Decision SpikeFormer: Spike-Driven Transformer for Decision Making

Wei Huang Qinying Gu * Nanyang Ye *

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

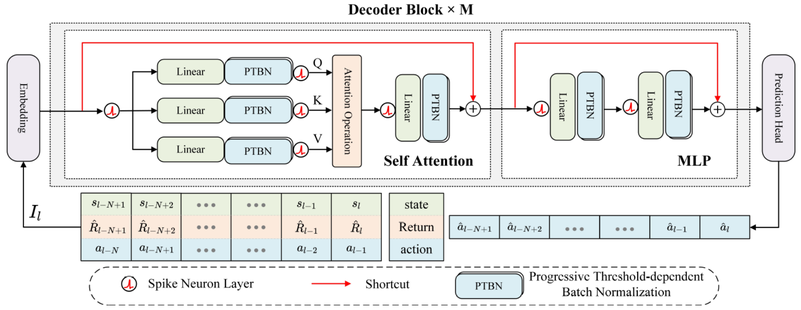
CVPR Nashville JUNE 11-15, 2025

Abstract:

- Offline reinforcement learning (RL) enables policy training solely on pre-collected data, avoiding direct environment interaction---a crucial benefit for energy-constrained embodied AI applications. Although Artificial Neural Networks (ANN)-based methods perform well in offline RL, their high computational and energy demands motivate exploration of more efficient alternatives. Spiking Neural Networks (SNNs) show promise for such tasks, given their low power consumption. In this work, we introduce DSFormer, the first spike-driven transformer model designed to tackle offline RL via sequence modeling.

Method:

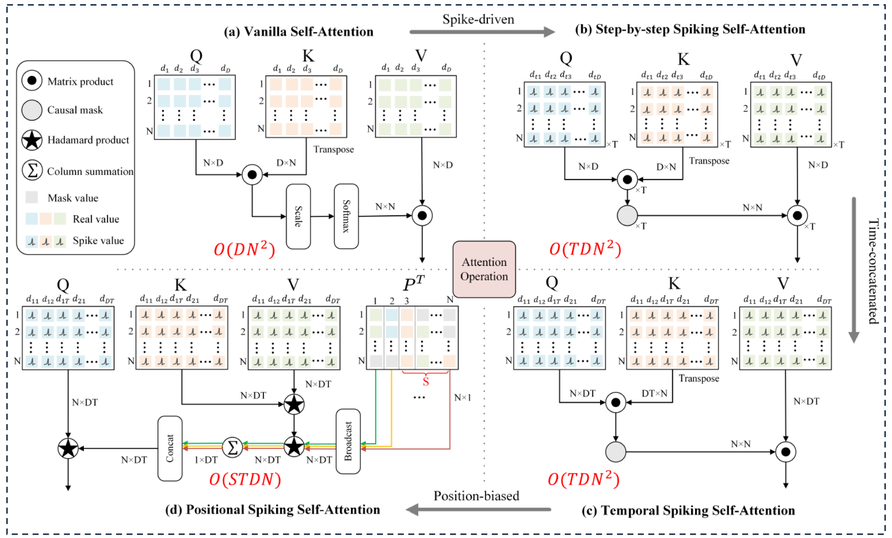
Architecture



- Framework: Autoregressive Transformer with Embedding Layer, M Decoder Blocks, and Prediction Head.
- Decoder Design: Uses spike-driven Self-Attention (TSSA, PSSA), MLP, PTBN (replacing BatchNorm), and Membrane Shortcut.
- Input: At step l, $I_l = (a_{l-N}, \hat{R}_{l-N+1}, s_{l-N+1}, \ldots, a_{l-1}, \hat{R}_l, s_l)$ becomes $X_l \in R^{N\times D}$, repeated T times.
- Output: Self-Attention ($Y = X + Self - Attention(X)$) and MLP ($X = Y + MLP(Y)$) process X; final output is normalized and predicts actions.

Spike-Driven Self-Attention

- TemporalnSpiking Self-Attention(TSSA): temporal dependency capture.
- Positional Spiking Self-Attention(PSSA): positional dependency capture.



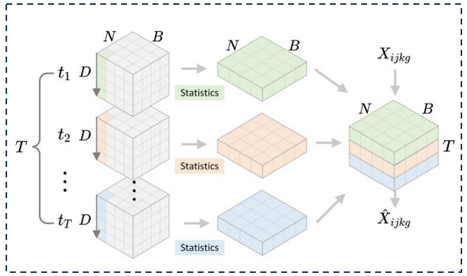
PTBN

- Threshold-dependent Layer Normalization (tdLN): contradicts SNN's spiking nature.

$$\hat{X}_{ijkg} = \frac{\alpha U_{\text{th}}\,(X_{ijkg} - \mu_{ijk})}{\sqrt{\sigma_{ijk}^2 + \epsilon}}, \quad Y_{ijkg} = \lambda \hat{X}_{ijkg} + \beta$$

- Progressive Threshold-dependent Batch Normalization (PTBN): gradually transitions from tdLN to a SNN-compatible tdBN during training.

$$\text{PTBN}(x) = \theta\,\text{tdLN}(x) + (1-\theta)\,\text{tdBN}(x)$$



Design of tdLN.

Experiments:

Results on MuJoCo

MuJoCo Tasks	BC	CQL	DT	FCNet	SpikeGPT	SpikeBert	TSSA	PSSA
halfcheetah-m-e	35.8	62.4	86.8±1.3	91.2±0.3	23.6±4.5	24.3±6.0	91.3±0.2	**91.5±0.3**
walker2d-m-e	6.4	98.7	108.1±0.2	108.8±0.1	22.6±4.8	92.5±22.4	108.6±0.2	**108.9±0.1**
hopper-m-e	**111.9**	111.0	107.6±1.8	110.5±0.5	32.7±5.4	84.1±8.8	111.0±0.7	110.9±0.2
halfcheetah-m	36.1	44.4	42.6±0.1	**42.9±0.4**	26.9±0.8	20.0±3.5	42.5±0.4	42.8±0.3
walker2d-m	6.6	79.2	74.0±1.4	**75.2±0.5**	16.4±10.2	22.9±10.4	72.4±4.5	**75.2±1.4**
hopper-m	29.0	58.0	67.6±1.0	57.8±6.0	25.1±6.4	31.4±4.9	64.6±2.1	**74.1±4.3**
halfcheetah-m-r	38.4	46.2	36.6 ±0.8	**39.8±0.8**	21.8±2.0	32.2±8.0	38.7±1.1	38.8±0.7
walker2d-m-r	11.3	26.7	66.6±3.0	63.5±7.5	16.7±3.3	21.2±6.4	66.0±3.3	**71.0±3.6**
hopper-m-r	11.8	48.6	82.7±7.0	85.8±1.7	51.5±7.1	30.1±8.6	85.8±3.6	**96.3±1.3**
MuJoCo mean ↑	31.9	63.9	74.7	75.1	26.4	39.9	75.7	**78.8**
Power (μJ) ↓	N/A	N/A	410.5	1022.03	**27.8**	806.7	96.1	88.8

Results on Adroit

Adroit Tasks	BC	CQL	DT	FCNet	SpikeGPT	SpikeBert	TSSA	PSSA
pen-e	85.1	107.0	110.4±20.9	108.0±11.3	30.5±10.3	46.2±19.5	104.6±13.2	**122.0±17.8**
door-e	34.9	101.5	95.5±5.7	102.9±2.9	65.3±16.9	96.4±4.6	105.0±0.3	**105.2±0.1**
hammer-e	125.6	86.7	89.7±24.6	121.1±6.1	51.1±18.7	71.3±16.5	126.4±0.4	**127.2±0.3**
relocate-e	101.3	95.0	15.3±3.6	50.0±6.0	0.7±0.9	0.3±0.5	106.3±2.6	**108.4±2.2**
pen-h	34.4	37.5	-0.2±1.8	57.7±11.1	29.8±11.7	20.0±16.7	**89.7±10.0**	75.7±25.1
door-h	0.5	**9.9**	0.1±0.0	0.4±0.5	0.1±0.0	0.2±0.0	0.4±0.1	0.2±0.0
hammer-h	1.5	**4.4**	0.3±0.0	1.2±0.0	0.3±0.0	0.3±0.0	0.4±0.1	0.2±0.0
relocate-h	0.0	0.2	**0.2±0.2**	0.0±0.0	0.1±0.0	0.0±0.0	0.0±0.0	0.0±0.0
pen-c	**56.9**	39.2	22.7±17.1	50.4±24.1	17.0±22.0	17.6±29.0	41.1±19.7	44.8±14.7
door-c	-0.1	**0.4**	0.1±0.0	-0.2±0.0	0.2±0.0	0.2±0.0	**0.4±0.8**	0.0±0.0
hammer-c	0.8	**2.1**	0.3±0.0	0.2±0.0	0.3±0.0	0.3±0.5	0.2±0.0	0.2±0.0
relocate-c	-0.1	-0.1	-0.3±0.0	-0.2±0.0	**0.1±0.0**	0.0±0.5	-0.2±0.0	-0.2±0.0
Adroit Mean ↑	36.7	40.3	27.8	41.0	16.3	21.1	47.9	**48.6**

Summary

- DSFormer's Design: DSFormer, the first spike-driven transformer for offline RL, uses TSSA and PSSA to capture temporal and positional dependencies, and PTBN to replace LayerNorm while preserving temporal dependencies.
- Performance: With SNN's threshold activation, DSFormer excels in long-range dependencies and sparse environments, outperforming SNN and ANN models on D4RL with 78.4% energy savings.