Spiking Neural Network-Based Radar Human Activity Recognition

Haotian Zhang*, Yu Zhou*, Xuyang Zheng§, Fei Luo‡, Tianwei Hou†, and Anna Li§
*School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China
†School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China
‡School of Computing and Information Technology, Great Bay University, Dongguan, China
§School of Computing and Communications, Lancaster University, Lancaster, United Kingdom
Email: {23721032, 23725079, twhou}@bjtu.edu.cn, luofei@gbu.edu.cn, {x.zheng25, a.li16}@lancaster.ac.uk

***Abstract*—Radar-based human activity recognition (HAR) has become an important technique in a wide range of applications, from smart homes to healthcare monitoring. Compared with vision-based sensing, radar is privacy-preserving and robust to adverse lighting conditions, making it well suited for indoor environments. However, achieving both high accuracy and energy efficiency remains challenging due to the complexity of radar signals and the subtle variability of human motions. In this paper, we propose *spiking-transformer-TopoLoss* (STT), a topology-regularized spiking transformer model for micro-Doppler-based HAR. STT integrates block-wise spatiotemporal attention with a TopoLoss constraint to enhance both representation capacity and robustness. Specifically, the STAtten module captures long-range spatiotemporal dependencies, while Leaky Integrate-and-Fire neurons maintain event-driven sparsity for efficient computation. Meanwhile, TopoLoss encourages spatially organized intermediate representations, helping suppress spurious activations and preserve the continuity and periodicity of micro-Doppler signatures. Experimental results on a nine-class micro-Doppler kitchen dataset demonstrate that STT achieves 98.99% overall accuracy. Evaluations on public datasets further show that STT remains competitive in both accuracy and computational efficiency, indicating its potential for robust and scalable radar-based HAR in real-world applications.**

***Index Terms*—Micro-Doppler, Spiking neural networks, Spatial-temporal attention, Topological loss, Human activity recognition**

I. INTRODUCTION

Radar-based human activity recognition (HAR) has attracted growing interest due to its broad applicability in medical healthcare, smart homes, and autonomous driving [1]. Compared with vision-based sensors, radar does not capture detailed facial appearance or identity cues, which mitigates privacy leakage in sensitive indoor settings [2]. Despite these advantages, radar-based HAR remains challenging due to the complex temporal and spatial dynamics embedded in radar measurements, which vary across activities. Robust methods are therefore required to extract discriminative features and achieve accurate classification. In recent years, deep learning has addressed these challenges by enabling automatic feature learning from radar data, leading to significant gains in recognition accuracy and robustness [3].

*Haotian Zhang and Yu Zhou contributed equally to this work.
Corresponding author: Anna Li.

Spiking Neural Networks (SNNs) are attractive for radar sensing due to their event-driven, low-power computation. A. Safa *et al.* [4] first combined SNNs with FMCW radar by converting radar signals into the event domain to reduce computational cost. Building on this, M. Arsalan *et al.* [5] processed raw ADC data using SNN-based operations that mimic Fourier transforms, achieving lower latency and power consumption. K. Zheng *et al.* [6] further introduced a neuromorphic radar framework where SNNs process radar outputs with competitive accuracy under constrained computation. More recently, Y. Wu *et al.* [7] proposed hybrid architectures integrating Gaussian Mixture Models (GMM) with SNNs to better capture spatiotemporal features in complex indoor environments. However, despite these efforts, conventional SNNs still suffer from information loss due to their sparse binary activations, which limits the preservation of fine-grained patterns required for challenging recognition tasks [8].

To alleviate the representation bottleneck of conventional SNNs, Spiking Transformers have been introduced to enhance long-range dependency modeling [9]. However, Spiking-Transformer-based activity recognition on micro-Doppler spectrograms remains under-explored, and directly applying standard spiking transformer architectures may not fully address the structural characteristics of micro-Doppler patterns [10]. In particular, micro-Doppler signatures often exhibit strong local continuity in the time-frequency domain, and models that do not explicitly encourage structured organization in intermediate representations may become sensitive to noise and subtle intra-class variations [11]. These considerations motivate incorporating an inductive bias that promotes organized and structurally coherent representations during learning.

To this end, we adopt the TopoLoss constraint to encourage topographic organization in neural network representations [12]. TopoLoss introduces a topology-inspired regularization that promotes smoothness and locality by encouraging nearby neurons to learn similar response patterns, analogous to topographic organization observed in biological systems. In practice, TopoLoss reshapes learned representations into a 2D map and regularizes this map to be consistent with a smoothed version of itself, thereby discouraging abrupt discontinuities and suppressing noisy activations. Such topographic regularization has been applied to both vision models and language mod-

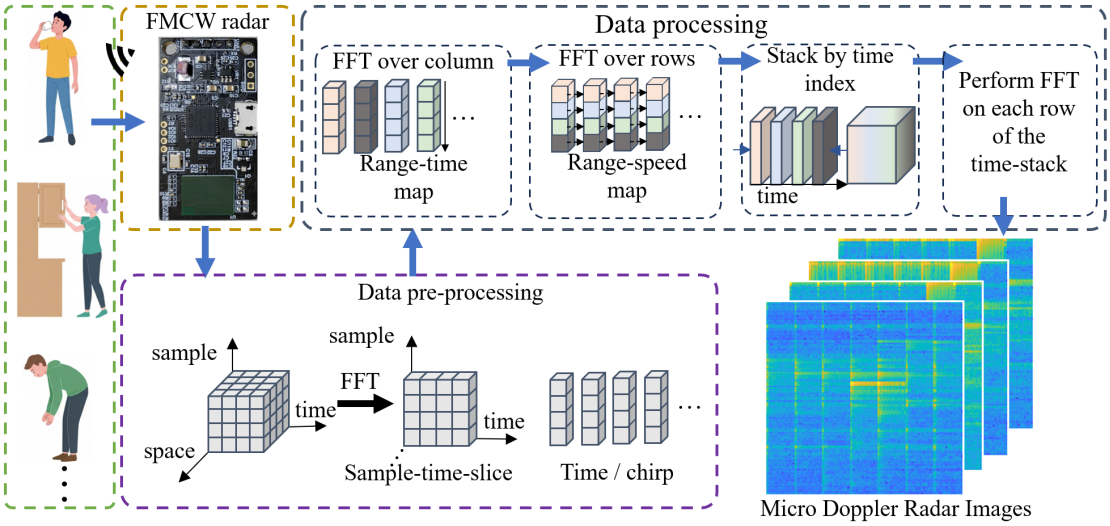

Fig. 1. FMCW Radar Signal Processing Pipeline.

els to form TopoNets, showing improved performance with minimal inference overhead. We hypothesize that TopoLoss is particularly suitable for micro-Doppler spectrogram analysis because it can promote structured, locally consistent features in the time-frequency domain.

In this paper, we propose a spiking transformer network called *spiking-transformer-TopoLoss* (STT)[1], a novel framework designed to enhance the efficiency and accuracy of radar-based HAR, enabling the model to focus on the most relevant features while capturing the necessary spatial and temporal dependencies. Our contributions are summarized as follows:

1) We propose STT, a spiking transformer architecture that integrates a spatiotemporal attention encoder with a brain-inspired regularization term, TopoLoss, for micro-Doppler-based human activity recognition.
2) We demonstrate that STT achieves superior performance on our micro-Doppler spectrogram dataset, reaching an overall accuracy of 98.99%.
3) We further evaluate STT on public datasets and analyze efficiency–accuracy trade-offs. In particular, STT (dim=256) maintains competitive accuracy with fewer parameters and lower computational cost, while STT (dim=768) achieves 93.79% accuracy on CIFAR-10.

II. METHODOLOGY

A. Signal Processing

In an FMCW-based mmWave sensing system, the transmitted waveform sweeps linearly in frequency over time within each chirp. The transmitter and receiver operate simultaneously, enabling continuous transmission and echo capture [13]. The instantaneous frequency of the transmitted chirp is

$$f_T(t) = f_c + \frac{B}{T_c}t, \tag{1}$$

[1]The source code is available at: https://github.com/zy-StarXH/STT

where f_c is the carrier frequency, B is the chirp bandwidth, and T_c is the chirp duration.

Accordingly, the transmitted signal is

$$s_T(t) = A_T \cos\left(2\pi f_c t + \pi \frac{B}{T_c} t^2 + \phi_0\right), \tag{2}$$

where A_T is the transmit amplitude and ϕ_0 is the initial phase.

For a target at distance d, the received echo is delayed and attenuated due to propagation loss and scattering:

$$s_R(t) = \alpha A_T \cos\left(2\pi f_c (t-\tau_d) + \pi \frac{B}{T_c}(t-\tau_d)^2 + \phi_0\right), \tag{3}$$

where α is the attenuation coefficient and $\tau_d = 2d/c$ is the round-trip delay with the speed of light c.

The transmitted and received signals are mixed to produce an intermediate-frequency (IF) beat signal. After low-pass filtering, it can be approximated as

$$s_{\text{IF}}(t) = s_T(t) \cdot s_R(t) \approx A_{\text{IF}} \cos(2\pi f_b t + \varphi_b), \tag{4}$$

where $f_b = (B/T_c)\tau_d$ is the beat frequency proportional to range, and φ_b is the phase offset from motion and reflection.

As shown in Fig. 1, a Fast Fourier Transform (FFT) is first applied along the *fast-time* axis to generate range profiles. The complex samples from the target range bin are then stacked across chirps (the *slow-time* axis), and a Short-Time Fourier Transform (STFT) is applied to produce the micro-Doppler spectrogram. The resulting 2D Doppler maps are used as inputs to the spiking neural network.

B. Spiking Neural Network

SNNs represent the third generation of neural network models, distinguished by their biological plausibility and event-driven information processing paradigm [14]. Unlike traditional artificial neural networks with continuous-valued activations, SNNs communicate through discrete spike trains. Information can be encoded in spike timing, firing rate, or

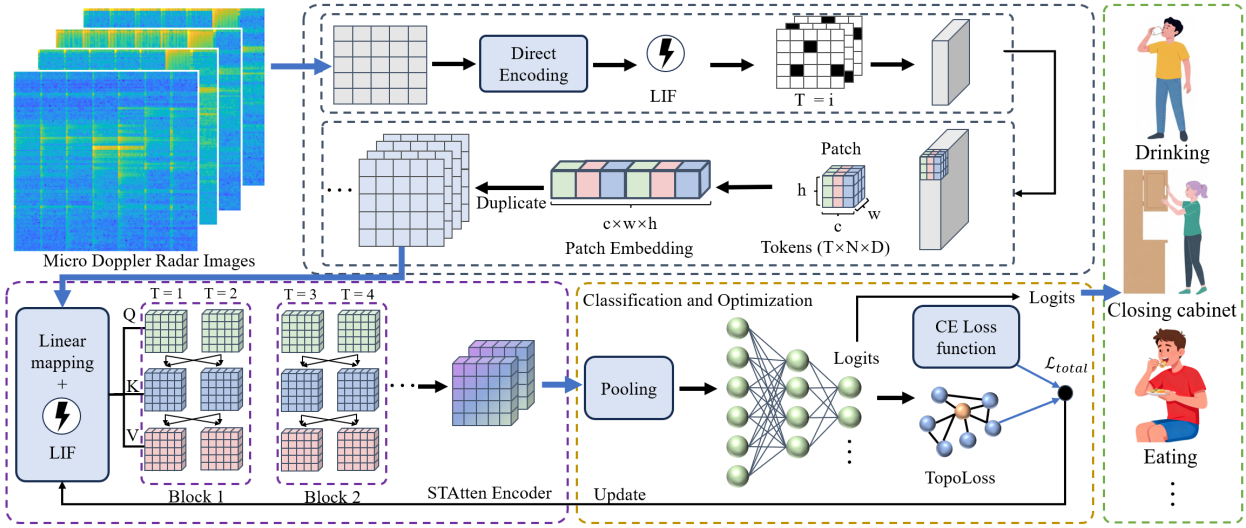

Fig. 2. Overview of the Proposed STT Architecture.

population activity, resembling the signaling mechanism of biological neurons. The core component of an SNN is the spiking neuron, and the leaky integrate-and-fire (LIF) model is widely adopted due to its balance between computational efficiency and biological realism [15]. These neurons integrate synaptic inputs to update their membrane potential and emit a spike when the potential crosses a predefined threshold. The dynamic behavior of the membrane potential $V_j(t)$ for neuron j can be described by the following iterative update equation:

$$V_j(t) = \beta V_j(t-1) + \sum_i w_{ji} x_i(t), \tag{5}$$

where β is the decay factor reflecting the leakage of the membrane, w_{ji} denotes the synaptic weight from input neuron i, and $x_i(t)$ represents the incoming binary spike input at time t. When $V_j(t)$ exceeds a firing threshold V_{th}, the neuron generates an output spike and the membrane potential is reset. An advantage of SNNs is energy efficiency, largely attributed to the sparsity of spike events, which makes them suitable for low-power hardware implementations [16]. However, conventional SNNs often struggle to capture long-range spatiotemporal dependencies and to preserve fine-grained patterns, limiting their performance on complex tasks such as micro-Doppler activity recognition. To address these limitations, Spiking-Transformer models combine spiking mechanisms with Transformer architectures to enhance feature extraction while maintaining the event-driven nature of SNNs.

C. TopoLoss Function

While Spiking Transformers improve long-range dependency modeling, they may neglect the underlying geometric structure of the data. To address this and enhance the structural continuity of feature representations and encourage local topographic organization, a TopoLoss function is introduced. Inspired by the brain's self-organizing cortical mechanism [12], this loss function encourages spatial smoothness by comparing each feature map with its blurred version, suppressing noisy activations while preserving meaningful spatial relations.

Let $X \in \mathbb{R}^{H\times W\times D}$ denote the feature activation map of a network layer, where H and W are spatial dimensions and D is the channel depth. A smoothed version $\tilde{X}$ is obtained through a downsampling–upsampling operation:

$$\tilde{X} = \mathcal{U}_\phi\big(\mathcal{D}_\phi(X)\big), \tag{6}$$

where $\mathcal{D}_\phi(\cdot)$ and $\mathcal{U}_\phi(\cdot)$ represent the downsampling and upsampling operators with scale factor ϕ controlling the level of smoothing.

The overall training objective combines the classification loss $\mathcal{L}_{\text{cls}}$ with the topological regularization term $\mathcal{L}_{\text{topo}}$, which measures the structural consistency between X and $\tilde{X}$:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda_{\text{topo}}\,\mathcal{L}_{\text{topo}}, \tag{7}$$

where λ_{topo} balances the influence of the topological constraint. The loss is applied to the final classification head and added to Cross-Entropy loss during training. The parameters include the scale, h, and w. Our tests show that model performance is sensitive to these values. Increasing the scale while keeping a ratio of h and w helps improve accuracy. TopoLoss smooths internal representations and improves robustness by preserving the continuity of micro-Doppler patterns.

D. The Proposed STT Model

The proposed STT model aims to improve spike-based feature learning by integrating advanced spatiotemporal attention with brain-inspired topographic constraints, as shown in Fig. 2. We first introduce the baseline architecture, STAtten, and then describe how we incorporate the TopoLoss term.

STAtten is a representative model that combines spiking mechanisms with a Transformer-style attention encoder [17]. It consists of three main components. The first component is a spike feature preprocessor that starts from an encoding of the input. It includes multiple 2D convolution and pooling layers,

TABLE I
THE DESCRIPTION OF NINE DIFFERENT HUMAN ACTIVITIES.

No.	Activity	Description	Samples
(a)	Bending	Bending from standing position	986
(b)	Closing cabinet	Closing a cabinet door	1020
(c)	Drinking	Drinking from a cup or bottle	967
(d)	Eating	Eating food while seated	929
(e)	Falling	Falling to the floor from standing	1049
(f)	Opening cabinet	Opening a cabinet door	990
(g)	Sit-to-stand	Transition from sitting to standing	870
(h)	Squat-to-stand	Squatting down then rising	1009
(i)	Stand-to-sit	Transition from standing to sitting	923

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT MODELS.

Models	Accuracy (%)	Kappa (%)	Size (MB)	Params (M)	FLOPs (G)
ViTSmall	100.00	100.00	82.66	21.649	4.25
Spiking-ResNet18	96.97	96.59	42.69	11.18	14.53
Spiking-ResNet34	94.84	94.19	81.28	21.29	29.34
SEW-ResNet18	97.98	97.73	42.69	11.18	14.53
SEW-ResNet34	82.94	80.81	81.28	21.29	29.34
STAtten	98.54	98.36	426.00	37.19	326.57
STT (dim=256)	98.54	98.36	47.74	4.15	50.88
STT (dim=768)	98.99	98.86	426.00	37.19	326.57

which convert the input into multi-scale feature maps. After each convolution, LIF neurons perform temporal integration and sparse thresholding to generate spike trains. This spike-based representation enhances time–frequency features and structural patterns while suppressing noise. The resulting features are then patched and reshaped into spatiotemporal tokens, preserving the original spatial and temporal relationships.

The second component is a block-wise spatiotemporal attention module. Queries, keys, and values are generated via 1×1 convolutions (learnable linear projections) followed by spiking neurons. For a binary input tensor X, we obtain

$$Q = \text{LIF}(W_Q X), \quad K = \text{LIF}(W_K X), \quad V = \text{LIF}(W_V X). \tag{8}$$

Instead of attending to all time steps simultaneously, STAtten adopts chunked attention that focuses on short temporal segments. This design captures short-term motion details without losing spatial patterns. For a temporal block indexed by b, the block-wise attention is computed as

$$\text{STAtten}(X[b]) = \text{LIF}(\text{Attn}(Q[b], K[b], V[b])), \tag{9}$$

where $\text{Attn}(\cdot)$ denotes the attention operation and $[b]$ indicates the corresponding temporal block.

The third component is the classification head. Spatial and temporal features from the last encoder block are aggregated into a single feature vector, which is passed through a spiking neuron before a linear classifier produces final class prediction.

To further enhance the STAtten baseline, STT introduces TopoLoss as an additional training objective applied to the logits produced by the classification module. Unlike standard objectives that focus solely on classification accuracy, TopoLoss serves as a topographic inductive bias by encouraging spatially adjacent neurons to learn similar response patterns. By enforcing smoothness in the learned topographic organization and penalizing abrupt discontinuities, the model suppresses noise while preserving the structural integrity of micro-Doppler signatures. Consequently, combining STAtten with TopoLoss enables STT to learn robust, spatially organized representations without compromising the event-driven efficiency of the original architecture.

III. EXPERIMENTAL SETUP

We employed a custom-designed 77 GHz mmwave radar module for indoor human motion recognition. Developed for low-cost and compact deployment, the system is integrated onto a single printed circuit board and costs approximately $20. The radar features four on-chip transmitting and four receiving antennas, controlled by an STM32 microcontroller that handles chirp scheduling and USB-based data transfer. The analog front-end integrates voltage-controlled oscillators, phase-locked loops, mixers, and analog-to-digital converters for signal generation and acquisition. The radar emits linearly frequency-modulated chirps with a duration of $T = 85\,\mu$s and a bandwidth of $B = 2.585\,$GHz. Each chirp is sampled at 3.636 MSPS, producing 256 fast-time samples per chirp.

To evaluate the system under realistic conditions, we constructed a dataset in a furnished kitchen environment with typical appliances and clutter. Unlike controlled lab setups, our environment includes items such as cabinets and refrigerators that introduce occlusion and multipath effects. The radar was mounted on top of a laptop screen approximately 1.2 m above the ground. It was positioned to directly face the activity zone, enabling the system to effectively capture human motion signatures within a range of 1.5 to 2 m. The laptop supplied power to the radar and handled real-time data acquisition. All participants provided written informed consent before data collection. The study was conducted in accordance with institutional ethical guidelines. Finally, a total of 9 kitchen-related activities were recorded, covering both routine tasks and safety-critical actions such as falling. These activities are summarized in Table I.

The dataset comprises 8,753 annotated samples collected from nine participants, whose heights ranged from 162 to 183 cm. We split the dataset into training, validation, and test sets with a 19:5:1 ratio. The division is based on random splitting of image samples, which includes data across different subjects and time periods, ensuring that the model is tested on various patterns across all participants. Each participant performed multiple repetitions at different distances and angles, ensuring diversity in motion patterns and radar perspectives. Raw radar signals are then processed to generate micro-Doppler spectrograms.

IV. RESULTS AND DISCUSSION

A. Overall Performance

We validated the performance of the proposed STT on our kitchen dataset. All experiments were conducted on an

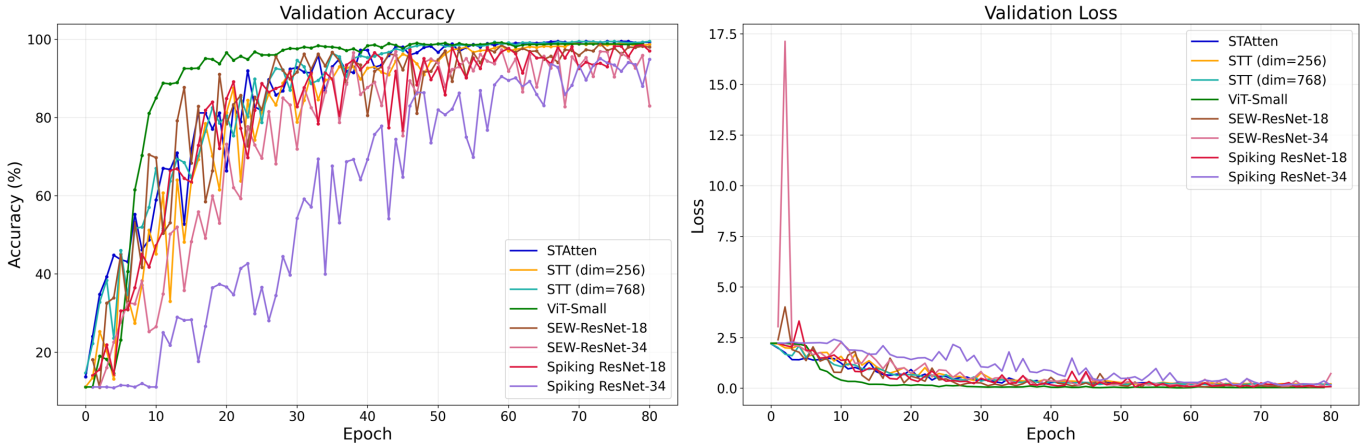

Fig. 3. The validation accuracy and loss of STT on our kitchen dataset.

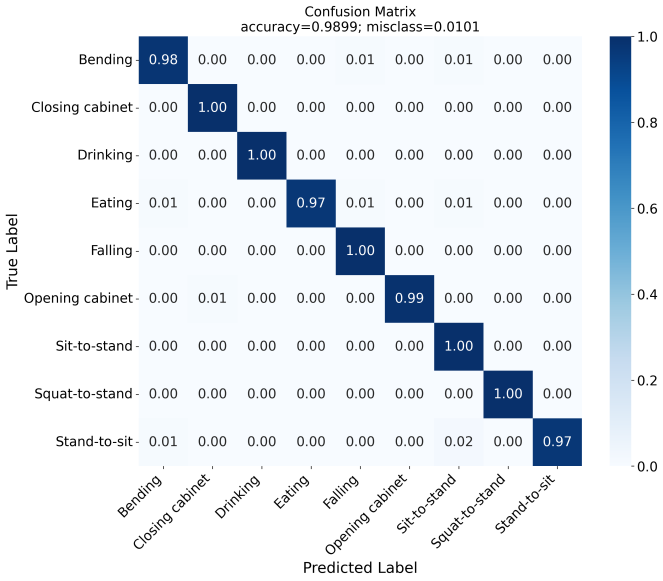

Fig. 4. Normalized Confusion Matrix of the Proposed STT (dim=768) Model.

NVIDIA V100-SXM2-32GB GPU. The models were trained for 80 epochs using the AdamW optimizer with an initial learning rate of 1×10^{-3}.

As presented in Table II, the STT (dim=768) achieves an outstanding accuracy of 98.99% and a Kappa coefficient of 98.86% with 37.19M parameters. For resource-constrained scenarios, the lightweight STT (dim=256) offers a competitive alternative, achieving 98.54% accuracy with only 4.15M parameters. Fig. 3 illustrates the validation accuracy and loss curves. It can be observed that the proposed STT (dim=768) converges stably, demonstrating a steady decrease in loss and a rapid increase in accuracy comparable to ViT-Small.

Fig. 4 further details the classification performance of STT (dim=768). The model achieves 100.00% accuracy on distinct actions such as Closing cabinet, Drinking, and Falling. Although minor misclassifications occur in actions like Eating (97.00%) and Stand-to-sit (97.00%) due to high similarity with other motions, the overall recognition capability remains highly robust.

B. Ablation Study

To validate the effectiveness of the TopoLoss constraint, we compare STT (dim=768) with the baseline STAtten. Table II shows that STT achieves an accuracy of 98.99%, surpassing STAtten (98.54%) by 0.45%. While Fig. 3 reveals larger initial training fluctuations for STT due to the joint optimization of classification and topographic constraints, this process mirrors biological synaptic pruning and refinement [18]. Ultimately, the model stabilizes at a higher accuracy, confirming that the spatial organization enforced by TopoLoss enhances feature robustness and noise resistance.

C. Comparison with State-of-the-Art Models

We further compare the properties and classification abilities of STT with representative SNNs [19], [20] and ViT [21], and the results are presented in Table II. Among SNN baselines, while SEW-ResNet34 drops to 82.94%, our STT (dim=768) attains 98.99%, outperforming the baseline STAtten by 0.45%. Although ViT_Small reaches 100.00% accuracy, our STT model delivers comparable high performance while leveraging the event-driven properties of SNNs. ViT reaches 100% likely because of the dataset difficulty and the strong training setup for Transformers. However, our STT model is more energy-efficient because it uses the event-driven features of SNNs. Furthermore, the lightweight STT (dim=256) matches STAtten's accuracy (98.54%) with only 4.15M parameters, demonstrating a superior balance between classification capability and model complexity.

D. Validation on CIFAR-10 Dataset

In addition to the evaluation on our kitchen dataset, we conducted experiments on the public CIFAR-10 dataset [22] to assess generalization. We use CIFAR-10 because it is a standard dataset for testing Spiking Neural Networks. This experiment helps us prove that our STT model works well

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT MODELS ON CIFAR-10 DATASET.

Models	Accuracy (%)	Kappa (%)	Size (MB)	Params (M)	FLOPs (G)
Spiking-ResNet34	88.6	86.3	81.25	21.28	9.28
SEW-ResNet34	92.73	91.2	81.25	21.28	9.28
STT(dim=256)	87.35	85.9	29.51	2.57	2.49
STT(dim=768)	93.79	91.3	426.01	37.25	29.77

on general image tasks, not just radar data. As shown in Table III, the proposed STT (dim=768) achieves an accuracy of 93.79%, outperforming SEW-ResNet34 (92.73%) and Spiking-ResNet34 (88.6%). Beyond recognition performance, we analyzed computational costs. While STT (dim=768) uses 37.25M parameters to ensure superior capability, the lighter STT (dim=256) offers a highly efficient alternative. It reduces the parameter count to 2.57M while maintaining competitive accuracy, making it suitable for resource-constrained devices.

E. Discussion of Main Findings

The experimental results demonstrate the feasibility of our approach. The STT (dim=768) model achieves the highest accuracy on both micro-Doppler and public datasets, demonstrating efficiency of the model. Our approach significantly outperforms traditional spiking neural networks on micro-Doppler dataset, suggesting that spatial-temporal attention mechanism is effective for activities recognition. The time-frequency patterns for micro-Doppler map aligns well with the spatial-temporal mechanism and the topoLoss function, resulting in an outstanding 98.99% accuracy. These results show that STT effectively recognizes micro-Doppler activities while maintaining its generality across different datasets.

V. CONCLUSION

This paper proposes STT, a spiking transformer model tailored for micro-Doppler-based human activity recognition. STT is designed to address limitations of conventional SNN-based radar pipelines, including information loss in sparse spike representations and reduced ability to preserve fine-grained time-frequency continuity that is critical for distinguishing activities with subtle motion differences. By introducing TopoLoss as a topographic regularizer, STT encourages spatially organized intermediate representations, which helps suppress noise while preserving the structural integrity and continuity of micro-Doppler patterns. Extensive experiments validate the effectiveness of our approach. STT achieves 98.99% accuracy on our kitchen dataset. Moreover, evaluations on public datasets further demonstrate its generalization capability and practical efficiency–accuracy trade-offs, supporting its suitability for radar-based HAR scenarios that require both high precision and energy-aware inference. Future research will focus on hybrid spike-continuous coding and hardware-aware attention mechanisms to optimize energy efficiency and real-time performance for radar-based activity recognition.

REFERENCES

[1] H. Kong, C. Huang, J. Yu, and X. Shen, "A survey of mmWave radar-based sensing in autonomous vehicles, smart homes and industry," IEEE Commun. Surv. Tutor., vol. 27, no. 1, pp. 1–1, 2024.

[2] A. Li, E. Bodanese, S. Poslad, T. Hou, K. Wu, and F. Luo, "A Trajectory-Based Gesture Recognition in Smart Homes Based on the Ultrawideband Communication System," IEEE Internet Things J., vol. 9, no. 22, pp. 22861-22873, 2022.

[3] J. Park, R. J. Javier, T. Moon, and Y. Kim, "Micro-Doppler based classification of human aquatic activities via transfer learning of convolutional neural networks," Sensors, vol. 16, no. 12, 2016.

[4] A. Safa, F. Corradi, L. Keuninckx, I. Ocket, A. Bourdoux, F. Catthoor, and G.G.E Gielen, "Improving the accuracy of spiking neural networks for radar gesture recognition through preprocessing," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 6, pp. 2869–2881, 2023.

[5] M. Arsalan, A. Santra, and V. Issakov, "Spiking neural network-based radar gesture recognition system using raw ADC data," IEEE Sens. Lett., vol. 6, no. 6, pp. 1–4, 2022.

[6] K. Zheng, K. Qian, T. Woodford, and X. Zhang, "NeuroRadar: A Neuromorphic Radar Sensor for Low-Power IoT Systems," in Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems, 2023, pp. 223–236.

[7] Y. Wu, L. Wu, T. Hu, Z. Xiao, M. Xiao, and L. Li, "An efficient radar-based gesture recognition method using enhanced GMM and hybrid SNN," IEEE Sens. J., vol. 25, no. 7, pp. 12511–12524, 2025.

[8] Y. Guo, Y. Chen, X. Liu, W. Peng, Y. Zhang, X. Huang, and Z. Ma, "Ternary spike: Learning ternary spikes for spiking neural networks," arXiv [cs.CV], 2023.

[9] M.-I. Stan and O. Rhodes, "Learning long sequences in spiking neural networks," Sci. Rep., vol. 14, no. 1, p. 21957, 2024.

[10] S. Shen, D. Zhao, L. Feng, Z. Yue, J. Li, T. Li, G. Shen, and Y. Zeng, "STEP: A unified spiking transformer evaluation platform for fair and reproducible benchmarking," arXiv [cs.NE], 2025.

[11] M. Czerkawski, C. Ilioudis, C. Clemente, C. Michie, I. Andonovic, and C. Tachtatzis, "A novel micro-Doppler coherence loss for deep learning radar applications," in 2021 18th European Radar Conference (EuRAD), 2022, pp. 305–308.

[12] D. Mayukh, D. Mainak, and R. M. N Apurva, "TopoNets: High performing vision and language models with brain-like topography," Int Conf Learn Represent, vol. 2025, pp. 81669–81689, 2025.

[13] Y. Sun, T. Fei, F. Schliep, and N. Pohl, "Gesture classification with handcrafted micro-Doppler features using a FMCW radar," in 2018 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), 2018, pp. 1–4.

[14] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," Neural Netw., vol. 111, pp. 47–63, 2019.

[15] K. Yamazaki, V.-K. Vo-Ho, D. Bulsara, and N. Le, "Spiking neural networks and their applications: A review," Brain Sci., vol. 12, no. 7, p. 863, 2022.

[16] D.-A. Nguyen, X.-T. Tran, and F. Iacopi, "A review of algorithms and hardware implementations for spiking neural networks," J. Low Power Electron. Appl., vol. 11, no. 2, p. 23, 2021.

[17] D. Lee, Y. Li, Y. Kim, S. Xiao, and P. Panda, "Spiking Transformer with Spatial-Temporal Attention," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025, pp. 13948–13958.

[18] M. M. Riccomagno and A. L. Kolodkin, "Sculpting neural circuits by axon and dendrite pruning," Annu. Rev. Cell Dev. Biol., vol. 31, no. 1, pp. 779–805, 2015.

[19] H. Zheng, Y. Wu, L. Deng, Y. Hu, and G. Li, "Going deeper with directly-trained larger spiking neural networks," Proc. Conf. AAAI Artif. Intell., vol. 35, no. 12, pp. 11062–11070, 2021.

[20] W. Fang, Z. Yu, Y. Chen, T. Masquelier, T. Huang, and Y. Tian, "Incorporating learnable membrane time constant to enhance learning of spiking neural networks," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 2661–2671.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv [cs.CV], 2020.

[22] A. Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., University of Toronto, 2009.