SpikeNeRF: Learning Neural Radiance Fields from Continuous Spike Stream

Supplementary Material

1. Synthetic Data Details

We generate synthetic spike data in NeRF using six scenes (chair, ficus, hotdog, lego, materials, and mic). To synthesize the spike stream, we initially resize the original images from Blender to 400 × 400. Employing the spike generator provided by [1], we simulate spike streams for each viewpoint.

For varying illumination conditions, we manipulate the intensity parameter in the simulator within the range of 16 to 64 (refer to Fig. 6 of the main manuscript and Table S1). Beyond generating spike streams from static viewpoint images, we also render 16 high-resolution images captured by the camera. These images are then input into the spike generator to replicate the dynamic recording process of the spike camera. The resulting spike streams for each view are of size 400 × 400 × 256.

Each scene encompasses 100 sets of images and their corresponding event data. The input images from Blender and the generated spike streams are visually depicted in Fig. S1.

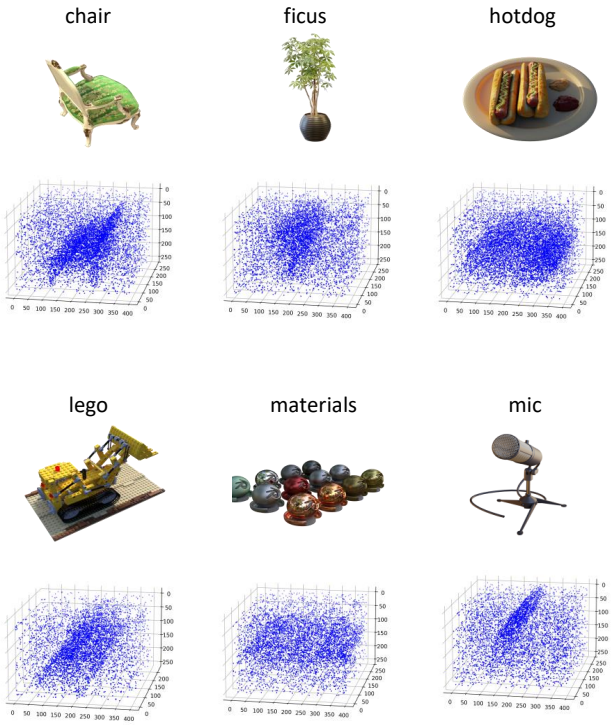

Figure S1. The input images and the visualization of the generated spike stream in synthetic scenes.

2. Real-world Data Details

The spike camera is capable of capturing spike streams with a spatial resolution of 250 × 400 and a temporal resolution of 20,000 Hz.

For each viewpoint, we simultaneously capture ideal and non-ideal conditions of spike data. Initially, we minimize noise by providing the spike camera with ideal light intensity and motion. Utilizing the Spk2img method, we reconstruct high-quality images and employ COLMAP for pose estimation. This process enables us to obtain the pose and camera parameters for spike data under typical conditions. Conducting handheld captures, we gather data from five real-world scenarios, each showcasing texture details under distinct illumination conditions. Each dataset consists of approximately 35 images from diverse viewpoints, accompanied by their corresponding spike data. The recorded spike streams are visually depicted in Fig. S2. The spike numbers of both synthetic spike data and real-world spike data are shown in Table S1.

3. More Details of Threshold Variation Simulation

The threshold variation of each spiking neuron can be modeled by Eq. S1, and the spike stream can be generated by

$$\hat{S}(x,y) = \mathrm{SN}(I(T_i(r))) \cdot R(x,y), \tag{S1}$$

where $\mathrm{SN}(\cdot)$ denotes the spiking neuron, $R(x,y)$ is the nonuniformity matrix and can be obtained by capturing a uniform light scene and recording the intensity.

Choosing the pixel $(x_m,\ y_m)$ which is closest to the average response value as the reference pixel, $R(x,y)$ is then obtained by calculating the ratio of the reference pixel's response value to the response values of other pixels: $R(x,y) = \frac{(L_2+L_d(x_m,y_m))T_2(x_m,y_m)}{(L_2+L_d(x,y))T_2}$, where L_2 and T_2 are variables to be calibrated. As referred to [1], fixed pattern noise includes dark current noise and response nonuniformity noise, the equivalent light intensity value for the dark signal, L_d can be calculated by capturing two uniformly illuminated scenes:

$$\begin{cases} C\Delta V = \alpha L_d T_d \\ C\Delta V = \alpha(L_1 + L_d)T_1 \end{cases}, \tag{S2}$$

where the first line of the equation represents capturing the scene brightness at zero (obtained by covering the lens in a dark room), while the second line represents capturing the scene brightness at L_1 (recorded using a photometer for L_1 value). T_d and T_1 respectively represent the spike emission

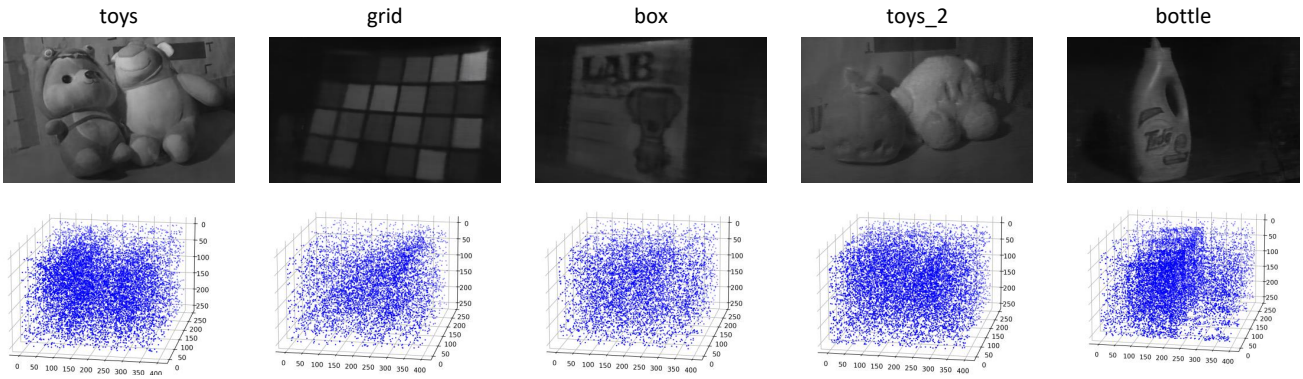

Figure S2. The scenes and the visualization of the recorded spike stream in real-world settings.

Table S1. The mean spike count for each view across synthetic and real-world datasets.

Synthetic spike data						
# SpikeNum/Scene	Lego	Chair	Hotdog	Ficus	Materials	Mic
Train	268,185	262,640	282,359	190,502	196,675	169,442
Test	264,969	264,822	255,235	197,473	210,756	167,507
Val	270,027	257,689	274,341	192,482	198,026	169,398
Real-world spike data						
# SpikeNum/Scene	Toy	Grid	Box	Toy_2	Bottle	-
-	585,016	422,037	441,255	725,272	691,757	-

intervals of the corresponding spike streams.

By solving Eq. S2, the additional brightness of the scene due to the equivalent dark current can be calculated:

$$L_d = \frac{L_1 T_1}{T_d - T_1}. \tag{S3}$$

Another spike streams with brightness L_2 needs to be captured to nullify the unknown photoelectric conversion constant α:

$$C\Delta V = \alpha(L_2 + L_d)T_2. \tag{S4}$$

Due to mismatches in capacitance and voltage between pixel circuits, different pixels exhibit varying responses to scene brightness, leading to fixed pattern noise. The corresponding error matrix can be defined as follows:

$$R(x, y) = \frac{(C + \delta C(x, y))\,(\Delta V + \delta V(x, y))}{C\Delta V}. \tag{S5}$$

Substituting Eq. S3 and Eq. S4 into the above equation yields a specific value for $R(x, y)$:

$$R(x, y) = \frac{(L_2 + L_d(x_m, y_m))T_2(x_m, y_m)}{(L_2 + L_d(x, y))T_2}. \tag{S6}$$

According to Eq. S6, the threshold variation of spiking neurons can be simulated based on the real-world spike distribution.

4. Additional Quantitative Results

The detailed quantitative results on six synthetic scenarios are shown in Table S2 and Table S3. Table S2 illustrates the superior performance of SpikeNeRF across all synthetic scenarios. This observation suggests that our method excels in learning a more precise 3D representation of the scene within the proposed framework.

In Table S3, we present the outcomes for the synthetic spike dataset under various light intensities. These diverse light conditions are achieved by adjusting the intensity parameter in the spike simulator, with settings for low (16), medium (32), and strong (64) illumination. The spike numbers, representing the count of generated spike data for a view, are detailed in the table. A higher spike number indicates a stronger light intensity. The results consistently reveal the superior performance of our final model compared to other configurations.

5. Additional Qualitative Results

The qualitative results on synthetic scenarios and real-world spike data are shown in Fig. S3, Fig. S4, and Fig. S5.

In Fig. S3 and Fig. S4, our SpikeNeRF effectively leverages the inherent relationship between spike streams and scenes to learn a sharp NeRF. Consequently, our results remain resilient to the noise inherent in spike data, yielding

Table S2. Detailed quantitative result on six synthetic scenes. We use **bold** to mark the best results. The results on the left pertain to measurements across the entire image, while those on the right specifically focus on measurements within the object region.

Novel View	Lego			Chair			Hotdog		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
TFP(32)+NeRF	15.27/15.57	0.154/0.785	0.588/0.065	14.94/15.21	0.126/0.844	0.662/0.057	16.18/16.56	0.191/0.853	0.641/0.059
TFP(256)+NeRF	14.44/16.41	0.127/0.794	0.652/0.088	13.92/15.47	0.089/0.836	0.735/0.079	15.15/17.34	0.175/0.870	0.685/0.059
TFI+NeRF	13.77/16.63	0.120/0.793	0.670/0.093	13.20/15.83	0.085/0.835	0.753/0.090	14.48/17.56	0.172/0.871	0.700/0.063
Spk2img+NeRF	13.42/14.05	0.066/0.726	0.724/0.129	12.76/13.23	0.055/0.794	0.778/0.102	14.20/14.92	0.123/0.810	0.728/0.087
Ours	**18.72/19.38**	**0.251/0.880**	**0.517/0.050**	**19.11/19.76**	**0.201/0.917**	**0.591/0.045**	**19.76/20.25**	**0.274/0.930**	**0.581/0.030**
Novel View	Ficus			Materials			Mic		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
TFP(32)+NeRF	18.44/19.17	0.096/0.876	0.758/0.050	/	/	/	/	/	/
TFP(256)+NeRF	16.55/20.45	0.067/0.894	0.814/0.049	11.52/20.63	0.026/0.855	0.823/0.130	17.85/25.14	0.080/0.937	0.817/0.098
TFI+NeRF	15.72/20.29	0.061/0.891	0.829/0.052	15.92/20.70	0.138/0.849	0.726/0.139	16.48/25.02	0.070/0.934	0.837/0.103
Spk2img+NeRF	16.33/17.60	0.037/0.848	0.872/0.070	16.62/17.78	0.092/0.795	0.782/0.166	18.94/21.86	0.049/0.901	0.889/0.121
Ours	**20.89/22.18**	**0.132/0.910**	**0.725/0.044**	**21.97/23.14**	**0.244/0.903**	**0.595/0.077**	**24.21/27.43**	**0.149/0.953**	**0.693/0.053**

impressive results in novel view synthesis.

Fig. S5 showcases the results obtained from five real-world spike sequences. When confronted with spike data noise, other methods exhibit limitations, introducing more noise into their output. In contrast, our SpikeNeRF excels in predicting accurate details and light intensities compared to alternative methods.

6. Additional Ablation Results

Comparison to finetuned/retrained Spk2imgNet. We conduct separate retraining and fine-tuning of Spk2imgNet on the dataset provided by RSIR, and the results are presented in Table S4. We show the results focused solely on measurements within the object region. Since the Spk2imgNet network is primarily designed for high-intensity lighting conditions, and the lighting situations in the simulated and real-world scenarios addressed in this paper are more complex, training Spk2imgNet with a noisy dataset makes it more challenging for the network to learn the correspondence between spikes and light intensity, resulting in no improvement in performance. RSIR processes spike data using a cyclic iterative approach, where, in the original study, spikes of length 32 or 64 are inputted in each iteration, and the loop performs optimally after 4-8 iterations. Therefore, we design three configurations for comparison, as shown in the table, where 'w' represents the length of each input spike, and 'c' represents the number of iterations. According to SSIM and LPIPS metrics, it can be observed that among these configurations, RSIR(w=32, c=8) performs best.

Ablation of the nonuniformity matrix. We conduct an ablation experiment of nonuniformity matrix R in Table S5. In Table 2 of our paper, we also compare our proposed loss $\mathcal{L}_s$ with $\mathcal{L}_i$, $\mathcal{L}_s$+$\mathcal{L}_i$, and $\mathcal{L}_i^*$.

Different sequence lengths. In Table S6, we conduct experiments using sequences of different lengths. We observed that PSNR might be more significantly influenced by the contrast, with the highest results achieved using w=64 under light intensity 16. SSIM and LPIPS metrics focus more on measuring structural and perceptual information, respectively. As the length of the spike sequence increases, the SSIM and LPIPS metrics improve under different light intensities. SpikeNeRF can converge under low illumination and with a short sequence length.

7. Supplementary Video

We provide a supplementary video to show the video results. For synthetic scenes, we show the results of TFI+NeRF, TFP+NeRF, Spk2img+NeRF and our SpikeNeRF. For the real scenes, we show the comparison of the results of all mentioned methods in the toy and toy_2 scenes. It is obvious that the results of our SpikeNeRF have less noise, better contrast, and sharper object texture details compared to other methods on both synthetic scenes and real scenes.

References

[1] Lin Zhu, Yunlong Zheng, Mengyue Geng, Lizhi Wang, and Hua Huang. Recurrent spike-based image restoration under general illumination. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 8251–8260, 2023. 1

Table S3. Quantitative evaluation of different light intensities on synthetic dataset.

Method	Loss	Light intensity (16)			
		PSNR ↑	SSIM ↑	LPIPS	#SpikeNum
TFP(32)+NeRF	MSE	15.27/15.57	0.154/0.785	0.588/0.065	268,185
TFP(256)+NeRF	MSE	14.44/16.41	0.127/0.794	0.652/0.088	
TFI+NeRF	MSE	13.77/16.63	0.120/0.793	0.670/0.093	
Spk2img+NeRF	MSE	13.42/14.05	0.066/0.726	0.724/0.129	
Ours	$\mathcal{L}_i{}^*$	13.77/16.63	0.120/0.793	0.670/0.093	
	$\mathcal{L}_i$	18.55/**19.76**	0.237/0.878	0.527/0.051	
	$\mathcal{L}_s$+$\mathcal{L}_i$	18.50/19.59	0.237/0.876	0.524/0.051	
	$\mathcal{L}_s$	**18.72**/19.38	**0.251/0.880**	**0.517/0.050**	
Method	Loss	Light intensity (32)			
		PSNR ↑	SSIM ↑	LPIPS	#SpikeNum
TFP(32)+NeRF	MSE	17.81/18.55	0.191/0.827	0.562/0.077	414,475
TFP(256)+NeRF	MSE	15.82/18.68	0.170/0.835	0.616/0.077	
TFI+NeRF	MSE	15.07/18.53	0.156/0.825	0.634/0.083	
Spk2img+NeRF	MSE	14.59/15.39	0.109/0.768	0.651/0.093	
Ours	$\mathcal{L}_i{}^*$	15.07/18.53	0.156/0.825	0.634/0.083	
	$\mathcal{L}_i$	21.21/22.81	0.272/0.890	0.492/0.055	
	$\mathcal{L}_s$+$\mathcal{L}_i$	21.76/23.59	0.279/0.913	0.486/0.053	
	$\mathcal{L}_s$	**22.05/23.66**	**0.300/0.926**	**0.477/0.050**	
Method	Loss	Light intensity (64)			
		PSNR ↑	SSIM ↑	LPIPS	#SpikeNum
TFP(32)+NeRF	MSE	20.63/21.86	0.239/0.872	0.528/0.067	720,032
TFP(256)+NeRF	MSE	17.28/21.27	0.189/0.852	0.596/0.081	
TFI+NeRF	MSE	16.16/20.56	0.187/0.852	0.602/0.077	
Spk2img+NeRF	MSE	16.37/17.73	0.169/0.837	0.595/0.073	
Ours	$\mathcal{L}_i{}^*$	16.16/20.56	0.187/0.852	0.602/0.077	
	$\mathcal{L}_i$	21.28/21.98	0.334/0.874	0.463/0.068	
	$\mathcal{L}_s$+$\mathcal{L}_i$	**24.09/24.85**	0.382/0.920	0.442/0.051	
	$\mathcal{L}_s$	23.89/24.46	**0.411/0.929**	**0.428/0.049**	

Table S4. Compasrison to Spk2imgNet and RSIR.

Method	Light intensity (16)			Light intensity (32)			Light intensity (64)		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Spk2imgNet-finetuned	13.38	0.712	0.121	14.78	0.749	0.100	16.92	0.793	0.091
Spk2imgNet-retrained	13.60	0.718	0.114	14.98	0.756	0.095	17.16	0.806	0.081
RSIR(w=32,c=8)	14.58	0.754	0.087	16.49	0.906	0.072	19.58	0.848	0.071
RSIR(w=64,c=4)	14.71	0.757	0.088	16.42	0.794	0.083	19.51	0.846	0.071
RSIR(w=256,c=1)	14.69	0.757	0.088	16.46	0.803	0.076	19.48	0.846	0.071
Ours	**19.38**	**0.880**	**0.050**	**23.66**	**0.926**	**0.050**	**24.46**	**0.929**	**0.049**

Table S5. Ablation of the nonuniformity matrix.

Method	Light intensity (16)			Light intensity (32)			Light intensity (64)		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
TFP	16.41	0.794	0.088	18.68	0.835	0.077	21.27	0.852	0.081
TFI	16.63	0.793	0.093	18.53	0.825	0.083	20.56	0.852	0.077
Ours(w/o R, w=256)	16.23	0.819	0.052	19.66	0.886	0.048	23.67	0.910	0.058
Ours(w=256)	**19.38**	**0.880**	**0.050**	**23.66**	**0.926**	**0.050**	**24.46**	**0.929**	**0.049**

Table S6. Comparisons on different sequence lengths.

Method	Light intensity (16)			Light intensity (32)			Light intensity (64)		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Ours(w=32)	16.94	0.767	0.145	25.29	0.894	0.059	23.74	0.909	0.054
Ours(w=64)	**20.59**	0.876	0.053	25.92	0.918	0.053	23.18	0.915	0.056
Ours(w=128)	19.77	0.879	0.052	**25.93**	0.925	**0.050**	23.85	0.922	0.052
Ours(w=256)	19.38	**0.880**	**0.050**	23.66	**0.926**	**0.050**	**24.46**	**0.929**	**0.049**

Figure S3. Quantitative results on synthetic spike data.

Figure S4. Quantitative results on synthetic spike data.

Figure S5. Quantitative results on real-world spike data.