

*L*²*DGS*: Low-Light Dynamic Gaussian Splatting (Supplementary Material)

Ashish Kumar A. N. Rajagopalan
Indian Institute of Technology Madras, India

ashish.k.research@gmail.com, raju@ee.iitm.ac.in

All figures and tables in the supplementary are labeled with prefix S. The supplementary material is structured as follows:

1. OCD-Net Architecture
2. BAFE-Net Architecture
3. Datasets
4. Additional comparisons with competing methods
5. Qualitative Ablations
6. Societal Impact
7. Discussion on Loss functions and B_1 and B_2
8. Optimization time
9. Clarification on Human Subjects
10. Additional Comparisons with GS4D

This supplementary material is also accompanied by a ‘.mp4’ video file.

S1. OCD-Net

The Occlusion-Disocclusion Network (OCD-Net) takes the features from the motion field, $f(\mu_i^t)$, as input to estimate the occlusion-disocclusion scale factors, $a_j(t)$ for each of the $(k+1)^2$ spherical harmonics coefficients, which is multiplied with the SH coefficients of l_i^c to yield l_i^t (Eq. 3 in main paper), which depends on both time-and-view. OCD-Net comprises two Multi-Layer Perceptrons (MLPs), with ReLU activation applied at the output layer. In our experiments, we set $k = 3$. The block diagram of OCD-Net is illustrated in Fig. S1. The intuition behind this design

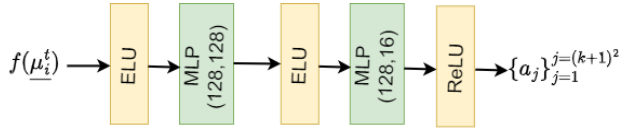


Figure S1. Block diagram of OCD-Net.

choice is that the input feature encodes spatiotemporal cues where the first MLP + ELU effectively extracts features corresponding to G_m^t , thereby acting as \mathcal{F}_1 . The second MLP + ReLU then serves as \mathcal{F}_2 by transforming the G_m^t features (output from the first MLP) into $a_j(t)$. Hence, OCD-Net implicitly models the composition of \mathcal{F}_1 and \mathcal{F}_2 . The two

functions are not redundant but rather contribute complementary information to the final mapping.

S2. BAFE-Net

The rasterized 2D features, \hat{B}_1 and \hat{B}_2 , inherently lack contextual information about the surrounding contents of the scene. To address this, BAFE-Net, a Convolutional Neural Network (CNN), adjusts these features by incorporating local neighborhood information. BAFE-Net enhances the BAF features by making them aware of their spatial neighborhood. The architecture of BAFE-Net is illustrated in Fig. S2. The outputs B_1 and B_2 are constrained to the range $[1, 10]$ to restrict the solution space in the well-lit domain.

S3. Datasets

S3.2 Low-Light Simulated Synthetic Data

We synthesize low-light videos using the method proposed in [12]. For this purpose, we utilize commonly used well-lit and high-visibility dynamic videos, including those from the iPhone dataset [3] and HyperNeRF [4], which are widely adopted in the NeRF and Gaussian Splatting (GS) domains. Only videos that appear realistic after the low-light transformation are selected for inclusion. Fig. S3 illustrates characteristics of the synthetic data.

S3.1 Proposed *L*²*DyV* Real Dataset

The *L*²*DyV* dataset (which we shall release) encompasses a wide range of complexities in low-light scenes, captured from diverse and unconstrained camera poses. It features significant variability in visibility and extremely low illumination levels. Key characteristics of each scene are illustrated in Table S1. Overall, the *L*²*DyV* dataset exhibits the following features:

1. Videos with diverse camera motion and object motion.
2. Scenes with multiple levels of visibility; some regions are well-lit and clearly visible, while other regions within the same scene may have significantly lower visibility.
3. Presence of both single and multiple dynamic objects, spanning different object classes and scales.

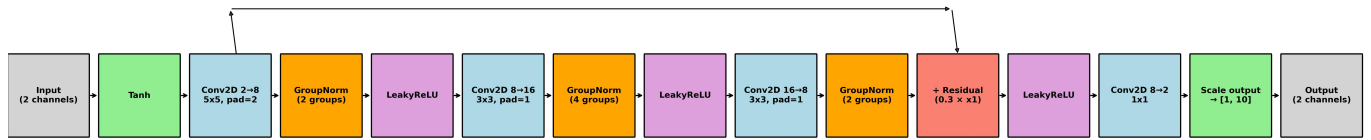


Figure S2. Block diagram of BAFE-Net.

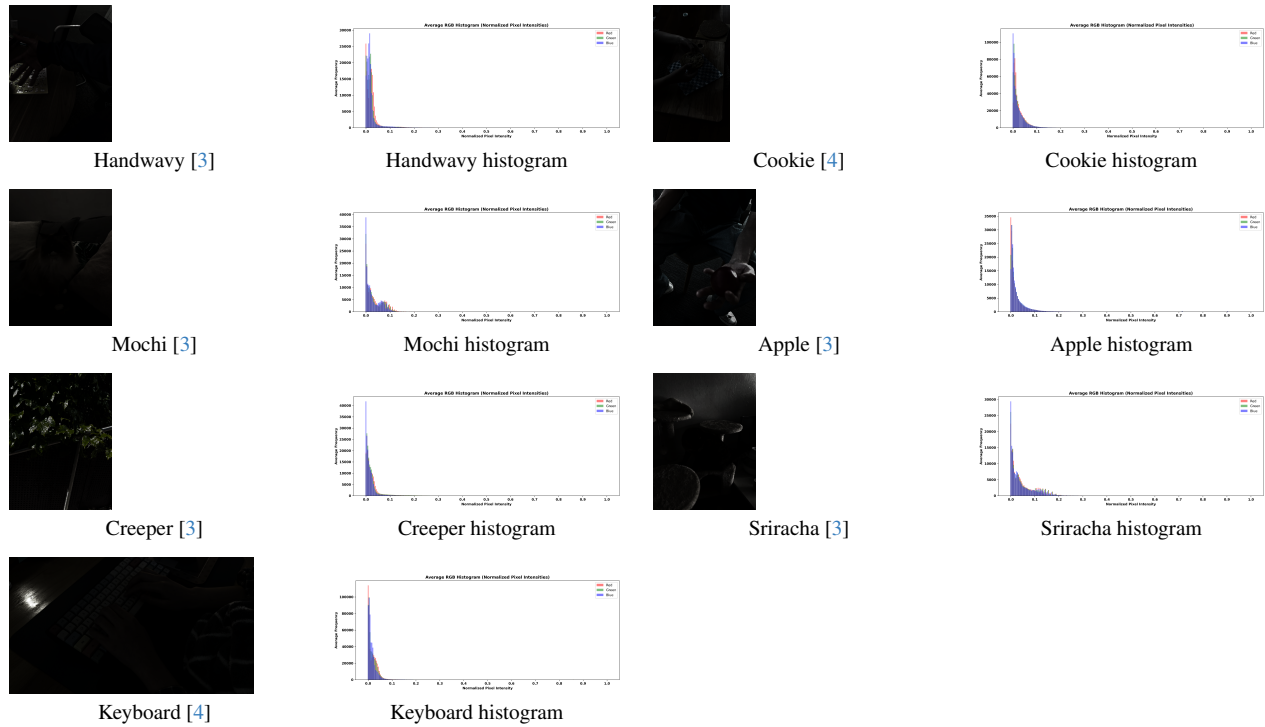

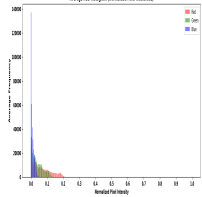
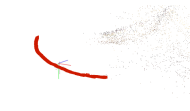

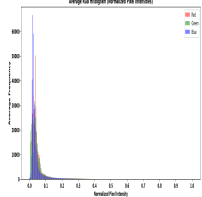


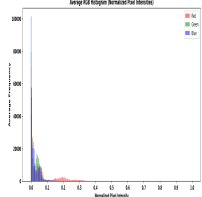


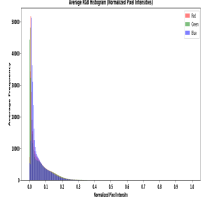


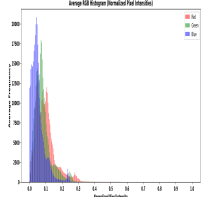


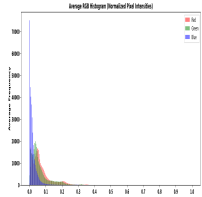


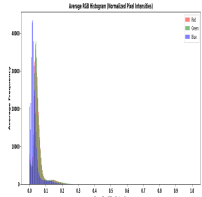
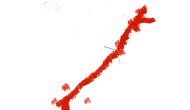


Figure S3. This figure shows a synthetic frame for each video and the RGB histogram averaged across all frames for each video. The majority of the pixel intensity values are below 0.1, demonstrating extremely low-light conditions and hence low visibility and loss of details such as color, geometry.

4. Inclusion of both rigid and deformable dynamic objects.
5. Dynamic object motion varies not only across scenes but also within each scene.
6. Cases of occlusion and disocclusion involving dynamic objects.
7. Presence of dynamic shadows within the scenes.
8. Scenes that include transparent objects.
9. Scenes that include time-varying intensity in static part of the scene due to light source motion.
10. Videos of varying lengths, ranging from 100 to 300 frames.

S.No	Video Name	Sample Image	RGB Histogram	Camera Poses	Salient Features
1	Skating				<ul style="list-style-type: none"> Multiple dynamic objects Static as well as dynamic shadows Visibility transition of a dynamic object Outdoor 300 frames
2	Deer				<ul style="list-style-type: none"> Visibility variations Multiple deformable dynamic objects Dynamic and static shadows Outdoor 270 frames
3	File				<ul style="list-style-type: none"> Significant outdoor visibility/brightness Reflective surface Color distortions Indoor 249 frames
4	Cat				<ul style="list-style-type: none"> Self occlusion–disocclusion Static shadows Outdoor 150 frames
5	Multicar				<ul style="list-style-type: none"> Multiple dynamic objects Multiple dynamic light sources Time-dependent brightness in static regions Outdoor 300 frames
6	Deer-2				<ul style="list-style-type: none"> Uneven visibility Multiple dynamic objects Shadows in low-light Outdoor 300 frames
7	Cat-2				<ul style="list-style-type: none"> Uneven visibility Small dynamic objects Multiclass dynamic objects (cat and car) Outdoor 272 frames


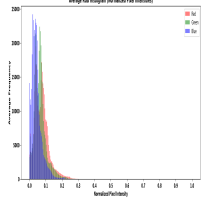


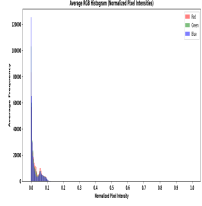


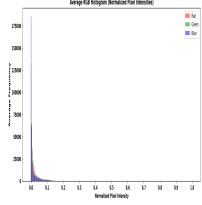


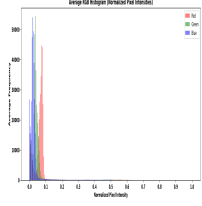
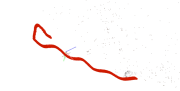

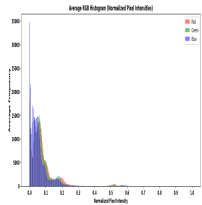

S.No	Video Name	Sample Image	RGB Histogram	Camera Poses	Salient Features
8	Car				<ul style="list-style-type: none"> • Temporally varying brightness in static regions • Dynamic light source • Rigid dynamic object • Outdoor • 260 frames
9	Whiteboard				<ul style="list-style-type: none"> • Transparent objects • Distant light source reflections • Low visibility • Indoor • 290 frames
10	Activity				<ul style="list-style-type: none"> • Low visibility • Complex camera trajectory • Indoor • 298 frames
11	Kids				<ul style="list-style-type: none"> • Multiple dynamic objects • Multiple actions • Outdoor • 100 frames
12	Bus				<ul style="list-style-type: none"> • Dynamic transparent object • Dynamic object scale variation • Occlusion-disocclusion of dynamic and static regions • Shadows • Outdoor • 200 frames

Table S1. Proposed L^2DyV Dataset: This dataset presents significant challenges across a wide range of challenging scenarios for dynamic scene reconstruction and illumination enhancement. Camera poses are estimated using COLMAP [6, 7]. The RGB histogram represents the distribution of pixel intensities across all frames in a video for a given scene. Histogram channels: red (R), green (G), blue (B).

S4. Qualitative comparisons with competing methods

S4.1 L^2DyV Dataset

Figures S7 through S13 contain comparisons on real-world data from L^2DyV dataset. Each figure consists of four rows: the top two rows show the well-lit rendered outputs from a test view using different methods, while the bottom two rows present the corresponding depth maps. In the depth maps, blue represents closer regions (smaller depth values), and red represents farther regions (larger depth values). The leftmost image in each figure is the corresponding low-light test view, which was not seen during training. Note that we have included an additional method, LushNeRF [5], for comparison. The reason we excluded it from the main paper is that it relies on [12] for low-light to well-lit transformation, unlike other baselines that contain their own enhancement modules. Across all these complex scenes, our method is the only one that successfully reconstructs the scene while preserving motion, geometry, and photometric consistency. Unlike other approaches, our method L^2DGS avoids saturation in well-lit areas and effectively handles the challenges posed by low-light dynamic environments. These include occlusion and disocclusion of both dynamic and static objects, temporally varying illumination due to moving light sources, and the presence of reflections. Competing methods mainly focus on increasing the scene brightness without considering the underlying scene dynamics.

S4.2 Low-light simulated dataset

Figures S14 through S19 contain comparisons on the synthetic data. Synthetic data has extremely low visibility and detail loss, as is evident from Fig. S3. Each figure consists of four rows: the top two rows show the well-lit rendered outputs from a test view using different methods, while the bottom two rows present the corresponding depth maps. In the depth maps, blue represents closer regions (smaller depth values), and red represents farther regions (larger depth values). The ground truth for the depth map is left blank as the ground truth of the depth map is not available. Note that we do not compare with GSDK [10] due to its dependence on camera metadata. The first image in each figure is the corresponding low-light test view, which was not seen during training. Across all these complex scenes, our method is the only one that consistently reconstructs the scene while preserving motion, geometry, and photometric consistency. While other approaches may enhance brightness better than L^2DGS , they often fail to accurately recover scene geometry. In contrast, L^2DGS enhances brightness in a manner that is consistently aware of geometry, motion, and photometric integrity across diverse scenes.

S5. Qualitative Ablations

In Fig. S4, we present qualitative results for different combinations of loss functions. The SSIM loss (the second term in $\mathcal{L}_{\text{photo}}$, as defined in Eq. 10 of the main paper) helps preserve the local structure of the scene; its absence results in noticeably blurry outputs. Both \mathcal{L}_{exp} and \mathcal{L}_L contribute to enhancing overall scene visibility. Although \mathcal{L}_D also improves visibility, it fails to reconstruct the fine details, as highlighted in the green bounding box (BB). The loss term \mathcal{L}_B is responsible for transforming the well-lit domain to the low-light domain. When omitted, the model struggles to preserve sharp edges, as is evident in the blue BB. BAFE-Net, a Convolutional Neural Network (CNN), leverages the inherent ability of CNNs to exploit local neighborhood information of \hat{B}_1 and \hat{B}_2 , ultimately estimating B_1 and B_2 , which capture the local structural cues of the well-lit scene. Our final loss \mathcal{L} which combines the effects of multiple regularizers produces best results that are well-lit, have improved visibility, and preserve both edge sharpness (blue BB) and fine details (green BB). It is also worth noting that maximizing R_w does not significantly affect reconstruction quality.

S6. Societal Impact

L^2DGS involves improving the visibility and quality of visual content captured in low-light environments, as well as synthesizing novel views. Just like any other advancement in the field, our method has both positive as well as negative societal impact. Some of the positive social impacts include

1. Improved scene/activity analysis: A better visibility ensures that the scene content and dynamics can be better analyzed, and depending on the application, the performance of the particular subject can be measured (for example, our skating scene, where the person’s pose can be analyzed).
2. Photography and videography (entertainment purpose): Photos and videos can be captured even in low-light environments and subsequently enhanced, or novel views and time instances can be synthesized for entertainment purposes.
3. Wildlife analysis: Analyzing animal behavior and interactions in low-light captured data, enhancing it to a well-lit version, and enabling novel spatio-temporal video synthesis can greatly support and improve the study of animal behavior and interactions.

Possible negative social impact include:

1. Manipulation of media: Enhanced visuals can be altered to create misleading narratives or fake videos.
2. Privacy concerns: Enhanced video footage could be misused for surveillance purposes, potentially leading to over-monitoring of citizens and infringement on privacy rights.

Method	Cookie									Sriracha								
	Dynamic			Static			Overall			Dynamic			Static			Overall		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
AlethNeRF [1]	13.03	0.36	0.09	13.89	0.54	0.43	13.71	0.51	0.52	14.23	0.30	0.08	12.60	0.58	0.39	12.80	0.53	0.47
LLNeRF [8]	11.40	0.47	0.05	11.92	0.50	0.23	11.82	0.50	0.28	14.08	0.30	0.05	13.51	0.53	0.27	13.59	0.49	0.32
LushNeRF [5]	5.86	0.12	0.10	7.44	0.25	0.44	7.11	0.22	0.53	12.17	0.20	0.05	11.36	0.56	0.24	11.40	0.50	0.28
Luminance-GS [2]	11.94	0.42	0.05	13.23	0.43	0.24	12.95	0.43	0.29	11.56	0.17	0.04	14.06	0.65	0.20	13.52	0.57	0.24
LITA-GS [11]	12.01	0.62	0.03	13.11	0.45	0.27	13.24	0.55	0.25	11.14	0.41	0.04	10.97	0.74	0.16	11.29	0.71	0.21
L^2DGS	18.81	0.78	0.04	18.52	0.79	0.17	18.57	0.79	0.20	18.14	0.63	0.03	11.93	0.77	0.13	12.49	0.75	0.15

Method	Keyboard								
	Dynamic			Static			Overall		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
AlethNeRF [1]	6.18	0.03	0.12	6.60	0.07	0.48	6.51	0.06	0.59
LLNeRF [8]	16.45	0.56	0.05	15.74	0.62	0.21	15.87	0.61	0.26
LushNeRF [5]	6.52	0.10	0.10	8.59	0.28	0.41	8.09	0.24	0.52
Luminance-GS [2]	15.54	0.53	0.05	13.60	0.44	0.20	13.92	0.45	0.25
LITA-GS [11]	14.48	0.55	0.05	12.83	0.55	0.20	13.67	0.55	0.19
L^2DGS	16.68	0.76	0.04	16.40	0.78	0.15	16.44	0.77	0.19

Table S2. Additional comparisons across synthetic simulated low-light scenes. L^2DGS consistently outperforms all competing methods by a significant margin in dynamic regions.

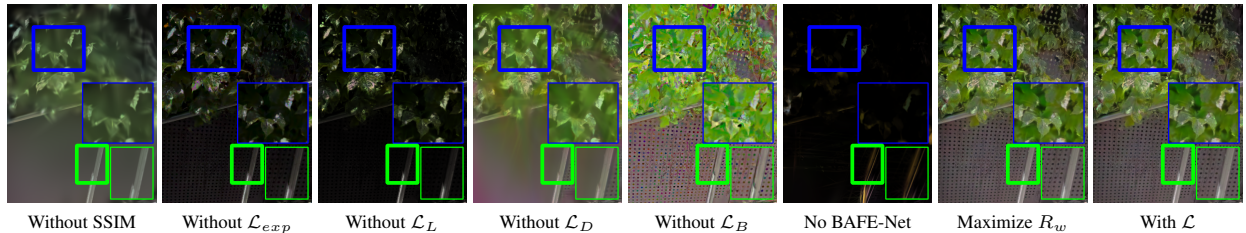


Figure S4. **Qualitative ablation study of individual loss components.** Each component contributes to the visual quality of the synthesized view. The final loss \mathcal{L} enhances structural sharpness and preserves lighting and details.

S7. Discussion on Loss functions and B_1 and B_2

Our self-supervised approach relies on careful selection and design of the loss functions to constrain the solution space. The loss in Eq. 10 is responsible for ensuring consistency of the rendered low-light image with the input low-light image for the same view and time. Eq. 12 imposes depth smoothness, except at locations with high image gradients. Eq. 11 encourages a moderate value for the average of R_w within a local neighborhood while Eq. 13 encourages L_w to be high (close to 1) so as to produce an overall well-lit image. Both illumination (L_w) and reflectance (R_w) possess well-defined physical interpretations (Retinex theory). The values that R_w and L_w take must be such that their product, after being transformed to low-light, respects the consistency condition imposed by Eq. 10. In our framework, we introduce B_1 and B_2 (the outputs of BAFE-Net) to systematically facilitate the transformation from well-lit to low-light domain. We design the loss in Eq. 14 such that the variation in B_1 is encouraged to follow the variation in R_w , and similarly, the variation in B_2 is guided to follow that of L_w . This introduces a desirable mutual dependency wherein B_1 is aligned with reflectance and B_2 with illumination to produce a geometry and photometry-aware domain transformation (Eq. 9). This intended behavior is vi-

usually confirmed in Fig. 2, where the estimated components B_1 and B_2 indeed align with the expected structural and photometric variations. In Row 3 of Fig. 3, the background is partially occluded by the tree. Notably, our method not only enhances visibility in this challenging region but also preserves the shadows of the tree.

It should also be noted that, in the Gaussian splatting framework, each Gaussian is parameterized by its mean, rotation, scale, opacity, and spherical harmonics coefficients. Since the ground-truth values for these parameters are seldom known, these attributes are learned jointly, so that the rendered images during training closely match the input views, and enable novel-view synthesis. For the synthetic dataset, we have also given objective metrics such as PSNR, SSIM and LPIPS values for the reconstructed views. In Fig. 2 of the main paper, we have shown an example to illustrate that the estimated B_1 and B_2 are indeed meaningful and exhibit alignment with the expected structural and photometric patterns. In our work, the proposed loss functions synergistically come together to solve for the unknowns. The results (main Table 1, supple Table S3, supple video timestamp 3:19-4:30, supple Figs. S4-S16) amply demonstrate how our method outperforms all existing methods in reconstructing dynamic well-lit views from low-light inputs.

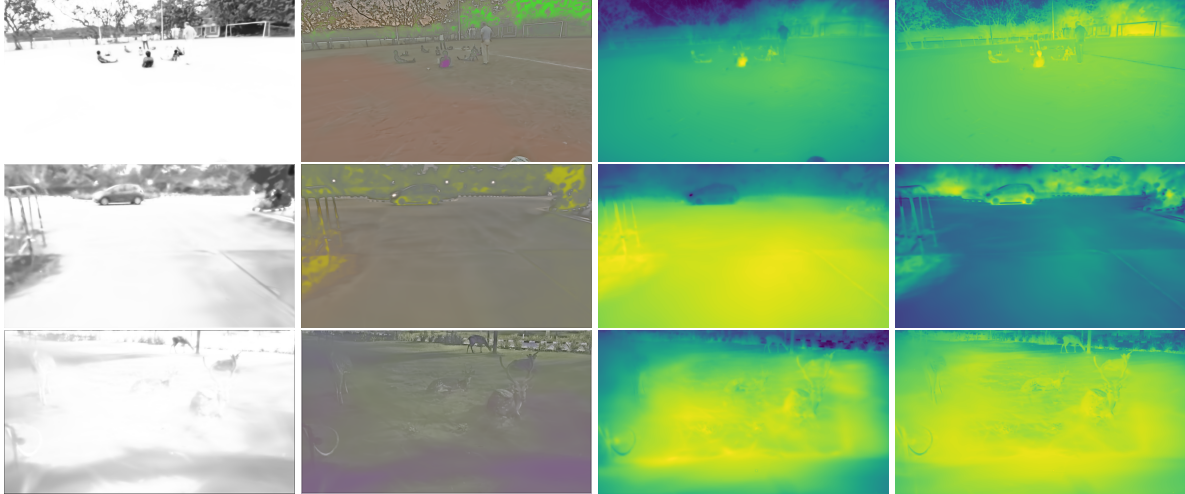


Figure S5. Visualization of intermediate outputs: We visualize $L_w(\underline{v}, t)$, R_w , B_1 , and B_2 maps (from left to right) for 3 scenes (in each row). B_1 and B_2 follows the constraints mentioned in Eq. 14, which allows B_1 and B_2 to follow gradients similar to R_w and $L_w(\underline{v}, t)$, respectively.

S8. Optimization Time

Table S3 compares the per-scene average optimization time (on NVIDIA RTX 3090 GPU) of different methods on our L^2DyV dataset.

Method	Training Time (hours)
LLNeRF [8]	7
AlethNeRF [1]	11.5
LuShNeRF [5]	9.5
GS-DK [10]	1.2
Luminance GS [2]	1.2
LITA-GS [11]	2.25
L^2DGS	1.5

Table S3. NeRF-based methods [1, 5, 8] rely on implicit scene representations and require volumetric rendering, which is computationally intensive. In contrast, Gaussian Splatting (GS) based methods (such as [2, 10, 11], and L^2DGS) adopt explicit scene representations, typically resulting in faster optimization times. It may be noted that our method, L^2DGS , alone can handle dynamic scenes.

S9. Clarifications on Human Subjects

Our proposed real video dataset L^2DyV contains dynamic objects such as a car, a bus, a deer, a cat, and a human. Our proposed method L^2DGS is agnostic to the nature of objects involved in the scene and their characteristics. While some videos contain humans, our method does not rely on features such as skin tone, body type, or motion. All the videos have been casually recorded. Hence, we do not dis-

cuss the diversity of these features in our work. L^2DGS and all competing methods are trained independently for each scene/video (synthetic as well as real); therefore, showcasing a fair comparison. Humans have been included primarily to show the effectiveness of the proposed method across complex dynamic scenes. Our work does not raise any ethical concerns.

S10. Additional Comparisons with 4DGS [9]

For completeness, we also report the metric values for 4DGS [9] in Table S4. Our implementation is built on top of 4DGS. Since the original 4DGS does not include a low-light-to-well-lit transformation module, it directly synthesizes low-light novel views when trained on low-light input videos. We have included the results of 4DGS [9] on the real data in the accompanying ‘.mp4’.

For 4DGS [9] trained on well-lit scenes, PSNR (w.r.t. well-lit GT) is 28.18/33.74/30.68 and SSIM is 0.84/0.93/0.89 for dynamic/static/overall regions, averaged over Apple, Creeper, Mochi, and Sriracha.

Scene	Dynamic			Static			Overall		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Apple	5.06	0.07	0.13	6.35	0.06	0.39	5.98	0.06	0.53
Creeper	6.98	0.04	0.36	5.22	0.05	0.19	6.29	0.04	0.55
Handwavy	5.86	0.04	0.09	5.75	0.06	0.44	5.77	0.06	0.53
Mochi	7.03	0.15	0.15	6.41	0.29	0.20	6.65	0.23	0.35
Sriracha	7.96	0.04	0.07	4.64	0.14	0.32	5.06	0.13	0.39

Table S4. Quantitative values for GS4D [9] (on well-lit reconstruction) across scenes for dynamic, static, and overall settings. Note that $L^2 DGS$ has PSNR values greater than 10dB for all cases (evident from Table 1 of main paper and Table S2).



Figure 6. This figure illustrates the inverse of the synthetic low-light generation process: scaling the low-light image (first column) by $\frac{L_{\text{well}}}{L_{\text{low}}}$ (second column) to qualitatively compare with the well-lit image (third column). Note that this experiment requires ground-truth well-lit images and is not intended for fair comparison.

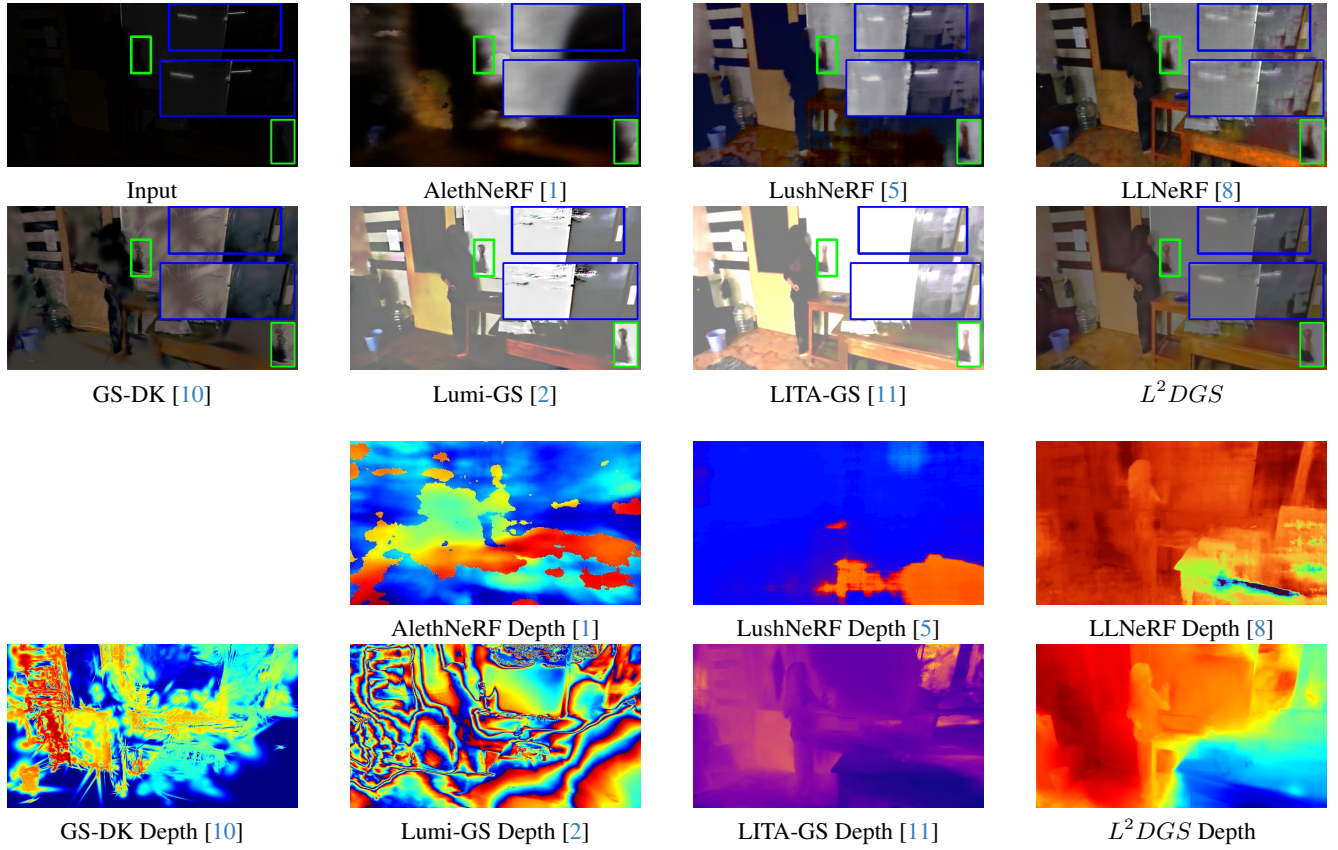


Figure S7. Comparisons on Whiteboard: Our method successfully reconstructs both the scene behind the transparent glass and the reflection of the distant light source (highlighted in blue). Additionally, it accurately captures the dynamic scene elements (highlighted in green). In contrast, competing methods fail to reconstruct details in both the foreground and background regions. Please zoom in for better visibility.

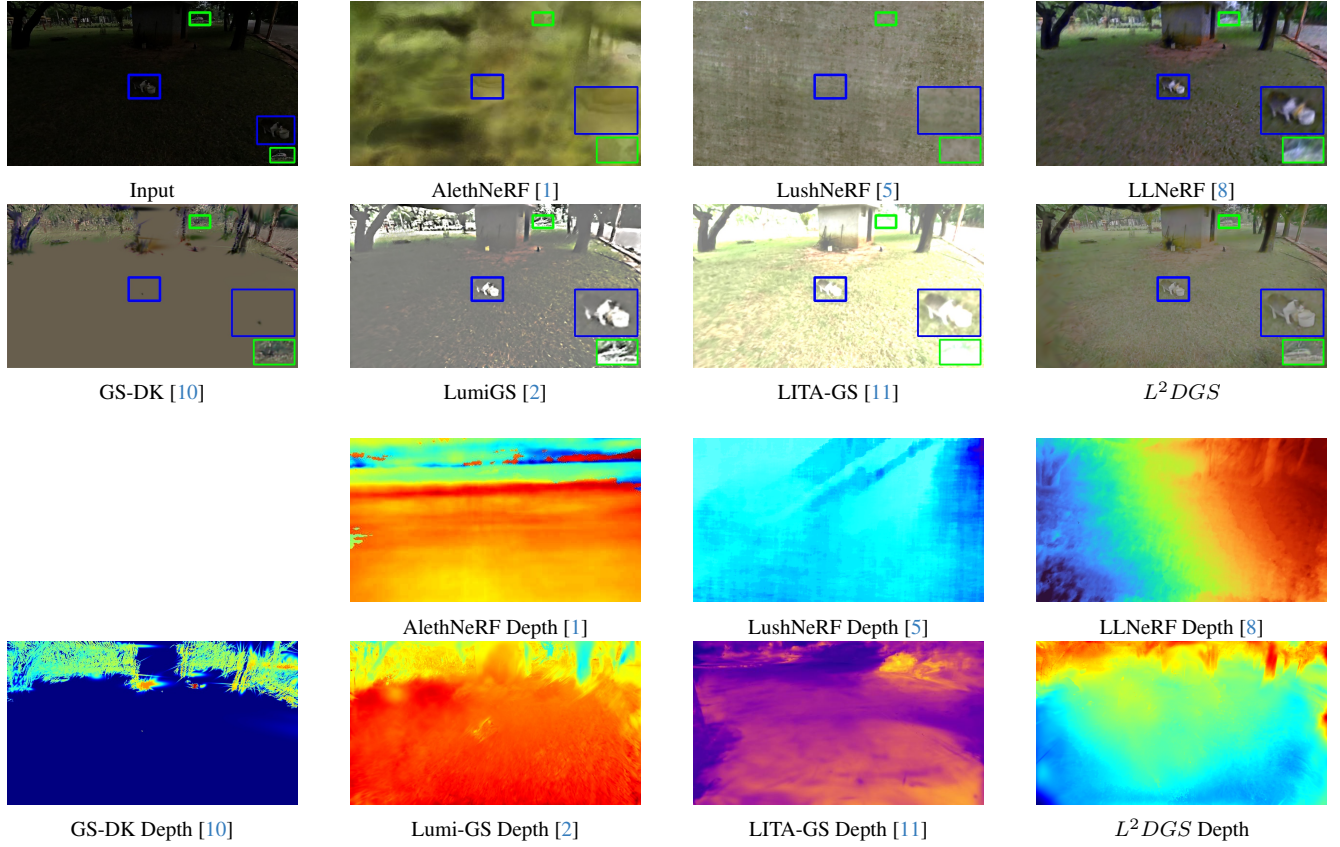


Figure S8. Comparisons on Cat-2: AlethNeRF, LushNeRF, and GS-DK fail to reconstruct the scene accurately. LLNeRF produces a blurry reconstruction of the cat (highlighted with a blue bounding box) and background, while Luminance-GS and LITA-GS saturate the scene. Both the methods also fail to capture the moving car in the background (highlighted in green). In contrast, L^2DGS successfully reconstructs the scene in a geometry-motion-and-intensity-aware manner.

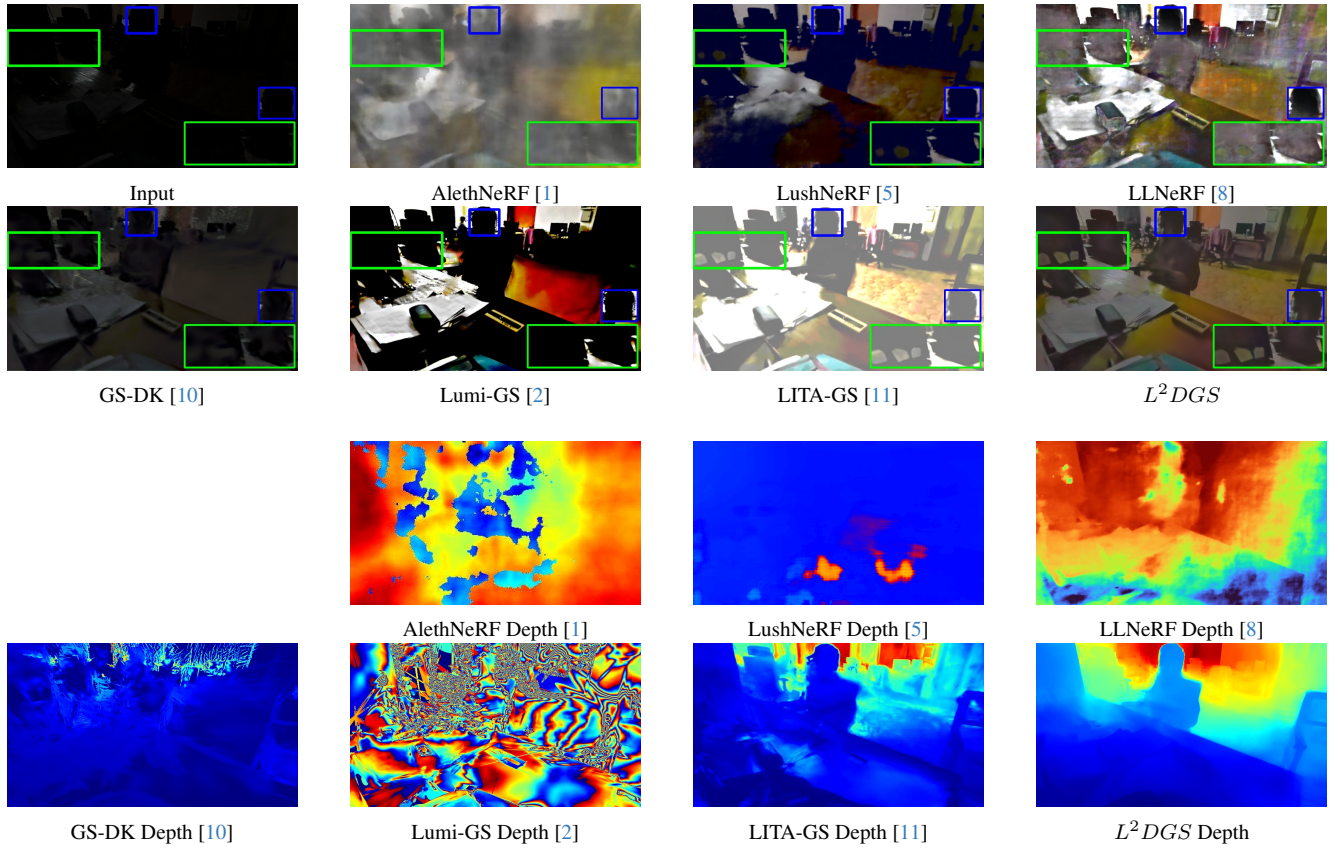


Figure S9. Comparisons on Activity: L^2DGS effectively reconstructs both the foreground and background, while competing methods either fail entirely or introduce noticeable artifacts. Our method is the only one that successfully enhances and reconstructs the background, specifically the room and the complex motion face (highlighted with a blue bounding box).

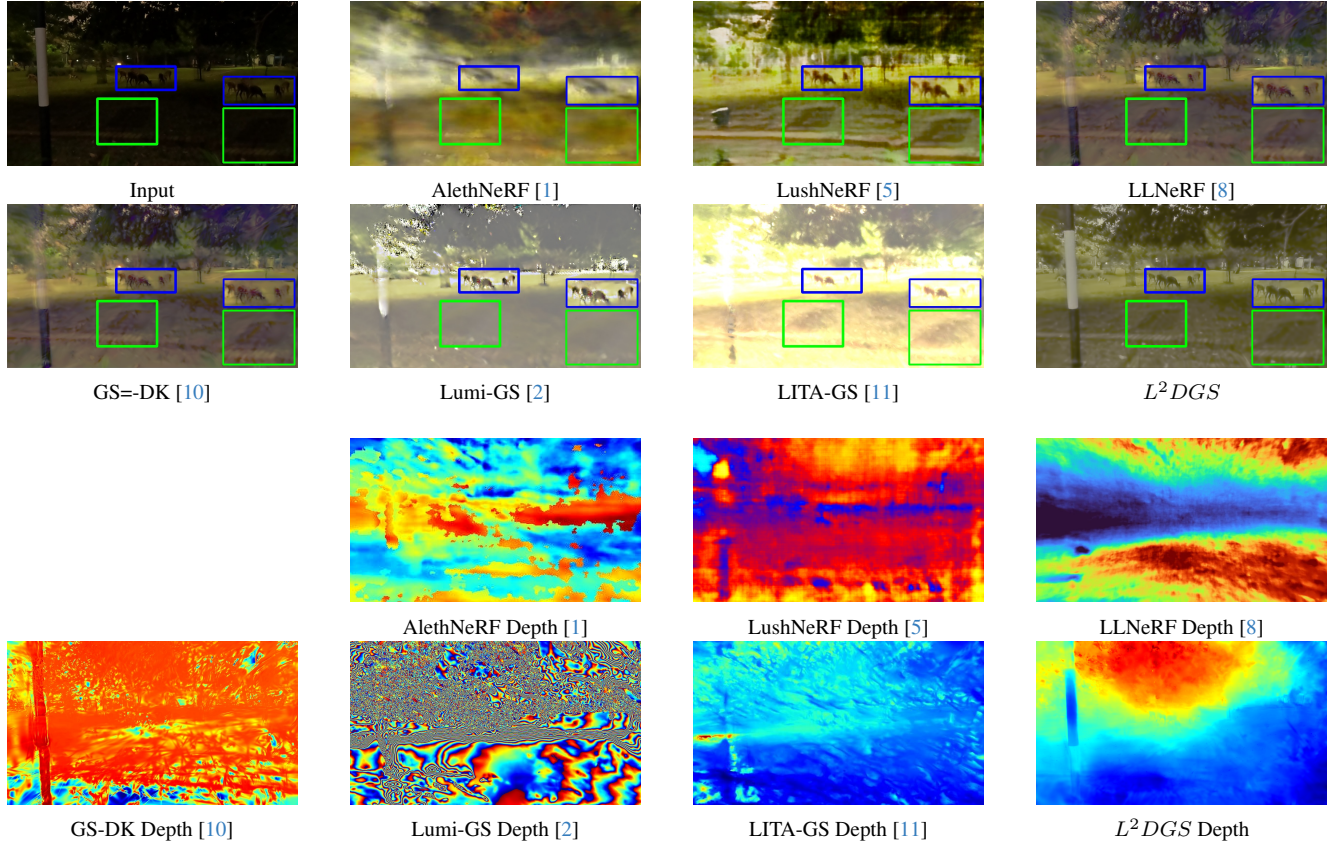


Figure S10. Comparisons on Deer-2: The scene contains multiple deer, and only L^2DGS is able to accurately reconstruct them (highlighted in blue). Additionally, L^2DGS preserves the shadow of a static object (highlighted in green) located in a low-light region, demonstrating the effectiveness of our scene-aware framework. In contrast, competing methods either fail to reconstruct the scene or merely enhance the brightness without preserving scene content.

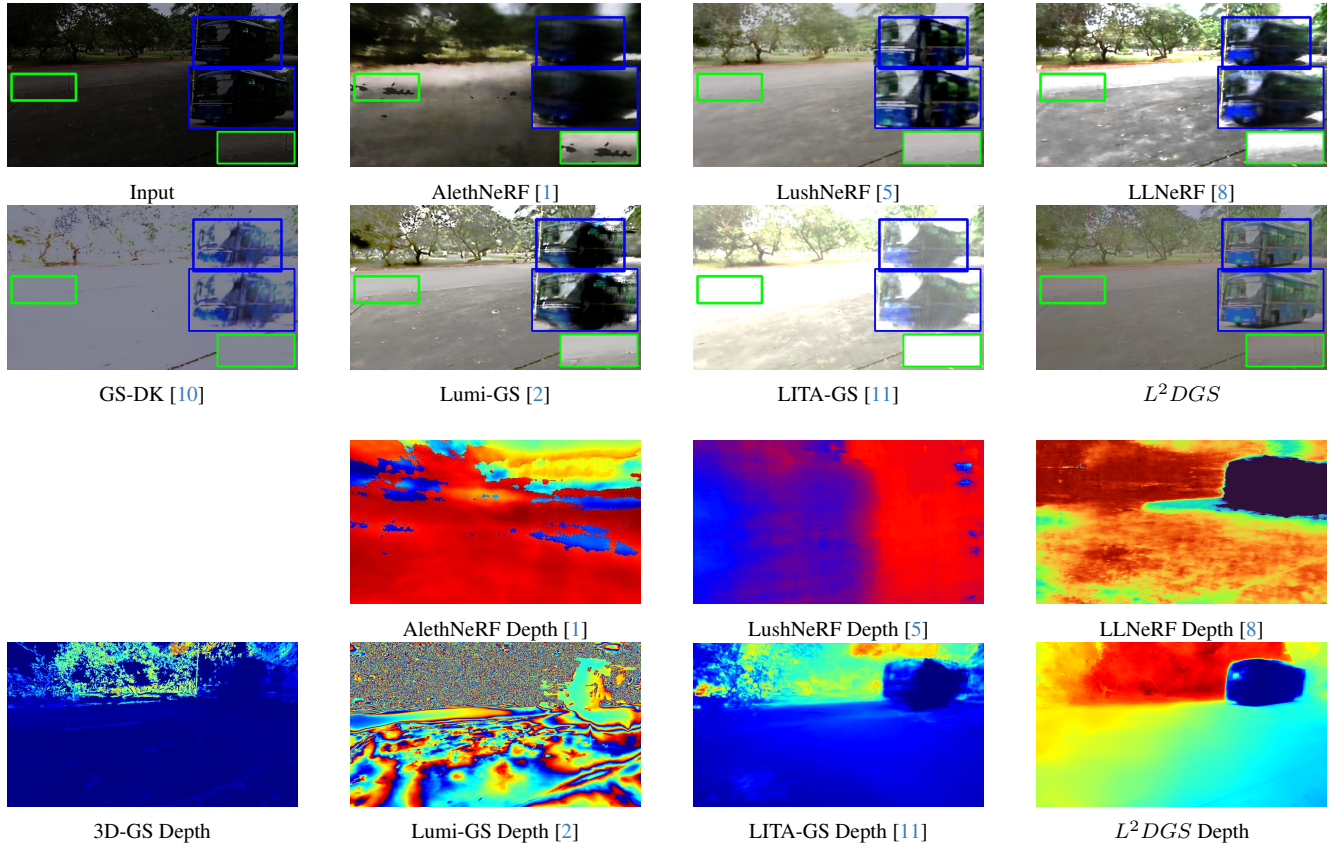


Figure S11. Comparisons on Bus: This scene features a bus approaching the camera, resulting in dynamic self-occlusion and disocclusion, as well as significant scale variations. AlethNeRF and GS-DK fail to reconstruct the scene. LLNeRF enhances the overall brightness, but saturates the regions near the bus and fails to reconstruct the bus itself (highlighted in blue). Luminance-GS achieves better brightness but still fails to reconstruct the bus (in blue) and does not capture the background accurately (in green). In contrast, our method, although slightly dimmer, successfully reconstructs the dynamic object (in blue) and preserves the scene geometry (in green), demonstrating its motion- and geometry-aware capability.

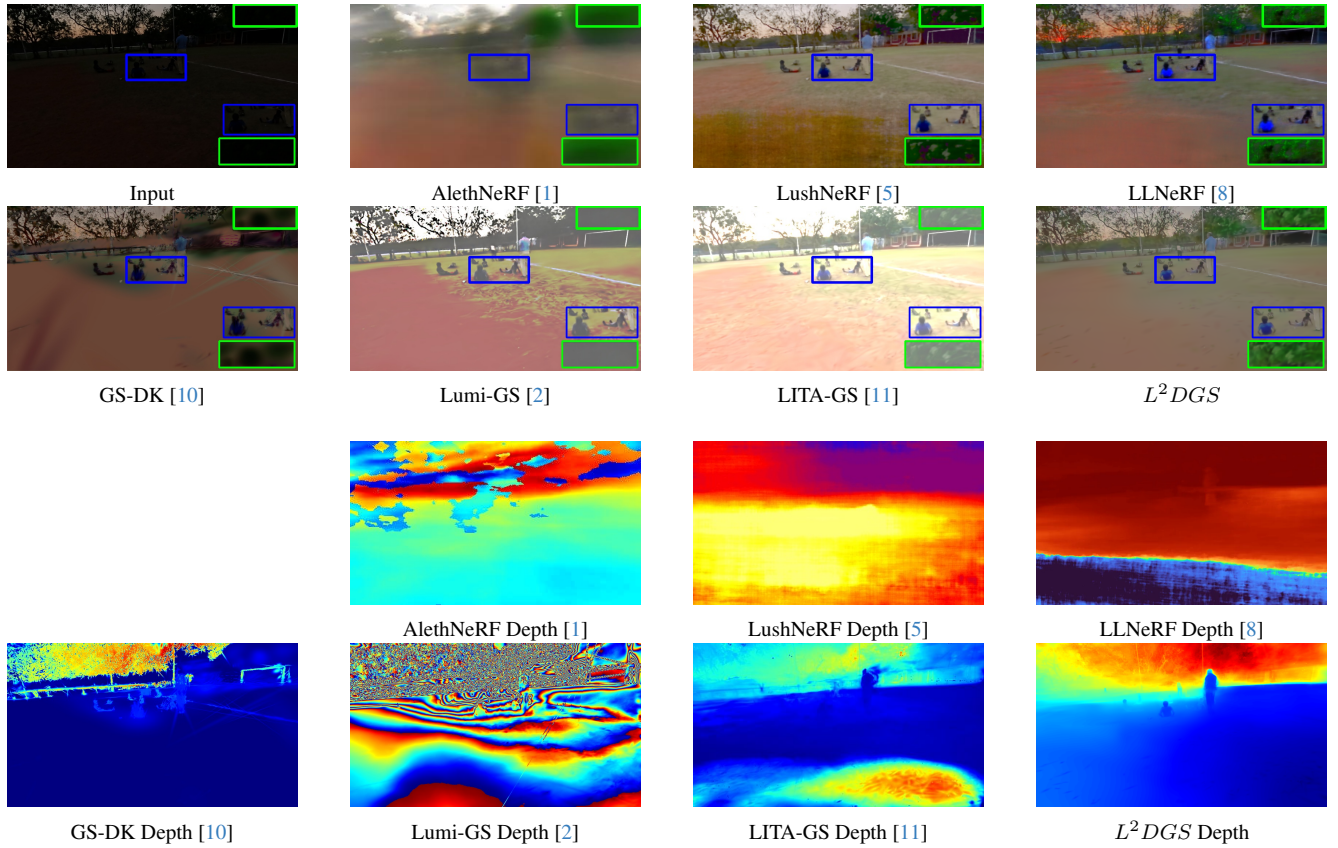


Figure S12. Comparisons on Kids: Our method successfully reconstructs the dynamic scene without introducing color artifacts (highlighted in blue) and avoids color distortions in the background (highlighted in green). Additionally, it prevents saturation in well-lit regions, preserving natural scene appearance across varying visibility levels.

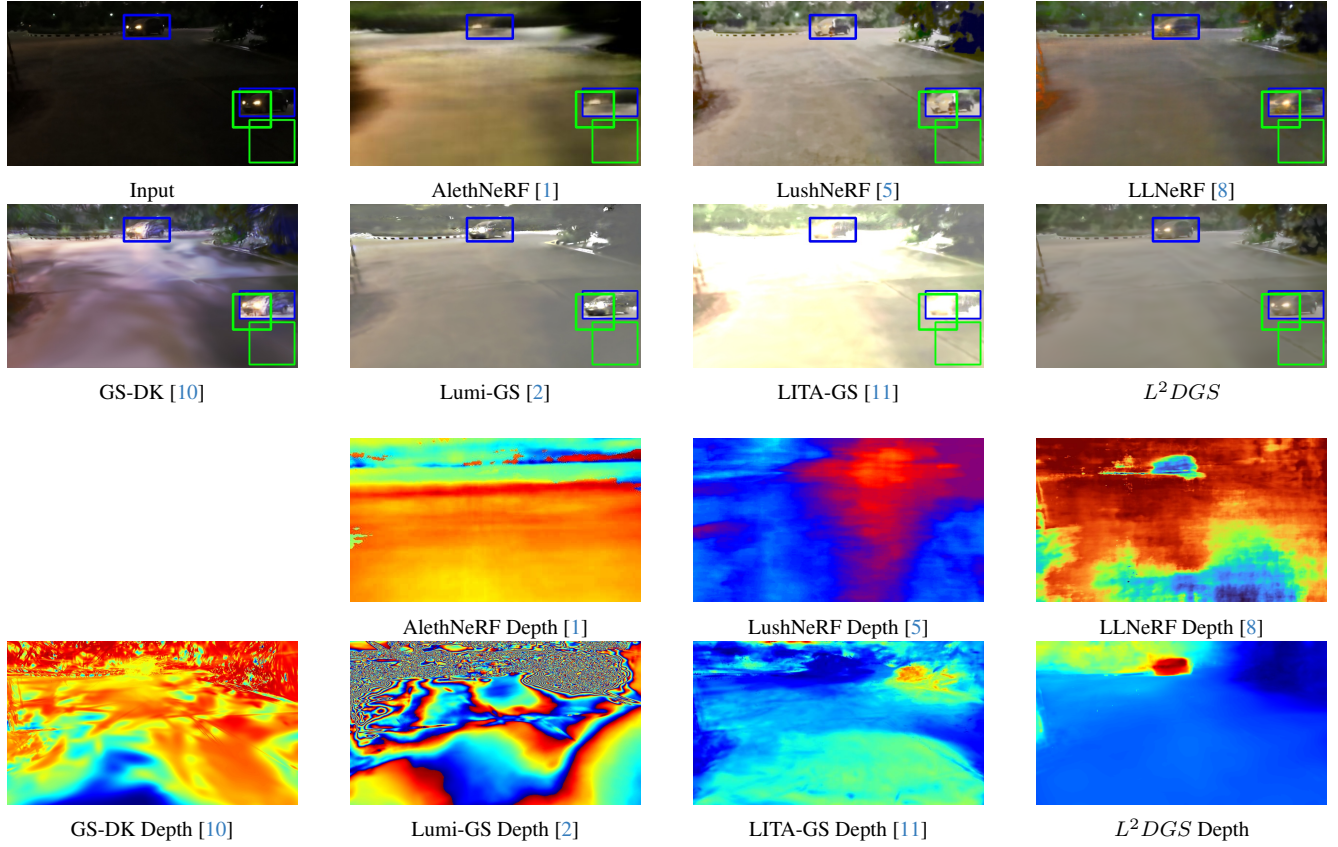


Figure S13. Comparisons on Car: The presence of a moving car with its headlights on introduces time-varying intensity on the ground, necessitating a model capable of learning time-dependent color variations. Our method not only reconstructs the dynamic scene accurately (highlighted in blue), but also captures the temporal changes in color and the occlusion–disocclusion effects caused by the car’s motion. Notably, the region beneath the car appears black due to occlusion, which our method preserves correctly.

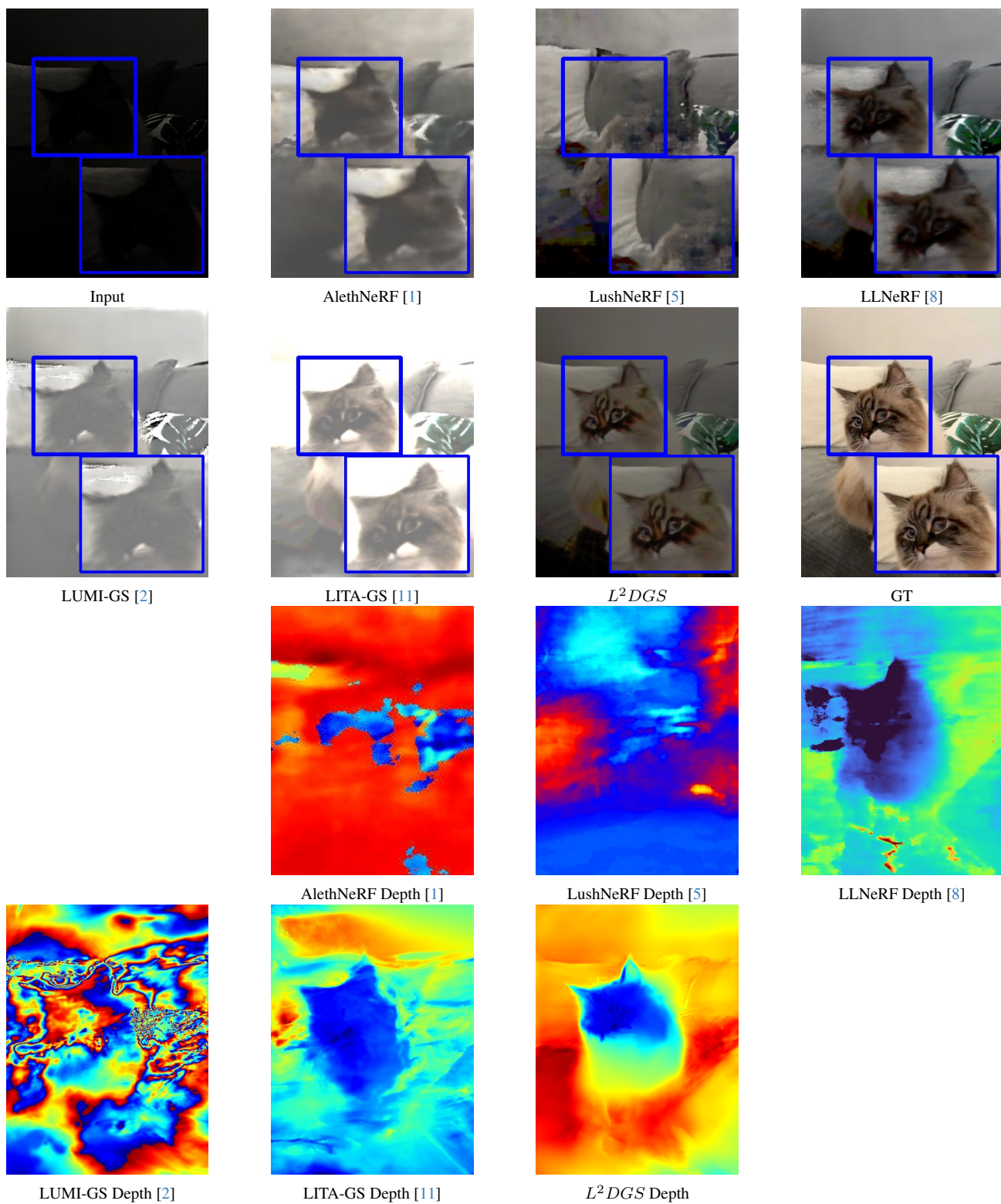


Figure S14. Comparisons on Mochi: Our method is able to recover the cat face and facial textures as well as the depth-variation on the dynamic object while all competing methods fail. LITA-GS [11] introduces saturation.

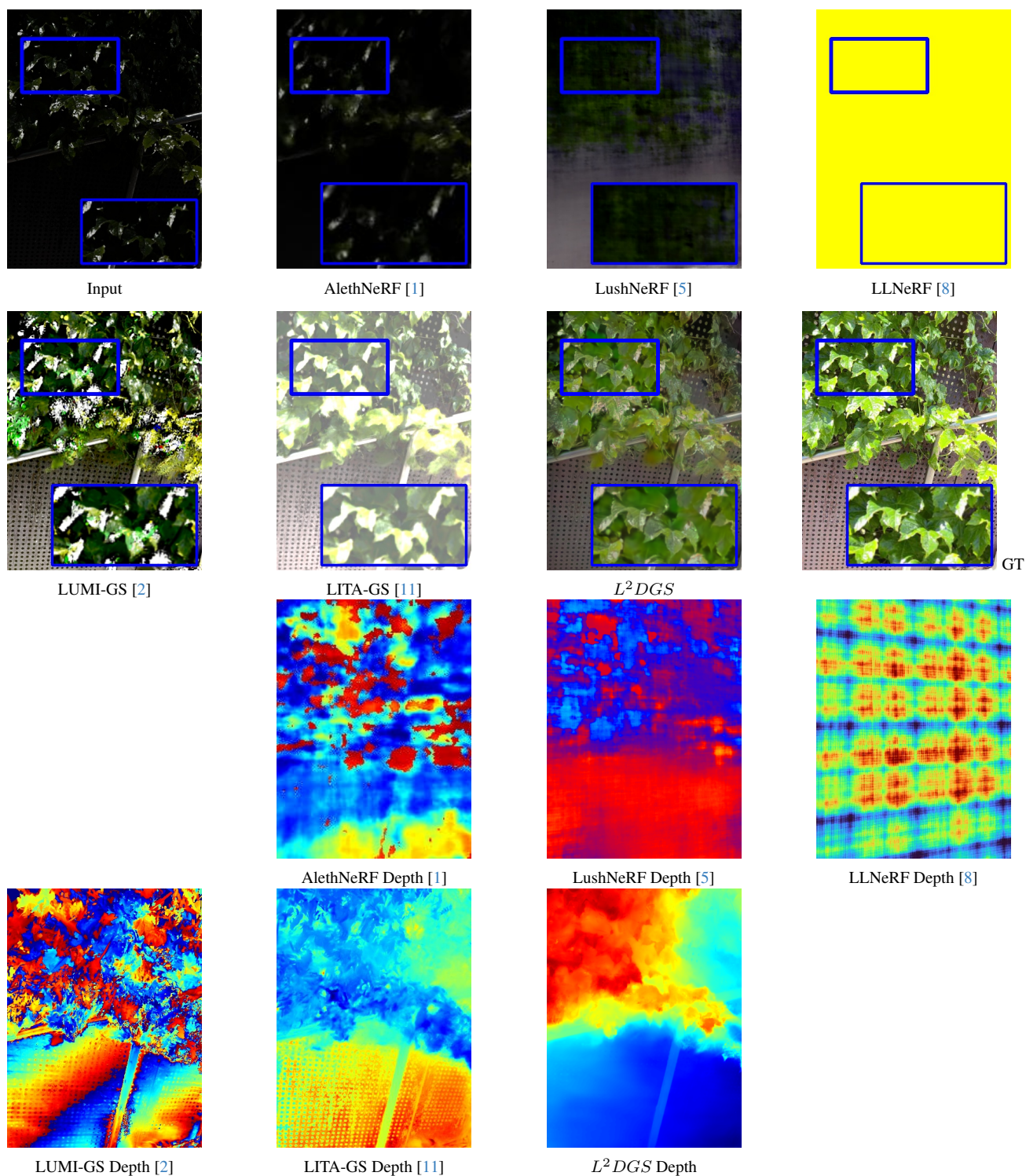
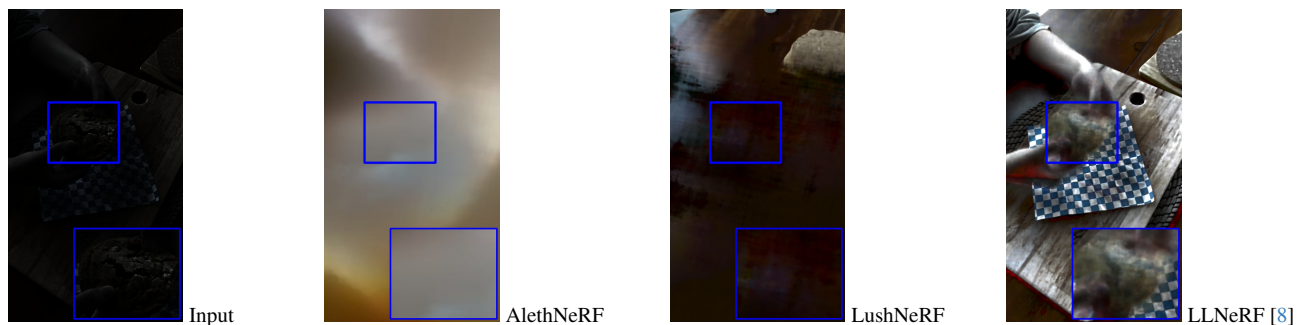


Figure S15. Comparisons on Creeper: LushNeRF and LLNeRF fail to reconstruct the scene. Luminance GS tends to amplify the low-light region occluded by leaves (highlighted in blue). In contrast, L^2DGS effectively handles the occluded leaf region and successfully reconstructs a well-lit scene.



[1]

[5]



[2]



[11]

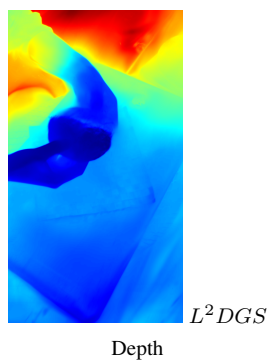
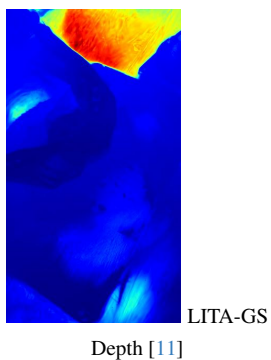
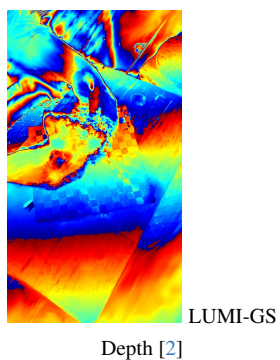
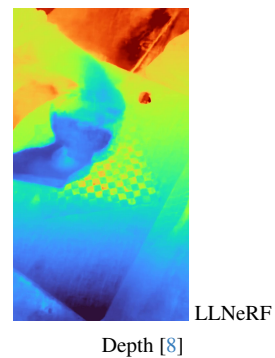
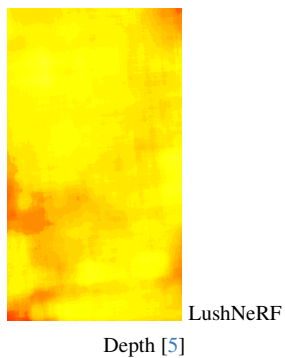
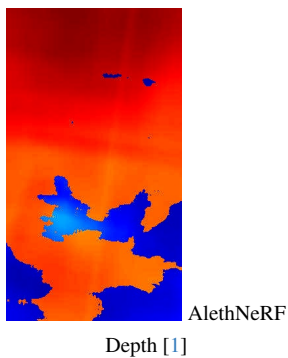
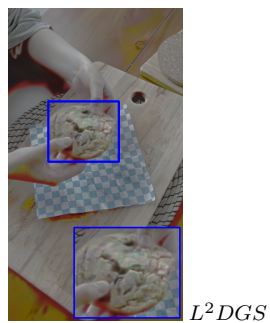


Figure S16. Comparisons on Cookie: $L^2 DGS$ is able to reconstruct the scene (see depth map) while competing methods, though, enhance the scene but fail to learn the underlying geometry. LITA-GS [11] introduces saturation.

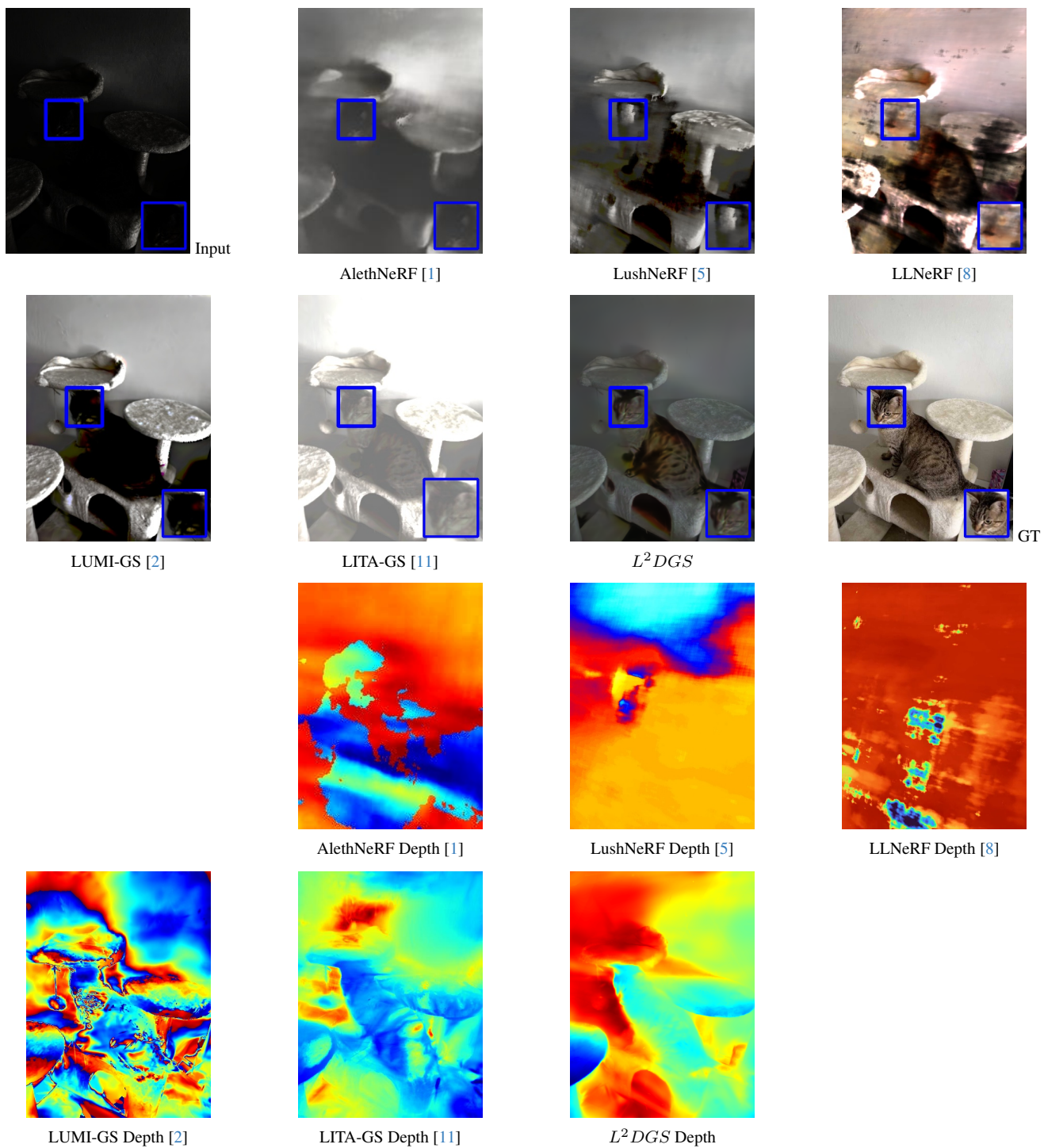


Figure S17. Comparisons on Sriracha: L^2DGS is able to learn the scene geometry and also able to recover the details of the cat.

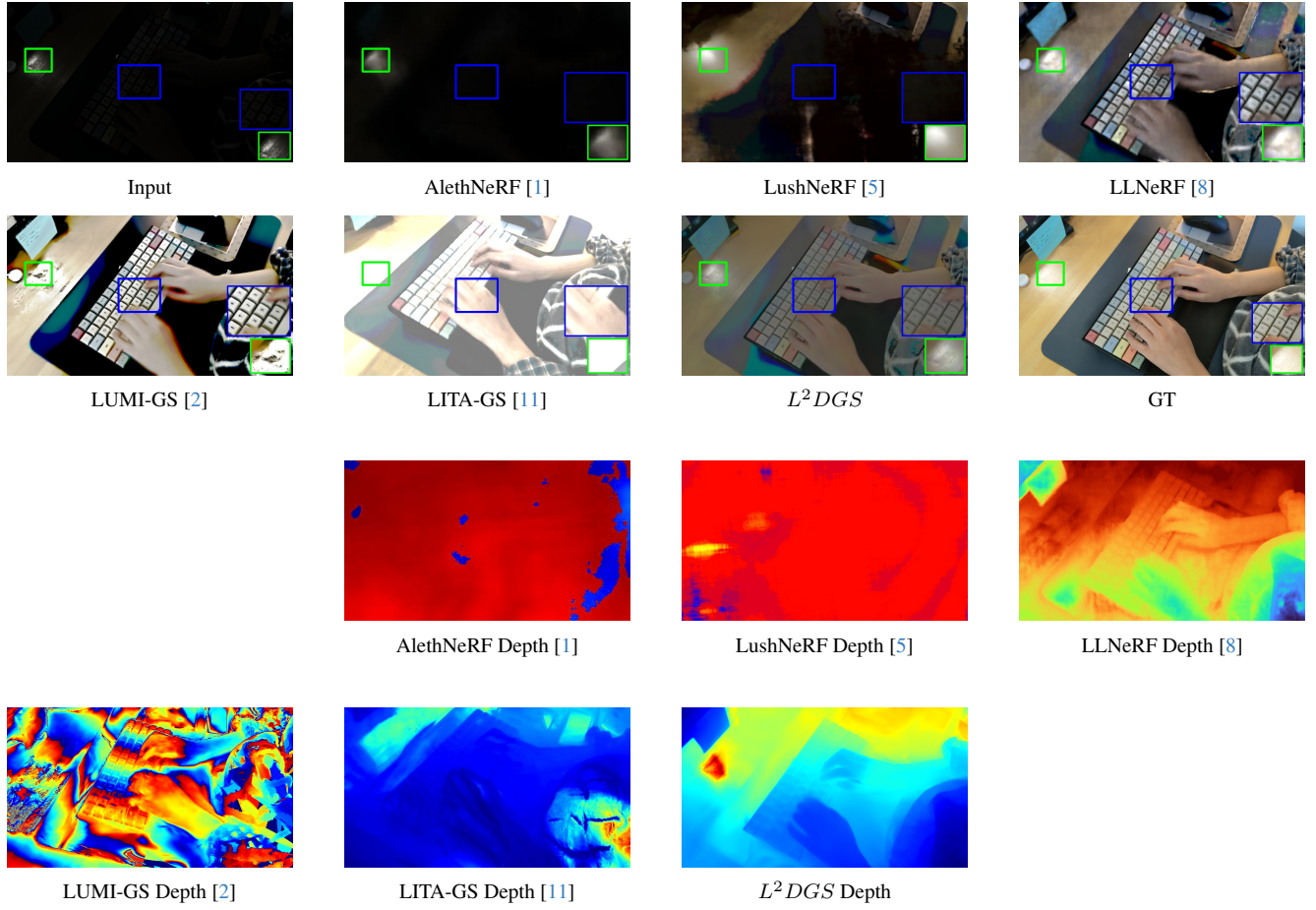


Figure S18. Comparisons on Keyboard: AlethNeRF and LushNeRF fail to reconstruct the scene accurately. While Luminance GS enhances overall brightness, it saturates the region highlighted in green and struggles to learn the underlying scene geometry. LLNeRF has difficulty recovering fine details, such as the letters printed on the keyboard (highlighted in blue). In contrast, our method, L^2DGS , preserves both scene geometry and appearance; note the clear depth recovery of the fingers.

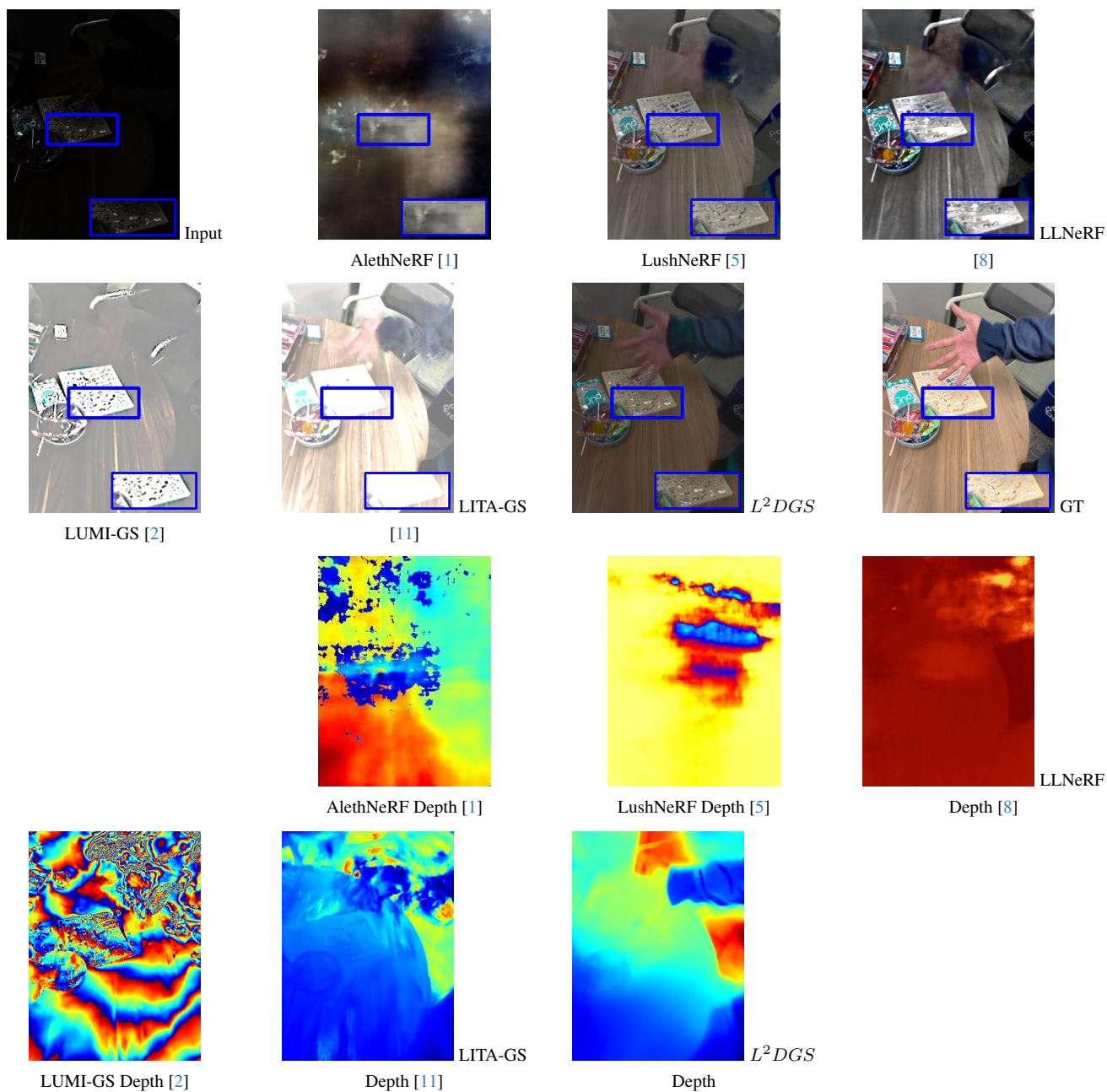


Figure S19. Comparisons on Handway: L^2DGS is able to reconstruct both static and dynamic parts of the scene. It also avoids saturation in the region highlighted in blue.

References

- [1] 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*, pages 1435–1444, 2024. [6](#), [7](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#)
- [2] 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. [6](#), [7](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#)
- [3] Hang Gao, Ruilong Li, Shubham Tulsiani, Bryan Russell, and Angjoo Kanazawa. Monocular dynamic view synthesis: A reality check. *Advances in Neural Information Processing Systems*, 35:33768–33780, 2022. [1](#), [2](#)
- [4] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228*, 2021. [1](#), [2](#)
- [5] Zefan Qu, Ke Xu, Gerhard Petrus Hancke, and Rynson WH Lau. Lush-nerf: Lighting up and sharpening nerfs for low-light scenes. *arXiv preprint arXiv:2411.06757*, 2024. [5](#), [6](#), [7](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#)
- [6] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [4](#)
- [7] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016. [4](#)
- [8] 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 (ICCV)*, pages 12632–12641, 2023. [6](#), [7](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#)
- [9] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20310–20320, 2024. [7](#), [8](#)
- [10] 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. In *Computer Graphics Forum*, page e15213. Wiley Online Library, 2024. [5](#), [7](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#)
- [11] 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 Computer Vision and Pattern Recognition Conference (CVPR)*, pages 21580–21589, 2025. [6](#), [7](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#)
- [12] Shangchen Zhou, Chongyi Li, and Chen Change Loy. Led-net: Joint low-light enhancement and deblurring in the dark. In *European conference on computer vision*, pages 573–589. Springer, 2022. [1](#), [5](#)