Supplementary Material for "Spiking NeRF: Representing the Real-World Geometry by a Discontinuous Representation"

Anonymous submission

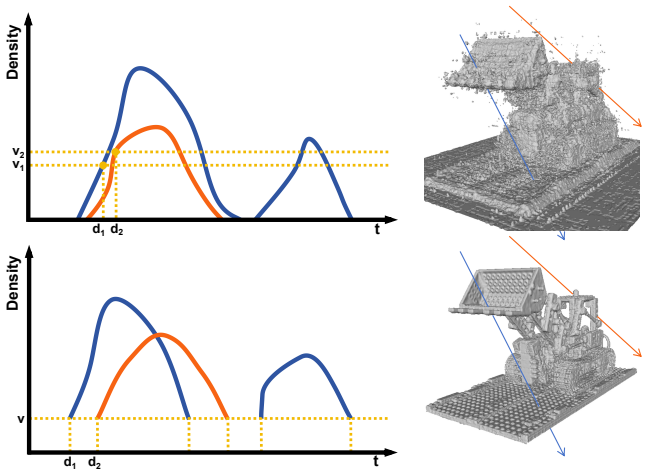

Figure 1: The blue line and orange line in the right represent two different light. The figures on the left represent the distribution of the density field through which light passes. t is the sampling distance from the sampling point to the camera origin. d_1 and d_2 represent t corresponding to the position of surface points along the direction of light rays (i.e., the depth corresponding to the direction of light). v_1 and v_2 represent the density corresponding to the position of surface points along the direction of light rays. For continuous fields, using the same threshold to extract geometric information cannot guarantee that the depth corresponding to the light is all accurate.

This supplementary material consists of two parts, the additional analysis and the additional results.

Additional Analysis

Optimal Threshold Perturbation

For a well-trained NeRF, $\int_0^T \sigma(t) \cdot \exp\left(-\int_0^t \sigma(t)\,\mathrm{d}t\right) \cdot t\,\mathrm{d}t$ can represent the depth of the corresponding light direction (Zhang et al. 2021; Guo et al. 2022). However, it depends on the direction. The final geometric information is extracted from the network using a threshold, independent of the direction. As shown in Fig. 1, when light from different directions passes through a density field, the corresponding density distribution is different, resulting in different density values corresponding to depth. So using the same threshold to extract geometric information cannot guarantee that the depth corresponding to the light is all accurate.

The Proof of Proposition 1

Define m is the index of the first non zero σ (i.e., the extracted depth $d_{\mathrm{v}} = t_m$), and m' is the index of the largest σ within m+1 and N (i.e., $V_{max} = \sigma_{m'}$). Δt_i is the sampling interval between sampling point $i-1$ and point i. The w_i refers to the weight at each sampling point as defined in (Mildenhall et al. 2021). We define $\beta_i = e^{-\sigma_i \Delta t_i}$. Then we have:

$$\sum_i^N w_i = 1 - \prod_i^N \beta_i. \tag{1}$$

Let $c = \prod_{i=1}^N \beta_i$. Since σ_1, ..., σ_{m-1} are all 0, we have:

$$0 < \beta_m e^{-\sigma_{m'}(t_N - t_m)} < c < \beta_m \beta_{m'} \tag{2}$$

According to previous analysis, for well trained NeRF, it can be considered that the depth d is $\int_0^T \sigma(t) \cdot \exp\left(-\int_0^t \sigma(t)\,\mathrm{d}t\right) \cdot t\,\mathrm{d}t$. Its Riemann sum form is $\sum_{i=1}^N w_i t_i$.

$$\begin{aligned} d - t_m &= \sum_{i=1}^N w_i t_i - t_m \\ &= \sum_{i=m}^N w_i t_i - t_m \left(\sum_{i=1}^N w_i + c\right) \\ &\geq (1 - w_m - c)(t_{m+1} - t_m) - t_m c \\ &\geq (t_{m+1} - t_m)\beta_m - t_{m+1}\beta_m\beta_{m'} \\ &\geq \left(\Delta t_{m+1} - T e^{-\sigma_{m'}\Delta t_{m'}}\right) e^{-\sigma_m \Delta t_m}. \end{aligned} \tag{3}$$

For a well-trained NeRF, the spiking threshold V_{th} equals to the non zero minimum value of its density field. Because the value before the last layer varies continuously with respect to the spatial point, the density value of the first point that a ray encounters with a non zero density equals to the non zero minimum value of its density field. Then $V_{\mathrm{th}} = \sigma_m$. So we have:

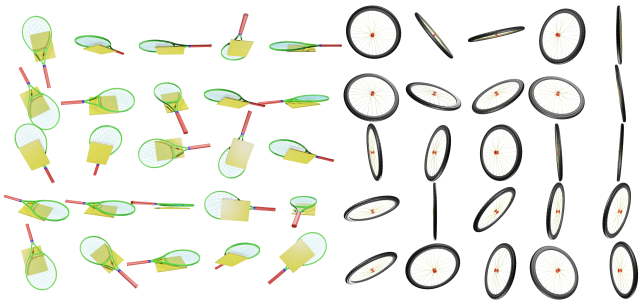

Figure 2: Self created scenes. Left: A tennis racket with a board. Right: A wheel.

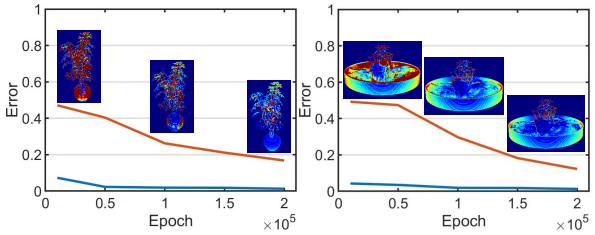

Figure 3: The relationship between the upper bound and the average depth error during training. We show two more scenes from Blender dataset (Mildenhall et al. 2021). The red curve represents upper bound while the blue curve represents the average depth error during training.

$$\begin{aligned} d - t_m &\geq \left(\Delta t_{m+1} - T e^{-\sigma_{m'} \Delta t_{m'}}\right) e^{-\sigma_m \Delta t_m} \\ &= \left(\Delta t_{m+1} - T e^{-V_{\max} \Delta t_{m'}}\right) e^{-V_{\text{th}} \Delta t_m}. \end{aligned} \tag{4}$$

$$\begin{aligned} d - t_m &= \sum_{i=1}^{N} w_i t_i - t_m \\ &= \sum_{i=m}^{N} w_i t_i - t_m \left(\sum_{i=1}^{N} w_i + c\right) \\ &\leq (t_N - t_m)(1 - w_m - c) - t_m c \\ &\leq (t_N - t_m)\beta_m - t_N \beta_m e^{-\sigma_{m'}(t_N - t_m)} \\ &\leq \left(t_N - t_N e^{-\sigma_{m'} t_N}\right) \beta_m \\ &= T\left(1 - e^{-V_{\max} T}\right) e^{-V_{\text{th}} \Delta t_m}. \end{aligned} \tag{5}$$

Since we do not explore the influence of sampling point positions, we have written Δt_m,Δt_{m+1} and $\Delta t_{m'}$ in the equation as Δt in the paper for the convenience of observation.

Then we have:

$$\left(\Delta t - T e^{-V_{\max} \Delta t}\right) e^{-V_{\text{th}} \Delta t} < d - d_{\text{v}} < T\left(1 - e^{-V_{\max} T}\right) e^{-V_{\text{th}} \Delta t}. \tag{6}$$

For a general setting, T is greater than 100 times Δt. And Δt is approximately 0.01. So $\left|\Delta t - T e^{-V_{\max} \Delta t}\right| <= T\left(1 - e^{-V_{\max} T}\right)$ for $V_{\max} > 4$. Therefore, for a fixed $V_{\text{th}} > 4$, which is basically valid, a small $V_{\max}$ can decrease the error.

Other Bounded Functions

IF spiking neuron. The IF spiking neuron has a bound. However, according to the analysis in **Preliminary**, it is ultimately difficult to maintain accuracy using IF spiking neuron.

FIF with a hard bound.

$$o_{t+1} = \begin{cases} 0 & u_{t+1}^{\text{pre}} < V_{\text{th}} \\ 1 & \text{otherwise}, \end{cases} \tag{7}$$

$$y_{t+1} = \min(o_{t+1} \cdot V_{\text{th}}, B). \tag{8}$$

Here, B is the upper bound of FIF spiking neuron. Although there is a bound, increasing the bound requires additional operations (e.g., additional loss term or a strategy to increase B), If B is not updated, the network will only output 0 when V_{th} exceeds B.

We use $k\tanh()$ in the paper because k can be updated through the original loss.

Analysis of Surrogate Gradient

The surrogate gradient in (Li et al. 2022) is:

$$\frac{\partial L}{\partial V_{\text{th}}} = \sum \frac{\partial L}{\partial y_t}(1 - o_t). \tag{9}$$

Due to $t = 1$ in our paper. We have:

$$\frac{\partial L}{\partial V_{\text{th}}} = \frac{\partial L}{\partial y}(1 - o). \tag{10}$$

Since the value of $1 - o$ can only be 0 or 1, it can lead to unstable training.

Additional Results

Introduction of Our Dataset

We additionally generate our own dataset containing path-traced images of two objects. They are all rendered from viewpoints uniformly sampled on a full sphere. We render 100 views of each scene as input and 200 for testing, all at 800 × 800 pixels (see Fig. 2).

Validation for Upper Bound of Proposition 1

We show two more scenes for the upper bound of proposition 1 in Fig. 3. It can be seen that the average depth error decreases with the upper bound and the average depth error keeps being less than upper bound during training.

References

Guo, Y.-C.; Kang, D.; Bao, L.; He, Y.; and Zhang, S.-H. 2022. NeRFReN: Neural Radiance Fields with Reflections. In *CVPR*.

Li, W.; Chen, H.; Guo, J.; Zhang, Z.; and Wang, Y. 2022. Brain-Inspired Multilayer Perceptron with Spiking Neurons. In *CVPR*.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM*.

Zhang, X.; Srinivasan, P. P.; Deng, B.; Debevec, P.; Freeman, W. T.; and Barron, J. T. 2021. NeRFactor: Neural Factorization of Shape and Reflectance Under an Unknown Illumination. *TOG*.