SpikeGS: Learning 3D Gaussian Fields from Continuous Spike Stream

Supplementary Material

1 Different lighting intensity experiments

We conducted experiments on three synthetic datasets (ficus, lego and materials) with varying lighting intensities to further demonstrate the generalization of our method under different illumination conditions (We set our model N to 256 for the experiment). For each scene, we performed experiments under three lighting intensity conditions: extremely low lighting, moderately low lighting, and original lighting intensity. The qualitative results are shown in Figure. 1 and Figure. 2. As can be observed, our method outperforms existing methods under all lighting conditions. Even in extremely low-light scenarios, our method is able to reconstruct complete scene structures and render fine texture details. In contrast, other methods either fail to reconstruct the complete scene structure or produce very blurred texture details, and even generate significant noise at relatively higher lighting intensities.

The quantitative comparison results are shown in Table 1. For each lighting condition in each scene, we calculated three metrics: PSNR, SSIM, and LPIPS (using the ground truth RGB images as reference, consistent with the full paper). As shown, our method outperforms existing methods on nearly all metrics.

2 Complete visualization results of the ablation experiments

According to the experimental setup described in the paper, we have provided the complete visual results for both the synthetic and real datasets in the supplementary materials. We conducted ablation experiments on 6 scenes from the synthetic dataset and 4 scenes from the real dataset. As shown in Fig. 3 and Fig. 4, when only using L_1 as the loss function, the texture details of the scenes appear overly smooth. Additionally, when the noise embedding pipeline is removed or when using estimated light intensity directly (bypassing the spike generation pipeline) as input, the synthesized views exhibit a significant amount of noise. In contrast, the complete framework demonstrates high robustness to noise and is capable of recovering fine texture details(The quantitative comparison of ablation experiments is presented in Table 3 of the full paper. here, we only showcase the complete visual results(qualitative comparison)).

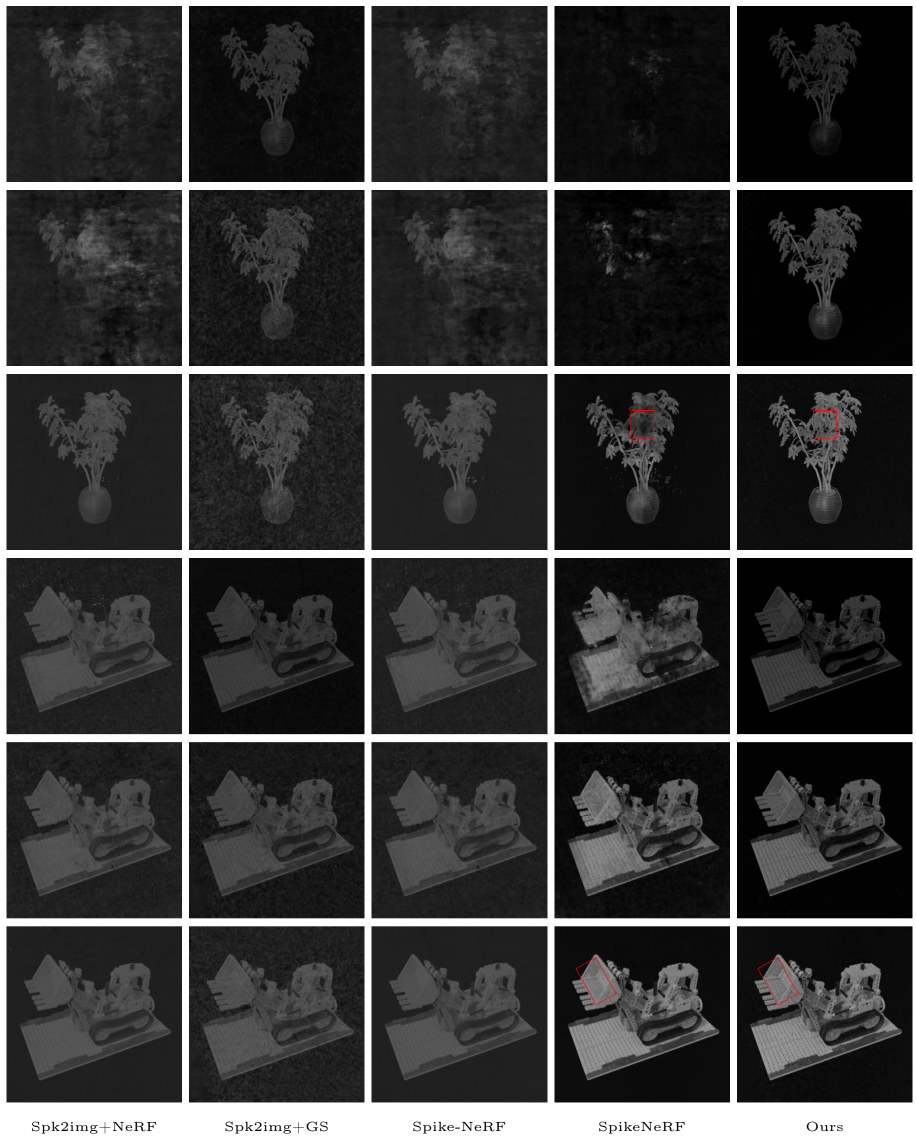

Fig. 1: Qualitative results on different light intensities. In the figure, every three rows represent one scene (The names of the two scenes are "ficus" and "lego"), with the first, second, and third rows corresponding to extreme low light intensity, medium low light intensity, and original light intensity, respectively. It is evident from the figure that our model consistently reconstructs the complete scene structure and fine details under all lighting conditions. In contrast, other methods often fail to reconstruct accurately and struggle to recover fine scene details under low lighting conditions, and they also produce significant noise at relatively higher lighting intensities.

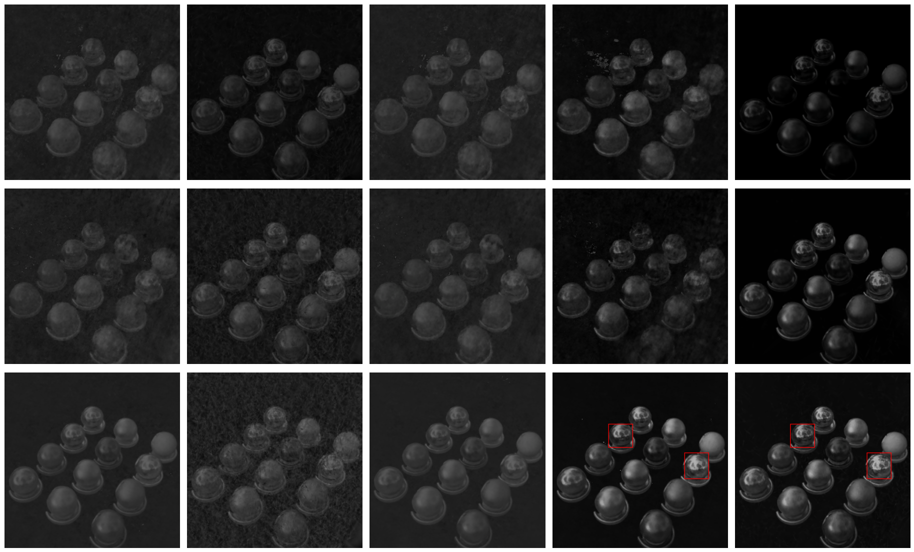

Fig. 2: Qualitative results on different light intensities. This figure is a continuation of Fig. 1, with the tested scene named "materials." The light intensity settings are consistent with Fig. 1. It can be observed that our model consistently recovers fine reflective details under different lighting intensities. In contrast, other methods struggle to recover fine details under lower lighting intensities and produce significant noise under relatively higher lighting intensities.

Table 1: Different Lighting Intensity Experiments. The terms "Light intensity (Low)," "Light intensity (Med)," and "Light intensity (Orig)" in the table correspond to extremely low lighting, moderately low lighting, and original lighting intensity, respectively. We calculated the average metrics for the three scenes (ficus, lego, and materials) under each lighting condition, and the results are shown below. Each color shading indicates the best and second-best result.

Method	Light intensity(Low)			Light intensity(Med)			Light intensity(Orig)			Time↓
	PSNR↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	
Spk2img+NeRF(200K)	15.73	.0786	.4162	15.83	.0788	.4582	16.92	.1020	.2445	>3 hours
Spk2img+GS(30K)	17.72	.1342	.2921	16.93	.1038	.4920	15.97	.0989	.5308	∼5 mins
Spike-NeRF [1](200K)	15.76	.0791	.4011	15.83	.0792	.4433	16.28	.1019	.2422	>3 hours
SpikeNeRF [2](200K)	17.33	.1126	.3578	18.11	.1352	.3782	20.7	.1882	.1677	>10 hours
Ours(30K)	18.25	.8021	.1977	19.89	.7035	.1574	21.0	.2010	.1875	∼40 mins

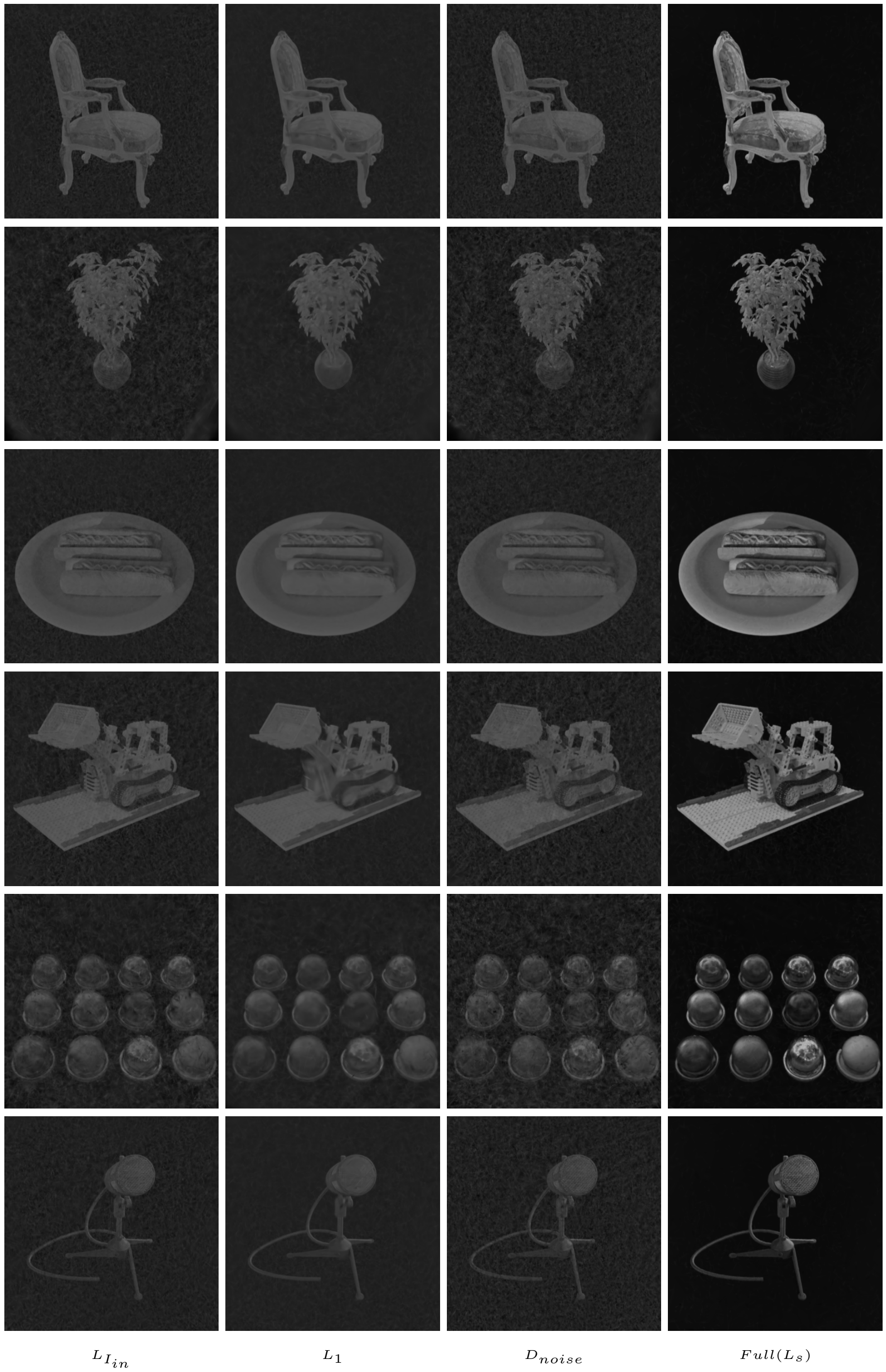

Fig. 3: Qualitative comparison of ablation experiments on the synthetic dataset. As shown in the figure, images rendered with $L_{I_{in}}$ and D_{noise} contain noticeable noise, while images rendered with L_1 are overly smooth and lack detail.

Fig. 4: Qualitative comparison of ablation experiments on the real dataset. As shown in the figure, images rendered with $L_{I_{in}}$ and D_{noise} contain noticeable noise, while images rendered with L_1 are overly smooth and lack detail.

References

1. Guo, Y., Bai, Y., Hu, L., Liu, M., Guo, Z., Ma, L., Huang, T.: Spike-nerf: Neural radiance field based on spike camera. arXiv preprint arXiv:2403.16410 (2024)
2. Zhu, L., Jia, K., Zhao, Y., Qi, Y., Wang, L., Huang, H.: Spikenerf: Learning neural radiance fields from continuous spike stream. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6285–6295 (2024)