

GaussHDR: High Dynamic Range Gaussian Splatting via Learning Unified 3D and 2D Local Tone Mapping

Supplementary Material

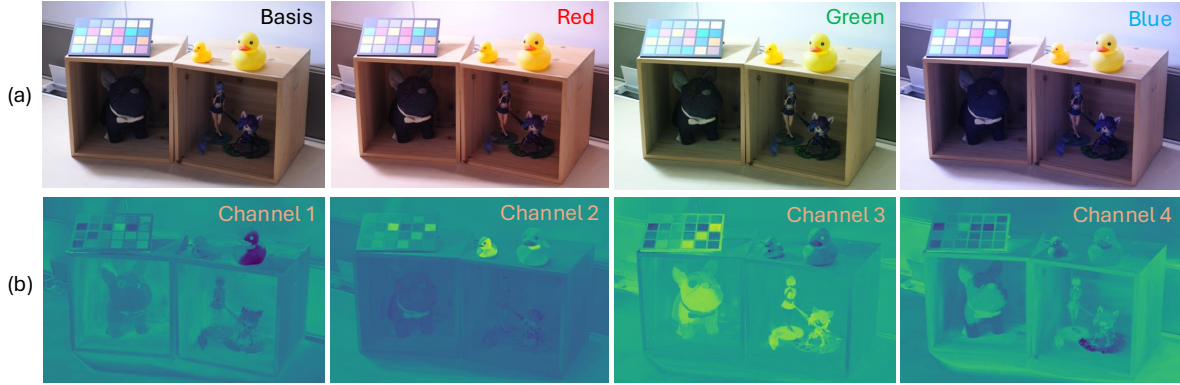


Figure 1. (a) By multiplying a factor to different RGB channels, we can control the white balance. (b) We visualize the 4-dimension context features channel by channel, which shows that different channels focus on capturing the tone-mapping characteristics of various regions.

1. Additional Implementation Details

Datasets. The original image resolutions of HDR-NeRF [2] synthetic and real datasets are 800×800 and 3216×2136 , respectively. We operate on the $\frac{1}{2}$ scale for synthetic scenes, *i.e.*, a resolution of 400×400 , and the $\frac{1}{4}$ scale for real scenes, *i.e.*, a resolution of 804×534 . For the four additional HDR-Plenoxels [3] real scenes, *desk* and *plant* scenes contain five different exposure times $\{t_1, t_2, t_3, t_4, t_5\}$, while *character* and *coffee* scenes include only three exposure times $\{t_1, t_3, t_5\}$. Consequently, we compute the averaged LDR-OE (t_1, t_3, t_5) metrics across the four scenes and the averaged LDR-NE (t_2, t_4) metrics over the *desk* and *plant* scenes. The original resolution of HDR-Plenoxels datasets is 5952×4480 . We operate on the $\frac{1}{6}$ scale, leading to a resolution of 992×746 .

Training Details. When employing 3DGS [5] and ScaffoldGS [8] representations, we adhere to the same training parameters as those specified in [5] and [8], respectively. For our local tone mapper and uncertainty model, the learning rate is initially set to 5×10^{-4} and exponentially decays to 5×10^{-5} .

2. Additional Results

2.1. Performance Comparison

The HDR irradiance, tone mapper, and context features are coupled during the scene-specific optimization. However, once we have obtained the HDR radiance field, we can still control the white balance by multiplying a factor to different channels, as shown in Fig. 1(a). Similarly, we can ignore

the context features and switch to other known global tone mappers for different styles, since the context features are only applicable to the learned local tone mapper.

2.2. Performance Comparison

We have presented the quantitative comparison results under the Exp-3 setting in the main paper. Here, we provide the results under Exp-1 setting, which strictly follows HDR-NeRF [2] by randomly selecting one exposure from $\{t_1, t_3, t_5\}$ for each view and keeping it fixed during training. The quantitative results for the real and synthetic datasets are listed in Tab. 2 and Tab. 3, respectively. Per-scene quantitative comparison outcomes are depicted in Tabs. 4 to 7. Additionally, we offer more examples of LDR and HDR qualitative results in the supplementary material, as demonstrated in Fig. 2 and Fig. 3, respectively. We also present three video demos corresponding to three different scenes: *bathroom* from HDR-NeRF [2] synthetic datasets, *flower* from HDR-NeRF real datasets, and *character* from HDR-Plenoxels [3] datasets. All the results further demonstrate that our method achieves state-of-the-art performance by improving both HDR and LDR learning.

2.3. Ablation on Uncertainty Learning

In the main paper, to investigate the effect of uncertainty learning, we perform experiments using a mixed loss function defined as $\mathcal{L} = \beta \mathcal{L}_{3d} + (1 - \beta) \mathcal{L}_{2d}$, which simply combines \mathcal{L}_{3d} and \mathcal{L}_{2d} through weighted summation. Then, we plot the performance variations of I_{3d}^* and I_{2d}^* with respect to the hyper-parameter β . Here, we show the results of more scenes, as depicted in Fig. 4, indicating that different scenes

Table 1. Additional ablation results on HDR-NeRF[2] datasets. Pos. indicates using Gaussian position as context feature for local tone-mapping. LDR PSNR denotes the average metric over all 5 exposures. All experiments are under the Exp-1 setting.

	Method	HDR-NeRF [2] Real Scenes	HDR-NeRF [2] Synthetic Scenes	
		LDR PSNR	LDR PSNR	HDR PSNR
(a)	Ours-3DGS (with raw Pos.)	32.68	40.21	36.91
(b)	Ours-3DGS (with fourier-encoded Pos.)	32.95	40.63	37.09
(c)	Ours-3DGS	33.58	41.32	37.41
(d)	Ours-Scaffold-GS (direct learnable weights)	34.03	42.01	38.12
(e)	Ours-Scaffold-GS (3-layer MLPs, 128 nodes)	34.75	42.43	38.67
(f)	Ours-Scaffold-GS	34.69	42.57	38.60

Table 2. Quantitative comparisons on HDR-NeRF [2] and HDR-Plenoxels [3] real datasets. Metrics are averaged over all scenes. LDR-OE and LDR-NE denote the LDR results with exposure $\{t_1, t_3, t_5\}$ and $\{t_2, t_4\}$, respectively. The training exposure setting is Exp-1.

Method	HDR-NeRF [2] Real Scenes						HDR-Plenoxels [3] Real Scenes					
	LDR-OE (t_1, t_3, t_5)			LDR-NE (t_2, t_4)			LDR-OE (t_1, t_3, t_5)			LDR-NE (t_2, t_4)		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
HDR-NeRF [2]	31.63	0.948	0.069	31.43	0.943	0.069	-	-	-	-	-	-
HDR-GS [1] \dagger	33.56	0.964	0.024	31.18	0.960	0.027	30.18	0.943	0.044	27.63	0.916	0.056
Ours (3DGS)	34.18	0.965	0.019	32.63	0.962	0.021	30.99	0.947	0.041	28.08	0.923	0.050
HDR-Scaffold-GS [8] *	34.09	0.967	0.016	31.88	0.964	0.019	31.05	0.944	0.045	28.13	0.916	0.059
Ours (Scaffold-GS)	35.37	0.972	0.014	33.68	0.969	0.016	32.24	0.954	0.031	28.98	0.932	0.041

\dagger We re-implement HDR-GS [1] under Exp-1 setting for fair comparison.

* We replace the scene representation in HDR-GS from 3DGS [5] to Scaffold-GS [8] to establish a baseline for our method utilizing Scaffold-GS.

exhibit varying balances between \mathcal{L}_{3d} and \mathcal{L}_{2d} . However, our uncertainty-based modulation can robustly achieve optimal results across diverse scenes.

?? in the main paper can be viewed as a regularization term that encourages the model to learn meaningful uncertainties or variances. In this part, we also try to directly employ learnable weights for the dual LDR results without any regularization. As shown in Tab. 1(d), this approach results in a performance drop, again highlighting the effectiveness of our uncertainty modeling.

2.4. Ablation on Tone-mapper MLPs

Since RGB channels may have different tone-mapping characteristics in real-world scenarios. Hence, using channel-specific MLPs for tone mapper is helpful, which has been verified in the HDR-NeRF [2] paper. Here, we also experiment with larger tone-mapper MLPs with deeper layers and more hidden nodes, as listed in Tab. 1(e). We can see that larger MLPs do not bring further improvements, but will increase the training and inference cost.

2.5. Ablation on Context Feature

In the main paper, we claim that utilizing pixel positions in image space for 2D local tone mapping is infeasible, as we need to distinguish the same pixel position across different views. However, we can directly leverage Gaussian positions as tone-mapping context features and render them into image space. We conduct ablation experiments for this with 3DGS [5] representation, since Scaffold-GS [8] utilizes anchor context features to predict Gaussian positions. As listed in Tab. 1(a,b), it will cause a degraded

performance with either raw Gaussian positions or Fourier-encoded ones. Using Gaussian positions implies that various Gaussians that are far apart must possess distinct tone-mapping characteristics. However, in fact, different spatial locations may exhibit similar characteristics.

We also visualize the 4-dimension context feature channel by channel in Fig. 1(b), which indicates that different channels focus on capturing the tone-mapping characteristics of various regions.

3. Limitations

Despite the significant enhancements in HDR reconstruction and LDR fitting capabilities provided by GaussHDR, some limitations remain. First, our method relies on COLMAP [10] to extract the initial point cloud and compute camera poses. In other words, we must use tripod-mounted cameras to capture multi-exposure images at each sampled view. However, a more general application scenario could involve utilizing a hand-held camera to capture a monocular video around the scene, where each frame (or view) has a single exposure level, making it difficult to match images captured under different exposures in COLMAP. Therefore, a potential direction for future work is to develop a COLMAP-free version of GaussHDR. Second, it is promising to introduce depth priors to enhance geometry reconstruction by utilizing off-the-shelf depth models [4, 6, 7, 9, 11, 12]. Finally, our method focuses on static scenes and lacks the ability to perform HDR reconstruction in dynamic environments, which is also an area worth exploring.

Table 3. Quantitative comparisons on HDR-NeRF [2] synthetic datasets. Metrics are averaged over all scenes. LDR-OE and LDR-NE denote the LDR results with exposure $\{t_1, t_3, t_5\}$ and $\{t_2, t_4\}$, respectively. HDR denotes the HDR results. The training setting is Exp-1.

Method	LDR-OE (t_1, t_3, t_5)			LDR-NE (t_2, t_4)			HDR		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
HDR-NeRF [2]	39.07	0.973	0.026	37.53	0.966	0.024	36.40	0.936	0.018
HDR-GS [1] \dagger	39.35	0.977	0.012	38.01	0.976	0.013	22.58	0.840	0.075
Ours (3DGS)	41.51	0.984	0.008	41.03	0.983	0.009	37.41	0.969	0.017
HDR-Scaffold-GS [8] *	42.21	0.986	0.005	40.44	0.986	0.006	26.11	0.915	0.062
Ours (Scaffold-GS)	42.94	0.988	0.004	42.02	0.988	0.005	38.60	0.975	0.011

\dagger We re-implement HDR-GS [1] under Exp-1 setting for fair comparison. Note that the authors of HDR-GS utilize HDR ground truth (GT) for supervision during training on synthetic datasets, whereas our re-implementation does not include this supervision.

* We replace the scene representation in HDR-GS from 3DGS [5] to Scaffold-GS [8] to establish a baseline for our method utilizing Scaffold-GS.

Table 4. Per-scene quantitative comparisons on HDR-NeRF [2] real datasets. LDR-OE and LDR-NE denote the LDR results with exposure $\{t_1, t_3, t_5\}$ and $\{t_2, t_4\}$, respectively. The training exposure setting is Exp-1.

	Method	Box			Computer			Flower			Luckycat		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
LDR-OE	HDR-NeRF [2]	31.54	0.953	0.068	32.42	0.950	0.077	29.81	0.948	0.069	32.85	0.938	0.062
	HDR-GS [1] \dagger	34.39	0.973	0.016	34.41	0.967	0.020	31.62	0.956	0.037	33.82	0.961	0.022
	Ours (3DGS)	35.87	0.976	0.012	34.99	0.967	0.020	31.84	0.958	0.024	34.03	0.961	0.020
	HDR-Scaffold-GS [8] *	34.66	0.974	0.012	34.55	0.965	0.017	32.70	0.967	0.019	34.43	0.963	0.016
	Ours (Scaffold-GS)	36.24	0.979	0.010	35.50	0.970	0.015	34.12	0.971	0.014	35.61	0.970	0.015
LDR-NE	HDR-NeRF [2]	31.40	0.944	0.079	31.21	0.931	0.098	30.05	0.949	0.058	33.13	0.948	0.051
	HDR-GS [1] \dagger	30.17	0.967	0.021	32.78	0.966	0.023	30.21	0.954	0.039	31.55	0.954	0.026
	Ours (3DGS)	32.57	0.972	0.014	34.18	0.966	0.022	30.52	0.956	0.027	33.24	0.956	0.022
	HDR-Scaffold-GS [8] *	31.87	0.970	0.015	33.16	0.964	0.019	30.68	0.964	0.021	31.80	0.957	0.019
	Ours (Scaffold-GS)	32.93	0.974	0.013	34.95	0.968	0.017	32.35	0.970	0.016	34.48	0.965	0.016

\dagger We re-implement HDR-GS [1] under Exp-1 setting for fair comparison.

* We replace the scene representation in HDR-GS from 3DGS [5] to Scaffold-GS [8] to establish a baseline for our method utilizing Scaffold-GS.

References

- [1] Yuanhao Cai, Zihao Xiao, Yixun Liang, 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. 2, 3, 4
- [2] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In *CVPR*, pages 18377–18387, 2022. 1, 2, 3, 4, 6
- [3] Kim Jun-Seong, Kim Yu-Ji, Moon Ye-Bin, and Tae-Hyun Oh. Hdr-plenoxels: Self-calibrating high dynamic range radiance fields. In *ECCV*, pages 384–401, 2022. 1, 2, 4
- [4] Bingxin Ke, Anton Obukhov, and Shengyu Huang. Repurposing diffusion-based image generators for monocular depth estimation. In *CVPR*, pages 9492–9502, 2024. 2
- [5] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM TOG*, 42(4):1–14, 2023. 1, 2, 3, 4
- [6] Jinfeng Liu, Lingtong Kong, Bo Li, Zerong Wang, Hong Gu, and Jinwei Chen. Mono-vifi: A unified learning framework for self-supervised single and multi-frame monocular depth estimation. In *ECCV*, pages 90–107, 2024. 2
- [7] Jinfeng Liu, Lingtong Kong, Jie Yang, and Wei Liu. Towards better data exploitation in self-supervised monocular depth estimation. *IEEE Robotics and Automation Letters*, 9(1):763–770, 2024. 2
- [8] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, and Dahua Lin. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *CVPR*, pages 20654–20664, 2024. 1, 2, 3, 4
- [9] Luigi Piccinelli, Yung-Hsu Yang, Christos Sakaridis, Mattia Segu, Siyuan Li, Luc Van Gool, and Fisher Yu. Unidepth: Universal monocular metric depth estimation. In *CVPR*, pages 10106–10116, 2024. 2
- [10] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, pages 4104–4113, 2016. 2
- [11] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR*, pages 10371–10381, 2024. 2
- [12] Wei Yin, Chi Zhang, Hao Chen, Zhipeng Cai, Gang Yu, Kaixuan Wang, Xiaozhi Chen, and Chunhua Shen. Metric3d: Towards zero-shot metric 3d prediction from a single image. In *ICCV*, pages 9009–9019, 2023. 2

Table 5. Per-scene quantitative comparisons on HDR-Plenoxels [3] real datasets. LDR-OE and LDR-NE denote the LDR results with exposure $\{t_1, t_3, t_5\}$ and $\{t_2, t_4\}$, respectively. The training exposure setting is Exp-1.

	Method	Character			Coffee			Desk			Plant		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
LDR-OE	HDR-GS [1] \dagger	34.96	0.978	0.027	27.79	0.943	0.049	28.78	0.922	0.045	29.18	0.930	0.054
	Ours (3DGS)	36.73	0.980	0.025	28.35	0.947	0.046	29.14	0.925	0.041	29.76	0.934	0.051
	HDR-Scaffold-GS [8] *	36.81	0.979	0.029	28.31	0.946	0.043	28.64	0.919	0.049	30.44	0.930	0.059
	Ours (Scaffold-GS)	38.24	0.983	0.018	29.31	0.955	0.032	29.91	0.933	0.035	31.48	0.945	0.040
LDR-NE	HDR-GS [1] \dagger	-	-	-	-	-	-	27.08	0.912	0.050	28.17	0.920	0.061
	Ours (3DGS)	-	-	-	-	-	-	27.30	0.917	0.044	28.85	0.929	0.055
	HDR-Scaffold-GS [8] *	-	-	-	-	-	-	27.03	0.911	0.053	29.23	0.920	0.065
	Ours (Scaffold-GS)	-	-	-	-	-	-	27.41	0.922	0.038	30.55	0.941	0.044

\dagger We re-implement HDR-GS [1] under Exp-1 setting for fair comparison.

* We replace the scene representation in HDR-GS from 3DGS [5] to Scaffold-GS [8] to establish a baseline for our method utilizing Scaffold-GS.

Table 6. Per-scene quantitative comparisons on HDR-NeRF [2] synthetic datasets (Part 1). LDR-OE and LDR-NE denote the LDR results with exposure $\{t_1, t_3, t_5\}$ and $\{t_2, t_4\}$, respectively. HDR denotes the HDR results. The training exposure setting is Exp-1.

	Method	Bathroom			Bear			Chair			Diningroom		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
LDR-OE	HDR-NeRF [2]	36.26	0.949	0.037	42.91	0.990	0.010	32.45	0.905	0.081	41.23	0.986	0.010
	HDR-GS [1] \dagger	38.06	0.963	0.020	41.61	0.989	0.005	35.07	0.952	0.024	38.26	0.979	0.014
	Ours (3DGS)	41.12	0.975	0.008	44.44	0.992	0.003	37.05	0.968	0.014	39.63	0.981	0.017
	HDR-Scaffold-GS [8] *	41.21	0.977	0.008	44.56	0.992	0.003	36.36	0.966	0.016	44.87	0.994	0.002
	Ours (Scaffold-GS)	42.08	0.981	0.006	45.40	0.993	0.002	37.65	0.971	0.012	45.21	0.994	0.002
LDR-NE	HDR-NeRF [2]	33.44	0.926	0.046	41.19	0.987	0.012	30.78	0.886	0.083	37.99	0.979	0.013
	HDR-GS [1] \dagger	36.03	0.963	0.024	41.69	0.988	0.005	34.05	0.951	0.026	34.62	0.976	0.020
	Ours (3DGS)	41.10	0.977	0.008	43.85	0.992	0.003	36.53	0.967	0.015	38.78	0.980	0.021
	HDR-Scaffold-GS [8] *	40.61	0.979	0.008	42.10	0.992	0.004	35.89	0.965	0.017	41.00	0.993	0.003
	Ours (Scaffold-GS)	41.71	0.982	0.007	44.43	0.993	0.002	37.14	0.970	0.013	43.33	0.993	0.003
HDR	HDR-NeRF [2]	33.97	0.925	0.048	43.22	0.991	0.008	34.14	0.924	0.069	38.57	0.981	0.015
	HDR-GS [1] \dagger	15.48	0.739	0.125	30.99	0.961	0.027	20.47	0.751	0.115	18.86	0.860	0.074
	Ours (3DGS)	34.87	0.943	0.024	41.47	0.987	0.007	36.75	0.961	0.019	34.82	0.967	0.028
	HDR-Scaffold-GS [8] *	22.28	0.873	0.061	29.79	0.970	0.021	24.27	0.842	0.163	26.60	0.952	0.037
	Ours (Scaffold-GS)	36.22	0.952	0.017	42.23	0.988	0.005	37.79	0.966	0.017	38.61	0.983	0.007

\dagger We re-implement HDR-GS [1] under Exp-1 setting for fair comparison.

* We replace the scene representation in HDR-GS from 3DGS [5] to Scaffold-GS [8] to establish a baseline for our method utilizing Scaffold-GS.

Table 7. Per-scene quantitative comparisons on HDR-NeRF [2] synthetic datasets (Part 2). LDR-OE and LDR-NE denote the LDR results with exposure $\{t_1, t_3, t_5\}$ and $\{t_2, t_4\}$, respectively. HDR denotes the HDR results. The training exposure setting is Exp-1.

	Method	Dog			Desk			Sofa			Sponza		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
LDR-OE	HDR-NeRF [2]	37.77	0.981	0.016	37.84	0.972	0.023	38.29	0.977	0.014	34.49	0.958	0.034
	HDR-GS [1] \dagger	42.28	0.991	0.004	39.81	0.980	0.007	41.75	0.989	0.004	37.99	0.972	0.016
	Ours (3DGS)	43.06	0.992	0.004	42.28	0.987	0.004	43.04	0.990	0.003	41.43	0.986	0.007
	HDR-Scaffold-GS [8] *	43.31	0.993	0.003	41.72	0.987	0.004	43.26	0.992	0.003	42.43	0.990	0.004
	Ours (Scaffold-GS)	43.76	0.994	0.002	43.04	0.990	0.003	43.44	0.992	0.003	42.91	0.991	0.004
LDR-NE	HDR-NeRF [2]	36.52	0.976	0.018	35.26	0.960	0.029	38.35	0.976	0.014	33.41	0.950	0.038
	HDR-GS [1] \dagger	40.66	0.990	0.004	38.47	0.980	0.007	41.55	0.989	0.004	36.99	0.975	0.015
	Ours (3DGS)	42.24	0.991	0.004	42.12	0.987	0.004	42.34	0.991	0.003	41.25	0.988	0.006
	HDR-Scaffold-GS [8] *	40.52	0.992	0.004	40.46	0.987	0.004	41.91	0.991	0.004	41.02	0.990	0.004
	Ours (Scaffold-GS)	42.12	0.993	0.003	42.76	0.989	0.003	42.03	0.992	0.003	42.60	0.992	0.003
HDR	HDR-NeRF [2]	37.72	0.980	0.016	43.38	0.993	0.007	39.05	0.976	0.017	32.33	0.939	0.049
	HDR-GS [1] \dagger	23.02	0.926	0.037	29.11	0.773	0.082	26.60	0.928	0.046	16.12	0.782	0.095
	Ours (3DGS)	36.36	0.973	0.016	43.98	0.993	0.006	36.70	0.966	0.016	34.32	0.964	0.023
	HDR-Scaffold-GS [8] *	24.33	0.934	0.040	31.40	0.929	0.044	28.42	0.947	0.032	21.81	0.870	0.095
	Ours (Scaffold-GS)	37.52	0.978	0.012	43.83	0.993	0.005	36.99	0.969	0.013	35.64	0.972	0.014

\dagger We re-implement HDR-GS [1] under Exp-1 setting for fair comparison.

* We replace the scene representation in HDR-GS from 3DGS [5] to Scaffold-GS [8] to establish a baseline for our method utilizing Scaffold-GS.

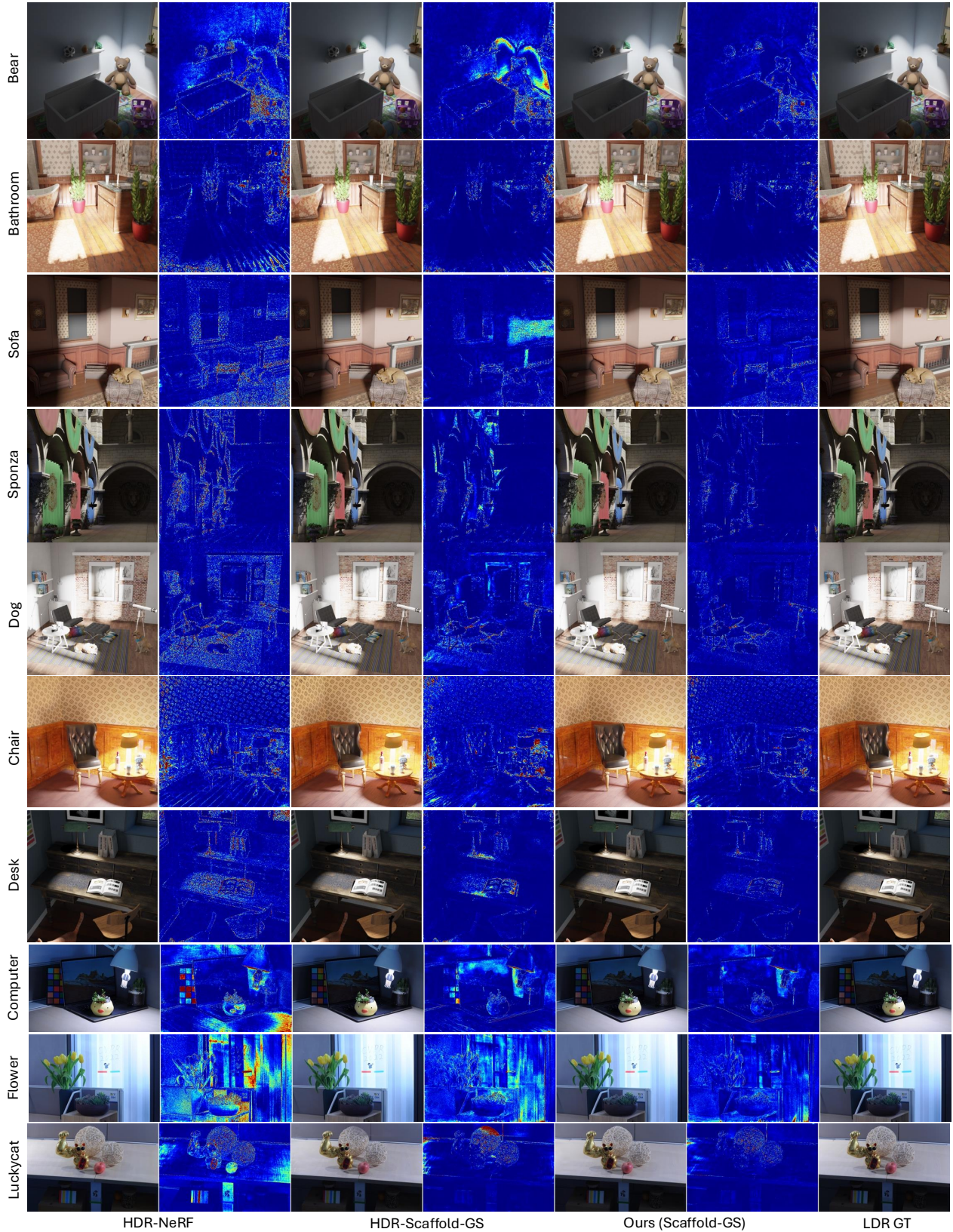


Figure 2. Qualitative LDR comparisons. Error maps in column 2, 4 and 6 show the MSE error compared to the ground truth, where color from blue to red indicates the error from small to large. Our method can reduce LDR fitting errors in some regions.

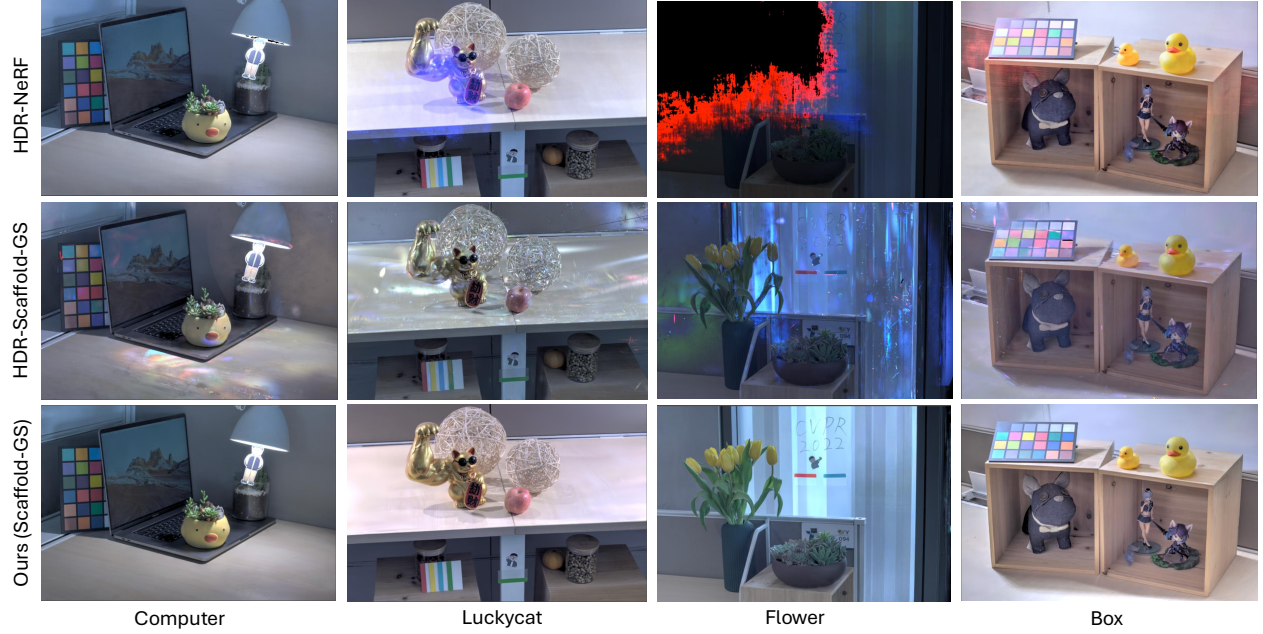


Figure 3. Qualitative HDR comparisons. Our method leads to stable HDR reconstruction results compared to the baselines.

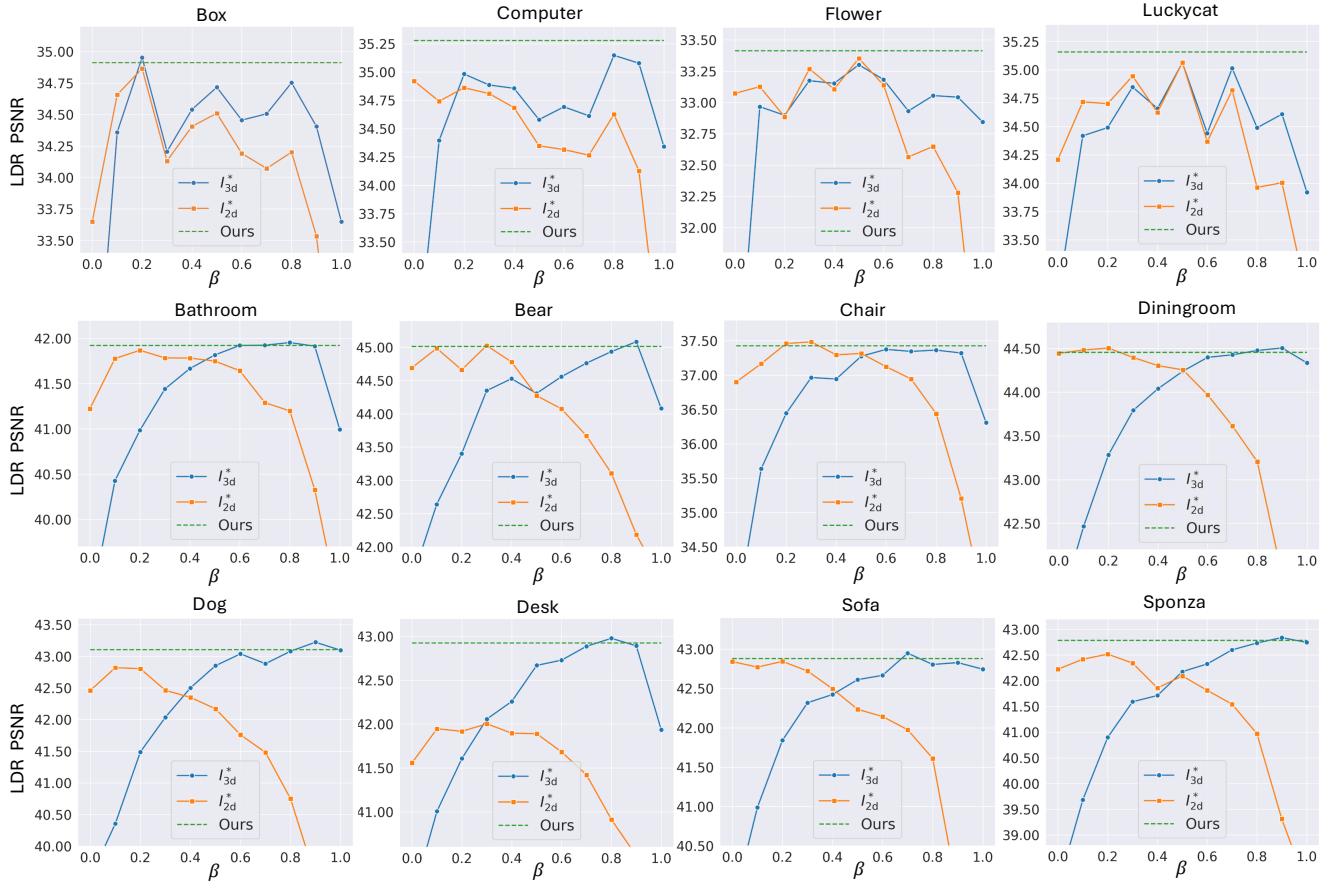


Figure 4. Performance variations of I_{3d}^* and I_{2d}^* with respect to β when simply using a loss combination $\mathcal{L} = \beta \mathcal{L}_{3d} + (1 - \beta) \mathcal{L}_{2d}$. Different scenes exhibit varying optimal values of β . In contrast, our uncertainty-based modulation (green dash line) can robustly achieve optimal results across diverse scenes without the selection of hyper-parameter β . Results include all of HDR-NeRF [2] real and synthetic scenes.