SSPT: Spiking Serialized Point Transformer for Couinaud Segmentation in 3D Medical Data

Yan Li, Ruiyang Li (✉), Yue Qiu, Qixuan Liu, Jialun Pei, Shi Qiu, and Pheng-Ann Heng

Department of Computer Science and Engineering
Institute of Medical Intelligence and XR
The Chinese University of Hong Kong
liruiruiyang@outlook.com

Abstract. Couinaud segmentation is a fundamental step in surgical planning for anatomical liver resection. Anatomically, Couinaud liver segments are defined by the vascular system. However, in medical imaging, the critical issue is that the defining landmarks, i.e., vessels, are extremely sparse within the vast liver parenchyma. Current methods employ handcrafted dense sampling strategies to exploit more vascular data points, leading to high energy consumption and a heavy dependency on manual data rebalancing. We propose Spiking Serialized Point Transformer (SSPT), a Spiking Neural Network (SNN) architecture, to tackle these challenges efficiently and effectively. In particular, SSPT devises Leaky Integrate-and-Fire (LIF) neurons to enable a 3D transformer framework, serializing unordered point cloud data into structured temporal spike trains. This spike-based design enables temporally consistent aggregation of critical anatomy (e.g., vessels) context from sparse medical imaging data, while avoiding manually designed sampling strategies and their associated biases. Experimental results on MSD8 and LiTS datasets demonstrate that SSPT delivers high-fidelity segmentation (up to 92.14% Dice) and robust boundary precision (down to 2.24 mm ASD) with a minimal energy footprint of 10.97mJ. Moreover, our experimental results demonstrate that spiking dynamics can inherently capture sparse anatomical topologies, presenting a robust and energy-efficient solution for offline preoperative liver surgery planning. Code is available at https://github.com/Yan0918/SSPT.

Keywords: Spiking Neural Networks · Couinaud Segmentation · Point Cloud · Serialized Attention · Spatio-temporal Encoding.

1 Introduction

Anatomical hepatectomy and segment-oriented surgical planning rely heavily on the Couinaud system, which divides the liver into eight functional segments based on the branching of the portal and hepatic veins [2]. Unlike organs with visible physical borders, the boundaries between Couinaud segments are "virtual planes" defined solely by vascular topology. Thus, automated Couinaud

segmentation methods [5] need to implicitly learn complex and often sparse vessel-related signals from liver CT or MRI data, rather than explicitly identifying organ contours. Recent unified and synergistic hepatic segmentation studies further support this anatomical dependency, showing that vascular topology provides key priors for Couinaud segment delineation [13, 12].

While voxel-based architectures [14, 4] have long dominated medical volume segmentation including Couinaud segmentation [3], their fixed-grid nature often results in significant computational redundancy when processing high-resolution medical data, where essential structural information is relatively sparse. Point-based architectures [23] have emerged as a viable alternative due to their inherent flexibility in handling irregular anatomical geometries. However, these methods frequently incorporate handcrafted sampling priors or vessel-boosted rebalancing to mitigate landmark sparsity. While effective in certain contexts, such reliance on manual manipulation can introduce sampling biases and limit the model's robustness when applied to raw, uncurated clinical data distributions. Hence, there is an imperative need for developing a Couinaud segmentation method capable of inherent anatomical perception, i.e., the ability to delineate Couinaud boundaries by autonomously identifying sparse structural information without manual manipulation.

We propose the Spiking Serialized Point Transformer (SSPT) to achieve endogenous anatomical perception while maintaining high efficiency. The core of our framework is the Spiking Serialized Transformer Block (SSTB), which serves as the primary computational unit for hierarchical feature extraction. In particular, our SSTB is supported by the Temporal Scanning Dispatcher (TSD), which transforms static 3D geometry to dynamic spiking activity via two-stage process. Specifically, the TSD first maps unstructured 3D coordinates into a 1D spatial sequence using space-filling curves to maintain local geometric proximity; then, it dispatches these points into T discrete temporal steps through a scanning strategy, thereby initiating a structured spatiotemporal spike flow. In addition, the SSTB employs the Spiking Conditional Positional Encoding (SCPE), leveraging spatio-temporal parallel convolutions to compensate for spatial information loss. Moreover, we design the Spiking Serialized Attention (SSA) that converts conventional dense dot-products into energy-efficient, pulse-driven accumulation. By conducting extensive experiments on *LiTS* [1] and *MSD8* [15] datasets, our proposed SSPT framework showcases an effective yet efficient Couinaud segmentation solution for precise surgical planning.

2 Method

2.1 Overview of SSPT

Overall Architecture. SSPT is a hierarchical spiking transformer model designed for energy-efficient, high-precision liver Couinaud segmentation. The architecture adopts a hierarchical U-Net structure for feature extraction and reconstruction. For computational efficiency, SSPT utilizes space-filling curves

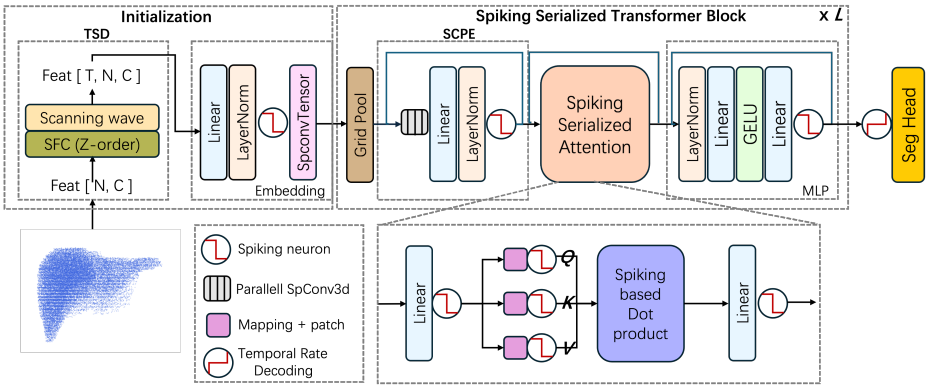

Fig. 1: The overall architecture of the Spiking Serialized Point Transformer (SSPT). The framework follows a hierarchical encoder-decoder design where the Spiking Serialized Transformer Block (SSTB) serves as the core computational unit. TSD modulates spatial points into temporal spikes, while SSA enables linear-complexity attention.

(SFC) to achieve linear complexity $O(N)$. Event-driven spiking dynamics replace floating-point multiplications with sparse accumulations (AC) to enable energy-efficient inference.

Overall Workflow. The SSPT workflow initiates with the Temporal Scanning Dispatcher (TSD), which dispatches serialized point clouds into discrete temporal sequences. The architecture follows a hierarchical spiking structure where the Spiking Serialized Transformer Block (SSTB) serves as the core computational unit. Within each SSTB, Spiking Conditional Positional Encoding (SCPE) leverages spatio-temporal parallel sparse convolution to process T time steps simultaneously while compensating for spatial information loss incurred during serialization. Subsequently, Spiking Serialized Attention (SSA) models long-range dependencies using a localized window mechanism on the serialized sequence. In the decoding phase, symmetric upsampling operators fuse multi-scale hierarchical features via skip connections to recover anatomical details. Finally, Temporal Rate Decoding (TRD) generates dense point-wise predictions by averaging spike activity across T steps. The model is supervised by CE loss plus Dice loss to optimize segmentation accuracy while mitigating class imbalance across anatomical segments.

2.2 Temporal Scanning Dispatcher

The core challenge in processing medical point clouds with Spiking Neural Networks (SNNs) is the *spatio-temporal mismatch* [24] between static 3D geometry and continuous spiking dynamics. Naive encoding via constant repetition fails to exploit the neural membrane potential's integration capability. We address this by decoupling the process into spatial serialization and temporal dispatching.

First, a Z-order curve [17] maps 3D coordinates into a 1D sequence. By utilizing bit-interleaving, the Z-order curve preserves spatial locality with significantly lower computational overhead than recursive space-filling curves like Hilbert curves. Second, the Temporal Scanning Dispatcher (TSD) partitions this serialized sequence and dispatches the resulting segments across T successive time steps. This rhythmic input enables spiking neurons to progressively integrate localized features into a spatiotemporally coherent representation. By shifting from static repetition to structured dispatching, the framework maintains high anatomical fidelity while adhering to the ultra-low-energy constraints of SNN architectures.

Scanning Mechanism. Let the point cloud be serialized into a sequence S based on the Z-order rank. At each time step $t \in \{1, \ldots, T\}$, TSD defines an active sliding window $\mathcal{W}_t$ to modulate the input:

$$X_i(t) = \begin{cases} x_i, & \text{if } \operatorname{rank}(p_i) \in \mathcal{W}_t \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where x_i represents the input features of point p_i. This scanning strategy ensures that even within a compressed temporal window (T), the point dispatcher maintains full spatial coverage while eliciting temporal dependencies across the serialized sequence. By sequentially highlighting topologically coherent neighborhoods, TSD establishes a *temporal consensus*, allowing Leaky Integrate-and-Fire (LIF) [21] neurons to integrate structured anatomical features rather than stochastic noise. This rhythmic integration reduces the entropy of spike patterns and enhances the stability of membrane potentials across the hierarchical stages.

2.3 Spiking Serialized Attention

The SSA module achieves $O(N \cdot w)$ complexity and enhanced inference efficiency by synergizing space-filling serialization with spiking dynamics.

Localized Window Serialization. To align unstructured 3D points with the spatiotemporal indexing required by SNNs, features $\mathcal{X} \in \mathbb{R}^{T \times N \times C}$ are first reordered using the Z-order ranks $\mathcal{O}$ generated by the TSD:

$$\mathcal{X}_{sorted} = \operatorname{Gather}(\mathcal{X}, \mathcal{O}) \tag{2}$$

The 1D sequence is then partitioned into non-overlapping localized windows of size w. By restricting the attention scope to these windows, the module ensures linear scaling relative to input size N, effectively bypassing the $O(N^2)$ bottleneck of global transformers.

Binary Spike Attention Core. Within each window, features are projected into Query (Q), Key (K), and Value (V) spaces. Crucially, the Query is modulated by LIF neurons to generate binary spike trains $Q_S \in \{0, 1\}^{N \times C}$. This allows the similarity score e_{ij} to be calculated via a conditional Accumulation (AC) process:

$$A_i = \sum_{j=1}^{w} \operatorname{Softmax}\left(\frac{\sum_{c=1}^{C} q_{i,c} \cdot k_{j,c}}{\sqrt{d_k}} + \mathcal{B} \right) v_j, \tag{3}$$

where $q_{i,c} \in \{0, 1\}$ represents the presence of a spike at the i-th point's c-th channel, $k_{j,c}$ is the c-th channel of the j-th key, and $\mathcal{B}$ introduces a 3D geometric prior via Relative Positional Encoding (RPE) [20]. Through this mechanism, the binary spikes act as discrete triggers that replace power-hungry floating-point multiplications with hardware-friendly additions during the $Q \cdot K^T$ similarity mapping. While the subsequent feature aggregation of Softmax maintains high-precision floating-point computation to preserve anatomical fidelity, the use of AC in the high-dimensional similarity space significantly reduces the primary computational energy budget. This hybrid approach ensures the model maintains robust structural perception for hepatic segment identification while reducing computational cost for offline preoperative planning scenarios.

The binary gating mechanism of LIF neurons serves as a dynamic threshold, prioritizing high-entropy structural regions such as vessel bifurcations and hepatic segment boundaries. By capturing anatomical topology at these critical junctions, spiking activity effectively filters redundant spatial information which enables the model to achieve high-precision segmentation.

3 Experiments

3.1 Experimental Setup

Datasets and Metrics. We evaluate the proposed SSPT on two representative liver datasets: (1) *LiTS* [1], comprising 131 CT scans for general liver and lesion segmentation; and (2) *MSD8* [15] with 191 cases annotated for intricate hepatic vascular structures. For Couinaud labels, we use the annotations released by APVN [22] for LiTS and those released by GLC-UNet [16] for MSD8. To ensure a rigorous comparison, all methods share an identical data sampling pipeline based on uniform farthest-point sampling (FPS) over the liver point cloud, without any vessel-boosted rebalancing or handcrafted sampling priors. Following APVN [22], we crop paired CT/label volumes to the Couinaud-annotated liver ROI, apply CT windowing and intensity normalization, and convert retained voxels into point samples with 3D coordinates, intensity, and labels for point-based methods. Voxel-based baselines use the same ROI-constrained normalized volumes, ensuring consistency in data source, anatomical scope, intensity normalization, and spatial preprocessing. Evaluation metrics include the pointwise *Accuracy* (Acc), the *Dice Similarity Coefficient* (Dice), and the *Average Symmetric Surface Distance* (ASD). Following APVN [22], point-wise predictions are mapped back to NIfTI label volumes using the stored voxel-index coordinates, evaluated within the annotated foreground mask, and used to compute ASD for labels 1–8 with physical voxel spacing. Theoretical energy consumption is estimated based on a 45nm CMOS process [7].

Implementation Details. For fair energy comparison, we use comparable input scales: 65,000 sampled points for point-based methods and $16 \times 64 \times 64$ volumetric inputs for voxel-based methods. All models are trained for 200 epochs using AdamW with a OneCycleLR scheduler, an initial learning rate of 5×10^{-5}, and a weight decay of 0.05. All experiments are conducted on a single NVIDIA

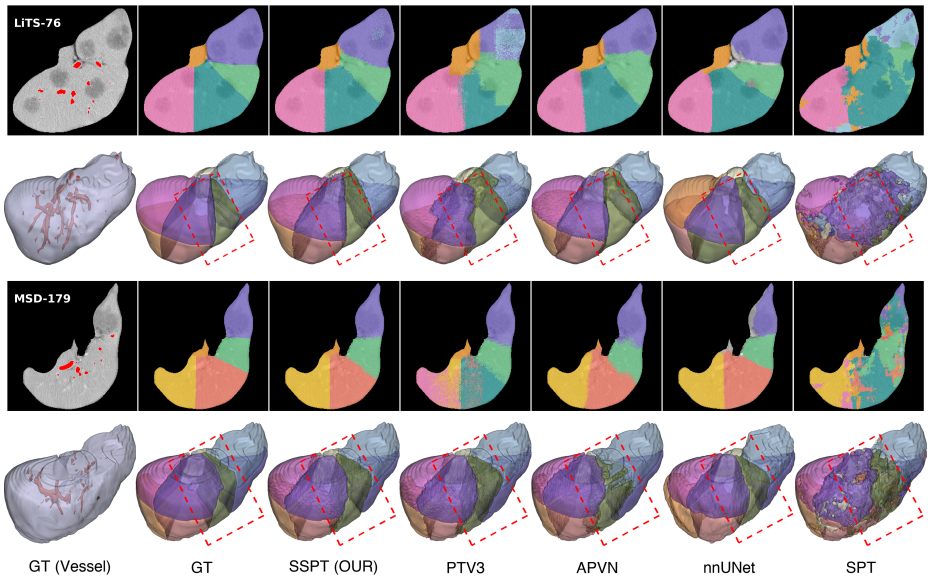

Fig. 2: Comparison of segmentation results. The first column displays key vessels in 2D and 3D views as anatomical references. Red dashed boxes highlight the precision of SSPT near segmentation planes. While PTv3 [20] (Column 4) yields high quantitative scores, 2D slices reveal its inability to define sharp boundaries. APVN [22] and nnU-Net [8] (Columns 5,6) exhibit an overall shift of the segmentation interface. SPT [18] (Columns 7) shows fragmented results due to early overfitting under the fixed-size point input setting.

GeForce RTX 3090 GPU. For practical efficiency evaluation, latency is measured as the average inference time per volume with batch size 1, excluding disk I/O, result serialization, and metric computation; GPU memory is reported as the peak CUDA allocated memory during timed inference.

3.2 Comparison with State-of-the-Art Methods

The top part of Table 1 compares SSPT with representative voxel-based, point-based, and hybrid methods on *LiTS* [1] and *MSD8* [15], including APVN as the publicly available state-of-the-art Couinaud segmentation baseline. SSPT achieves strong performance on both datasets, with 92.14% Dice / 2.44 mm ASD on *LiTS* and 90.42% Dice / 2.24 mm ASD on *MSD8*.

Fig. 2 visualizes the boundary fidelity of slices near the segmentation planes. Although PTv3 [20] demonstrates competitive Dice and Accuracy, its 2D segmentations reveal a clear weakness in delineating interfaces. Both APVN [22] and nnU-Net [8] exhibit an overall spatial shift of the segmentation boundary, failing to align with the ground truth. Point-based models with limited capacity,

Table 1: Top: quantitative comparison on *LiTS* and *MSD8*. GPU memory and latency are measured on a single RTX 3090; Energy denotes theoretical compute energy based on MAC/AC operations. Bottom: segment-wise Dice comparison of APVN [22] and SSPT (S1–S8 denote Couinaud liver segments). Subscripts denote standard deviations.

Category	Methods	LiTS					MSD8					Energy(mJ)
		Acc ↑	Dice ↑	ASD ↓	GPU(GB)	Latency(s/vol)	Acc ↑	Dice ↑	ASD ↓	GPU(GB)	Latency(s/vol)	
Voxel-based	nnU-Net[8]	$79.36_{\pm 8.66}$	$88.45_{\pm 5.79}$	$6.34_{\pm 3.27}$	**2.87**	19.294	$78.74_{\pm 9.06}$	$87.80_{\pm 6.22}$	$5.28_{\pm 5.56}$	**2.87**	4.822	1318.07
	3D UX-Net[9]	$61.34_{\pm 14.11}$	$76.82_{\pm 10.80}$	$8.13_{\pm 6.30}$	15.65	81.386	$65.40_{\pm 7.89}$	$78.84_{\pm 6.91}$	$7.34_{\pm 7.32}$	4.80	14.101	3290.75
	Swin-Unetr[6]	$66.22_{\pm 10.49}$	$78.47_{\pm 8.43}$	$6.34_{\pm 6.14}$	15.46	16.172	$72.91_{\pm 7.79}$	$83.94_{\pm 6.36}$	$5.79_{\pm 10.32}$	2.67	4.742	76.45
Hybrid	APVN(SOTA)[22]	$81.99_{\pm 6.80}$	$89.95_{\pm 4.35}$	$4.17_{\pm 2.84}$	6.72	17.264	$77.32_{\pm 7.56}$	$86.30_{\pm 4.84}$	$6.40_{\pm 3.42}$	6.08	4.438	1200.45
Point-based	PointNet++[10]	$55.60_{\pm 12.41}$	$67.35_{\pm 9.86}$	$22.72_{\pm 7.22}$	7.82	26.032	$63.23_{\pm 17.23}$	$70.14_{\pm 16.42}$	$20.65_{\pm 8.65}$	6.92	6.625	15.78
	PTv3[20] (ANN)	$83.12_{\pm 8.93}$	$91.08_{\pm 6.66}$	$3.84_{\pm 5.39}$	6.82	27.805	$82.26_{\pm 9.67}$	$89.63_{\pm 6.51}$	$3.14_{\pm 4.48}$	5.91	6.933	572.47
	SPT[18]	$64.81_{\pm 19.21}$	$75.46_{\pm 18.43}$	$14.94_{\pm 8.00}$	6.85	110.818	$67.91_{\pm 10.91}$	$80.38_{\pm 8.89}$	$16.72_{\pm 7.74}$	6.79	28.728	68.61
Ours	**SSPT (SNN)**	$\mathbf{85.55}_{\pm 4.37}$	$\mathbf{92.14}_{\pm 2.55}$	$\mathbf{2.44}_{\pm 1.21}$	5.98	**14.436**	$\mathbf{84.54}_{\pm 4.61}$	$\mathbf{90.42}_{\pm 2.71}$	$\mathbf{2.24}_{\pm 1.22}$	6.47	**3.79**	**10.97**

Dataset	Method	S1	S2	S3	S4	S5	S6	S7	S8
LiTS	APVN [22]	$89.94_{\pm 8.15}$	$91.67_{\pm 9.55}$	$86.77_{\pm 16.86}$	$82.29_{\pm 15.69}$	$80.61_{\pm 12.74}$	$92.55_{\pm 10.05}$	$96.15_{\pm 4.86}$	$93.11_{\pm 12.44}$
	SSPT	$90.05_{\pm 11.22}$	$93.41_{\pm 6.60}$	$90.41_{\pm 11.51}$	$89.11_{\pm 8.85}$	$94.89_{\pm 5.94}$	$92.48_{\pm 8.13}$	$94.52_{\pm 6.69}$	$95.38_{\pm 5.17}$
MSD8	APVN [22]	$88.16_{\pm 6.92}$	$88.66_{\pm 10.86}$	$83.33_{\pm 12.81}$	$73.27_{\pm 17.49}$	$76.85_{\pm 11.93}$	$90.45_{\pm 4.17}$	$97.81_{\pm 4.43}$	$89.09_{\pm 4.56}$
	SSPT	$88.49_{\pm 5.77}$	$96.88_{\pm 4.93}$	$93.30_{\pm 9.17}$	$84.67_{\pm 9.18}$	$77.88_{\pm 9.92}$	$90.85_{\pm 4.34}$	$98.20_{\pm 2.36}$	$89.96_{\pm 5.77}$

such as SPT [18], suffer from early overfitting during training due to the fixed-size point input setting. In contrast, the event-driven thresholding in SSPT filters boundary noise and maintains structural alignment, confirming that our method preserves high precision while significantly reducing energy consumption.

Comparison with the serialized ANN-based PTv3 [20] isolates the effects of spiking dynamics. At an equivalent scale of 12.0M parameters, SSPT achieves a lower ASD than PTv3 [20] (2.44 mm vs. 3.84 mm). By replacing floating-point Multiply-Accumulate (MAC) operations with pulse-driven Accumulate (AC) operations, the inference energy is reduced from 575.68 mJ to 10.97 mJ, a 98.1% decrease. Furthermore, SSPT demonstrates superior efficiency relative to APVN [22]. Despite utilizing significantly fewer parameters (**12.0M vs. 35.5M**), SSPT improves Accuracy on the *MSD8* [15] dataset by **7.22%**. The energy cost of 10.97 mJ is merely **0.91%** of that required by APVN [22] (1200.45 mJ). These results confirm that endogenous spiking perception via TSD and SSA is a more effective strategy for complex Couinaud segmentation than the explicit multi-representation stacking employed by hybrid SOTA architecture [22].

In practical inference, SSPT also achieves favorable latency among the strongest baselines. On MSD8, SSPT runs at 3.79 s/volume, which is 45.3% faster than PTv3 and 14.6% faster than APVN, with only a moderate increase in peak GPU memory; on LiTS, SSPT is also faster than PTv3 and APVN while using less GPU memory. To further assess segment-level robustness, the bottom part of Table 1 reports segment-wise Dice scores for APVN and SSPT. SSPT achieves consistently higher Dice on most Couinaud segments, particularly on challenging middle segments (S3–S5), indicating more balanced segment-wise performance.

3.3 Efficiency and Firing Dynamics

Following widely used theoretical energy estimation protocols in SNN studies [18], we evaluate compute energy under a 45nm CMOS technology model.

Table 2: Ablation study of SSPT on *LiTS* and *MSD8*, covering serialization, time steps, and temporal input strategies. Subscripts denote standard deviations.

Category	Variant / Setting	GPU (GB) ↓	LiTS			MSD8			Energy ↓
			Acc ↑	Dice ↑	ASD ↓	Acc ↑	Dice ↑	ASD ↓	
I. Serialized strategy	Random	**10.8**	$73.10_{\pm 11.18}$	$83.96_{\pm 7.98}$	$9.44_{\pm 8.15}$	$72.02_{\pm 9.11}$	$80.54_{\pm 10.57}$	$12.44_{\pm 4.97}$	10.97
	k-NN Graph	21.7	$69.81_{\pm 14.90}$	$80.82_{\pm 14.22}$	$10.64_{\pm 7.48}$	$67.91_{\pm 6.36}$	$80.38_{\pm 7.01}$	$11.72_{\pm 16.71}$	61.48
	Space filling curve(Z-order)	12.2	$\mathbf{85.55}_{\pm 4.37}$	$\mathbf{92.14}_{\pm 2.55}$	$\mathbf{2.44}_{\pm 1.21}$	$\mathbf{84.54}_{\pm 4.61}$	$\mathbf{90.42}_{\pm 2.71}$	$\mathbf{2.24}_{\pm 1.22}$	**10.97**
II. Timestep (T)	$T = 1$ (Lightweight)	**6.4**	$85.34_{\pm 4.43}$	$92.08_{\pm 2.56}$	$2.54_{\pm 1.14}$	$83.71_{\pm 5.36}$	$\mathbf{90.83}_{\pm 6.62}$	$2.63_{\pm 4.19}$	**5.12**
	$T = 2$ **(Balanced)**	12.2	$85.55_{\pm 4.37}$	$92.14_{\pm 2.55}$	$2.44_{\pm 1.21}$	$84.54_{\pm 4.61}$	$90.42_{\pm 2.71}$	$2.24_{\pm 1.22}$	10.97
	$T = 3$ (Limit)	22.8	$\mathbf{85.56}_{\pm 4.83}$	$\mathbf{92.18}_{\pm 2.84}$	$\mathbf{2.43}_{\pm 1.19}$	$\mathbf{84.60}_{\pm 9.67}$	$90.73_{\pm 6.51}$	$\mathbf{2.20}_{\pm 4.48}$	15.50
III. Strategy	Repeat (Constant)	12.2	$83.34_{\pm 4.90}$	$89.08_{\pm 4.22}$	$2.84_{\pm 3.76}$	$82.71_{\pm 6.79}$	$89.83_{\pm 4.36}$	$2.63_{\pm 1.37}$	15.56
	Interlaced (Pattern)	12.2	$85.37_{\pm 4.63}$	$92.05_{\pm 2.72}$	$2.59_{\pm 1.15}$	$83.22_{\pm 7.56}$	$\mathbf{90.65}_{\pm 4.84}$	$2.88_{\pm 3.42}$	10.97
	Scanning Wave (TSD)	**12.2**	$\mathbf{85.55}_{\pm 4.37}$	$\mathbf{92.14}_{\pm 2.55}$	$\mathbf{2.44}_{\pm 1.21}$	$\mathbf{84.54}_{\pm 4.61}$	$90.42_{\pm 2.71}$	$\mathbf{2.24}_{\pm 1.22}$	**10.97**

For ANN baselines, energy is estimated from the total number of MAC operations. For SNN models, the first layer is computed with MAC operations, while subsequent spiking layers are computed with firing-rate-weighted AC operations. Total energy per inference is calculated from Synaptic Operations (SOPs), distinguishing between MAC operations ($E_{MAC} = 4.6\,pJ$ [7]) and AC operations ($E_{AC} = 0.9\,pJ$ [7]). The calculation follows:

$$E_{Total} = E_{First} + \sum_{l=2}^{L} \zeta_l \cdot SOPs^{(l)} \cdot E_{AC} \tag{4}$$

where ζ_l is the firing rate at stage l. The measured inference cost of SSPT is **10.97 mJ** per forward pass. Compared to hybrid and voxel-based architectures, SSPT saves 99.1% energy compared to APVN [22] (1203.35 mJ) and 99.2% compared to nnU-Net [8] (1318.08 mJ). Even compared to the smaller-scale PointNet++ [10] (15.86 mJ), SSPT reduces energy by 30.8%.

Efficiency is attributed to the sparsity of spiking activity. The weighted average firing rate is recorded at **3.30%** across 145 spiking layers. While specific CPE layers maintain firing rates near 20% to preserve geometric detail, the majority of attention and MLP layers operate at lower activity levels. This event-driven sparsity shifts the computing workload from dense MACs to sparse ACs [11, 19].

3.4 Ablation Studies

Table 2 evaluates the impact of topological encoding and temporal steps. Category I confirms that Z-order curves yield the highest Accuracy (85.55%) and the lowest ASD (**2.24 mm** on *MSD8* [15]). Compared to k-NN graphs (21.74 GB), the Z-order strategy reduces memory to 12.2 GB while preserving spatial locality. Random serialization, which ignores topological proximity, results in a significantly higher ASD of 12.44 mm on *MSD8* [15].

Temporal analysis in Category II suggests that $T = 2$ provides the optimal balance. While $T = 1$ shows a higher boundary error (**2.63 mm** ASD on *MSD8* [15]), $T = 3$ increases memory consumption to 22.8 GB without proportional gains in Dice scores. Category III compares input strategies, where the Scanning Wave (TSD) achieves the lowest ASD of 2.24 mm on *MSD8* [15]. Unlike the Repeat or Interlaced patterns, the rhythmic input of TSD better stabilizes

membrane potentials for sparse signal detection. This confirms that topologically consistent scanning waves are essential for delineating complex vascular structures.

4 Conclusion

This work presents SSPT, a spiking framework for 3D medical point cloud segmentation. Evaluation on *LiTS* [1] and *MSD8* [15] datasets demonstrates that SSPT achieves a landmark ASD of **2.24 mm** on complex vascular structures, surpassing conventional voxel-based and hybrid architectures in anatomical precision. The transition to pulse-driven AC operations reduces inference energy to 10.97 mJ, a 98.1% reduction from ANN-based counterparts. The 3.30% sparse firing rate effectively preserves complex topology while minimizing computational redundancy. These results confirm that spiking perception is a viable, high-precision solution for offline preoperative Couinaud segmentation and surgical planning. While efficient, our model relies on initial sampling that may bypass fine vessels. To improve boundary precision, future research will investigate adaptive sampling and multi-modal fusion in clinical settings.

Acknowledgments. The authors acknowledge support from the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T45-401/22-N and by the Hong Kong Innovation and Technology Fund, under Project GHP/252/23SZ.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., et al.: The liver tumor segmentation benchmark (lits). Medical image analysis **84**, 102680 (2023)
2. Bismuth, H.: Surgical anatomy and anatomical surgery of the liver. World journal of surgery **6**(1), 3–9 (1982)
3. Couinaud, C.: Liver anatomy: portal (and suprahepatic) or biliary segmentation. Digestive surgery **16**(6), 459–467 (1999)
4. Fan, W., Fang, H., Li, R., Lin, Y., An, C., Luo, X.: Anatomy-aware frequency-attention transformer networks for liver couinaud ct/mr segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 55–65. Springer (2025)
5. Gotra, A., Sivakumaran, L., Chartrand, G., Vu, K.N., Vandenbroucke-Menu, F., Kauffmann, C., Kadoury, S., Gallix, B., de Guise, J.A., Tang, A.: Liver segmentation: indications, techniques and future directions. Insights into imaging **8**(4), 377–392 (2017)
6. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)

7. Horowitz, M.: 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)
8. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
9. Lee, H.H., Bao, S., Huo, Y., Landman, B.A.: 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. arXiv preprint arXiv:2209.15076 (2022)
10. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
11. Qiu, X., Yao, M., Zhang, J., Chou, Y., Qiao, N., Zhou, S., Xu, B., Li, G.: Efficient 3d recognition with event-driven spike sparse convolution (2024). https://doi.org/10.48550/arXiv.2412.07360
12. Qiu, Y., Li, Y., Tong, Y., Liu, Q., Yang, Q., Qiu, S., Heng, P.A., Fu, C.W.: Toward synergistic learning for liver vessel and couinaud segmentations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer (2026)
13. Qiu, Y., Qiu, S., Heng, P.A., Fu, C.W.: Uniliver: Gradient-conditioned unified model for multi-target hepatic segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer (2026)
14. Shaker, A., Maaz, M., Rasheed, H., Khan, S., Yang, M., Khan, F.: " unetr++: Delving into efficient and accurate 3d medical image segmentation", ieee transactions on medical imaging (2024)
15. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
16. Tian, J., Liu, L., Shi, Z., Xu, F.: Automatic couinaud segmentation from ct volumes on liver using glc-unet. In: Machine Learning in Medical Imaging. pp. 274–282. Springer (2019). https://doi.org/10.1007/978-3-030-32692-0_32
17. Wang, P.S.: Octformer: Octree-based transformers for 3d point clouds. ACM Transactions on Graphics (TOG) **42**(4), 1–11 (2023)
18. Wu, P., Chai, B., Li, H., Zheng, M., Peng, Y., Wang, Z., Nie, X., Zhang, Y., Sun, X.: Spiking point transformer for point cloud classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 21563–21571 (2025)
19. Wu, P., Chai, B., Zheng, M., Li, W., Hu, Z., Chen, J., Zhang, Z., Li, H., Sun, X.: Efficient spiking point mamba for point cloud analysis (2025). https://doi.org/10.48550/arXiv.2504.14371
20. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point transformer v3: Simpler faster stronger. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4840–4851 (2024)
21. Wu, Y., Deng, L., Li, G., Zhu, J., Shi, L.: Spatio-temporal backpropagation for training high-performance spiking neural networks. CoRR **abs/1706.02609** (2017), http://arxiv.org/abs/1706.02609
22. Zhang, X., Ali, S., Liu, T., Zhao, X., Cui, Z., Han, M., Ma, S., Zhu, J., Kang, Y., Wang, L., et al.: Robust and smooth couinaud segmentation via anatomical structure-guided point-voxel network. Computers in Biology and Medicine **182**, 109202 (2024)

23. Zhang, X., Liu, Y., Ali, S., Zhao, X., Sun, M., Han, M., Liu, T., Zhai, P., Cui, Z., Zhang, P., et al.: Anatomical-aware point-voxel network for couinaud segmentation in liver ct. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 465–474. Springer (2023)
24. Zhou, C., Yu, L., Zhou, Z., Zhang, H., Ma, Z., Zhou, H., Tian, Y.: Spikingformer: Spike-driven residual learning for transformer-based spiking neural network (2023). https://doi.org/10.48550/arXiv.2304.11954