DECOUPLED SPIKING MAMBA FOR HIGH-FIDELITY NEURAL RADIANCE FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural Radiance Fields (NeRF) built on Artificial Neural Networks (ANNs) dominate 3D reconstruction and rendering but incur a severe energy bottleneck. Spiking Neural Networks (SNNs) replace Multiply-Accumulate (MAC) with sparse Accumulate (AC) operations for low-power 3D representation, yet SNN-based NeRF methods, despite narrowing the metric gap to ANN-based NeRF, still suffer from geometric fragmentation and floater artifacts. Fragmentation arises because NeRF queries a Multi-Layer Perceptron (MLP) independently per point without neighborhood context, and it coincides with unstable spike firing in deep spiking layers; floater artifacts arise from density ambiguity in unconstrained volume rendering, which the limited capacity of SNNs amplifies. To address these issues, this paper proposes DSMamba-NeRF (Decoupled Spiking Mamba for Neural Radiance Fields), an end-to-end high-fidelity spiking architecture. A Decoupled Spiking Mamba (DSMamba) module tackles fragmentation by splitting voxel channels into a global branch that aggregates neighborhood topology via multidirectional state space scanning and a local branch that preserves high-frequency boundaries through 3D spiking convolutions. A Surface-aware spiking Rendering and Refinement (SRF) mechanism replaces unconstrained density with a spiking VolSDF that imposes explicit geometric constraints near the zero level set to suppress floater artifacts. An Adaptive Ternary Spike Neuron (ATSN) with Spike Firing Adaptation (SFA) expands information capacity and modulates the membrane threshold, ensuring stable gradients and an accumulate dominated inference pipeline, with multiplications confined to the compact decoding heads. On synthetic and real-world scenes, DSMamba-NeRF alleviates both defects, attains state-of-the-art rendering metrics among SNN-based NeRF methods, and approaches ANN-based NeRF in visual quality at low energy cost.

1 INTRODUCTION

Neural Radiance Fields (NeRF) (Mildenhall et al., 2021; Guo et al., 2024a; Pumarola et al., 2021; Cui et al., 2024; Chen et al., 2024) advance 3D reconstruction and novel view synthesis through end-to-end differentiable training and photorealistic rendering, with applications in single-image view synthesis and digital human modeling (Lin et al., 2023; Hong et al., 2022; Li et al., 2024), virtual reality and robotic perception (Deng et al., 2022; Li et al., 2022), and medical imaging (Corona-Figueroa et al., 2022; Chen et al., 2023). However, NeRF based on Artificial Neural Networks (ANNs) depends on dense floating-point Multiply-Accumulate (MAC) operations (Kang & Yoo, 2022; Al-Rikabi & Renczes, 2024) and pointwise independent queries (Gunawardena et al., 2022), which creates an energy bottleneck for resource-constrained mobile and edge devices (Abbas et al., 2018; Satyanarayanan, 2017).

To overcome this bottleneck, recent studies bring Spiking Neural Networks (SNNs) (Yao et al., 2023a; Bal & Sengupta, 2025; Zhao et al., 2024; Liao et al., 2024) into NeRF, replacing floating-point activations with event-driven spikes (Wan et al., 2025; Guo et al., 2024b) and MAC with sparse Accumulate (AC) operations. Subsequent methods (Yao et al., 2023b; Liao et al., 2024; Wan et al., 2025) narrow the gap to ANN-based NeRF on global metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), yet such image-level averages are insensitive to localized degradation that occupies few pixels in high-frequency details, and visual comparison with ANN-based NeRF exposes two key problems.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

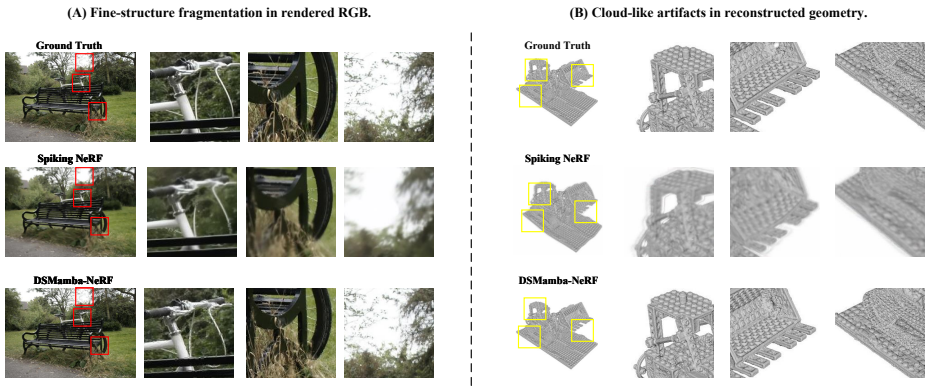


Figure 1: Geometric fragmentation (A, rendered RGB on *Bicycle*) and floater artifacts (B, reconstructed geometry on *Lego*) in the SNN-based Spiking NeRF, both mitigated by DSMamba-NeRF.

The first is geometric fragmentation, where bicycle spokes, brake cables, and leaf edges appear broken or missing, as shown in Figure 1(A). NeRF queries a Multi-Layer Perceptron (MLP) independently per sampled point and stitches isolated predictions through implicit smooth interpolation (Turk et al., 2001), which severs neighborhood correlation and undermines global geometric consistency. Unstable spike firing in SNNs weakens this correlation further, amplifying fragmentation.

The second is density ambiguity. The transmittance-weighted integral along a ray does not determine density uniquely, because distinct density profiles yield nearly identical colors and leave nonzero density far from the true surface (Zhang et al., 2020; Wang et al., 2021; Yariv et al., 2021). When a capacity-limited SNN fits this underdetermined field (Guo et al., 2023; 2024c;d), spike noise amplifies the ambiguity that continuous MLP outputs would otherwise mask, producing blurred boundaries and floater artifacts, as shown in Figure 1(B).

To address these problems, this paper proposes Decoupled Spiking Mamba for Neural Radiance Fields (DSMamba-NeRF), an end-to-end architecture for high-fidelity, low-power novel view synthesis, illustrated in Figure 2. The main contributions are as follows.

- **Decoupled Spiking Mamba (DSMamba) Module**: A channel-decoupled dual-branch module that first integrates spike-compatible state space modeling into the NeRF rendering pipeline. The global branch aggregates large-scale geometric topology via multidirectional spiking state space scanning, the local branch preserves high-frequency boundaries through

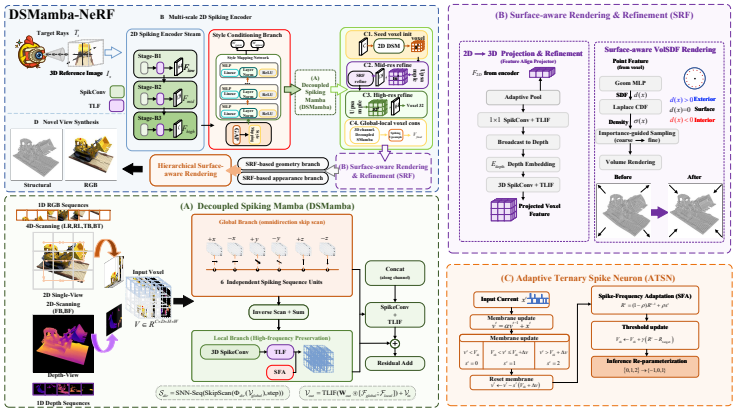


Figure 2: Overall architecture of DSMamba-NeRF. The channel-decoupled DSMamba module aggregates global topology and local high-frequency boundaries, SRF performs SDF-constrained spiking rendering and refinement, and every spiking layer is instantiated with ATSN.

3D spiking convolutions, and their cooperation in the spiking domain builds cross-point neighborhood correlation that alleviates geometric fragmentation from pointwise queries.

- **Surface-aware spiking Rendering and Refinement (SRF)**: SRF replaces the unconstrained NeRF density with a spiking VolSDF distribution under explicit Signed Distance Function (SDF) constraints, concentrating density near the zero level set to suppress density ambiguity and floater artifacts, while high-frequency spiking features from the DSMamba local branch are projected onto the surface neighborhood to preserve thin structures and sharp edges.
- **Adaptive Ternary Spike Neuron (ATSN)**: Spike Firing Adaptation (SFA) modulates the membrane threshold of ternary spike neurons against spike silence and saturation, stabilizing gradients in end-to-end training and supporting an accumulate dominated inference pipeline.
- **Extensive Experimental Evaluation**: On the NeRF synthetic (Mildenhall et al., 2021) and real-world datasets, DSMamba-NeRF attains state-of-the-art PSNR, SSIM, and Learned Perceptual Image Patch Similarity (LPIPS) among SNN-based NeRF methods and narrows the gap to ANN-based NeRF at low energy cost, and qualitative results on Mip-NeRF 360 (Barron et al., 2022) confirm these gains.

2 Related Work

2.1 Spiking Neural Networks for 3D Representation

Recent work extends SNNs from 2D vision to 3D implicit reconstruction (Wang et al., 2023; Ran et al., 2023) and other 3D perception tasks, replacing dense floating-point ANN computation with sparse AC operations. For SNN-based NeRF, the model of Yao et al. (2023b), SpikingNeRF, first aligns the SNN temporal dimension with radiance rays and resolves temporal length and padding effects on membrane potentials, yet its weak global modeling leaves geometric fragmentation and floater artifacts. Spiking NeRF (Liao et al., 2024) proposes the first SNN-ANN hybrid with bounded spikes for discontinuous density, demonstrating spiking volume rendering, but still regresses density with ANNs and trails ANN baselines. Spik-NeRF (Wan et al., 2025) enlarges information capacity with ternary spike coding and distills a pretrained ANN teacher, yet its two-stage pipeline raises training cost, bounds performance by the teacher, and leaves ANN dependence unresolved.

Beyond the rendering pipeline, the fixed-threshold Leaky Integrate-and-Fire (LIF) neuron (Yao et al., 2022) suffers from spike silence and saturation in deep 3D networks, which attenuates cross-layer information and breaks gradient backpropagation, while the local receptive field of spiking convolutions prevents long-range 3D modeling (Tal & Schwartz, 1997) and aggravates fragmentation. End-to-end spiking 3D modeling therefore requires moving beyond ANN distillation and hybrid rendering.

2.2 Efficient Neural Radiance Fields and State Space Models for 3D

Explicit feature structures accelerate NeRF by reducing MLP queries. TensoRF (Chen et al., 2022) compresses the radiance field into low-rank tensors, and Instant NGP (Müller et al., 2022) and Nerfacto (Tancik et al., 2023) use multi-resolution hash grids for near-real-time training and rendering. These methods still rely on pointwise independent queries (Gao et al., 2026), in which each sampled feature depends only on its local grid cell and lacks explicit global geometric modeling.

Structured State Space Models (SSMs) such as Mamba (Dao & Gu, 2024; Gu & Dao, 2024; Lahoti et al., 2026; Zhu et al., 2025) capture long-range dependencies with linear-time sequence modeling, and PointMamba (Liang et al., 2024) and Mamba3D (Han et al., 2024) apply them to point cloud understanding through serialized scanning. In NeRF, where high-frequency textures and sharp edges are essential, omnidirectional flattening encodes voxels into a 1D sequence, destroying inter-voxel topology and over-smoothing surfaces. Porting Mamba state propagation to hash grids also fails topologically, because multi-resolution hash encoding maps adjacent coordinates to discontiguous addresses and strips state propagation of geometric meaning. Reconciling the global modeling of SSMs with 3D local structural integrity thus remains open.

2.3 SURFACE-CONSTRAINED NEURAL RENDERING

Conventional NeRF renders with unconstrained density and extracts surfaces as iso-surfaces, so without explicit geometric regularization its reconstructions show blurred edges, floater artifacts, and distorted thin structures. VolSDF (Yariv et al., 2021) and NeuS (Wang et al., 2021) therefore introduce differentiable mappings through the Laplace Cumulative Distribution Function (CDF) and the logistic sigmoid, imposing SDF constraints that stabilize surface evolution and improve reconstruction. Neuralangelo (Li et al., 2023) combines multi-resolution hash encoding with progressive refinement for high-fidelity surface reconstruction at scale. Gu et al. (2025) embed a Full-precision Integrate-and-Fire (FIF) spiking neuron into Nerfacto as an auxiliary density thresholder, with Round-Robin alternating training and a learnable truncation threshold that sharpens density and reduces surface artifacts, yet it retains dense floating-point MAC operations, leaving substantial energy headroom for edge deployment. This paper therefore embeds SDF constraints stably into an SNN-based NeRF pipeline and adds controllable high-frequency guidance that compensates for the smoothing bias of SDF.

3 METHODOLOGY

The appendix reviews NeRF volume rendering, VolSDF, the ternary spike coding adopted from Spik-NeRF, and the full system pipeline with its component definitions. Sections 3.1 to 3.3 present DSMamba, SRF, and ATSN, and Section 3.4 defines the optimization objective. All spiking layers in Sections 3.1 and 3.2 use ATSN.

3.1 CHANNEL-DECOUPLED SPIKING MAMBA

Spiking convolutions suffer from limited receptive fields in 3D space, while direct 1D scanning of voxels with SSMs disrupts local topology and over-smooths high-frequency edges despite their linear-complexity long-range modeling. DSMamba resolves this conflict by partitioning voxel channels into a global modeling branch and a local preserving branch, which jointly capture contours and boundaries at low power in the spiking domain.

3.1.1 MULTIDIRECTIONAL SKIP SCAN

Given an input voxel feature $\mathcal{V}_{\text{in}} \in \mathbb{R}^{C\times D\times H\times W}$, a ratio γ splits the channels into $\mathcal{V}_{\text{global}} \in \mathbb{R}^{\gamma C\times D\times H\times W}$ and $\mathcal{V}_{\text{local}} \in \mathbb{R}^{(1-\gamma)C\times D\times H\times W}$. For $\mathcal{V}_{\text{global}}$, a Reference-Centered 3D Dynamic Scan performs mirrored concatenation, flattening, and skip sampling along the six signed directions of depth (D), height (H), and width (W), yielding a 1D sequence for each $\text{dir} \in \{d^+, d^-, h^+, h^-, w^+, w^-\}$. A spiking state space unit built on ATSN (Section 3.3) scans each sequence, and the inverse scan Φ_{dir}^{-1} restores every directional output to $D\times H\times W$ before element-wise summation:

$$
\begin{aligned}
x_k^{\text{dir}} &= \big[\Pi_{s,o}\,\text{vec}\,\Phi_{\text{dir}}(\mathcal{V}_{\text{global}})\big]_k, \qquad \alpha_k^{\text{dir}} = \alpha\,\mathbf{1}\big[s_{k-1}^{\text{dir}} = 0\big],\\
u_k^{\text{dir}} &= \alpha_k^{\text{dir}} u_{k-1}^{\text{dir}} + \mathbf{W}_{\text{in}}^{\text{dir}} x_k^{\text{dir}}, \qquad s_k^{\text{dir}} = \mathcal{Q}\big(u_k^{\text{dir}}; V_{th}, \Delta v\big),\\
\mathcal{F}_{\text{global}} &= \sum_{\text{dir}} \Phi_{\text{dir}}^{-1}\big(\Pi_{s,o}^{\top}\mathbf{W}_{\text{out}}^{\text{dir}} s^{\text{dir}}\big)
\end{aligned}
\tag{1}
$$

where Φ_{dir} denotes mirrored concatenation and flattening along direction dir, $\Pi_{s,o}$ keeps one position every s steps from offset o and $\Pi_{s,o}^{\top}$ returns the outputs to their voxel indices, $\mathcal{Q}$ is the ternary firing function of Equation (4), and α is the membrane leak. The state is carried across blocks of M scanned positions with gradients truncated at block boundaries, which bounds activation memory and keeps the scan linear in the voxel count. Because a firing event clears the residual potential, α_k^{dir} switches between α and zero according to the input, which is the binary limit of the continuous selection used by selective state space models. Appendix A.3 details the scan, the block size, and the parameter sharing rules.

3.1.2 3D SPIKING CONVOLUTION AND RESIDUAL FUSION

Since 1D flattening weakens 3D neighborhood relations, the local branch applies an ATSN-based $3 \times 3 \times 3$ spiking convolution to $\mathcal{V}_{\text{local}}$, whose ternary encoding replaces dense MAC operations with sparse additions while preserving 3D topology, and $\{-1, 0, 1\}$ is the symmetric inference form of the asymmetric training levels $\{0, 1, 2\}$. A $1 \times 1 \times 1$ spiking convolution with an identity residual then projects the two branches, concatenated along channels, back to C channels, and Appendix A.4 states both operators.

The fused field therefore gives SRF spatially coherent voxel features that combine large-scale topology and boundary detail.

3.2 SURFACE-AWARE SPIKING RENDERING AND REFINEMENT

SRF first replaces unconstrained volume density with SDF constraints in the spiking domain, then uses deterministic 2D boundaries to guide 3D voxel refinement against surface degradation from quantization noise.

3.2.1 SPIKING VOLSDF

To keep deep features within the active firing range, a threshold-dependent Layer Normalization rescales their magnitudes to the ATSN threshold through a learnable coefficient, as stated in Appendix A.4. An ATSN-based spiking MLP predicts the signed distance $d(\mathbf{x}) \in \mathbb{R}$, and the Laplace CDF with scale η maps it to the density $\sigma(\mathbf{x})$, concentrating learning near the zero level set $d(\mathbf{x}) = 0$.

3.2.2 SPATIAL REFINEMENT FEATURE

Large-scale voxel decoding must upsample decoupled channels and inject 2D details from the DS-Mamba local branch without breaking spike discreteness, which interpolation and deconvolution violate by introducing floating-point values. SRF therefore adopts a ternary channel-decoupled lossless upsampling scheme that replaces interpolation with tensor operations. Adaptive pooling projects multi-scale 2D spiking features $\mathcal{F}_{\text{2D}}$ into 3D frustum space, and a learnable depth positional matrix $\mathbf{E}_{\text{depth}} \in \mathbb{R}^{1 \times D \times 1 \times 1}$ reduces the depth ambiguity of ray broadcasting:

$$\mathcal{V}_{\text{proj}} = \text{Broadcast}(\text{ATSN}(\mathbf{W}_{\text{align}} \circledast \mathcal{F}_{\text{2D}})) + \mathbf{E}_{\text{depth}} \tag{2}$$

A 3D spiking convolution then propagates features along depth to yield $\mathcal{F}_{\text{SRF}}$. At the intermediate resolution, voxels are mean pooled along depth, a Spiking 2D Mamba (SDM) calibrates in-plane consistency with four-directional scanning, and frustum expansion reinjects the result so that deterministic 2D boundaries and in-plane continuity reduce 3D distortion. With $\mathcal{V}_l^{\text{up}}$ denoting the upsampled voxel feature at resolution level l, the refinement reads:

$$\begin{aligned} \mathcal{V}_{\text{planar}} &= \text{SDM}(\text{Mean}_z(\mathcal{V}_l^{\text{up}} + \mathcal{F}_{\text{SRF}})), \\ \mathcal{V}_l^{\text{refined}} &= \mathcal{V}_l^{\text{up}} + \mathcal{F}_{\text{SRF}} + \text{Broadcast}(\mathcal{V}_{\text{planar}}) \end{aligned} \tag{3}$$

SRF therefore pairs an SDF-constrained density field with in-plane boundary evidence, which suppresses floater artifacts while preserving thin structures.

3.3 ADAPTIVE TERNARY SPIKE NEURON

The Ternary Spike Neuron (TSN) of Spik-NeRF (Wan et al., 2025) leaves two trainability issues unresolved. Its fixed threshold triggers spike silence and saturation on 3D features with large dynamic range, destabilizing gradients in deep networks, and its ternary firing function exhibits two non-differentiable jumps that a single-threshold surrogate cannot cover. ATSN therefore augments the TSN with SFA, which modulates the threshold by layer-wise firing rate deviation, alongside a dual-threshold sigmoid surrogate gradient for stable end-to-end training.

At time step t, the TSN updates the membrane potential v^t from the residual u^{t-1} and the current input. With the SFA-modulated threshold V_{th}^t and step margin Δv, asymmetric ternary firing reads:

$$v^t = \alpha u^{t-1} + Wx^t + b,$$
$$s^t = \begin{cases} 2, & v^t > V_{th}^t + \Delta v \\ 1, & V_{th}^t \le v^t \le V_{th}^t + \Delta v \\ 0, & v^t < V_{th}^t \end{cases} \tag{4}$$

where x^t is the input spike train of the preceding layer, α is the membrane leak, and each firing clears the residual potential, $u^t = v^t \mathbf{1}[s^t = 0]$. Setting $\Delta v = 0.5\, V_{th}^0$ keeps the two firing levels separable as SFA updates V_{th}^t and prevents synchronous margin compression from destabilizing firing statistics.

A dual-threshold sigmoid surrogate places a smooth derivative at both jumps, V_{th}^t and $V_{th}^t + \Delta v$, where automatic differentiation transmits no valid gradient:

$$\frac{\partial s^t}{\partial v^t} \approx \frac{1}{2}\left[\sigma'_\beta(v^t - V_{th}^t) + \sigma'_\beta(v^t - V_{th}^t - \Delta v)\right] \tag{5}$$

where $\sigma'_\beta(x) = \beta\sigma(\beta x)(1 - \sigma(\beta x))$ and β controls steepness. The dual-threshold form propagates gradients through both jumps, whereas a single-threshold surrogate vanishes beyond the first jump and breaks down in highly activated regions.

SFA updates the threshold at each iteration by the deviation between the exponentially averaged firing rate $\bar{R}^t = \lambda \bar{R}^{t-1} + (1 - \lambda)\mathbb{E}_{\text{batch},t}[s^t/2]$ and the target R_{target}:

$$V_{th}^{t+1} = \text{Clip}\big(V_{th}^t + \gamma_{\text{sfa}}(\bar{R}^t - R_{\text{target}}),\, V_{\text{min}},\, V_{\text{max}}\big) \tag{6}$$

High firing rates raise the threshold to suppress saturation while low rates lower it to reactivate silent neurons, yielding layer-wise self-stabilization for deep ternary spiking networks.

3.4 End-to-End Optimization

A composite objective supervises color, geometry, and spike statistics:

$$\begin{aligned} \mathcal{L}_{\text{total}} = \mathcal{L}_{\text{color}} &+ \lambda_{\text{lpips}}\mathcal{L}_{\text{lpips}} + \lambda_{\text{eik}}\mathcal{L}_{\text{eik}} \\ &+ \lambda_{\text{spike}}\mathcal{L}_{\text{spike}} + \lambda_{\text{tv}}\mathcal{L}_{\text{tv}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} \end{aligned} \tag{7}$$

$\mathcal{L}_{\text{color}}$ combines Mean Squared Error (MSE) with an L_1 edge term, and $\mathcal{L}_{\text{lpips}}$ is an AlexNet perceptual loss on rendered patches. $\mathcal{L}_{\text{eik}}$ constrains the SDF gradient norm toward unity through forward finite differences, because the SNN step function blocks second-order automatic differentiation. $\mathcal{L}_{\text{spike}}$ penalizes deviation of the batch-averaged firing rate from R_{target}, $\mathcal{L}_{\text{tv}}$ smooths the voxel grid by total variation, and $\mathcal{L}_{\text{reg}}$ regularizes the style vectors. A three-stage curriculum activates the Eikonal, LPIPS, and L_1 terms in sequence so that geometry converges before appearance is refined.

4 Experiments

DSMamba-NeRF is evaluated on synthetic and real-world datasets following established protocols (Mildenhall et al., 2021), with training images resized to 400×400 pixels. The synthetic dataset contains eight object scenes with 100 training and 200 test views each, and the real-world dataset contains eight smartphone-captured scenes from the Local Light Field Fusion (LLFF) benchmark with 20 to 60 images per scene, split 8:2 into training and test sets. Mip-NeRF 360 further tests generalization, and PSNR, SSIM, and LPIPS serve as evaluation metrics.

Implementation Details. DSMamba-NeRF is optimized with AdamW for 300k iterations using 1,024 rays per batch and a cosine-decayed learning rate from 5×10^{-4} to 1×10^{-6}. SDM uses four-directional 2D scans, and DSMamba uses six-directional axial 3D scans with a skip step of 2. Hierarchical voxel grids expand from 8^3 to 256^3, with $\gamma = 0.5$ and $\eta \in [0.02, 0.5]$. ATSN neurons use $T = 2$, surrogate gradient sharpness $\beta = 4$, and SFA hyperparameters $R_{\text{target}} = 0.35$, $\gamma_{\text{sfa}} = 0.01$, $V_{\text{min}} = 0.1$, and $V_{\text{max}} = 2.0$. Coarse-to-fine sampling draws 96 coarse points and then

Table 1: Per-scene PSNR (dB), SSIM, and LPIPS on the scenes in Figure 3 and the scenes with the largest and smallest PSNR gains over Spik-NeRF on each dataset. **Bold** marks the best SNN-based result.

		Synthetic					Real-world (LLFF)				
Metric	Method	Chair	Lego	Ship	Drums	Ficus	Flower	Horns	Room	Trex	Orchids
PSNR↑	NeRF (ANN)	34.06	31.39	29.39	25.51	29.17	27.44	28.06	31.41	27.25	21.17
	Spiking NeRF	31.96	29.37	27.89	24.46	24.04	26.22	26.31	30.16	25.34	20.46
	Spik-NeRF	33.37	30.84	28.71	25.24	26.09	26.97	27.11	30.87	26.22	20.91
	DSMamba-NeRF	**33.89**	**31.13**	**29.01**	**25.31**	**27.01**	**27.14**	**27.47**	**31.07**	**26.79**	**20.99**
SSIM↑	NeRF (ANN)	0.977	0.959	0.872	0.935	0.959	0.848	0.871	0.935	0.891	0.731
	Spiking NeRF	0.959	0.942	0.841	0.903	0.893	0.785	0.798	0.905	0.820	0.638
	Spik-NeRF	0.969	0.949	0.854	0.922	0.927	0.823	0.833	0.917	0.865	0.685
	DSMamba-NeRF	**0.973**	**0.953**	**0.863**	**0.926**	**0.942**	**0.829**	**0.841**	**0.922**	**0.874**	**0.692**
LPIPS↓	NeRF (ANN)	0.016	0.022	0.088	0.057	0.022	0.081	0.079	0.047	0.067	0.124
	Spiking NeRF	0.035	0.040	0.125	0.089	0.117	0.131	0.167	0.097	0.138	0.227
	Spik-NeRF	0.021	0.026	0.107	0.064	0.062	0.098	0.126	0.069	0.091	0.172
	DSMamba-NeRF	**0.019**	**0.025**	**0.097**	**0.061**	**0.054**	**0.092**	**0.121**	**0.053**	**0.083**	**0.165**

Table 2: Average quality over the eight scenes of each dataset, edge region quality on the real-world scenes, and measured energy per point query over all operations. **Bold** marks the best SNN-based result; N/A denotes not reported.

	Synthetic			Real-world (LLFF)			Edge region		Energy
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	LPIPS↓	(μJ)↓
NeRF (ANN)	31.11	0.952	0.033	26.85	0.849	0.084	24.97	0.132	1.77
Spiking NeRF	28.94	0.925	0.066	25.55	0.780	0.163	21.05	0.458	N/A
Spik-NeRF	30.21	0.939	0.046	26.24	0.816	0.121	22.23	0.273	0.727
DSMamba-NeRF	**30.54**	**0.945**	**0.042**	**26.48**	**0.824**	**0.110**	**23.61**	**0.193**	**0.252**

320 fine points during training and 192 fine points at evaluation. All models are trained on four NVIDIA A100 40 GB GPUs, and every reported number is the average of five runs per scene.

4.1 Comparison with State-of-the-Art Methods

The quantitative comparison covers the SNN baselines Spiking NeRF (Liao et al., 2024) and the state-of-the-art Spik-NeRF (Wan et al., 2025) together with the ANN baseline NeRF (Mildenhall et al., 2021), so every reported margin is measured inside the spiking regime and against the standard ANN reference. Spiking NeRF uses binary spikes and Spik-NeRF uses ternary spikes with teacher distillation, but both rely on unconstrained volume density rendering.

Table 1 and Table 4 in the appendix show that DSMamba-NeRF outperforms Spik-NeRF on all sixteen scenes, by 0.07 dB (*Drums*) to 0.92 dB (*Ficus*) on the synthetic dataset and by 0.08 dB (*Orchids*) to 0.57 dB (*Trex*) on the real-world dataset, and exceeds the binary Spiking NeRF by 1.60 dB and 0.93 dB on average. Under sparse accumulate dominated inference, its remaining gap to the ANN baseline NeRF narrows to 0.20 dB on *Drums* and 0.18 dB on *Orchids*. SSIM and LPIPS also improve over Spik-NeRF on every scene (averages in Table 2, per-scene values in Tables 1 and 4). Appendix A.5.6 analyzes the scenes that retain the largest gap to the ANN baseline.

Figure 3 confirms the advantage of surface-aware rendering. By constraining the spatial divergence of volume density, DSMamba-NeRF removes floater artifacts and produces smooth surfaces, as the *Ship* scene shows, and the Mip-NeRF 360 visualizations in the appendix further confirm robust boundary definition across diverse environments.

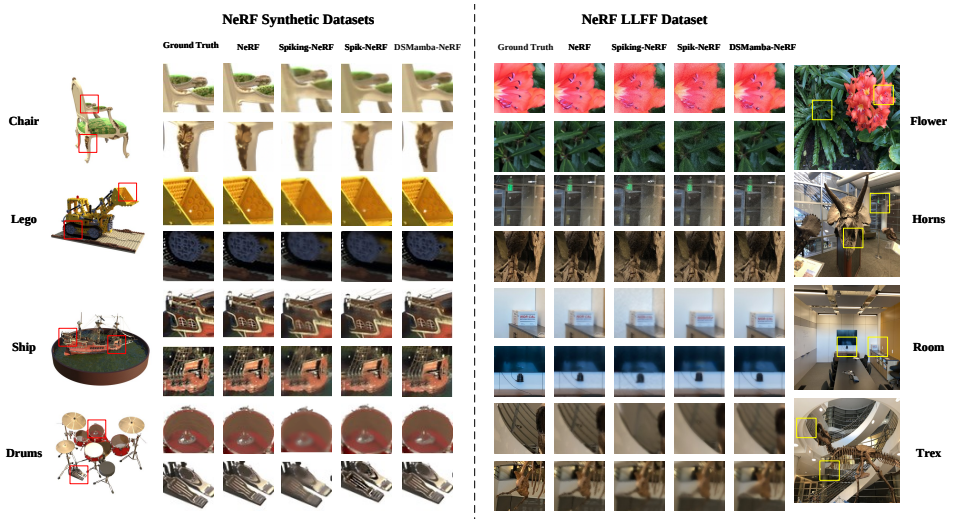


Figure 3: Qualitative comparison on NeRF Synthetic (left: *Chair*, *Lego*, *Ship*, *Drums*) and LLFF (right: *Flower*, *Horns*, *Room*, *Trex*).

Global averages dilute localized error, so the edge region columns of Table 2 report PSNR and LPIPS inside a ground-truth Canny edge mask dilated by 3 pixels, which covers 15% of each real-world image. Every method degrades inside the mask, but DSMamba-NeRF loses only 2.87 dB, against 4.50 dB for Spiking NeRF and 4.01 dB for Spik-NeRF, and remains closest to the ANN baseline. Its margin over Spik-NeRF widens from 0.24 dB to 1.38 dB in PSNR and from 0.011 to 0.080 in LPIPS on thin structures and boundaries.

Operation counts are measured by instrumenting the implementation during inference, normalized per network query on one sampled point, and converted with the neuromorphic convention of 3.2 pJ per MAC and 0.1 pJ per AC (Horowitz, 2014), excluding memory traffic and the shared volume rendering stage. Per point query DSMamba-NeRF consumes 0.252 μJ against 0.727 μJ for Spik-NeRF and 1.77 μJ for NeRF (Table 2), a reduction of 2.9 times and 7.0 times at the highest average PSNR among spiking methods. Its spiking feature pipeline accounts for 0.022 μJ with 0.22 M ternary ACs, against 0.049 μJ for Spik-NeRF. The margin comes from the explicit voxel grid, which replaces dense MLP queries with sparse AC lookups, and from the half-width global branch at $\gamma = 0.5$, and Appendix A.5.2 reports the measurements per ray and per frame.

4.2 Ablation Study & Detailed Analysis

Table 3 evaluates the contribution of ATSN, SRF, and DSMamba. The baseline replaces ATSN with binary spiking neurons in the same 3D spiking radiance field, so consecutive rows differ only in the module under test.

Replacing the binary neurons with ATSN raises PSNR by 0.31 dB on *Flower* and 0.25 dB on *Horns* and reduces LPIPS on both, because SFA bounds the firing rate against the saturation and silence

Table 3: Ablation study of DSMamba-NeRF on the *Flower* and *Horns* real-world scenes. **Bold** marks the best result.

Model	Flower			Horns		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Binary SNN	26.45	0.781	0.126	26.37	0.808	0.154
+ ATSN	26.76	0.805	0.103	26.62	0.824	0.137
+ ATSN + SRF	26.94	0.819	0.097	27.09	0.830	0.129
Full	**27.14**	**0.829**	**0.092**	**27.47**	**0.841**	**0.121**

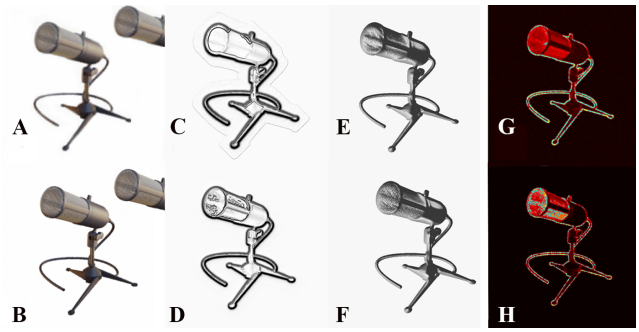


Figure 4: Ablation of SRF on the *Mic* scene. Top: baseline without SRF. Bottom: DSMamba-NeRF with SRF. (A, B) Rendered RGB images. (C, D) Depth maps. (E, F) Structural maps. (G, H) Error maps.

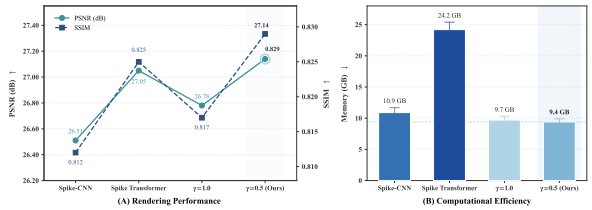


Figure 5: Ablation of SNN backbones and the channel-decoupling ratio γ on the *Flower* scene. (A) PSNR and SSIM. (B) Memory footprint.

of fixed thresholds and expands effective information capacity. Adding SRF on top of ATSN raises PSNR by 0.18 dB on *Flower* and 0.47 dB on *Horns* and further reduces LPIPS by 0.006 on *Flower* and 0.008 on *Horns*. SRF replaces unconstrained density with the spiking VolSDF formulation, removing the diffuse density that appears as floater artifacts. Adding DSMamba on top of ATSN and SRF raises PSNR by a further 0.20 dB on *Flower* and 0.38 dB on *Horns* and lifts SSIM to 0.829 and 0.841, as its global and local branches jointly reduce fragmentation in thin structures such as plant stems and sharp edges.

Figure 4 visualizes SRF on the *Mic* scene. Without SRF, the depth map shows an artifact halo in free space and the honeycomb mesh of the microphone is largely missing, whereas SRF removes the halo and preserves the mesh; Appendix A.5.7 analyzes each panel.

The ablation covers three controlled axes, namely the SNN backbone together with the channel-decoupling ratio γ, the neuron type together with the threshold mechanism, and the three SRF sub-components. On the *Flower* scene, spiking convolutions have limited receptive fields and fracture plant stems and leaves, with a low SSIM of 0.812 (Figure 5(A)). Spiking Transformers capture finer details, but their $O(N^2)$ complexity costs 24.2 GB and triggers Out of Memory failures at high resolution. With linear-complexity multidirectional scanning and $\gamma = 0.5$, DSMamba preserves global topology and local details while cutting memory to 9.4 GB, 61% below the Transformer backbone (Figure 5(B)). Appendix A.5.5 reports the temporal resolution T and the individual contributions of SFA and SRF.

5 Conclusion

DSMamba-NeRF is a high-fidelity spiking architecture that approaches ANN-based NeRF in rendering quality at low energy cost while mitigating geometric fragmentation and floater artifacts. DSMamba decouples voxel channels to aggregate global topology alongside local boundaries, SRF enforces SDF constraints on density, and ATSN adapts the membrane threshold for stable end-to-end training. The model attains state-of-the-art metrics among SNN-based NeRF methods with larger gains on thin structures and boundaries. Two directions remain open, namely spatially adaptive channel allocation between the two branches and extension to dynamic 3D video reconstruction.

AI USE STATEMENT

This work strictly follows the ICLR 2027 AI Policy for Authors. Generative AI tools were used only to edit the text for readability, namely spelling and grammar checking and polishing for coherence between paragraphs, so that the research content, the problem addressed, the novelty of this work, and the contribution of the proposed model to solving that problem are presented clearly. We reviewed all AI-assisted work. We take responsibility for the final content of this paper, including text, claims, and artifacts.

ETHICS STATEMENT

This work uses only the public NeRF Synthetic, LLFF, and Mip-NeRF 360 benchmarks, which contain no human subjects and no personal data.

REPRODUCIBILITY STATEMENT

Sections 3.1 to 3.4 specify the DSMamba, SRF, and ATSN modules and the optimization objective. Section 4 lists the datasets and splits, the optimizer and schedule, the voxel grid resolutions, the ATSN and SFA hyperparameters, the sampling counts, and the hardware, and it states the per-operation energy costs and the counting procedure behind the reported energy. The appendix provides the background formulations, the per-scene results of the remaining scenes, and the temporal resolution, SFA, and SRF ablations.

REFERENCES

Nasir Abbas, Yan Zhang, Amir Taherkordi, and Tor Skeie. Mobile edge computing: A survey. *IEEE Internet of Things Journal*, 5(1):450–465, 2018. doi: 10.1109/JIOT.2017.2750180.

H. Al-Rikabi and B. Renczes. Floating-point quantization analysis of multi-layer perceptron artificial neural networks. *Journal of Signal Processing Systems*, 96(4):301–312, 2024. ISSN 1939-8115. doi: 10.1007/s11265-024-01911-0. URL https://doi.org/10.1007/s11265-024-01911-0.

Malyaban Bal and Abhronil Sengupta. P-SPIKESSM: HARNESSING PROBABILISTIC SPIKING STATE SPACE MODELS FOR LONG-RANGE DEPENDENCY TASKS. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Sf4ep9Udjf.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5460–5469, 2022. doi: 10.1109/CVPR52688.2022.00539.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pp. 333–350, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-19823-6. doi: 10.1007/978-3-031-19824-3_20. URL https://doi.org/10.1007/978-3-031-19824-3_20.

Jiahao Chen, Yipeng Qin, Lingjie Liu, Jiangbo Lu, and Guanbin Li. Nerf-hugs: Improved neural radiance fields in non-static scenes using heuristics-guided segmentation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19436–19446, 2024. doi: 10.1109/CVPR52733.2024.01838.

Z. Chen, L. Yang, J. H. Lai, and X. Xie. CuNeRF: Cube-based neural radiance field for zero-shot medical image arbitrary-scale super resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 21128–21138, 2023. doi: 10.1109/ICCV51070.2023.01937. URL https://doi.ieeecomputersociety.org/10.1109/ICCV51070.2023.01937.

Abril Corona-Figueroa, Jonathan Frawley, Sam Bond Taylor, Sarath Bethapudi, Hubert P. H. Shum, and Chris G. Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3843–3848, 2022. doi: 10.1109/EMBC48229.2022.9871757.

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 Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i2.27908. URL https://doi.org/10.1609/aaai.v38i2.27908.

Tri Dao and Albert Gu. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=ztn8FCR1td.

Nianchen Deng, Zhenyi He, Jiannan Ye, Budmonde Duinkharjav, Praneeth Chakravarthula, Xubo Yang, and Qi Sun. Fov-nerf: Foveated neural radiance fields for virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3854–3864, 2022. doi: 10.1109/TVCG.2022.3203102.

Kyle Gao, Yina Gao, Hongjie He, Dening Lu, Linlin Xu, and Jonathan Li. Neural radiance fields in 3d vision: A comprehensive review. *Computational Visual Media*, pp. 1–55, 2026. doi: 10.26599/CVM.2026.9450535.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Yi Gu, Zhaorui Wang, and Renjing Xu. Sharpening your density fields: Spiking neuron aided fast geometry learning. In Beibei Wang and Alexander Wilkie (eds.), *36th Eurographics Symposium Rendering, EGSR 2025 - Symposium Track, Copenhagen, Denmark, June 25-27, 2025*. Eurographics Association, 2025. doi: 10.2312/SR.20251197. URL https://doi.org/10.2312/sr.20251197.

Nishan Gunawardena, Jeewani Anupama Ginige, and Bahman Javadi. Eye-tracking technologies in mobile devices using edge computing: A systematic review. *ACM Comput. Surv.*, 55(8), December 2022. ISSN 0360-0300. doi: 10.1145/3546938. URL https://doi.org/10.1145/3546938.

Lidong Guo, Xuefei Ning, Yonggan Fu, Tianchen Zhao, Zhuoliang Kang, Jincheng Yu, Yingyan (Celine) Lin, and Yu Wang. Rad-nerf: Ray-decoupled training of neural radiance field. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 113742–113771. Curran Associates, Inc., 2024a. doi: 10.52202/079017-3611. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/ce31378e9f41d8907e97dab172b6c559-Paper-Conference.pdf.

Yi-Jia Guo, Yuanxi Bai, Liwen Hu, Mianzhi Liu, Ziyi Guo, Lei Ma, and Tiejun Huang. Spike-nerf: Neural radiance field based on spike camera. *2024 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2024b. URL https://api.semanticscholar.org/CorpusID:268680823.

Yufei Guo, Weihang Peng, Yuanpei Chen, Liwen Zhang, Xiaode Liu, Xuhui Huang, and Zhe Ma. Joint a-snn: Joint training of artificial and spiking neural networks via self-distillation and weight factorization. *Pattern Recognition*, 142:109639, 2023. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2023.109639. URL https://www.sciencedirect.com/science/article/pii/S0031320323003400.

Yufei Guo, Yuanpei Chen, Xiaode Liu, Weihang Peng, Yuhan Zhang, Xuhui Huang, and Zhe Ma. Ternary spike: Learning ternary spikes for spiking neural networks. In *AAAI*, pp. 12244–12252, 2024c. URL https://doi.org/10.1609/aaai.v38i11.29114.

Yufei Guo, Weihang Peng, Xiaode Liu, Yuanpei Chen, Yuhan Zhang, Xin Tong, Zhou Jie, and Zhe Ma. Enof-snn: Training accurate spiking neural networks via enhancing the output feature. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 51708–51726. Curran Associates, Inc., 2024d. doi: 10.52202/079017-1638. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/5c6f928e3fc5f32ee29a1d916b68e6f5-Paper-Conference.pdf.

Xu Han, Yuan Tang, Zhaoxuan Wang, and Xianzhi Li. Mamba3d: Enhancing local features for 3d point cloud analysis via state space model. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, pp. 4995–5004, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400706868. doi: 10.1145/3664647.3681173. URL https://doi.org/10.1145/3664647.3681173.

Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. Headnerf: A real-time nerf-based parametric head model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20374–20384, June 2022.

Mark Horowitz. 1.1 computing's energy problem (and what we can do about it). In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 10–14, 2014. doi: 10.1109/ISSCC.2014.6757323.

Seungyeop Kang and Sungjoo Yoo. Ternarynerf: Quantizing voxel grid-based nerf models. In *2022 IEEE International Workshop on Rapid System Prototyping (RSP)*, pp. 8–14, 2022. doi: 10.1109/RSP57251.2022.10039009.

Aakash Lahoti, Kevin Li, Berlin Chen, Caitlin Wang, Aviv Bick, J Zico Kolter, Tri Dao, and Albert Gu. Mamba-3: Improved sequence modeling using state space principles. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=HwCvaJOiCj.

Chaojian Li, Sixu Li, Yang Zhao, Wenbo Zhu, and Yingyan Lin. Rt-nerf: Real-time on-device neural radiance fields towards immersive ar/vr rendering. In *Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design*, ICCAD '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392174. doi: 10.1145/3508352.3549380. URL https://doi.org/10.1145/3508352.3549380.

Dongze Li, Kang Zhao, Wei Wang, Bo Peng, Yingya Zhang, Jing Dong, and Tieniu Tan. Ae-nerf: Audio enhanced neural radiance field for few shot talking head synthesis. In *AAAI*, pp. 3037–3045, 2024. URL https://doi.org/10.1609/aaai.v38i4.28086.

Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H. Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8456–8465, 2023. doi: 10.1109/CVPR52729.2023.00817.

Dingkang Liang, Xin Zhou, Wei Xu, Xingkui Zhu, Zhikang Zou, Xiaoqing Ye, Xiao Tan, and Xiang Bai. Pointmamba: A simple state space model for point cloud analysis. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 32653–32677. Curran Associates, Inc., 2024. doi: 10.52202/079017-1026. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/395371f778ebd4854b88521100af30ad-Paper-Conference.pdf.

Zhanfeng Liao, Yan Liu, Qian Zheng, and Gang Pan. Spiking nerf: Representing the real-world geometry by a discontinuous representation. In *AAAI*, pp. 13790–13798, 2024. URL https://doi.org/10.1609/aaai.v38i12.29285.

Kai-En Lin, Lin Yen-Chen, Wei-Sheng Lai, Tsung-Yi Lin, Yi-Chang Shih, and Ravi Ramamoorthi. Vision transformer for nerf-based view synthesis from a single input image. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 806–815, 2023. doi: 10.1109/WACV56688.2023.00087.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65(1):99–106, December 2021. ISSN 0001-0782. doi: 10.1145/3503250. URL https://doi.org/10.1145/3503250.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4), July 2022. ISSN 0730-0301. doi: 10.1145/3528223.3530127. URL https://doi.org/10.1145/3528223.3530127.

Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10318–10327, June 2021.

Yunlong Ran, Jing Zeng, Shibo He, Jiming Chen, Lincheng Li, Yingfeng Chen, Gimhee Lee, and Qi Ye. Neurar: Neural uncertainty for autonomous 3d reconstruction with implicit neural representations. *IEEE Robotics Autom. Lett.*, 8(2):1125–1132, 2023. doi: 10.1109/LRA.2023.3235686. URL https://doi.org/10.1109/LRA.2023.3235686.

Mahadev Satyanarayanan. The emergence of edge computing. *Computer*, 50(1):30–39, 2017. doi: 10.1109/MC.2017.9.

Doron Tal and Eric L. Schwartz. Computing with the leaky integrate-and-fire neuron: Logarithmic computation and multiplication. *Neural Computation*, 9(2):305–318, 1997. doi: 10.1162/neco.1997.9.2.305.

Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David Mcallister, Justin Kerr, and Angjoo Kanazawa. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*, SIGGRAPH '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701597. doi: 10.1145/3588432.3591516. URL https://doi.org/10.1145/3588432.3591516.

G. Turk, Huong Quynh Dinh, J.F. O'Brien, and G. Yngve. Implicit surfaces that interpolate. In *Proceedings International Conference on Shape Modeling and Applications*, pp. 62–71, 2001. doi: 10.1109/SMA.2001.923376.

Gang Wan, Qinlong Lan, LiZihan, Huimin Wang, Wu Yitian, wang zhen, Wanhua Li, and Yufei Guo. Spik-neRF: Spiking neural networks for neural radiance fields. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=047VzZEpnu.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 27171–27183. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/e41e164f7485ec4a28741a2d0ea41c74-Paper.pdf.

Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3295–3306, October 2023.

Man Yao, JiaKui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 64043–64058. Curran Associates, Inc., 2023a. doi: 10.52202/075280-2798. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/ca0f5358dbadda74b3049711887e9ead-Paper-Conference.pdf.

Xingting Yao, Fanrong Li, Zitao Mo, and Jian Cheng. GLIF: A unified gated leaky integrate-and-fire neuron for spiking neural networks. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=UmFSx2c4ubT.

Xingting Yao, Qinghao Hu, Tielong Liu, Zitao Mo, Zeyu Zhu, Zhengyang Zhuge, and Jian Cheng. Spiking nerf: Making bio-inspired neural networks see through the real world. *CoRR*, abs/2309.10987, 2023b. URL https://doi.org/10.48550/arXiv.2309.10987.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=GlEWs-V9boR.

H. Zhang, J. Yao, L. Yu, and Y. Zhang. The noise cancelation effects caused by spike-frequency adaptation in single neurons. *Nonlinear Dynamics*, 100(2):1825–1835, 2020. ISSN 1573-269X. doi: 10.1007/s11071-020-05559-w. URL https://doi.org/10.1007/s11071-020-05559-w.

Rui Zhao, Ruiqin Xiong, Jian Zhang, Xinfeng Zhang, Zhaofei Yu, and Tiejun Huang. Optical flow for spike camera with hierarchical spatial-temporal spike fusion. In *AAAI*, pp. 7496–7504, 2024. URL https://doi.org/10.1609/aaai.v38i7.28581.

Chunyu Zhu, Shangqi Deng, Xuan Song, Yachao Li, and Qi Wang. Mamba collaborative implicit neural representation for hyperspectral and multispectral remote sensing image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025. doi: 10.1109/TGRS.2025.3537638.

A Appendix

A.1 Volume Rendering and SDF-based Surface Representation

Given a camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, NeRF samples N points along the ray and computes the pixel color $C(\mathbf{r})$ through discrete volume rendering:

$$\begin{cases} C(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i c_i, \\ T_i = \prod_{j<i} (1 - \alpha_j), \\ \alpha_i = 1 - \exp(-\sigma_i \delta_i) \end{cases} \tag{8}$$

Conventional NeRF (Mildenhall et al., 2021) represents scenes with an unconstrained volume density σ that lacks explicit perception of physical surfaces, frequently producing floater artifacts. To achieve high-fidelity geometric reconstruction, VolSDF (Yariv et al., 2021) replaces unconstrained density with an SDF $d(\mathbf{x})$, abbreviated d, and maps it to volume density via the Laplace Cumulative Distribution Function (CDF) Ψ_η:

$$\sigma(\mathbf{x}) = \tfrac{1}{\eta}\, \Psi_\eta\big(-d\big) = \begin{cases} \dfrac{1}{2\eta}\, e^{-d/\eta}, & d > 0, \\ \dfrac{1}{\eta}\big(1 - \tfrac{1}{2}\, e^{\,d/\eta}\big), & d \le 0, \end{cases} \tag{9}$$

where $\eta > 0$ is a learnable inverse temperature controlling surface sharpness. The main paper uses η for this Laplace scale and reserves β for the surrogate gradient steepness. Although SDFs provide rigorous physical constraints, the frequent floating-point computation of distance fields and spatial gradients $\nabla d(\mathbf{x})$ in Artificial Neural Networks (ANNs) becomes increasingly costly as sampling density grows, motivating the transition of SDF surface evolution into the energy-efficient spiking domain.

A.2 Ternary Spike Coding from Spik-NeRF

The binary spike representation $s \in \{0, 1\}$ of conventional Leaky Integrate-and-Fire (LIF) neurons has limited capacity, restricting the expressiveness of Spiking Neural Network (SNN)-based 3D implicit representation. Spik-NeRF (Wan et al., 2025) first introduced ternary spike coding into SNN-based NeRF and raised the information capacity to $\log_2 3 \approx 1.585$ bit through the asymmetric training space $s \in \{0, 1, 2\}$ together with the symmetric inference shift $\hat{s} \in \{-1, 0, 1\}$, alleviating quantization information loss in 3D implicit representation.

To preserve multiplication-free inference under this ternary encoding, Spik-NeRF further introduced a reparameterization that equivalently rewrites the linear transformation as:

$$Ws^t + b = W\hat{s}^t + (b + W\mathbf{1}) \tag{10}$$

where $W\mathbf{1}$ is a constant bias obtained by summing weights along the input channel dimension. Since $\hat{s}^t \in \{-1, 0, 1\}$, $W\hat{s}^t$ degenerates into a selective accumulation among $\{-w_j, 0, +w_j\}$ depending on $\hat{s}^t$, which eliminates element-wise multiplication in the spiking feature pipeline and leaves the compact geometry and appearance heads as its only multiplicative stage. The Ternary Spike Neuron (TSN) of Spik-NeRF nonetheless adopts a single fixed firing threshold and provides no surrogate gradient adapted to the two non-differentiable jumps of the ternary firing function. The remaining trainability and stability issues motivate the adaptive design of the Adaptive Ternary Spike Neuron (ATSN) presented in the main paper.

A.3 Directional Spiking State Space Scan

This section completes the definition of the scan used in Equation (1). For a direction dir acting on axis $a \in \{D, H, W\}$, mirrored concatenation forms $\tilde{\mathcal{V}} = \mathrm{Cat}_a(\mathcal{V}_{\mathrm{global}}, \mathrm{Flip}_a(\mathcal{V}_{\mathrm{global}}))$ for the positive direction and the reversed order for the negative one, so every sequence starts at the reference view side and ends at its mirror. The two depth sequences therefore order voxels by increasing and decreasing distance to the reference camera, which supplies an occlusion consistent front to back prior for the SDF branch. Row-major flattening gives a sequence of length $2N$ with $N = DHW$, and the skip operator $\Pi_{s,o}$ retains positions $o, o + s, \ldots$, so the scanned length is

$$L = \left\lceil \frac{2N - o}{s} \right\rceil, \qquad s = 2, \tag{11}$$

where the offset o advances by one at every forward pass and wraps modulo s, so all voxel positions are covered across consecutive iterations. The inverse operator $\Pi_{s,o}^{\top}$ writes each output back to its sampled index, the mirrored half is discarded, and the six directional grids are summed.

The state recurrence of Equation (1) is evaluated on blocks of M scanned positions. The state enters a block from the last position of the previous block and gradients are truncated at the boundary, which bounds activation memory by $O(M)$ per direction while preserving the front to back ordering across the whole sequence. The state is diagonal with one scalar per channel, so the state dimension equals the global branch width γC, in contrast to the expanded state of continuous state space models. The two dimensional block shares one pair $(\mathbf{W}_{\text{in}}, \mathbf{W}_{\text{out}})$ across its four in-plane directions, the three dimensional block uses six independent pairs, and the planar refinement module shares one pair within the horizontal pair of directions and another within the vertical pair.

Each direction costs $O(LC_hC)$ accumulations with $C_h = \gamma C$, so the six directions together cost $O(6LC_hC)$ and remain linear in the voxel count because L is proportional to N/s, whereas spiking self-attention over the same sequence costs $O(N^2)$. Writing the firing and reset of Equation (4) as a state gate gives $\alpha_k = \alpha\mathbf{1}[s_{k-1} = 0]$, which corresponds to

$$\alpha_k = \exp(-\Delta_k), \qquad \Delta_k = \begin{cases} -\log \alpha, & s_{k-1} = 0, \\ \infty, & s_{k-1} \neq 0, \end{cases} \tag{12}$$

so the ternary neuron realizes the binary limit of the input dependent decay of selective state space models, at no additional parameters, while SFA shifts the operating point of this gate through the threshold update of Equation (6).

A.4 System Pipeline and Component Definitions

A single reference view is encoded by a pure spiking image encoder into three ternary spike feature maps at low, middle, and high levels. Global average pooling of the encoder output feeds a style mapping network that produces a geometry style vector and an appearance style vector, which modulate the decoding heads through affine scaling and shifting and therefore adapt the shared architecture to the radiometric statistics of each scene.

The highest-level two dimensional feature passes through the four-directional spiking Mamba block and is projected into an 8^3 seed voxel by the feature align projector, which applies adaptive pooling, a 1×1 spiking convolution, depth broadcasting, and the learnable depth embedding of Equation (2). Ternary spiking upsampling then lifts the grid from 8^3 to 16^3 and to 32^3, and each level adds the projected encoder feature of the matching scale. Planar refinement operates at the intermediate resolutions after depth mean pooling, and the six-directional three dimensional block of Section 3.1 operates on the 32^3 grid before the grid is expanded toward 256^3.

Auxiliary Operators. The local branch of Section 3.1 applies an ATSN-based 3D spiking convolution, and a $1 \times 1 \times 1$ spiking convolution with an identity residual fuses the two branches:

$$\mathcal{F}_{\text{local}} = \text{ATSN}(\mathbf{W}_{\text{3D}} \circledast \mathcal{V}_{\text{local}}), \qquad \mathcal{V}_{\text{out}} = \text{ATSN}(\mathbf{W}_{\text{out}} \circledast [\mathcal{F}_{\text{global}}; \mathcal{F}_{\text{local}}]) + \mathcal{V}_{\text{in}}, \tag{13}$$

where $\circledast$ denotes 3D convolution. The threshold-dependent Layer Normalization of Section 3.2 rescales feature magnitudes to the firing threshold:

$$\text{tdLN}(\mathbf{x}) = (\alpha_{\text{LN}} \cdot V_{th}) \frac{\mathbf{x} - \mu}{\sqrt{\text{Var}(\mathbf{x}) + \epsilon}} \odot \mathbf{w} + \mathbf{b}, \tag{14}$$

where α_{LN} is a learnable coefficient that aligns the normalized output with V_{th}.

Rendering draws 96 coarse samples along every ray and then 320 fine samples during training and 192 fine samples at evaluation, guided by the coarse weights. Each sample reads the voxel grid by trilinear interpolation, concatenates a positional encoding of its coordinate, and passes through the geometry head, which outputs the signed distance that the Laplace CDF maps to density. The appearance head receives the geometry feature, the appearance style vector, and the encoded viewing direction, and outputs radiance, after which Equation (8) composites the samples into pixel color.

A.5 Additional Experimental Results

A.5.1 Per-Scene Results on the Remaining Scenes

Table 4 reports per-scene PSNR, SSIM, and LPIPS on the six scenes not shown in Table 1.

A.5.2 Measured Energy at Three Aggregation Levels

Across three aggregation levels DSMamba-NeRF consumes 7.0 times less energy per network query, 10.6 times less per camera ray, and 9.0 times less per rendered frame than the ANN baseline (Table 5). Counters in the inference pipeline record the multiply accumulate and accumulate operations of every module, and the neuromorphic convention of 3.2 pJ per MAC and 0.1 pJ per AC converts these counts into energy (Horowitz, 2014). The accounting spans the spiking feature pipeline, the geometry and appearance heads, the reference encoder, and the voxel decoder at a rendering resolution of 400×400. Each ray issues 128 network queries for DSMamba-NeRF and Spik-NeRF and 193 for the ANN baseline. Spik-NeRF is measured without its distillation teacher, which supervises training and stays out of the deployed model.

DSMamba-NeRF holds the lowest energy at every level. The spiking feature pipeline accounts for 0.022 μJ of the 0.252 μJ spent on one query, and the geometry and appearance heads carry the

Table 4: Per-scene PSNR (dB), SSIM, and LPIPS on the three synthetic and three real-world scenes not shown in Table 1. **Bold** marks the best SNN-based result.

		Synthetic			Real-world (LLFF)		
Metric	Method	Hotdog	Materials	Mic	Fern	Leaves	Fortress
PSNR↑	NeRF (ANN)	36.87	29.35	33.11	26.21	21.95	31.33
	Spiking NeRF	34.48	27.77	31.54	25.05	20.77	30.09
	Spik-NeRF	35.94	28.81	32.69	25.64	21.43	30.74
	DSMamba-NeRF	**36.15**	**28.96**	**32.85**	**25.87**	**21.62**	**30.91**
SSIM↑	NeRF (ANN)	0.978	0.959	0.979	0.834	0.787	0.894
	Spiking NeRF	0.964	0.933	0.962	0.767	0.694	0.837
	Spik-NeRF	0.972	0.947	0.973	0.796	0.740	0.865
	DSMamba-NeRF	**0.974**	**0.952**	**0.975**	**0.812**	**0.752**	**0.872**
LPIPS↓	NeRF (ANN)	0.015	0.024	0.023	0.104	0.117	0.056
	Spiking NeRF	0.033	0.045	0.045	0.211	0.202	0.134
	Spik-NeRF	0.024	0.036	0.030	0.167	0.155	0.089
	DSMamba-NeRF	**0.019**	**0.030**	**0.027**	**0.153**	**0.136**	**0.078**

Table 5: Measured energy at three aggregation levels. The geometry and appearance heads enter from the second row.

Level	DSMamba-NeRF	Spik-NeRF	NeRF (ANN)
Query, spiking pipeline	0.022 μJ	0.049 μJ	N/A
Query, all operations	0.252 μJ	0.727 μJ	1.77 μJ
Ray	32.3 μJ	93.1 μJ	342 μJ
Frame	6.07 J	14.9 J	54.7 J

remainder. The reference encoder and the voxel decoder contribute 1.36 J and are counted once for the whole image, so this term is independent of the number of rays and the margin widens with resolution, while the cost of the ANN baseline follows the number of sampled points. Table 6 splits the per query cost by operation type for each method. Multiplications are confined to the geometry and appearance heads. One query executes 0.22 M accumulate operations and 0.072 M multiply accumulate operations, so multiplications cover 24.6 percent of the operations and 91.3 percent of the energy, and at the frame level the heads contribute 4.71 J of the 6.07 J total. Spik-NeRF follows the same structure with 93.3 percent of its per query energy in its heads, which cost 2.9 times the heads of DSMamba-NeRF. Ternary spiking heads of equal width would lower the per query cost to 0.029 μJ and the frame cost to 1.51 J, which bounds the gain available from this direction.

A.5.3 Enlarged Qualitative Comparisons on NeRF Synthetic and LLFF

Figure 6 and Figure 7 provide enlarged, high-resolution versions of Figure 3, enabling detailed inspection of thin structures, boundaries, and floater artifacts. Figure 6 presents the NeRF Synthetic scenes and Figure 7 presents the Local Light Field Fusion (LLFF) scenes. Compared with existing SNN-based NeRF models, DSMamba-NeRF preserves sharper edges and boundaries while maintaining rendering quality comparable to ANN baselines. Spiking NeRF often shows pixel fragmentation, blur, and spatial offsets near thin structures, and Spik-NeRF benefits from ANN teacher distillation but still produces blurred or distorted results in overlapping regions.

A.5.4 Rendering Performance on the Mip-NeRF 360 Dataset

Figure 8 presents qualitative reconstruction and rendering results on the Mip-NeRF 360 dataset (Barron et al., 2022), using the more challenging *Bicycle*, *Garden*, *Kitchen*, and *Stump* scenes to test robustness under complex indoor and outdoor geometry. Consistent with the NeRF LLFF comparison above, DSMamba-NeRF preserves sharp edges and topological consistency while maintaining

Table 6: Per query energy split by operation type. The distillation teacher of Spik-NeRF supervises training and stays out of the deployed model.

Method	Spiking accumulate	Geometry and appearance heads	Auxiliary branch	Total
DSMamba-NeRF	0.022 μJ	0.230 μJ	none	0.252 μJ
Spik-NeRF	0.049 μJ	0.678 μJ	none	0.727 μJ
NeRF (ANN)	N/A	1.77 μJ	none	1.77 μJ

Table 7: Ablation of spike firing strategies on the *Trex* scene. Adaptive SFA lowers the spike rate and raises PSNR over fixed thresholds.

Neuron Type	Threshold Mechanism	Spike Rate (%)	PSNR ↑	LPIPS ↓
Binary SNN (fixed)	Fixed (V_{th})	12.4	24.32	0.1318
Ternary SNN (fixed)	Fixed (V_{th})	28.9	25.67	0.0910
+ ATSN (ours, no SRF/DSMamba)	Adaptive	21.7	26.04	0.0885
DSMamba-NeRF (full)	**Adaptive (SFA)**	**20.5**	**26.79**	**0.0830**

rendering quality comparable to ANN baselines, and both SNN baselines exhibit the same failure modes observed on LLFF.

A.5.5 Temporal Resolution, SFA, and SRF Ablations

This study evaluates how the temporal step T affects information capacity and inference latency. As shown in Figure 9(a) and Figure 9(b), $T = 1$ excessively compresses temporal information and degrades reconstruction quality. Increasing T beyond 2 provides only minor accuracy gains but increases computation nearly linearly. Figure 9(c) shows that $T = 8$ incurs 6.9× inference latency and improves PSNR by only 0.11 dB relative to $T = 2$. Therefore, DSMamba-NeRF adopts $T = 2$ to balance accuracy and inference efficiency.

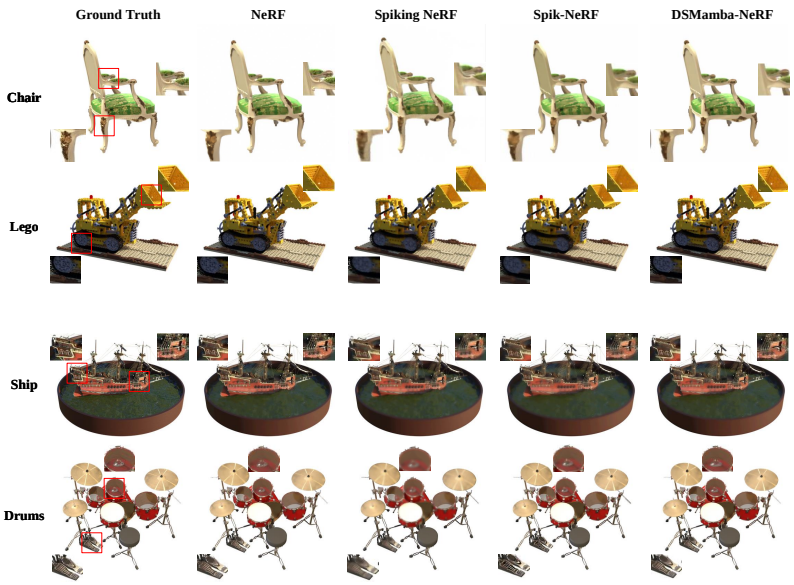


Figure 6: Enlarged qualitative comparison on the NeRF Synthetic dataset (*Chair*, *Lego*, *Ship*, *Drums*), a high-resolution version of the left panel of Figure 3.

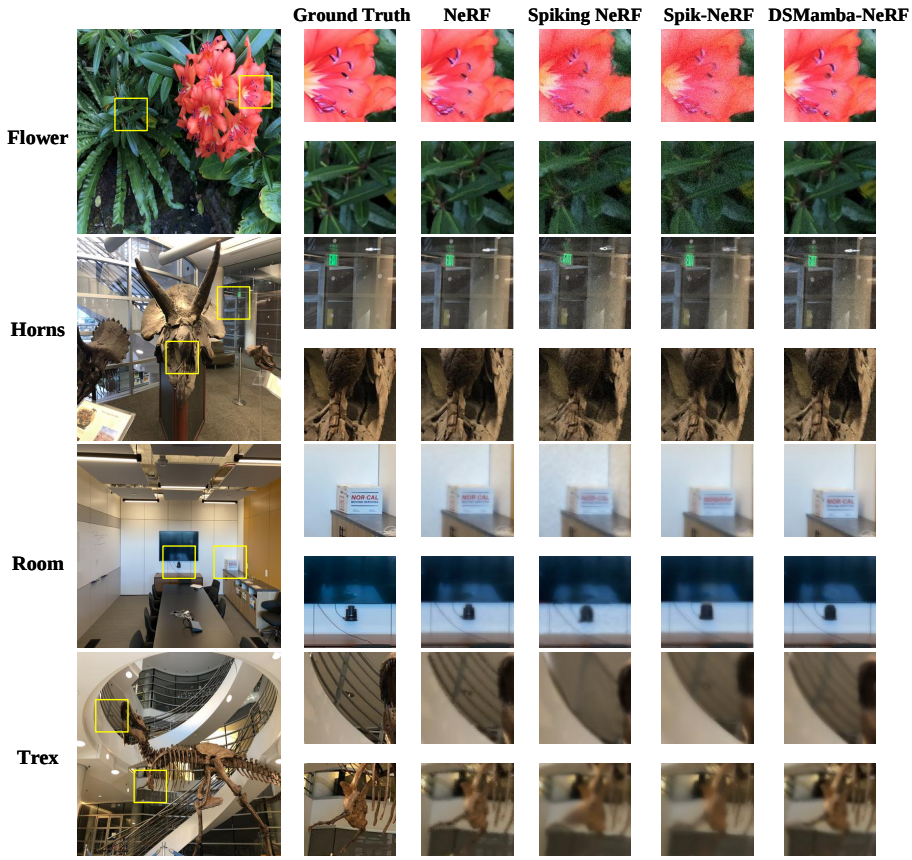


Figure 7: Enlarged qualitative comparison on the NeRF LLFF dataset (*Flower*, *Horns*, *Room*, *Trex*). The proposed method provides clear boundary definition.

Table 8: Ablation of SRF on the *Room* scene. Each SRF row adds one component to the Spiking NeRF baseline, and the last row further adds ATSN and DSMamba.

Configuration	SDF density	SDF-guided sampling	HF refinement	PSNR↑	SSIM↑
Baseline (Spiking NeRF)				30.16	0.905
+ Spiking VolSDF	✓			30.65	0.911
+ SDF-guided sampling	✓	✓		30.82	0.914
+ HF refinement (full SRF)	✓	✓	✓	30.94	0.919
DSMamba-NeRF (full)	✓	✓	✓	**31.07**	**0.922**

To evaluate Spike Firing Adaptation (SFA) in ATSN, this study keeps the baseline architecture fixed and varies only the neuron type and threshold mechanism, as shown in Table 7. Binary spikes have limited information capacity and yield lower PSNR. The fixed-threshold ternary baseline reaches a 28.9% spike rate but is prone to saturation in high-frequency regions under the wide dynamic lighting range of real-world scenes. ATSN uses adaptive SFA to regulate the membrane threshold and maintains a stable average firing rate of 21.7%, which is 7.2 percentage points lower than the fixed-threshold ternary baseline. With only the neuron replaced, ATSN improves PSNR by 0.37 dB and reduces LPIPS by 0.0025 over the fixed-threshold ternary baseline, showing that adaptive thresholding reduces gradient blocking and information saturation without ANN-to-SNN distillation. The additional 0.75 dB PSNR gain in full DSMamba-NeRF comes from the complementary effects of



Figure 8: Qualitative comparison on the Mip-NeRF 360 dataset. DSMamba-NeRF maintains topological consistency in complex outdoor scenes.

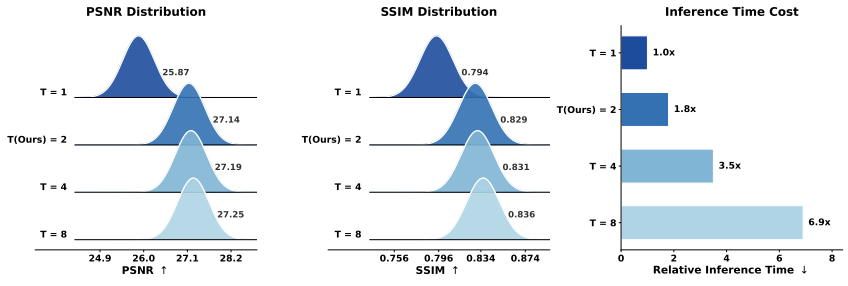


Figure 9: Ablation study of temporal resolution. Impact of time step T on reconstruction quality and computational efficiency on the *Flower* scene.

SRF and DSMamba, whose contributions are isolated in the ablation study of the main paper and in Table 8.

SRF is essential for suppressing floater artifacts and preserving fine surface details in DSMamba-NeRF. Table 8 isolates its three subcomponents on the *Room* scene. Replacing unconstrained density with SDF-induced density through spiking VolSDF yields the largest single gain, improving PSNR by 0.49 dB and confirming that explicit geometric constraints reduce floater artifacts. Adding SDF-guided hierarchical sampling further concentrates samples near the zero level set, improving PSNR by 0.17 dB and SSIM by 0.003. The high-frequency refinement branch projects spiking features

onto the SDF surface neighborhood and brings an additional gain of 0.12 dB PSNR and 0.005 SSIM, validating its ability to compensate for SDF smoothing near thin and sharp edges. Together, these SRF components improve PSNR by 0.78 dB and SSIM by 0.014 over the Spiking NeRF baseline. The remaining 0.13 dB PSNR gap to full DSMamba-NeRF is closed by ATSN and DSMamba, showing that ATSN, SRF, and DSMamba work together to raise both reconstruction accuracy and rendering quality.

A.5.6 Remaining Gaps and Failure Modes

The residual gap to the ANN baseline concentrates on scenes dominated by thin high-frequency geometry. On *Ficus* the gap reaches 2.16 dB PSNR and 0.032 LPIPS, the largest among the sixteen scenes, while the same scene shows the largest gain over Spik-NeRF at 0.92 dB, which places dense foliage and thin branches at the limit of ternary spiking features. *Horns* and *Trex* follow with gaps of 0.59 dB and 0.46 dB, where skeletal structures and specular background regions carry most of the residual error. The smallest gaps appear on *Chair* at 0.17 dB, *Orchids* at 0.18 dB, and *Drums* at 0.20 dB, all of which contain compact surfaces with moderate texture frequency. Two mechanisms explain this ordering. Ternary quantization bounds the dynamic range available to high-frequency residuals, and the SDF prior smooths geometry whose thickness approaches the voxel spacing, so scenes whose detail scale falls below one voxel retain the largest error.

A.5.7 Visual Analysis of Floater Artifact Suppression

To further verify the effectiveness of SRF in suppressing floater artifacts, this section performs fine-grained visual ablation on the *Mic* scene using rendered views, depth maps, structural maps, and error maps, as shown in Figure 4.

Existing SNN-based NeRF frameworks struggle with unconstrained continuous density fields, where discrete spike quantization amplifies noise and produces floater artifacts. As shown in Figure 4(A), diffuse noise surrounds the object. The depth map in Figure 4(C) shows an artifact halo in free space caused by unconstrained density integration under spiking jitter. The structural map in Figure 4(E) further shows noisy and uneven surfaces, with fine microscopic details such as the microphone honeycomb mesh largely missing. The error map in Figure 4(G) contains concentrated red contours, indicating that the lack of explicit geometric constraints reduces surface boundary accuracy.

SRF introduces spiking VolSDF to replace unconstrained density with an SDF-constrained representation. Since the SDF zero level set promotes topological continuity, SRF drives free space density toward zero and reduces invalid voxel interference. As shown in the bottom row, SRF suppresses quantization noise and improves surface convergence. The RGB rendering in Figure 4(B) is sharp and clean, while the depth map in Figure 4(D) removes floater artifacts and follows the object geometry. The structural map in Figure 4(F) preserves a smooth surface and the honeycomb pattern, and the reduced contour noise in Figure 4(H) further confirms that SRF mitigates density fitting errors and surface degradation.