.893
GSIR	0.706	0.543	0.804	0.716	0.586	0.867	0.839	0.867	0.883
Ours (3^3 SGMs)	0.734	0.561	0.826	0.665	0.586	0.906	0.883	0.910	0.923

Table 2. SSIM scores for Mip-NeRF360 scenes.

Method	bicycle	flowers	garden	stump	treehill	room	counter	kitchen	bonsai
NeRF++	0.455	0.466	0.331	0.416	0.466	0.335	0.351	0.260	0.291
Plenoxels	0.506	0.521	0.386	0.503	0.540	0.419	0.441	0.447	0.398
INGP-Base	0.487	0.481	0.312	0.450	0.489	0.301	0.342	0.254	0.227
INGP-Big	0.446	0.441	0.257	0.421	0.450	0.261	0.306	0.195	0.205
Mip-NeRF 360	0.289	0.345	0.164	0.254	0.338	0.211	0.203	0.126	0.177
3DGS	0.205	0.336	0.103	0.210	0.317	0.220	0.204	0.129	0.205
2DGS	0.218	0.346	0.115	0.222	0.329	0.223	0.208	0.133	0.214
GaussianShader	0.274	0.377	0.130	0.297	0.406	0.304	0.242	0.167	0.257
GSIR	0.259	0.371	0.158	0.258	0.372	0.279	0.260	0.188	0.264
Ours (3^3 SGMs)	0.249	0.368	0.144	0.329	0.400	0.235	0.229	0.146	0.226

Table 3. LPIPS scores for Mip-NeRF360 scenes.

Method	Trucks	Train	Avg.	Drjohnson	Playroom	Avg.
3DGS	25.18	21.09	23.13	28.76	30.04	29.40
2DGS	24.78	21.67	23.22	29.02	30.23	29.62
GaussianShader	20.13	23.56	21.84	22.18	19.88	21.03
GSIR	24.09	20.42	22.25	26.47	28.13	27.30
Ours (3^3 SGMs)	24.51	23.60	22.71	28.05	28.55	28.30

Table 4. PSNR scores for DB and T&T Dataset.

Method	Trucks	Train	Avg.	Drjohnson	Playroom	Avg.
3DGS	0.879	0.802	0.840	0.899	0.906	0.902
2DGS	0.867	0.803	0.835	0.901	0.907	0.904
GaussianShader	0.763	0.843	0.803	0.786	0.845	0.815
GSIR	0.833	0.742	0.787	0.863	0.869	0.866
Ours (3^3 SGMs)	0.858	0.778	0.818	0.886	0.885	0.885

Table 5. SSIM scores for DB and T&T Dataset.

Method	Trucks	Train	Avg.	Drjohnson	Playroom	Avg.
3DGS	0.148	0.218	0.183	0.244	0.241	0.242
2DGS	0.183	0.227	0.205	0.248	0.250	0.249
GaussianShader	0.271	0.191	0.231	0.336	0.359	0.344
GSIR	0.195	0.273	0.234	0.314	0.305	0.309
Ours (3^3 SGMs)	0.182	0.250	0.216	0.268	0.271	0.269

Table 6. LPIPS scores for DB and T&T Dataset.

positions, and SG parameters of light sources can be independently modified. This explicit emissive formulation enables physically plausible and interactive light editing, as shown in Figure 4 and the accompanying video.

Composition Our method models the scene's illumination as a combination of localized and environment lighting, allowing for complete illumination decomposition. This decomposition enables seamless integration of relightable content across different scenes. As demonstrated

		Normal (MAE)	Albedo			Novel View Synthesis		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Lego	InvRender	9.980	21.435	0.882	0.160	24.391	0.883	0.151
	TensoIR	5.980	25.240	0.900	0.145	34.700	0.968	0.037
	GS-IR	8.078	24.958	0.889	0.143	34.379	0.968	0.036
	Ours (3^3 SGMs)	7.595	25.155	0.911	0.138	37.377	0.975	0.023
Hotdog	InvRender	3.708	27.028	0.950	0.094	31.832	0.952	0.089
	TensoIR	4.050	30.370	0.947	0.093	36.820	0.976	0.045
	GS-IR	4.771	26.745	0.941	0.088	34.116	0.972	0.049
	Ours (3^3 SGMs)	4.269	31.546	0.949	0.0874	37.599	0.983	0.024
Armadillo	InvRender	1.723	35.573	0.959	0.076	31.116	0.968	0.057
	TensoIR	1.950	34.360	0.989	0.059	39.050	0.986	0.039
	GS-IR	2.176	38.572	0.986	0.051	39.287	0.980	0.039
	Ours (3^3 SGMs)	2.927	44.585	0.986	0.044	43.858	0.988	0.024
Ficus	InvRender	4.884	25.335	0.942	0.072	22.131	0.934	0.057
	TensoIR	4.420	27.130	0.964	0.044	29.780	0.973	0.041
	GS-IR	4.762	30.867	0.948	0.053	33.551	0.976	0.031
	Ours (3^3 SGMs)	3.681	32.702	0.959	0.049	37.681	0.990	0.008

Table 7. Per-scene results on TensoIR Synthetic dataset. For albedo reconstruction results, we follow NeRFactor and scale each RGB channel using a global scalar.

		Albedo			Novel View Synthesis		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Donut	GS-IR	25.224	0.949	0.029	35.171	0.975	0.046
	GSshader	24.138	0.941	0.031	34.171	0.971	0.047
	RelightGS	27.315	0.959	0.025	37.414	0.988	0.019
	Ours (3^3 SGMs)	28.515	0.969	0.0216	38.414	0.989	0.018
Figurine	GS-IR	25.224	0.949	0.029	35.171	0.975	0.046
	GSshader	28.563	0.905	0.059	37.709	0.961	0.048
	RelightGS	31.295	0.972	0.046	41.933	0.992	0.011
	Ours (3^3 SGMs)	30.895	0.963	0.048	40.633	0.986	0.012
Birdhouse	GS-IR	26.584	0.906	0.121	35.814	0.972	0.035
	GSshader	23.584	0.890	0.125	34.991	0.968	0.039
	RelightGS	28.011	0.933	0.094	38.287	0.986	0.012
	Ours (3^3 SGMs)	26.084	0.926	0.101	38.021	0.982	0.013
Table	GS-IR	25.718	0.938	0.063	38.439	0.972	0.036
	GSshader	25.213	0.912	0.064	38.123	0.969	0.038
	RelightGS	27.132	0.948	0.062	43.280	0.996	0.006
	Ours (3^3 SGMs)	29.752	0.958	0.055	44.581	0.994	0.006

Table 8. Per-scene results on ADT dataset. For albedo reconstruction results, we follow NeRFactor and scale each RGB channel using a global scalar.

in Figure 5 and the video, we incorporate a decomposed TensoIR-synthesized object into a real-world Mip-NeRF scene, achieving realistic lighting consistency.

Figure 2. Visualization of our albedo estimation results with other methods on the TensoIR synthetic and ADT dataset.

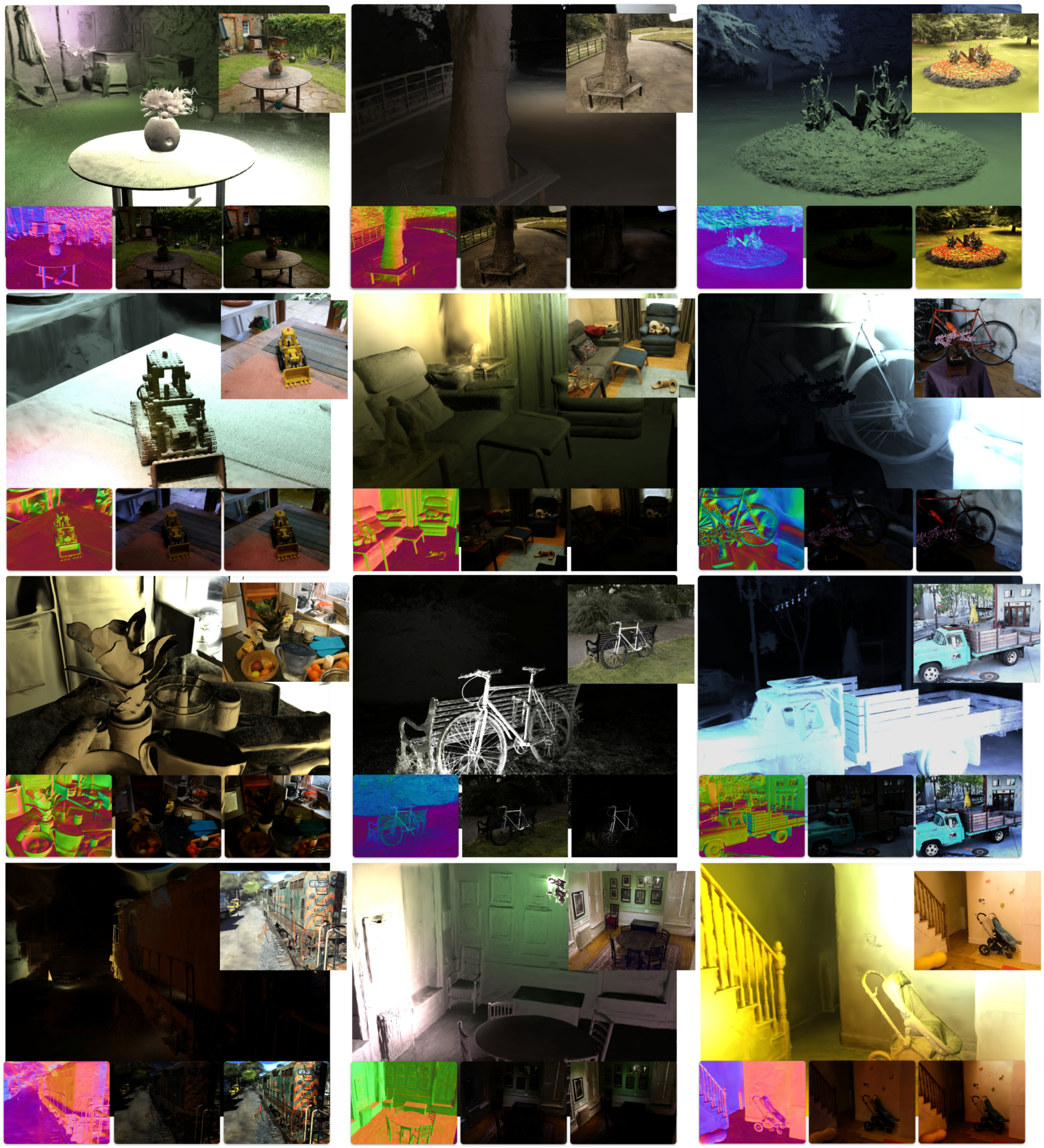

Figure 3. Visualization of our illumination decomposition results on the Mip-NeRF 360, DeepBlending, and T&T datasets. The main image shows the average of all effective pure localized light in the scene, while other smaller images include the normal map, environment light, localized light, and ambient occlusion.

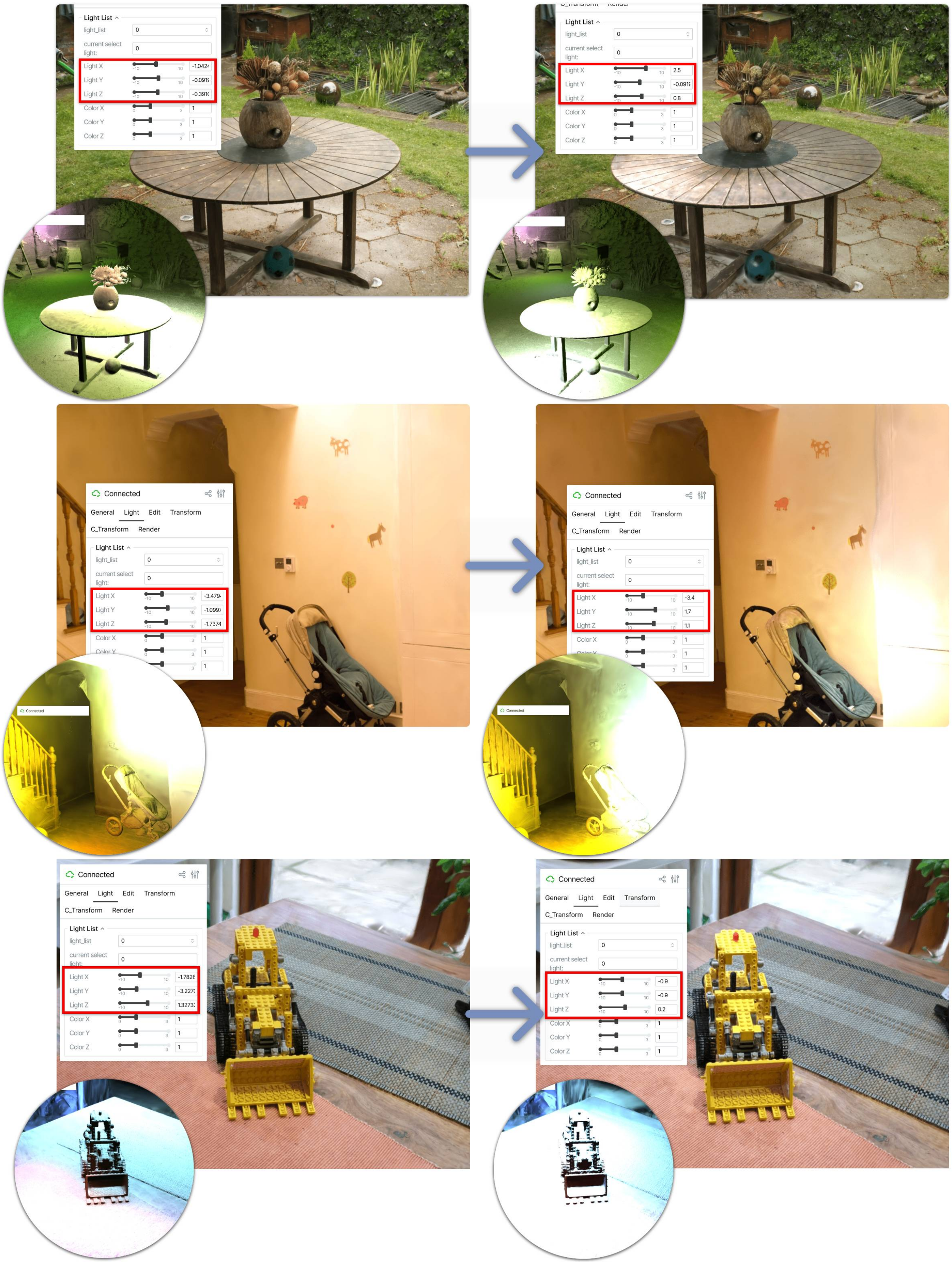

Figure 4. Visualization of light editing results.

Figure 5. Visualization of scene composition results.