

SpikeTrack: A Spike-driven Framework for Efficient Visual Tracking

Supplementary Material

In the supplementary materials, Sec. 1 presents a visual comparison of the tracking results. Sec. 2 presents more exploring experiments. Sec. 3 presents SpikeTrack’s performance on each attribute in the LaSOT Benchmark’s attribute challenge. Finally, Sec. 4 explains the method for estimating energy consumption and reports the spiking firing rate (SFR) for each layer of SpikeTrack-B256-T3.

1. Visualization

1.1. Visualization Retrieval Results

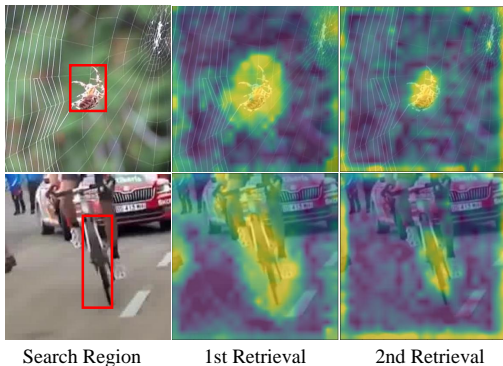


Figure 1. Visualization of the spike tensor obtained after two retrievals in MRM, based on spike firing rate.

To further demonstrate the effectiveness of recurrent retrieval, we visualize the retrieval results at each iteration based on the per-pixel channel spike firing rate. Fig. 1 demonstrate a progressive attention refinement process, with increasingly concentrated spike firing locations and reduced noise, validating the recurrent design’s effectiveness. This result also supports our hypothesis regarding the experimental observations in Fig. ??: excessive recurrent iterations narrow the attention excessively, causing useful information to be overlooked.

1.2. Visualization Tracking Results

As shown in Fig. 2, we present the tracking results of SpikeTrack, efficiency-oriented ANNs, and precision-oriented ANNs. The video sequences include challenging scenarios such as deformation, blur, and similar objects, demonstrating SpikeTrack’s ability to maintain accurate tracking over extended time spans.

Table 1. Template update parameter sensitivity analysis of SpikeTrack-B256-T3. Rows indicate update confidence and columns indicate update interval. Underlines indicate default unified settings of the SpikeTrack Family for short- and long-term datasets.

GOT	<u>25</u>	35	45	55	LaSOT	<u>40</u>	60	80
0.65	72.2	72.5	72.2	72.1	0.65	66.8	66.8	67.1
0.7	72.2	72.2	72.0	71.9	0.7	67.0	67.2	67.1
0.75	71.8	72.4	71.6	71.8	0.75	66.7	67.6	66.7
0.8	71.7	72.3	71.3	71.7	<u>0.8</u>	67.1	67.5	66.8

2. More Exploring Experiments

2.1. Updating Hyper-parameters

We conduct a sensitivity analysis about updating hyper-parameters. As shown in Tab. 1, within reasonable ranges, lower update intervals and thresholds favor short-term tracking, whereas long-term tracking exhibits the opposite trend, suggesting a stronger dependence on stable templates. Note that our default setting isn’t optimal since we use unified settings for all datasets (except LaSOT) to show true generalization.

3. Performance of Attribute Challenges on LaSOT

As shown in Fig. 3, we provide a detailed comparison of the scores achieved by SpikeTrack and other ANN tracking methods for each challenge attribute in LaSOT. These challenge attributes include: Illumination Variation (IV), Partial Occlusion (POC), Deformation (DEF), Motion Blur (MB), Camera Motion (CM), Rotation (ROT), Background Clutter (BC), Viewpoint Change (VC), Scale Variation (SV), Full Occlusion (FOC), Fast Motion (FM), Out-of-View (OV), Low Resolution (LR), Aspect Ratio Change (ARC).

4. Energy Consumption Estimation

4.1. Spike-driven Operators in SNNs

Spike-driven operators are the foundation of low-power neuromorphic computing, especially for SNNs. In spike-driven convolution and MLP computations, matrix multiplication between weight matrices and input spike matrices can be implemented on neuromorphic chips as address-based additions [1]. In spike-driven attention mechanisms, the Q_S , K_S , and V_S operations involve two matrix multiplications. Similar to Conv and MLP, selecting one side as the spike matrix and the other as the weight matrix transforms the computation into sparse additions. Table 2 shows a comparison of the energy consumption of ANN and Spike-driven SNN operators.



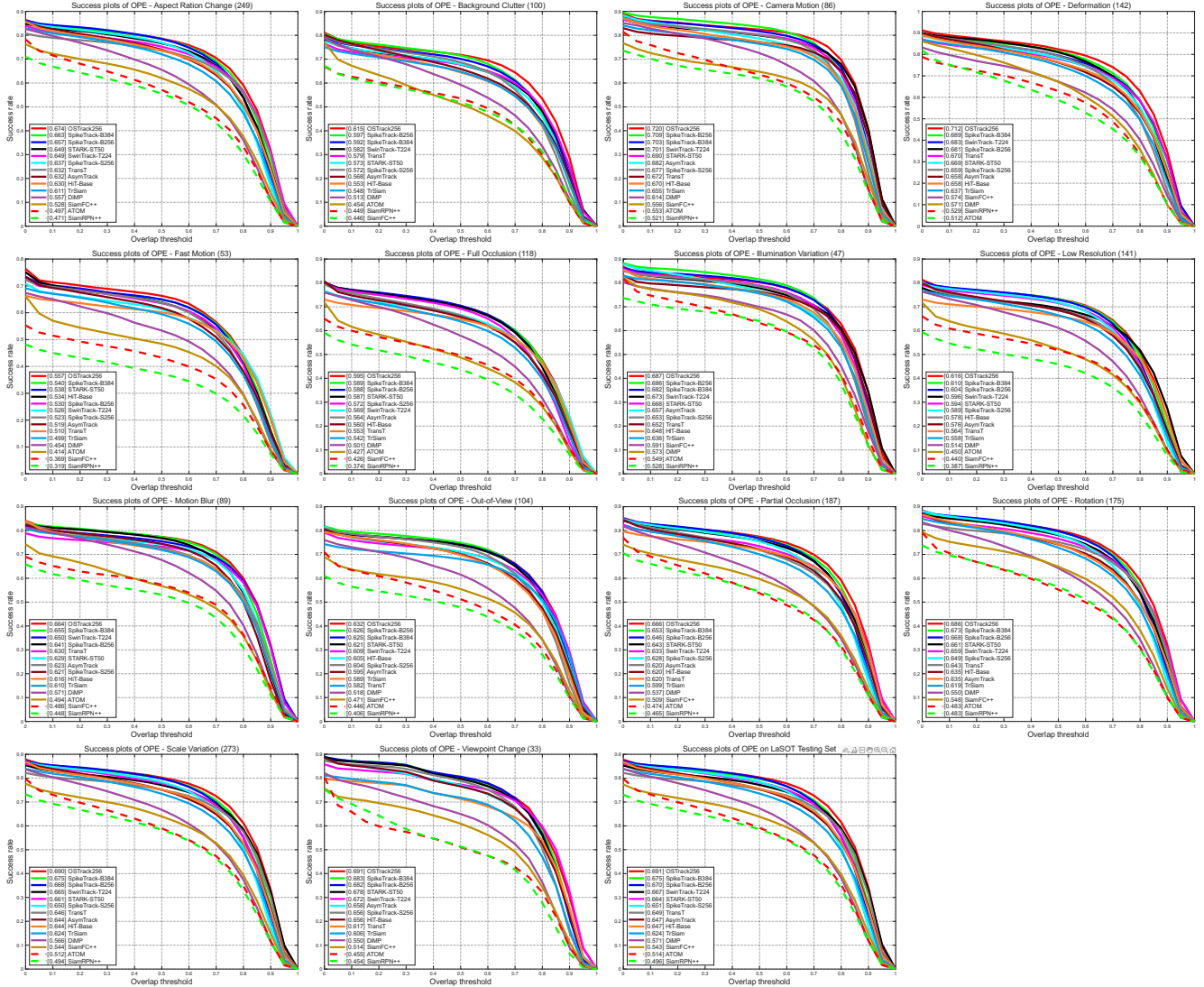
Figure 2. Visualization comparison of SpikeTrack and other ANN-based Trackers.

4.2. Energy Consumption of SpikeTrack

In this work, we use the same refined energy consumption evaluation method as recent SNN studies [3–8]. We first measure the spiking firing rate (the proportion of non-zero elements in the spike matrix, SFR) of each layer, then calculate each layer’s energy consumption as its FLOPs multiplied by E_{AC} and the corresponding SFR , and finally sum the energy consumption across all layers. In 45nm technology [2], the energy of a MAC and an AC are $E_{MAC} = 4.6pJ$ and $E_{AC} = 0.9pJ$, respectively.

The SFR of each layer is obtained by averaging the fir-

ing rates of the model over the large-scale benchmarks LaSOT and GOT-10K. In order to give readers an intuitive feeling about the spiking firing rate, we give the detailed spiking firing rates of SpikeTrack-B256-T3 in Tab. 3 (search branch) and Tab. 4 (template branch). Overall, the firing rates exhibit a decreasing trend from shallow to deep layers. The TemporalFusion-WeightMLP1 layer in the MRM module shows a relatively high firing rate, indicating that while the spatiotemporal membrane potential processing and recurrent retrieval introduced by the MRM module improve performance, they also increase energy consumption



to some extent.

Notably, the NI-LIF neurons [3] are trained with normalized integer activation values, with the max integer capped at 4. During inference, these integers are mapped to equivalent spikes, which can naturally result in certain layers exhibiting an *SFR* greater than 1.

More SpikeTrack variations of spiking firing rate and energy consumption calculation tables can be found in the Excel document in the supplementary material ZIP.

Table 2. Energy evaluation. FL_{Conv} and FL_{MLP} represent the FLOPs of the Conv and MLP models in the ANNs, respectively. R denote the spiking firing rate (the proportion of non-zero elements in the spike matrix) of the layer corresponding to the operator. T is timestep. E_{MAC} and E_{AC} are the energy consumption of performing one MAC and one AC operation, respectively. In the spike-driven attention, the scale operation is avoided by incorporating the scale factor into the neuron’s leakage factor. The first downsampling layer uses MAC convolution, consistent with other works [4–8].

		ANN-based Tracker	Spike-driven Tracker
Conv	First Conv	$E_{MAC} \cdot FL_{Conv}$	$E_{MAC} \cdot T \cdot R \cdot FL_{Conv}$
	Other Conv	$E_{MAC} \cdot FL_{Conv}$	$E_{AC} \cdot T \cdot R \cdot FL_{Conv}$
Self-Attention	Q, K, V	$E_{MAC} \cdot 3ND^2$	$E_{AC} \cdot T \cdot R \cdot 3FL_{Conv}$
	$f(Q, K, V)$	$E_{MAC} \cdot 2N^2D$	$E_{AC} \cdot T \cdot R \cdot 2ND^2$
	Scale	$E_M \cdot N^2$	-
	Softmax	$E_{MAC} \cdot 2N^2$	-
	Linear	$E_{MAC} \cdot FL_{MLP}$	$E_{AC} \cdot T \cdot R \cdot FL_{MLP0}$
MLP	Layer1	$E_{MAC} \cdot FL_{MLP1}$	$E_{AC} \cdot T \cdot R \cdot FL_{MLP1}$
	Layer2	$E_{MAC} \cdot FL_{MLP2}$	$E_{AC} \cdot T \cdot R \cdot FL_{MLP2}$

Table 3. Layer spiking firing rates of SpikeTrack-B256-T3 Search Branch.

			T1	T2	T3	
Stage 1	DownSampling		Conv	1	-	
	ConvFormer Spike Block	SepSpikeConv	PWConv1	1.4643	-	
			DWConv	1.3569	-	
		Channel Conv	PWConv2	1.2556	-	
			Conv1	1.5301	-	
		Conv2	0.2497	-		
Stage 2	DownSampling		Conv	1.0722	-	
	Memory Retrieval Module	Head- Q	Head- Q	0.5670	-	
			Q_S1	0.3496	0.3841	
		SepSpikeConv	PWConv	1.1416	1.1947	
			DWConv	0.3178	0.3532	
		Project1	Conv	0.4639	0.4746	
			Q_S2	0.6606	0.7157	
		TemporalFusion	WeightMLP1	1.7143	1.6588	
			WeightMLP2	0.5168	0.5279	
		Project2	Conv	0.3493	-	
			Linear1	0.6840	-	
		Channel MLP	Linear2	0.2197	-	
		ConvFormer Spike Block	SepSpikeConv	PWConv1	0.6638	-
				DWConv	0.9398	-
	Channel Conv		PWConv2	0.6144	-	
			Conv1	0.9048	-	
			Conv2	0.1488	-	
	Stage 3	DownSampling		Conv	0.7663	-
		Memory Retrieval Module	Head- Q	Head- Q	0.6318	-
				Q_S1	0.2818	0.3085
			SepSpikeConv	PWConv	0.8162	0.9758
				DWConv	0.3054	0.3400
			Project1	Conv	0.3661	0.3896
				Q_S2	0.3456	0.3681
TemporalFusion			WeightMLP1	0.7749	0.8173	
			WeightMLP2	0.3484	0.3827	
Project2			Conv	0.1389	-	
			Linear1	0.6861	-	
Channel MLP			Linear2	0.1752	-	
ConvFormer Spike Block			SepSpikeConv	PWConv1	0.6677	-
				DWConv	0.9180	-
		Channel Conv	PWConv2	0.5269	-	
			Conv1	0.6199	-	
			Conv2	0.0906	-	
ConvFormer Spike Block		SepSpikeConv	PWConv1	0.9326	-	
			DWConv	0.5476	-	
		Channel Conv	PWConv2	0.4813	-	
			Conv1	0.6739	-	
			Conv2	0.065	-	
Stage 4		DownSampling		Conv	0.8766	-
		Memory Retrieval Module	Head- Q	Head- Q	0.4919	-
	Q_S1			0.3510	0.3837	
	SepSpikeConv		PWConv	2.1617	2.3470	
			DWConv	0.3806	0.4072	

	Project1	Conv	0.4583	0.4769	0.5183
		Q_S2	0.7903	0.8414	0.8667
	TemporalFusion	WeightMLP1	2.8883	2.9805	3.0267
		WeightMLP2	0.2744	0.2848	0.2924
	Project2	Conv	0.7324	-	-
	Channel MLP	Linear1	0.5211	-	-
		Linear2	0.1942	-	-
TransFormer Spike Block	SepSpikeConv	PWConv1	0.5216	-	-
		DWConv	0.8427	-	-
		PWConv2	0.3655	-	-
		Head- QKV	0.5699	-	-
	SSA	Q_S	1.0672	-	-
		K_S	0.2967	-	-
		V_S	0.7571	-	-
		Linear	2.1776	-	-
	Channel MLP	Linear1	0.6194	-	-
		Linear2	0.1132	-	-
TransFormer Spike Block	SepSpikeConv	PWConv1	0.7469	-	-
		DWConv	0.7103	-	-
		PWConv2	0.4588	-	-
		Head- QKV	0.6988	-	-
	SSA	Q_S	0.7085	-	-
		K_S	0.2105	-	-
		V_S	0.3949	-	-
		Linear	1.3788	-	-
	Channel MLP	Linear1	0.7699	-	-
		Linear2	0.1025	-	-
TransFormer Spike Block	SepSpikeConv	PWConv1	0.7743	-	-
		DWConv	0.5849	-	-
		PWConv2	0.3319	-	-
		Head- QKV	0.7370	-	-
	SSA	Q_S	0.7367	-	-
		K_S	0.1592	-	-
		V_S	0.5306	-	-
		Linear	1.2955	-	-
	Channel MLP	Linear1	0.8817	-	-
		Linear2	0.0799	-	-
Memory Retrieval Module	SepSpikeConv	Head- Q	0.8867	-	-
		Q_S1	0.3233	0.3501	0.3357
		PWConv	1.2107	1.4589	1.424
		DWConv	0.3771	0.4003	0.3844
	Project1	Conv	0.4302	0.4495	0.4991
		Q_S2	0.5933	0.6266	0.6632
	TemporalFusion	WeightMLP1	1.7760	1.9401	2.0085
		WeightMLP2	1.0628	1.1578	1.1791
	Project2	Conv	0.4434	-	-
	Channel MLP	Linear1	0.9183	-	-
		Linear2	0.2099	-	-
TransFormer Spike Block	SepSpikeConv	PWConv1	0.9292	-	-
		DWConv	0.6615	-	-
		PWConv2	0.3422	-	-
		Head- QKV	0.8495	-	-
	SSA	Q_S	0.7161	-	-

Stage 5	TransFormer Spike Block	Channel MLP	K_S	0.1350	-	-
			V_S	0.4256	-	-
			Linear	1.0989	-	-
			Linear1	0.9172	-	-
			Linear2	0.0677	-	-
		SepSpikeConv	PWConv1	0.8158	-	-
			DWConv	0.6410	-	-
			PWConv2	0.345	-	-
			Head- QKV	0.7505	-	-
			Q_S	0.7999	-	-
	SSA	K_S	0.1222	-	-	
		V_S	0.3271	-	-	
		Linear	0.8896	-	-	
		Linear1	0.9085	-	-	
		Linear2	0.0744	-	-	
	TransFormer Spike Block	SepSpikeConv	PWConv1	0.7901	-	-
			DWConv	0.6209	-	-
			PWConv2	0.2965	-	-
			Head- QKV	0.7643	-	-
			Q_S	0.7730	-	-
		SSA	K_S	0.1993	-	-
			V_S	0.5176	-	-
			Linear	1.6896	-	-
			Linear1	0.8989	-	-
			Linear2	0.0871	-	-
	Memory Retrieval Module	SepSpikeConv	Head- Q	1.0431	-	-
			Q_S1	0.4587	0.4759	0.4673
			PWConv	1.5555	1.5786	1.4752
			DWConv	0.3697	0.3760	0.3707
			Project1	Conv	0.5233	0.5438
		TemporalFusion	Q_S2	0.3806	0.4307	0.4813
			WeightMLP1	1.2338	1.489	1.5396
WeightMLP2			0.9094	1.0829	1.1119	
Project2			Conv	0.2615	-	-
Channel MLP			Linear1	1.0764	-	-
TransFormer Spike Block	SepSpikeConv	Linear2	0.1975	-	-	
		PWConv1	1.1655	-	-	
		DWConv	0.5649	-	-	
		PWConv2	0.5767	-	-	
		Head- QKV	0.9361	-	-	
	SSA	Q_S	0.7536	-	-	
		K_S	0.0724	-	-	
		V_S	0.1860	-	-	
		Linear	0.5488	-	-	
		Channel MLP	Linear1	1.0615	-	-
TransFormer Spike Block	SepSpikeConv	Linear2	0.0453	-	-	
		PWConv1	0.8718	-	-	
		DWConv	0.2728	-	-	
		PWConv2	0.2346	-	-	
		Head- QKV	0.4639	-	-	
	SSA	Q_S	1.0436	-	-	
		K_S	0.0690	-	-	

	Channel MLP	V_S	0.0402	-	-
		Linear	0.2674	-	-
		Linear1	0.1904	-	-
		Linear2	0.0043	-	-
	Memory Retrieval Module	Head- Q	0.2893	-	-
		Q_S1	0.2288	0.2511	0.2457
		PWConv	1.4003	1.5141	1.4140
		DWConv	0.2876	0.3126	0.3128
		Project1	Conv	0.3465	0.3634
		Q_S2	0.5225	0.5534	0.5321
		TemporalFusion	WeightMLP1	2.3932	2.5136
			WeightMLP2	0.7827	0.8143
		Project2	Conv	0.5717	-
		Channel MLP	Linear1	0.3105	-
			Linear2	0.2024	-
Head	Location	Conv1	0.3149	-	-
		Conv2	0.2198	-	-
		Conv3	0.3037	-	-
		Conv4	0.6767	-	-
		Conv5	1.8021	-	-
	Offset	Conv1	0.3149	-	-
		Conv2	0.2182	-	-
		Conv3	0.3431	-	-
		Conv4	0.5206	-	-
		Conv5	0.6281	-	-
	Size	Conv1	0.3149	-	-
		Conv2	0.2271	-	-
		Conv3	0.3000	-	-
		Conv4	0.6038	-	-
		Conv5	0.5048	-	-

Table 4. Layer spiking firing rates of SpikeTrack-B256-T3 Template Branch.

			T1	T2	T3
Stage 1	DownSampling	Conv	1	1	1
	ConvFormer Spike Block	SepSpikeConv	PWConv1	1.458	1.4853
			DWConv	1.3534	1.3790
			PWConv2	1.2563	1.3188
		Channel Conv	Conv1	1.5297	1.5035
			Conv2	0.2500	0.2496
Stage 2	DownSampling	Conv	1.0752	1.0885	1.0787
	Memory Retrieval Module	Head- KV	0.5344	0.5987	0.5938
		V_S	0.3571	0.3961	0.3985
		K_S	0.2998	0.3259	0.3188
	ConvFormer Spike Block	SepSpikeConv	PWConv1	0.6596	0.6715
			DWConv	0.9408	0.9589
			PWConv2	0.6114	0.6465
		Channel Conv	Conv1	0.9114	0.9267
			Conv2	0.1480	0.1460
Stage 3	DownSampling	Conv	0.7826	0.7730	0.7694
	Memory Retrieval Module	Head- KV	0.6126	0.6834	0.6718
		V_S	0.3057	0.3386	0.3369

Stage 4	ConvFormer Spike Block	SepSpikeConv	K_S	0.3060	0.3400	0.3313		
			PWConv1	0.6688	0.6766	0.6739		
			DWConv	0.9315	0.9402	0.9380		
		Channel Conv	PWConv2	0.5135	0.5246	0.5213		
			Conv1	0.6318	0.6374	0.6359		
			Conv2	0.0906	0.0903	0.0888		
	ConvFormer Spike Block	SepSpikeConv	PWConv1	0.9604	0.9774	0.9793		
			DWConv	0.5758	0.5904	0.5886		
			PWConv2	0.4773	0.4883	0.4884		
		Channel Conv	Conv1	0.6978	0.6975	0.7002		
			Conv2	0.0694	0.0680	0.0678		
			DownSampling		Conv	0.9081	0.9001	0.8983
	Memory Retrieval Module	Head- KV	0.4755	0.5012	0.5137			
		V_S	0.4067	0.4387	0.4539			
		K_S	0.5307	0.5693	0.5842			
	TransFormer Spike Block	SepSpikeConv	PWConv1	0.4915	0.5026	0.5066		
			DWConv	0.8554	0.8717	0.8803		
			PWConv2	0.3502	0.3697	0.3749		
			Head- QKV	0.5477	0.5811	0.5833		
		SSA	Q_S	1.1002	1.2265	1.2395		
			K_S	0.2663	0.2994	0.3055		
			V_S	0.7438	0.7834	0.7932		
			Linear	2.1391	2.3258	2.3621		
		Channel MLP	Linear1	0.6040	0.5976	0.5933		
			Linear2	0.0985	0.0980	0.0974		
			TransFormer Spike Block	SepSpikeConv	PWConv1	0.7451	0.7327	0.7315
					DWConv	0.7205	0.7095	0.7085
	PWConv2	0.4378			0.4552	0.4575		
Head- QKV	0.7029	0.7192			0.7023			
SSA	Q_S	0.7605		0.7370	0.7062			
	K_S	0.2392		0.2287	0.2077			
	V_S	0.4056		0.4129	0.4003			
	Linear	1.6298		1.5722	1.3865			
Channel MLP	Linear1	0.7850		0.7871	0.7767			
	Linear2	0.1052		0.1089	0.1039			
TransFormer Spike Block	SepSpikeConv	PWConv1	0.7683	0.7498	0.7574			
		DWConv	0.5794	0.5768	0.5805			
		PWConv2	0.3307	0.3437	0.3452			
		Head- QKV	0.7291	0.7302	0.7277			
	SSA	Q_S	0.7556	0.7416	0.7483			
		K_S	0.1557	0.1395	0.1411			
		V_S	0.5425	0.5424	0.5408			
		Linear	1.3957	1.3158	1.3046			
	Channel MLP	Linear1	0.8744	0.8590	0.8692			
		Linear2	0.0735	0.0722	0.0723			
Memory Retrieval Module	Head- KV	0.8776	0.8964	0.9105				
	V_S	0.2765	0.3190	0.3222				
	K_S	0.2897	0.2831	0.2953				
TransFormer Spike Block	SepSpikeConv	PWConv1	0.8760	0.8442	0.8623			
		DWConv	0.6305	0.6225	0.6328			
		PWConv2	0.3151	0.3138	0.3219			
		Head- QKV	0.8052	0.8112	0.8115			
	SSA	Q_S	0.6392	0.6850	0.6665			

Stage 5	TransFormer Spike Block	Channel MLP	K_S	0.1438	0.1365	0.1354
			V_S	0.4303	0.4325	0.4365
			Linear	1.1670	1.2427	1.2174
			Linear1	0.9013	0.8820	0.8989
			Linear2	0.0616	0.0599	0.0602
	TransFormer Spike Block	SepSpikeConv	PWConv1	0.7866	0.7636	0.7793
			DWConv	0.6259	0.6263	0.6347
			PWConv2	0.3305	0.3450	0.3529
			Head- QKV	0.7390	0.7483	0.7388
		SSA	Q_S	0.7814	0.8434	0.8010
			K_S	0.1488	0.1365	0.1227
			V_S	0.3419	0.3411	0.3399
			Linear	0.9959	1.0392	0.9468
		Channel MLP	Linear1	0.8848	0.8524	0.8672
			Linear2	0.0808	0.0799	0.0751
	TransFormer Spike Block	SepSpikeConv	PWConv1	0.7143	0.6833	0.7108
			DWConv	0.5917	0.5949	0.5955
			PWConv2	0.2799	0.2859	0.2897
			Head- QKV	0.7315	0.7115	0.7250
		SSA	Q_S	0.8245	0.8770	0.8687
			K_S	0.1935	0.1864	0.1898
			V_S	0.4987	0.4917	0.5064
			Linear	1.8150	1.8574	1.8907
		Channel MLP	Linear1	0.8072	0.7669	0.8002
			Linear2	0.1164	0.1183	0.1128
	DownSampling		Conv	0.3914	0.3706	0.3886
		Memory Retrieval Module	Head- KV	1.0682	1.0843	1.0665
			V_S	0.1857	0.1970	0.1987
	K_S		0.2269	0.2089	0.2198	
	TransFormer Spike Block	SepSpikeConv	PWConv1	1.0331	1.0798	1.0915
			DWConv	0.4430	0.4637	0.4686
			PWConv2	0.3645	0.4348	0.4429
			Head- QKV	0.8320	0.8420	0.8145
		SSA	Q_S	0.9074	0.9703	1.0095
			K_S	0.0435	0.0389	0.0375
			V_S	0.1377	0.1324	0.1297
			Linear	0.5249	0.4881	0.4730
		Channel MLP	Linear1	0.9349	0.9175	0.9127
			Linear2	0.0181	0.01697	0.0161
	TransFormer Spike Block	SepSpikeConv	PWConv1	0.7944	0.7935	0.7932
			DWConv	0.1714	0.1775	0.1778
			PWConv2	0.1576	0.1701	0.1740
			Head- QKV	0.3908	0.3597	0.3534
		SSA	Q_S	0.5916	0.6136	0.6324
			K_S	0.0311	0.0302	0.0299
			V_S	0.0268	0.0255	0.0246
			Linear	0.0421	0.0446	0.0520
		Channel MLP	Linear1	0.3862	0.3311	0.3099
			Linear2	0.0091	0.0077	0.0068
	Memory Retrieval Module		Head- KV	0.8044	0.7835	0.7841
		V_S	0.3875	0.3687	0.3515	
		K_S	0.3411	0.3344	0.3173	

References

- [1] Charlotte Frenkel, David Bol, and Giacomo Indiveri. Bottom-up and top-down approaches for the design of neuromorphic processing systems: Tradeoffs and synergies between natural and artificial intelligence. *Proceedings of the IEEE*, 111(6): 623–652, 2023. [1](#)
- [2] 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)*, pages 10–14. IEEE, 2014. [2](#)
- [3] Zhenxin Lei, Man Yao, Jiakui Hu, Xinhao Luo, Yanye Lu, Bo Xu, and Guoqi Li. Spike2former: Efficient spiking transformer for high-performance image segmentation. In *AAAI*, pages 1364–1372, 2025. [2](#), [3](#)
- [4] Yimeng Shan, Zhenbang Ren, Haodi Wu, Wenjie Wei, Ruijie Zhu, Shuai Wang, Dehao Zhang, Yichen Xiao, Jieyuan Zhang, Kexin Shi, et al. Sdtrack: A baseline for event-based tracking via spiking neural networks. *arXiv preprint arXiv:2503.08703*, 2025. [4](#)
- [5] Jingjun Yang, Liangwei Fan, Jinpu Zhang, Xiangkai Lian, Hui Shen, and Dewen Hu. Fully spiking neural networks for unified frame-event object tracking. *arXiv preprint arXiv:2505.20834*, 2025.
- [6] Man Yao, Jiakui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer. *NeurIPS*, 36: 64043–64058, 2023.
- [7] 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 TPAMI*, 2025.
- [8] Shihao Zou, Qingfeng Li, Wei Ji, Jingjing Li, Yongkui Yang, Guoqi Li, and Chao Dong. Spikevideoformer: An efficient spike-driven video transformer with hamming attention and $\mathcal{O}(t)$ complexity. *arXiv preprint arXiv:2505.10352*, 2025. [2](#), [4](#)