SPIKEEEG-3D: SPIKING NEURAL NETWORKS FOR 3D VISUAL DECODING FROM EEG

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG-based 3D visual decoding is an important yet highly challenging task that aims to recover three-dimensional visual content from neural activity. Spiking neural networks (SNNs), with their brain-inspired and energy-efficient computation, have been increasingly studied for EEG decoding, yet their potential for EEG-to-3D visual decoding remains unexplored. In this work, we present SpikeEEG-3D, to the best of our knowledge, the first SNN framework for EEG-based 3D visual decoding, spanning the complete pipeline from EEG representation learning to colored point-cloud reconstruction. For EEG representation learning, SpikeEEG-3D employs Learnable Dynamics Leaky Integrate-and-Fire (LD-LIF) neurons, whose input gain and membrane time constant are jointly optimized to provide adaptive and diverse neuronal dynamics. We further propose Event-Enhanced Cross-Attention (EECA), which exploits spike-state transitions to facilitate information exchange between the static and dynamic EEG streams. For 3D reconstruction, we propose SpikeBit, which leverages bit-weighted virtual timesteps to provide finer-grained numerical representations while preserving spike-driven computation. Experiments on EEG-3D demonstrate excellent EEG classification performance and effective 3D reconstruction, with 20.8$\times$ and 3.1$\times$ lower energy consumption than Neuro-3D, respectively. These results demonstrate the feasibility of extending SNNs to EEG-based 3D visual decoding.

1 INTRODUCTION

Human visual perception transforms high-dimensional visual stimuli into distributed and time-varying neural activity Grill-Spector & Malach (2004); Gifford et al. (2022). Decoding such neural activity back into visual content can advance our understanding of how visual information is represented in the brain and support brain–computer interfaces Zhang et al. (2022). Electroencephalography (EEG) is particularly attractive for visual decoding due to its millisecond-level temporal resolution, portability, and ability to capture rapid neural dynamics Li et al. (2024); Song et al. (2024). Meanwhile, spiking neural networks (SNNs) provide a brain-inspired and energy-efficient computing paradigm Fang et al. (2021); Ye et al. (2026); Yao et al. (2025); Guo et al. (2024). Since EEG measures brain electrical activity while SNNs emulate event-driven neuronal computation, their combination has been increasingly explored for EEG decoding Kumar et al. (2022); Gong et al. (2023); Xu et al. (2024); Zhang et al. (2025). Recent advances have extended EEG-based visual decoding from 2D Singh et al. (2023); Zheng et al. (2026) to the more challenging 3D setting Guo et al. (2025); Yao et al. (2026). This raises a natural question: can SNNs be extended to EEG-to-3D visual decoding and effectively perform the task?

Existing EEG-to-3D visual decoding methods generally follow a two-stage paradigm of EEG representation learning and EEG-conditioned 3D reconstruction Guo et al. (2025); Jing et al. (2026). Extending SNNs to this task requires addressing the distinct demands of both stages. For EEG representation learning, 1) the neuronal dynamics of spiking neurons determine how temporal EEG information is integrated Kumar et al. (2022); Zhang et al. (2024). As illustrated in Fig. 1(a), different neuronal parameters can produce markedly different responses to the same EEG sequence. Directly adopting fixed parameters may constrain neuronal diversity and temporal representation learning. This is particularly limiting for EEG-to-3D decoding, which relies on informative temporal representations. Moreover, 2) a distinctive characteristic of EEG-to-3D decoding is the availability of both static and dynamic EEG responses, and jointly exploiting these two streams is important for

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.

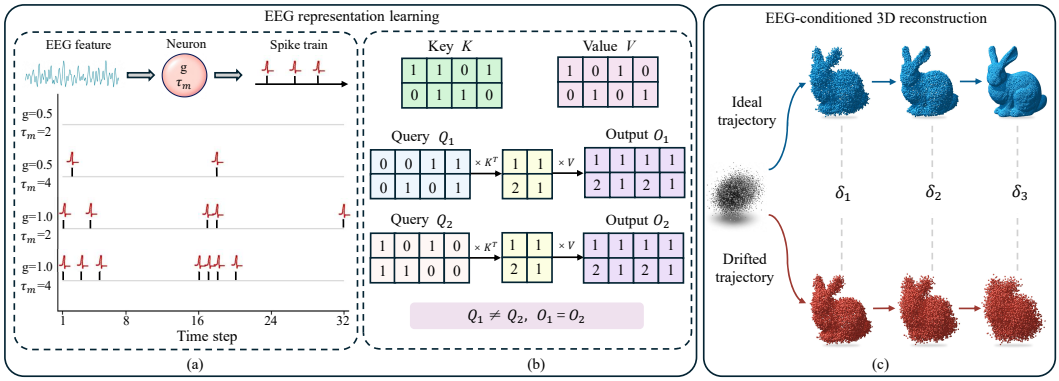


Figure 1: **Three challenges in extending SNNs to EEG-to-3D visual decoding.** (a) Different neuronal parameters lead to markedly different spike responses to the same EEG input. g and τ_m denote the input gain and membrane time constant, respectively. (b) In conventional spiking attention, different query patterns can produce identical outputs under the same key and value. (c) The firing of conventional spiking neurons can introduce approximation errors that may accumulate during iterative 3D reconstruction, causing trajectory drift.

capturing holistic object information and detailed 3D structure Guo et al. (2025); Yao et al. (2026). However, as illustrated in Fig. 1(b), in conventional spiking attention Zhou et al. (2022), different query patterns can produce identical interaction outputs under the same key and value, weakening cross-stream information exchange. For EEG-conditioned 3D reconstruction, 3) iterative 3D generation models repeatedly estimate updates over multiple steps Luo & Hu (2021); Zhou et al. (2021). However, conventional spiking neurons quantize continuous membrane potentials into binary spikes, which can introduce substantial approximation errors Guo et al. (2023); Luo et al. (2024). As illustrated in Fig. 1(c), such errors may propagate and accumulate throughout the iterative reconstruction process He et al. (2023), progressively deviating the generation trajectory from the desired result.

In this work, we propose SpikeEEG-3D, a unified spiking framework for EEG-based 3D visual decoding. As illustrated in Fig. 2, it consists of a Spiking EEG-Fusion Encoder for EEG representation learning and a Spiking Point Cloud Decoder for 3D reconstruction. The encoder extracts and fuses static and dynamic EEG features into geometry and appearance representations that condition shape and color generation in the decoder. For EEG representation learning, we first introduce Learnable Dynamics Leaky Integrate-and-Fire (LD-LIF) neurons, which jointly learn the input gain and membrane time constant to provide more task-adaptive and diverse neuronal dynamics. Furthermore, we introduce Event-Enhanced Cross-Attention (EECA) to strengthen the interaction between static and dynamic EEG streams. Inspired by the role of state transitions in neural dynamics Michel & Koenig (2018); Bair et al. (2002), EECA incorporates spike-state transitions as additional interaction cues, allowing differences in spike sequences to be better reflected during cross-stream fusion. For 3D reconstruction, we introduce the SpikeBit neuron, which leverages bit-weighted virtual timesteps to expand representation capacity. It further extends the representation into the negative domain with separate scaling factors for positive and negative responses. SpikeBit provides finer-grained numerical representations to reduce approximation errors during iterative reconstruction while retaining spike-driven computation. Our contributions are summarized as follows:

- We propose SpikeEEG-3D, to the best of our knowledge, the first SNN framework for EEG-based 3D visual decoding, covering EEG representation learning and EEG-conditioned 3D reconstruction.
- We introduce LD-LIF and EECA for EEG representation learning. LD-LIF jointly learns the input gain and membrane time constant to provide more task-adaptive and diverse neuronal dynamics, while EECA incorporates spike-state transitions to enhance interaction between static and dynamic EEG streams.
- We introduce the SpikeBit neuron for spiking 3D reconstruction. SpikeBit leverages bit-weighted virtual timesteps and extends the representation into the negative domain to reduce approximation errors while retaining spike-driven computation.

- SpikeEEG-3D achieves favorable classification performance and effective 3D reconstruction, demonstrating the feasibility of spike-driven EEG-to-3D visual decoding.

2 Related Work

2.1 EEG-based Visual Decoding

EEG-based visual decoding aims to infer perceived visual stimuli from scalp-recorded neural responses. Early studies mainly focus on visual category recognition Spampinato et al. (2017); Palazzo et al. (2020); Bagchi & Bathula (2022); Kumari et al. (2022), while recent works increasingly explore direct image reconstruction from EEG Singh et al. (2023); Bai et al. (2024); Li et al. (2024); Ferrante et al. (2024); Qian et al. (2024). EEG2IMAGE Singh et al. (2023) employs contrastive EEG representation learning and generative models for image reconstruction. DreamDiffusion Bai et al. (2024) and ATM Li et al. (2024) further leverage diffusion models and EEG–vision alignment to improve reconstruction quality. More recently, EEG visual decoding has been extended from 2D images to 3D objects. Neuro-3D Guo et al. (2025) introduces the EEG-3D benchmark and a diffusion-based framework for colored point-cloud reconstruction. BrainSSD Yao et al. (2026) explores static–dynamic EEG fusion for 3D perception, while BrainGS Jing et al. (2026) employs object-aware Gaussian Splatting for high-fidelity and view-consistent reconstruction. These advances highlight the rapid progress of artificial neural network (ANN)-based EEG visual decoding.

2.2 Spiking Neural Networks for EEG Decoding

SNNs have attracted increasing attention in EEG decoding due to their brain-inspired computation and energy efficiency. Early studies mainly focused on P300 detection and motor-imagery decoding Goel et al. (2008); Antelis et al. (2020), while later works have extended SNNs to broader EEG tasks. NeuroSense Tan et al. (2021) models spatio-temporal EEG spike patterns for short-term emotion recognition, and compact SNNs based on spectral EEG features have been developed for energy-efficient emotion classification Yan et al. (2022). EESNN Zhang et al. (2024) introduces a recurrent spiking convolutional architecture for cross-patient epileptic seizure detection while reducing energy consumption. SNNs combining LIF neurons, STDP learning, and neuromorphic devices have also been explored for SSVEP recognition Wang et al. (2023). Nevertheless, to our knowledge, SNNs have not yet been explored for the challenging task of EEG-based 3D visual decoding. The goal of this work is to develop an effective SNN-based framework for EEG-to-3D visual decoding.

3 Method

3.1 Overview

We propose SpikeEEG-3D, an SNN framework for reconstructing colored 3D objects from EEG signals. As illustrated in Fig. 2, SpikeEEG-3D consists of a Spiking EEG-Fusion Encoder and an EEG-conditioned Spiking Point Cloud Decoder. The encoder first extracts and integrates static and dynamic EEG representations, which are subsequently transformed into geometry and appearance features. These features serve as conditions for shape and color generation in the spiking decoder.

Spiking EEG-Fusion Encoder. Given static and dynamic EEG recordings $\mathbf{E}_{\mathrm{s}} \in \mathbb{R}^{T_{\mathrm{s}} \times C}$ and $\mathbf{E}_{\mathrm{d}} \in \mathbb{R}^{T_{\mathrm{d}} \times C}$, where C denotes the number of EEG channels and T_{s} and T_{d} denote the recording lengths, the Spiking EEG-Fusion Encoder first extracts stream-specific representations through two Spiking Transformer branches Zhou et al. (2022). Specifically, for each EEG stream $\mathbf{E}_b$, where $b \in \{\mathrm{s}, \mathrm{d}\}$ indexes the static and dynamic streams, the query $\mathbf{Q}_b$, key $\mathbf{K}_b$, and value $\mathbf{V}_b$ are generated through linear projections and spiking neurons, and then fed into spiking self-attention (SSA):

$$\mathrm{SSA}\left(\mathbf{Q}_b, \mathbf{K}_b, \mathbf{V}_b\right) = \mathrm{Linear}\left(\mathrm{SN}\left(\alpha\, \mathbf{Q}_b \mathbf{K}_b^{\top} \mathbf{V}_b\right)\right), \tag{1}$$

where α denotes the scaling factor and $\mathrm{SN}(\cdot)$ denotes the spiking neuron used in the encoder, whose LD-LIF formulation is detailed in Sec. 3.2. The attention output is subsequently processed by a spiking MLP block Zhou et al. (2022). Let $\mathbf{H}_s \in \mathbb{R}^{T_s \times C}$ and $\mathbf{H}_d \in \mathbb{R}^{T_d \times C}$ denote the feature representations produced by the static and dynamic Spiking Transformer branches, respectively.

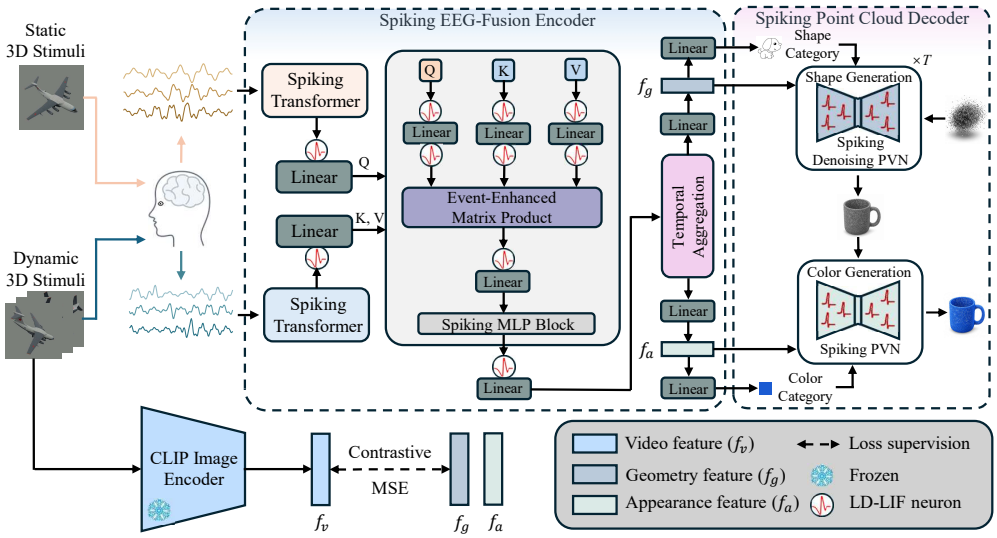


Figure 2: **Overview of SpikeEEG-3D for EEG-based 3D visual decoding.** Static and dynamic EEG signals are processed and fused by the Spiking EEG-Fusion Encoder, which learns geometry and appearance features under visual alignment and category supervision and predicts the corresponding shape and color. The geometry feature and predicted shape category guide shape generation, while reconstructed point cloud, appearance feature, and predicted color category are further guide color generation, producing the final colored point cloud in Spiking Point Cloud Decoder.

To enable subsequent cross-stream interaction, the two representations are transformed by spiking neurons and temporal linear projections to a common temporal length $\hat{T}$:

$$\mathbf{X}_s = \mathrm{Linear}_s(\mathrm{SN}(\mathbf{H}_s)), \qquad \mathbf{X}_d = \mathrm{Linear}_d(\mathrm{SN}(\mathbf{H}_d)), \tag{2}$$

where $\mathbf{X}_s, \mathbf{X}_d \in \mathbb{R}^{\hat{T}\times C}$. The static and dynamic representations $\mathbf{X}_s$ and $\mathbf{X}_d$ are then fed into the proposed Event-Enhanced Cross-Attention (EECA) for cross-stream interaction. EECA is detailed in Sec. 3.3. The EECA output is processed by a spiking MLP block and spike-activated linear transformations to obtain the fused representation $\mathbf{F} \in \mathbb{R}^{\hat{T}\times D_f}$. Then, we aggregate the temporal information in $\mathbf{F}$ using the mean, maximum, and standard deviation along the temporal dimension:

$$\mathbf{f}_{\mathrm{agg}} = [\mathrm{Mean}_t(\mathbf{F}); \mathrm{Max}_t(\mathbf{F}); \mathrm{Std}_t(\mathbf{F})], \tag{3}$$

where $\mathbf{f}_{\mathrm{agg}} \in \mathbb{R}^{3D_f}$. It is subsequently projected into geometry and appearance features $\mathbf{f}g$ and $\mathbf{f}a$:

$$\mathbf{f}_g = \mathrm{Linear}(\mathbf{f}_{\mathrm{agg}}), \qquad \mathbf{f}_a = \mathrm{Linear}(\mathbf{f}_{\mathrm{agg}}). \tag{4}$$

To supervise representation learning, both $\mathbf{f}_g$ and $\mathbf{f}_a$ are aligned with the visual representation $\mathbf{f}_v$ extracted by a frozen CLIP image encoder Radford et al. (2021) using contrastive and mean squared error (MSE) losses. Meanwhile, shape and color classifiers are applied to $\mathbf{f}_g$ and $\mathbf{f}_a$, producing predictions $\hat{\mathbf{y}}g$ and $\hat{\mathbf{y}}_a$, which are supervised by cross-entropy losses against the corresponding ground-truth categories. The overall representation learning objective is

$$\mathcal{L} = \sum_{\mathbf{f}\in\{\mathbf{f}_g,\mathbf{f}_a\}} [\lambda_{\mathrm{con}}\mathcal{L}_{\mathrm{con}}(\mathbf{f},\mathbf{f}_v) + \lambda_{\mathrm{mse}}\mathcal{L}_{\mathrm{mse}}(\mathbf{f},\mathbf{f}_v)] + \lambda_{\mathrm{cls}}\left[\mathrm{CE}(\hat{\mathbf{y}}_g, y_g) + \mathrm{CE}(\hat{\mathbf{y}}_a, y_a)\right], \tag{5}$$

where λ_{con}, λ_{mse}, and λ_{cls} are weighting coefficients. Subsequently, $\mathbf{f}_g$ and $\hat{\mathbf{y}}_g$ guide shape generation, while the reconstructed point cloud, $\mathbf{f}_a$, and $\hat{\mathbf{y}}_a$ are used for color generation.

Spiking Point Cloud Decoder. Given the learned geometry and appearance representations, the Spiking Point Cloud Decoder reconstructs the colored 3D object through sequential shape and color generation. We formulate shape generation as a conditional point-cloud diffusion process Guo et al. (2025). Let $\mathbf{P}_0 \in \mathbb{R}^{N\times 3}$ denote the target point cloud. The forward process gradually perturbs $\mathbf{P}_0$ with Gaussian noise:

$$\mathbf{P}_k = \sqrt{\bar{\alpha}_k}\mathbf{P}_0 + \sqrt{1-\bar{\alpha}_k}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{6}$$

where k is the diffusion step and $\bar{\alpha}_k$ follows the noise schedule. For the reverse process, we reformulate the Point-Voxel Network (PVN) Liu et al. (2019) as a Spiking Denoising PVN conditioned on $\mathbf{f}_g$ and $\hat{\mathbf{y}}_g$ to iteratively predict noise and reconstruct object geometry, supervised by an MSE loss against the ground-truth noise. Finally, the reconstructed point cloud, appearance feature $\mathbf{f}_a$, and color prediction $\hat{\mathbf{y}}_a$ are fed into a Spiking PVN to predict point-wise colors, supervised by an MSE loss against the ground-truth colors. Both networks adopt membrane-shortcut (MS) spiking architectures Hu et al. (2024) and the proposed SpikeBit neuron detailed in Sec. 3.4.

3.2 Learnable Dynamics Leaky Integrate-and-Fire Neuron

The neuronal dynamics of spiking neurons determine how temporal EEG information is integrated. Conventional neurons Nahmias et al. (2013) typically adopt preset neuronal parameters, which impose a fixed response configuration throughout representation learning. To provide more flexible neuronal dynamics for EEG-to-3D decoding, we introduce the LD-LIF neuron. Given an EEG-derived feature sequence $\mathbf{I} \in \mathbb{R}^{T\times D}$, where T and D denote the temporal length and feature dimension, respectively, the membrane potential of LD-LIF evolves as

$$u_t = \left(1 - \frac{1}{\tau_m}\right) h_{t-1} + gI_t, \tag{7}$$

where u_t is the membrane potential at time step t, h_{t-1} is the post-reset membrane potential from the preceding time step, τ_m denotes the membrane time constant, and g denotes the input gain. When u_t exceeds the firing threshold v_{th}, the neuron emits a spike and resets its membrane state:

$$s_t = \Theta(u_t - v_{\text{th}}), \qquad h_t = (1 - s_t)u_t + s_t u_{\text{reset}}, \tag{8}$$

where s_t is the output spike, $\Theta(\cdot)$ is the Heaviside function, and u_{reset} is the reset potential. Since $\Theta(\cdot)$ is non-differentiable, we use the ATan derivative as a surrogate gradient for backpropagation.

LD-LIF jointly learns the input gain g and membrane time constant τ_m. The input gain controls the contribution of the input to membrane accumulation, while the membrane time constant determines membrane-state retention. Jointly optimizing these parameters allows neuronal dynamics to better adapt to task-specific demands rather than following a preset integration rule, thereby enabling more flexible temporal representation learning from EEG. Compared with PLIF Fang et al. (2021), which learns only the membrane time constant, LD-LIF provides a broader learnable space for neuronal dynamics. LD-LIF is employed as the spiking neuron throughout the Spiking EEG-Fusion Encoder.

3.3 Event-Enhanced Cross-Attention

In conventional spiking attention, different query spike patterns can produce identical interaction outputs under the same key and value, making their differences insufficiently reflected during cross-stream interaction and thereby weakening information exchange between the two streams. To alleviate this limitation, we introduce Event-Enhanced Cross-Attention (EECA). Specifically, given the static and dynamic representations $\mathbf{X}_s, \mathbf{X}_d \in \mathbb{R}^{\hat{T}\times C}$, we construct the query from the static stream and the key and value from the dynamic stream:

$$\mathbf{Q} = \mathrm{SN}(\mathrm{Linear}_Q(\mathrm{SN}(\mathbf{X}_s))), \quad \mathbf{K} = \mathrm{SN}(\mathrm{Linear}_K(\mathrm{SN}(\mathbf{X}_d))), \quad \mathbf{V} = \mathrm{SN}(\mathrm{Linear}_V(\mathrm{SN}(\mathbf{X}_d))), \tag{9}$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \{0,1\}^{\hat{T}\times C}$. Inspired by the significant role of temporal state changes in neural activity Michel & Koenig (2018); Bair et al. (2002), EECA incorporates query-state transitions as additional interaction cues. Leveraging the binary nature of spike representations, these transitions can be naturally identified between adjacent time steps using an element-wise XOR operation:

$$\mathbf{e}_t = \begin{cases} \mathbf{0}, & t = 1, \\ \mathbf{Q}_t \oplus \mathbf{Q}_{t-1}, & t > 1, \end{cases} \tag{10}$$

where $\oplus$ denotes the element-wise XOR operation. The transition events are stacked as $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_{\hat{T}}]^\top$, and the EECA interaction is formulated as

$$\mathbf{Z} = \mathrm{Linear}\left(\mathrm{SN}\left(\alpha\, \mathbf{Q}\mathbf{K}^\top \left(\mathbf{V} + \mathbf{E} \odot \mathbf{V}\right)\right)\right), \tag{11}$$

where $\mathbf{Z}$ is the EECA output and $\odot$ denotes element-wise multiplication. For notational simplicity, $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$, and $\mathbf{E}$ are transposed in Eq. 11. We use query-state transitions as event cues because

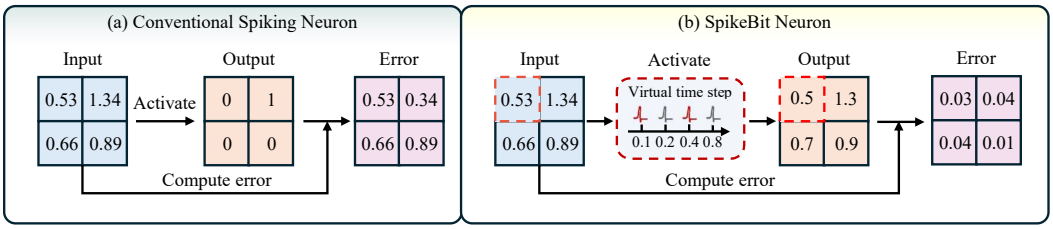


Figure 3: Illustration of the representations of (a) a conventional spiking neuron and (b) SpikeBit. SpikeBit leverages bit-weighted virtual timesteps to provide finer-grained outputs.

the query is derived from the relatively stable static EEG stream, making its state changes more indicative of meaningful shifts Guo et al. (2025). The term $\mathbf{E} \odot \mathbf{V}$ introduces a transition-guided modulation of the value representation, allowing differences in query spike sequences to be further reflected during cross-stream interaction and reducing the risk of indistinguishable interaction effects. This promotes information exchange between the static and dynamic EEG streams.

3.4 SpikeBit Neuron

Conventional spiking neurons quantize continuous membrane potentials into binary spikes, introducing substantial approximation errors Guo et al. (2023); Luo et al. (2024), as illustrated in Fig. 3(a). This is particularly problematic for iterative 3D reconstruction, where per-step errors may propagate and accumulate He et al. (2023). To reduce such approximation errors while retaining spike-driven computation, we introduce the SpikeBit neuron. As illustrated in Fig. 3, SpikeBit expands each neuronal timestep into B bit-weighted virtual timesteps, where each virtual timestep emits a binary spike corresponding to a different bit position. Given a membrane potential u, the spike emitted by the neuron at the r-th virtual timestep is defined as

$$b^{(r)}(u) = \left\lfloor \frac{\mathrm{round}\left[\mathrm{clip}\left(\frac{u}{\Delta}, 0, 2^B - 1\right)\right]}{2^r} \right\rfloor \bmod 2, \qquad r = 0, \ldots, B-1, \tag{12}$$

where $b^{(r)}(u) \in \{0, 1\}$ and Δ controls the numerical resolution. The virtual-timestep outputs are aggregated according to their bit weights: $\tilde{s}(u) = \Delta \sum_{r=0}^{B-1} 2^r b^{(r)}(u)$. Although each virtual timestep still communicates a binary spike, their bit-weighted combination provides richer numerical representations, thereby reducing the approximation error associated with binary quantization. To further improve numerical representation, we extend SpikeBit to the negative domain and introduce separate scaling factors for positive and negative responses. The final SpikeBit output is defined as

$$s(u) = \begin{cases} \gamma_+ \Delta \displaystyle\sum_{r=0}^{B-1} 2^r b^{(r)}(u), & u \geq 0, \\ -\gamma_- \Delta \displaystyle\sum_{r=0}^{B-1} 2^r b^{(r)}(-u), & u < 0, \end{cases} \tag{13}$$

where γ_+ and γ_- are scaling factors. During training, virtual-timestep outputs are aggregated and a straight-through estimator (STE) is used for backpropagation. During inference, the representation is unfolded into binary spikes across virtual timesteps, preserving spike-driven computation.

4 Experiments

4.1 Experimental Setup

We conduct experiments on the EEG-3D dataset Guo et al. (2025) and follow its official training and test split. We evaluate SpikeEEG-3D on EEG classification and 3D reconstruction tasks. For EEG classification, we report Top-1 and Top-5 accuracy for 72-way object classification and Top-1 and Top-2 accuracy for 6-way color classification. For 3D reconstruction, we adopt N-way Top-K accuracy to evaluate the semantic fidelity of reconstructed objects. Specifically, we use a point-cloud

Table 1: Comparison of EEG classification performance and energy consumption on the EEG-3D dataset. The best and second-best results are highlighted in bold and underline, respectively.

Method	Object Type		Color Type		Energy (mJ) ↓
	Top-1	Top-5	Top-1	Top-2	
Chance Level	1.39	6.94	16.67	33.33	–
DeepNet Schirrmeister et al. (2017)	3.70	9.90	20.95	49.71	0.241
EEGNetv4 Lawhern et al. (2018)	3.94	11.28	24.13	48.44	<u>0.133</u>
TSConv Song et al. (2024)	4.05	10.13	31.13	59.49	0.306
EEG Conformer Song et al. (2022)	4.05	10.30	18.27	35.81	0.643
ShallowFBCSPNet Schirrmeister et al. (2017)	4.46	11.11	17.19	34.26	0.236
Neuro-3D Guo et al. (2025)	5.91	<u>16.30</u>	39.93	61.40	2.304
BrainAlign Shi et al. (2025)	<u>6.11</u>	16.25	<u>40.70</u>	<u>63.61</u>	1.472
SpikeEEG-3D	**6.42**	**16.61**	**43.00**	**65.57**	**0.111**

Table 2: Quantitative comparison of 3D reconstruction performance and energy consumption on the EEG-3D dataset under different inference timesteps.

Method	Inference Timestep	$(2,1)\uparrow$	$(10,3)\uparrow$	CD ↓	F1 ↑	SSIM ↑	Energy (J) ↓
Neuro-3D Guo et al. (2025)	–	55.72	36.58	**5.954**	**72.725**	<u>0.7085</u>	327.798
SpikeEEG-3D w/ LIF	4	52.14	31.48	6.773	68.324	0.6572	110.415
SpikeEEG-3D w/ LIF	8	52.63	32.39	6.604	68.487	0.6503	173.406
SpikeEEG-3D w/ SpikeBit	4	<u>57.18</u>	<u>38.53</u>	6.199	68.842	0.6906	**102.098**
SpikeEEG-3D w/ SpikeBit	8	**58.58**	**40.70**	<u>6.028</u>	<u>70.916</u>	**0.7189**	<u>105.609</u>

classifier trained on Objaverse-derived data for object-category prediction Deitke et al. (2023). We report $(2,1)$ and $(10,3)$, denoting 2-way Top-1 and 10-way Top-3 accuracy, respectively. We further use Chamfer Distance (CD, $\times 10^{-2}$), F1 score (F1, %), and Structural Similarity Index Measure (SSIM) to evaluate reconstruction quality, and report energy consumption for computational efficiency. Detailed dataset information, implementation settings, metric calculations, spike-driven inference mechanisms, and energy consumption calculations are provided in Appendix A.1–A.3.

4.2 Main Results

Classification Tasks. We first evaluate the discriminative capability and energy efficiency of the learned EEG representations on object and color classification. As shown in Table 1, SpikeEEG-3D achieves the best performance across all four metrics, reaching 6.42%/16.61% Top-1/Top-5 accuracy for object classification and 43.00%/65.57% Top-1/Top-2 accuracy for color classification. It surpasses Neuro-3D Guo et al. (2025) by 3.07% in Color Top-1 accuracy. Moreover, SpikeEEG-3D achieves the lowest energy consumption at only 0.111 mJ, approximately 20.8× lower than Neuro-3D. We further provide t-SNE visualizations in Appendix A.4, showing more discriminative representations than the ANN-based method. These results demonstrate that SpikeEEG-3D learns discriminative EEG representations with high energy efficiency.

3D Reconstruction Tasks. We evaluate 3D reconstruction in Table 2. Using conventional LIF neurons in our model causes clear degradation across all metrics, whereas using SpikeBit substantially improves reconstruction. With 8 inference timesteps, SpikeEEG-3D achieves 58.58% and 40.70% on the $(2,1)$ and $(10,3)$ metrics, outperforming Neuro-3D Guo et al. (2025) by 2.86% and 4.12%, respectively. It also achieves the best SSIM of 0.7189 while maintaining CD and F1 close to Neuro-3D. Moreover, SpikeBit consumes only 105.609 J, reducing energy by 39.1% over LIF and 67.8% over Neuro-3D. As shown in Fig. 4, SpikeEEG-3D produces recognizable reconstructions across diverse object categories, recovering overall geometry for simple objects and preserving major structures for more complex shapes. The reconstructed point clouds also capture the dominant colors of the target objects. These results show that SpikeEEG-3D can achieve effective 3D visual reconstruction from EEG signals while retaining spike-driven computation.



Figure 4: **Qualitative 3D reconstruction results of SpikeEEG-3D.** Ground truth is shown in the top row, and the corresponding SpikeEEG-3D reconstructions are shown in the bottom row.

Table 3: Ablation study of LD-LIF and EECA on EEG classification, with LIF and standard spiking cross-attention as their respective baselines.

LD-LIF	EECA	Object Type		Color Type	
		Top-1	Top-5	Top-1	Top-2
		4.75	14.24	37.50	57.41
✓		5.50	**16.61**	42.02	63.83
	✓	5.90	16.38	41.84	63.48
✓	✓	**6.42**	**16.61**	**43.00**	**65.57**

Table 4: Ablation study of the learnable input gain g and membrane time constant τ_m in LD-LIF, without using EECA.

g	τ_m	Object Type		Color Type	
		Top-1	Top-5	Top-1	Top-2
		4.75	14.24	37.50	57.41
✓		5.27	14.53	39.82	58.10
	✓	5.15	15.51	39.99	61.29
✓	✓	**5.50**	**16.61**	**42.02**	**63.83**

Table 5: Ablation study of different enhancement cues in EECA. LD-LIF is used for all variants.

Enhancement Type	Cue Definition	Object Type		Color Type	
		Top-1	Top-5	Top-1	Top-2
None	–	5.50	**16.61**	42.02	63.83
Sustained Activation	$\mathbf{Q}_t \wedge \mathbf{Q}_{t-1}$	5.79	16.38	42.19	64.24
Transition + Sustained	$(\mathbf{Q}_t \oplus \mathbf{Q}_{t-1}) + (\mathbf{Q}_t \wedge \mathbf{Q}_{t-1})$	5.32	16.03	42.31	64.00
State Transition	$\mathbf{Q}_t \oplus \mathbf{Q}_{t-1}$	**6.42**	**16.61**	**43.00**	**65.57**

4.3 Ablation Studies

Effectiveness of LD-LIF and EECA. We evaluate LD-LIF and EECA on EEG classification in Table 3. Either component improves all four metrics over the baseline, while combining both achieves the best overall performance. Compared with the baseline, the complete model improves object Top-1/Top-5 accuracy by 1.67%/2.37% and color Top-1/Top-2 accuracy by 5.50%/8.16%, respectively. These results validate the effectiveness of both designs for EEG representation learning.

Effect of Learnable Neuronal Dynamics. We investigate the effects of the learnable input gain g and membrane time constant τ_m in Table 4. Learning either parameter individually improves performance over the fixed-parameter baseline, while jointly learning both achieves the best results across all four metrics. These results show that jointly adapting g and τ_m provides more flexible neuronal dynamics and improves EEG classification performance.

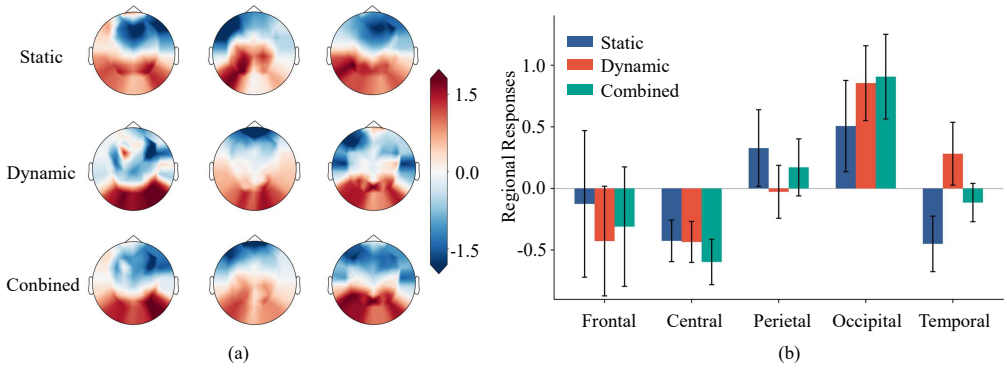


Figure 5: Spatial response analysis of SpikeEEG-3D. (a) EEG topographies under static, dynamic, and combined conditions for representative subjects. (b) Responses across five scalp regions, including the frontal, central, parietal, occipital, and temporal regions, under the three conditions.

Effect of State Transitions. We further investigate the effects of different enhancement cues in Table 5. Sustained activation provides only limited improvement over the setting without enhancement cues, while combining it with state transitions provides no further improvement. In contrast, state transitions achieve the best overall performance, improving Object Top-1 by 0.92% and Color Top-1/Top-2 by 0.98%/1.74% over the baseline. These results show that state transitions provide more effective cues for cross-stream interaction.

4.4 Brain-Region Response Analysis

To investigate the spatial response patterns captured by SpikeEEG-3D, we analyze EEG topographies under static, dynamic, and combined conditions using Class Activation Topography Song et al. (2022). As shown in Fig. 5(a), despite inter-subject variability, subjects exhibit a common posterior-dominant pattern, with stronger responses over parietal and occipital regions. This is consistent with the established role of posterior visual areas, particularly the occipital cortex, in visual processing and object representation Grill-Spector & Malach (2004); Malach et al. (1995).

We further examine responses across five scalp regions in Fig. 5(b). The occipital region shows the strongest positive response across all conditions, further reinforcing the link between the observed spatial pattern and visual processing in the occipital cortex Grill-Spector et al. (2001); Kourtzi & Kanwisher (2001). Beyond this common occipital response, distinct regional patterns emerge across stimulus conditions. Parietal responses, particularly under static and combined conditions, are consistent with the role of the dorsal visual cortex in visuospatial processing and 3D shape representation Erlikhman et al. (2018); Shikata et al. (2001). Notably, the temporal response becomes more positive under dynamic stimulation, which may reflect motion-sensitive occipito-temporal mechanisms involved in visual motion and 3D structure-from-motion processing Sarkheil et al. (2008); Orban et al. (1999). These patterns align with the functional organization of human visual processing, suggesting that SpikeEEG-3D captures neurophysiologically meaningful spatial information.

5 Conclusion

In this work, we presented SpikeEEG-3D, to the best of our knowledge, the first SNN framework for EEG-based 3D visual decoding, covering EEG representation learning and colored point-cloud reconstruction. To address the task-specific challenges, we introduced LD-LIF to learn adaptive neuronal dynamics, EECA to enhance information exchange between static and dynamic EEG streams through spike-state transitions, and SpikeBit for finer-grained numerical representations in iterative 3D reconstruction while retaining spike-driven computation. Experiments on EEG-3D demonstrate strong EEG classification performance, effective 3D reconstruction, and favorable energy efficiency. Brain-region response analysis further reveals spatial patterns consistent with the functional organization of human visual processing. Overall, our results demonstrate the effectiveness of extending SNNs to EEG-based 3D visual decoding and support more efficient neural visual decoding.

AI USE STATEMENT

In this work, we used generative AI tools solely for language polishing, including improving writing clarity and grammar. All AI-assisted edits were reviewed and revised by the authors. We take full responsibility for the final content of this work, including all text, claims, and results.

REFERENCES

Javier M Antelis, Luis E Falcón, et al. Spiking neural networks applied to the classification of motor tasks in eeg signals. *Neural networks*, 122:130–143, 2020.

Subhranil Bagchi and Deepti R Bathula. Eeg-convtransformer for single-trial eeg-based visual stimulus classification. *Pattern Recognition*, 129:108757, 2022.

Yunpeng Bai, Xintao Wang, Yan-Pei Cao, Yixiao Ge, Chun Yuan, and Ying Shan. Dreamdiffusion: High-quality eeg-to-image generation with temporal masked signal modeling and clip alignment. In *European Conference on Computer Vision*, pp. 472–488. Springer, 2024.

Wyeth Bair, James R Cavanaugh, Matthew A Smith, and J Anthony Movshon. The timing of response onset and offset in macaque visual neurons. *The Journal of neuroscience*, 22(8):3189–3205, 2002.

Mike Davies, Narayan Srinivasa, Tsung-Han Lin, Gautham Chinya, Yongqiang Cao, Sri Harsha Choday, Georgios Dimou, Prasad Joshi, Nabil Imam, Shweta Jain, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *Ieee micro*, 38(1):82–99, 2018.

Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13142–13153, 2023.

Gennady Erlikhman, Gideon P Caplovitz, Gennadiy Gurariy, Jared Medina, and Jacqueline C Snow. Towards a unified perspective of object shape and motion processing in human dorsal cortex. *Consciousness and cognition*, 64:106–120, 2018.

Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2641–2651. IEEE, 2021.

Matteo Ferrante, Tommaso Boccato, Stefano Bargione, and Nicola Toschi. Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models. *Computers in Biology and Medicine*, 178:108701, 2024.

Alessandro T Gifford, Kshitij Dwivedi, Gemma Roig, and Radoslaw M Cichy. A large and rich eeg dataset for modeling human visual object recognition. *NeuroImage*, 264:119754, 2022.

Piyush Goel, Honghai Liu, David Brown, and Avijit Datta. On the use of spiking neural network for eeg classification. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 12(4):295–304, 2008.

Peiliang Gong, Pengpai Wang, Yueying Zhou, and Daoqiang Zhang. A spiking neural network with adaptive graph convolution and lstm for eeg-based brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1440–1450, 2023.

Kalanit Grill-Spector and Rafael Malach. The human visual cortex. *Annu. Rev. Neurosci.*, 27(1): 649–677, 2004.

Kalanit Grill-Spector, Zoe Kourtzi, and Nancy Kanwisher. The lateral occipital complex and its role in object recognition. *Vision research*, 41(10-11):1409–1422, 2001.

Yufei Guo, Xiaode Liu, Yuanpei Chen, Liwen Zhang, Weihang Peng, Yuhan Zhang, Xuhui Huang, and Zhe Ma. Rmp-loss: Regularizing membrane potential distribution for spiking neural networks. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17345–17355. IEEE, 2023.

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. *Advances in Neural Information Processing Systems*, 37:51708–51726, 2024.

Zhanqiang Guo, Jiamin Wu, Yonghao Song, Jiahui Bu, Weijian Mai, Qihao Zheng, Wanli Ouyang, and Chunfeng Song. Neuro-3d: Towards 3d visual decoding from eeg signals. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23870–23880. IEEE, 2025.

Yefei He, Luping Liu, Jing Liu, Weijia Wu, Hong Zhou, and Bohan Zhuang. Ptqd: Accurate post-training quantization for diffusion models. *Advances in neural information processing systems*, 36:13237–13249, 2023.

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. IEEE, 2014.

Yifan Hu, Lei Deng, Yujie Wu, Man Yao, and Guoqi Li. Advancing spiking neural networks toward deep residual learning. *IEEE transactions on neural networks and learning systems*, 36(2):2353–2367, 2024.

Haodong Jing, Dongyao Jiang, Jixin Wang, Junhao Jia, Yanshu Li, Yongqiang Ma, and Nanning Zheng. More natural, more real: Object-aware gaussian splatting for 3d visual decoding from human brain. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19033–19044, 2026.

Zoe Kourtzi and Nancy Kanwisher. Representation of perceived object shape by the human lateral occipital complex. *Science*, 293(5534):1506–1509, 2001.

Neelesh Kumar, Guangzhi Tang, Raymond Yoo, and Konstantinos P Michmizos. Decoding eeg with spiking neural networks on neuromorphic hardware. *Transactions on Machine Learning Research*, 2022.

Nandini Kumari, Shamama Anwar, and Vandana Bhattacharjee. Automated visual stimuli evoked multi-channel eeg signal classification using eegcapsnet. *Pattern Recognition Letters*, 153:29–35, 2022.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Dongyang Li, Chen Wei, Shiying Li, Jiachen Zou, Haoyang Qin, and Quanying Liu. Visual decoding and reconstruction via eeg embeddings with guided diffusion. *arXiv preprint arXiv:2403.07721*, 2024.

Zhijian Liu, Haotian Tang, Yujun Lin, and Song Han. Point-voxel cnn for efficient 3d deep learning. *Advances in neural information processing systems*, 32, 2019.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2836–2844. IEEE, 2021.

Xinhao Luo, Man Yao, Yuhong Chou, Bo Xu, and Guoqi Li. Integer-valued training and spike-driven inference spiking neural network for high-performance and energy-efficient object detection. In *European Conference on Computer Vision*, pp. 253–272. Springer, 2024.

Rafael Malach, JB Reppas, RR Benson, KK Kwong, H Jiang, WA Kennedy, PJ Ledden, TJ Brady, BR Rosen, and RB Tootell. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18): 8135–8139, 1995.

Christoph M Michel and Thomas Koenig. Eeg microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage*, 180:577–593, 2018.

Mitchell A Nahmias, Bhavin J Shastri, Alexander N Tait, and Paul R Prucnal. A leaky integrate-and-fire laser neuron for ultrafast cognitive computing. *IEEE journal of selected topics in quantum electronics*, 19(5):1–12, 2013.

Guy A Orban, Stefan Sunaert, James T Todd, Paul Van Hecke, and Guy Marchal. Human cortical regions involved in extracting depth from motion. *Neuron*, 24(4):929–940, 1999.

Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and Mubarak Shah. Decoding brain representations by multimodal learning of neural activity and visual features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3833–3849, 2020.

Jing Pei, Lei Deng, Sen Song, Mingguo Zhao, Youhui Zhang, Shuang Wu, Guanrui Wang, Zhe Zou, Zhenzhi Wu, Wei He, et al. Towards artificial general intelligence with hybrid tianjic chip architecture. *Nature*, 572(7767):106–111, 2019.

Dongguan Qian, Hong Zeng, Wenjie Cheng, Yu Liu, Taha Bikki, and Jianjiang Pan. Neurodm: Decoding and visualizing human brain activity with eeg-guided diffusion model. *Computer Methods and Programs in Biomedicine*, 251:108213, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Pegah Sarkheil, Quoc C Vuong, Heinrich H Bülthoff, and Uta Noppeney. The integration of higher order form and motion by the human brain. *Neuroimage*, 42(4):1529–1536, 2008.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

Enze Shi, Huawen Hu, Qilong Yuan, Kui Zhao, Sigang Yu, and Shu Zhang. Brainalign: Eeg-vision alignment via frequency-aware temporal encoder and differentiable cluster assigner. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 98–108. Springer, 2025.

Elisa Shikata, Farsin Hamzei, Volkmar Glauche, René Knab, Christian Dettmers, Cornelius Weiller, and Christian Büchel. Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fmri study. *Journal of Neurophysiology*, 85(3):1309–1314, 2001.

Prajwal Singh, Pankaj Pandey, Krishna Miyapuram, and Shanmuganathan Raman. Eeg2image: image reconstruction from eeg brain signals. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022.

Yonghao Song, Bingchuan Liu, Xiang Li, Nanlin Shi, Yijun Wang, and Xiaorong Gao. Decoding natural images from eeg for object recognition. In *International conference on learning representations*, volume 2024, pp. 47648–47665, 2024.

Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. Deep learning human mind for automated visual classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4503–4511. IEEE, 2017.

Clarence Tan, Marko Šarlija, and Nikola Kasabov. Neurosense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal eeg patterns. *Neurocomputing*, 434:137–148, 2021.

Qinan Wang, Chun Zhao, Yi Sun, Rongxuan Xu, Chenran Li, Chengbo Wang, Wen Liu, Jiangmin Gu, Yingli Shi, Li Yang, et al. Synaptic transistor with multiple biological functions based on metal-organic frameworks combined with the lif model of a spiking neural network to recognize temporal information. *Microsystems & nanoengineering*, 9(1):96, 2023.

FeiFan Xu, Deng Pan, Haohao Zheng, Yu Ouyang, Zhe Jia, and Hong Zeng. Eescn: A novel spiking neural network method for eeg-based emotion recognition. *Computer methods and programs in biomedicine*, 243:107927, 2024.

Zhanglu Yan, Jun Zhou, and Weng-Fai Wong. Eeg classification with spiking neural network: Smaller, better, more energy efficient. *Smart Health*, 24:100261, 2022.

Zheyu Yang, Taoyi Wang, Yihan Lin, Yuguo Chen, Hui Zeng, Jing Pei, Jiazheng Wang, Xue Liu, Yichun Zhou, Jianqiang Zhang, et al. A vision chip with complementary pathways for open-world sensing. *Nature*, 629(8014):1027–1033, 2024.

Man Yao, Xuerui Qiu, Tianxiang Hu, Jiakui Hu, Yuhong Chou, Keyu Tian, Jianxing Liao, Luziwei Leng, Bo Xu, and Guoqi Li. Scaling spike-driven transformer with efficient spike firing approximation training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Yincheng Yao, Enze Shi, and Shu Zhang. Decoding 3d perception via brainssd: Synergistic fusion of eeg representations from static and dynamic visual streams. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 42711–42721, 2026.

Binghao Ye, Wenjuan Li, Dengfeng Xue, Bing Li, Weiming Hu, Dong Liang, and Kun Shang. Resolving the timestep scaling paradox in spiking neural networks with a timestep-scalable neuron model. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=U63l7Uyeel.

Herui Zhang, Haoran Wang, Jiayu An, Shitao Zheng, and Dongrui Wu. A lightweight spiking neural network for eeg-based motor imagery classification. *Neural Networks*, 191:107741, 2025.

Yi-Jun Zhang, Zhao-Fei Yu, Jian K Liu, and Tie-Jun Huang. Neural decoding of visual information across different neural recording modalities and approaches. *Machine Intelligence Research*, 19 (5):350–365, 2022.

Zongpeng Zhang, Mingqing Xiao, Taoyun Ji, Yuwu Jiang, Tong Lin, Xiaohua Zhou, and Zhouchen Lin. Efficient and generalizable cross-patient epileptic seizure detection through a spiking neural network. *Frontiers in Neuroscience*, 17:1303564, 2024.

Jiawen Zheng, Haonan Jia, MING LI, Yuhui Zheng, Yufeng Zeng, Yang Gao, and Chen Liang. Learning brain representation with hierarchical visual embeddings. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=IEq71qS8B7.

Chenlin Zhou, Liutao Yu, Zhaokun Zhou, Zhengyu Ma, Han Zhang, Huihui Zhou, and Yonghong Tian. Spikingformer: Spike-driven residual learning for transformer-based spiking neural network. *arXiv preprint arXiv:2304.11954*, 2023.

Chenlin Zhou, Han Zhang, Zhaokun Zhou, Liutao Yu, Liwei Huang, Xiaopeng Fan, Li Yuan, Zhengyu Ma, Huihui Zhou, and Yonghong Tian. Qkformer: Hierarchical spiking transformer using qk attention. *Advances in Neural Information Processing Systems*, 37:13074–13098, 2024.

Linqi Zhou, Yilun Du, and Jiajun Wu. 3d shape generation and completion through point-voxel diffusion. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5806–5815. IEEE, 2021.

Zhaokun Zhou, Yuesheng Zhu, Chao He, Yaowei Wang, Shuicheng Yan, Yonghong Tian, and Li Yuan. Spikformer: When spiking neural network meets transformer. *arXiv preprint arXiv:2209.15425*, 2022.

A Appendix

A.1 Dataset, Implementation, and Evaluation Metrics

Dataset. To the best of our knowledge, EEG-3D is currently the only publicly available dataset specifically designed for EEG-based 3D visual decoding. The dataset contains EEG recordings from 12 subjects viewing 72 categories of 3D objects. Each category contains 10 objects, of which 8 are used for training and 2 for testing. The objects are further grouped into 6 color categories according to their dominant color styles. For each object, EEG responses are recorded under both static-image and dynamic-video stimuli. The static stimulus lasts 0.5 s, while the dynamic stimulus consists of a 6 s rotating video providing multi-view observations of the object. EEG signals are recorded using 64 channels at a sampling rate of 1000 Hz. We follow the official training and test split of EEG-3D.

Implementation Details. Our implementation is based on PyTorch and runs on a system equipped with a 56-core Intel Xeon Gold 6330 CPU at 2.00 GHz and four NVIDIA GeForce RTX 4090 GPUs. We use the AdamW optimizer Loshchilov & Hutter (2017) with $\beta = (0.95, 0.999)$. For EEG discriminative tasks, we use an initial learning rate of 1×10^{-2}, as a relatively large learning rate facilitates discriminative feature learning in directly trained SNNs Zhou et al. (2023; 2024). For 3D reconstruction, the initial learning rate is set to 1×10^{-3}. The batch size is set to 48. Each video sequence is downsampled to $n = 4$ frames for feature extraction. For visual alignment, these four frames are individually encoded by a frozen CLIP image encoder, and the resulting features are averaged to obtain the visual representation $\mathbf{f}_v$. Each reconstructed point cloud contains $N = 8192$ points. Since 3D reconstruction does not involve an inherent temporal sequence, SpikeBit uses only one neuronal timestep and expands it into multiple virtual timesteps for inference. For conventional LIF neurons, the input is repeatedly presented over multiple neuronal timesteps, with the number of repeated timesteps serving as the inference timestep. For LD-LIF, we initialize $\tau_m = 2$ and $g = 1$. When either parameter is fixed in the ablation studies, it is kept at the same default value. For SpikeBit, we set $\Delta = 0.1$, $\gamma_+ = 1$, and $\gamma_- = 0.1$. Our experiments are conducted with three random seeds, and the results are averaged across them.

A.2 Evaluation Metric Calculations

Before computing geometric metrics, the predicted and ground-truth point clouds are independently centered and normalized to the unit sphere. Given a point cloud $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, we normalize each point as

$$\tilde{\mathbf{x}}_i = \frac{\mathbf{x}_i - \frac{1}{N}\sum_{j=1}^N \mathbf{x}_j}{\max_j \left\| \mathbf{x}_j - \frac{1}{N}\sum_{k=1}^N \mathbf{x}_k \right\|_2}. \tag{14}$$

This removes translation and scale differences before geometric comparison.

Semantic Fidelity. We use a DGCNN-based point-cloud classifier trained on Objaverse-derived data to evaluate whether the reconstructed geometry preserves object semantics. Let r denote the rank of the ground-truth category among all $C = 72$ categories. For an N-way Top-K evaluation, the ground-truth category is retained together with $N - 1$ sampled distractor categories, and a reconstruction is considered correct if the ground-truth category falls within the Top-K predictions. The expected score can be computed as

$$S_{N,K}(r) = \frac{\sum_{j=0}^{K-1} \binom{r-1}{j}\binom{C-r}{N-1-j}}{\binom{C-1}{N-1}}. \tag{15}$$

Accordingly, $(2, 1)$ corresponds to 2-way Top-1 accuracy and $(10, 3)$ to 10-way Top-3 accuracy. For $(2, 1)$, $S_{2,1}(r) = (72 - r)/71$, while $(10, 3)$ requires at most two of the nine sampled distractors

to rank above the ground-truth category. The dataset-level score is averaged over all reconstructed samples and reported in percentage. We use three DGCNN evaluators with fixed random seeds and average their class probabilities before determining the ground-truth category rank.

Chamfer Distance. For normalized predicted and ground-truth point clouds $P = \{\mathbf{p}_i\}_{i=1}^{N_P}$ and $G = \{\mathbf{g}_j\}_{j=1}^{N_G}$, we compute the bidirectional squared Chamfer Distance as

$$\mathrm{CD}(P, G) = \frac{100}{2}\left[\frac{1}{N_P}\sum_{\mathbf{p}\in P}\min_{\mathbf{g}\in G}\|\mathbf{p}-\mathbf{g}\|_2^2 + \frac{1}{N_G}\sum_{\mathbf{g}\in G}\min_{\mathbf{p}\in P}\|\mathbf{g}-\mathbf{p}\|_2^2\right]. \tag{16}$$

The two directions are averaged and multiplied by 100, corresponding to the CD ($\times 10^{-2}$) values reported in our tables. Lower values indicate better geometric reconstruction.

F1 Score. We compute the F1 score using a Euclidean-distance threshold of $\tau = 0.25$ in the normalized coordinate space. Precision and recall are defined as

$$\mathrm{Precision}_\tau = \frac{1}{N_P}\sum_{\mathbf{p}\in P}\mathbb{I}\left[\min_{\mathbf{g}\in G}\|\mathbf{p}-\mathbf{g}\|_2 \le \tau\right], \tag{17}$$

$$\mathrm{Recall}_\tau = \frac{1}{N_G}\sum_{\mathbf{g}\in G}\mathbb{I}\left[\min_{\mathbf{p}\in P}\|\mathbf{g}-\mathbf{p}\|_2 \le \tau\right]. \tag{18}$$

The final score is

$$\mathrm{F1}_\tau = 100 \times \frac{2\,\mathrm{Precision}_\tau\,\mathrm{Recall}_\tau}{\mathrm{Precision}_\tau + \mathrm{Recall}_\tau}. \tag{19}$$

If both precision and recall are zero, the F1 score is set to zero. Higher values indicate better geometric consistency.

Structural Similarity. To evaluate the rendered structural quality of reconstructed objects, we compute SSIM on grayscale geometry renderings rather than directly on 3D coordinates. The predicted and ground-truth point clouds are rendered into 224×224 images from four fixed viewpoints, $(0°, 18°)$, $(90°, 18°)$, $(180°, 18°)$, and $(270°, 18°)$. The same camera settings are used for all samples, and no color information is involved.

For each view, SSIM is computed as

$$\mathrm{SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \tag{20}$$

where $C_1 = (0.01)^2$ and $C_2 = (0.03)^2$. Local statistics are computed using a Gaussian filter with standard deviation 1.5. The final SSIM is obtained by averaging over all pixels and then over the four rendered views:

$$\mathrm{SSIM}_{\mathrm{MV4}} = \frac{1}{4}\sum_{k=1}^{4}\mathrm{SSIM}\left(I_k^{\mathrm{pred}}, I_k^{\mathrm{GT}}\right). \tag{21}$$

Higher SSIM indicates greater structural similarity between the reconstructed and ground-truth objects.

A.3 Neuromorphic Inference and Energy Evaluation

Neuromorphic Inference of LD-LIF. LD-LIF maintains the basic event-driven computation of conventional LIF neurons while introducing learnable neuronal dynamics. The input gain g and membrane time constant τ_m are optimized during training and remain fixed during inference. Accordingly, the membrane decay factor $\lambda = 1 - 1/\tau_m$ becomes a fixed neuronal parameter after training, while the input gain g can be folded into the corresponding layer weights. Therefore, LD-LIF does not require additional parameter optimization during inference and can be directly implemented on neuromorphic hardware.

Neuromorphic Inference of SpikeBit. For a B-bit SpikeBit neuron, a single neuronal timestep is unfolded into B virtual timesteps. At each virtual timestep, the transmitted event remains binary,

while the numerical significance of the event is determined by the corresponding bit position. The SpikeBit representation can be written as

$$s(u) = \Delta \sum_{r=0}^{B-1} 2^r \left(\gamma_+ b_r^+(u) - \gamma_- b_r^-(u)\right), \tag{22}$$

where $b_r^+, b_r^- \in \{0, 1\}$ denote the positive and negative binary events at the r-th virtual timestep. The sign of the membrane potential directly determines whether an event belongs to b_r^+ or b_r^-, and the corresponding scaling factor γ_+ or γ_- is selected accordingly.

Given the weight W of the following layer, applying the layer operation to the SpikeBit output gives

$$Ws(u) = W\Delta \sum_{r=0}^{B-1} 2^r \left(\gamma_+ b_r^+ - \gamma_- b_r^-\right) \tag{23}$$

$$= \sum_{r=0}^{B-1} \left[\alpha_r^+ W b_r^+ - \alpha_r^- W b_r^-\right], \tag{24}$$

where

$$\alpha_r^+ = \Delta 2^r \gamma_+, \qquad \alpha_r^- = \Delta 2^r \gamma_-. \tag{25}$$

Since b_r^+ and b_r^- are binary, the computations Wb_r^+ and Wb_r^- can still be performed in a spike-driven manner, with only an additional lightweight scaling by the corresponding coefficient and thus little extra energy overhead. In addition, compared with conventional spiking neurons, SpikeBit only requires an extra coefficient modulation at each virtual timestep when performing inference on neuromorphic hardware. This operation can be coordinated by the global timing mechanisms already available on mainstream neuromorphic platforms, such as Tianjic Pei et al. (2019), KA200 Yang et al. (2024), and Loihi Davies et al. (2018).

Energy Evaluation. For ANNs, convolutional, linear, and matrix multiplication operations are mainly implemented using multiply–accumulate (MAC) operations. Following the 45 nm CMOS energy model in Horowitz (2014), each MAC operation consumes approximately 4.6 pJ. Accordingly, the energy consumption of these operations is calculated as

$$E_{\text{Conv}}^{\text{ANN}} = F_{\text{out}} C_{\text{in}} C_{\text{out}} K^2 E_{\text{MAC}}, \tag{26}$$

$$E_{\text{Linear}}^{\text{ANN}} = D_{\text{in}} D_{\text{out}} E_{\text{MAC}}, \tag{27}$$

$$E_{\text{MatMul}}^{\text{ANN}} = MDN E_{\text{MAC}}, \tag{28}$$

where F_{out} denotes the number of spatial positions in the output feature map and K^2 denotes the 2D kernel size. For 3D convolutional layers in the reconstruction network, F_{out} denotes the number of output voxels and K^2 is replaced by $K_h K_w K_d$. D_{in} and D_{out} denote the input and output dimensions of a linear layer. For matrix multiplication between matrices of dimensions $M \times D$ and $D \times N$, the number of MAC operations is MDN. We use $E_{\text{MAC}} = 4.6$ pJ in our calculation.

For SNNs, binary spikes allow multiplication to be replaced by sparse accumulate (AC) operations. Under the same 45 nm CMOS setting, each AC operation consumes approximately 0.9 pJ Horowitz (2014). The corresponding energy consumption is calculated as

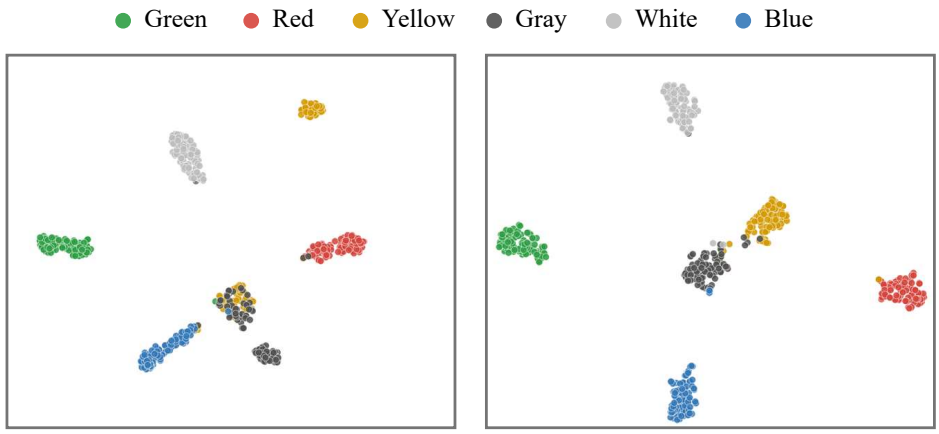
$$E_{\text{Conv}}^{\text{SNN}} = T \times fr \times F_{\text{out}} C_{\text{in}} C_{\text{out}} K^2 E_{\text{AC}}, \tag{29}$$

$$E_{\text{Linear}}^{\text{SNN}} = T \times fr \times D_{\text{in}} D_{\text{out}} E_{\text{AC}}, \tag{30}$$

$$E_{\text{MatMul}}^{\text{SNN}} = T \times fr \times MDN E_{\text{AC}}, \tag{31}$$

where T denotes the number of inference timesteps, fr is the average firing rate measured from spike activities during inference, and $E_{\text{AC}} = 0.9$ pJ. The matrix multiplication term is used to account for the spike-driven matrix products in the attention modules. Operations receiving continuous-valued inputs are counted as MACs instead of ACs.

For the EEG classification network, T corresponds to the temporal length of the input EEG sequence. For conventional LIF neurons in the reconstruction network, T denotes the number of neuronal timesteps over which the input is repeatedly processed. For SpikeBit, a single neuronal timestep is



(a) Neuro-3D (b) SpikeEEG-3D

Figure 6: t-SNE visualization of the learned appearance representations by Neuro-3D (left) and SpikeEEG-3D (right) on the EEG-3D dataset.

unfolded into B virtual timesteps, and we therefore set $T = B$ when calculating the spike-driven computation. In addition, we explicitly include the extra energy cost of applying the timestep-dependent coefficient at each virtual timestep as $E_{\text{scale}} = E_{\text{MAC}} B N^{\text{out}}$, where N^{out} denotes the total number of output elements involved in the coefficient modulation. This introduces only a small additional energy overhead.

A.4 t-SNE Visualization of Learned Representations

As shown in Fig. 6, SpikeEEG-3D exhibits more compact intra-class distributions and clearer separation among different categories compared with Neuro-3D, indicating more discriminative learned representations.

A.5 Class Activation Topography

Following Song et al. (2022), we use Class Activation Topography (CAT) to analyze the spatial EEG responses captured by SpikeEEG-3D. For each EEG sample, we first obtain a temporal class-activation map using Grad-CAM, which reflects the contribution of different EEG timesteps to the model prediction.

For each stimulus condition, let $\mathbf{X}^{(i)} \in \mathbb{R}^{C\times T}$ denote the EEG signal of the i-th sample and $\mathbf{m}^{(i)} \in \mathbb{R}^{T}$ its temporal activation weights. We first average the EEG signals and activation weights separately across M samples:

$$\bar{\mathbf{X}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{X}^{(i)}, \qquad \bar{\mathbf{m}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{m}^{(i)}. \tag{32}$$

The CAT response of channel c is then computed by temporally weighting the averaged EEG response:

$$R_c = \frac{\sum_{t=1}^{T}\bar{m}_t\bar{X}_{c,t}}{\sum_{t=1}^{T}\bar{m}_t}, \tag{33}$$

where R_c represents the activation-weighted response of the c-th EEG channel. The resulting channel responses are projected onto their electrode locations to obtain the scalp topographies.

Static and dynamic CAT responses are computed separately from their corresponding EEG streams. The combined CAT response is obtained by fusing the static and dynamic CAT responses. For the regional analysis in Fig. 5(b), channel-wise CAT responses are further aggregated within the frontal, central, parietal, occipital, and temporal scalp regions.