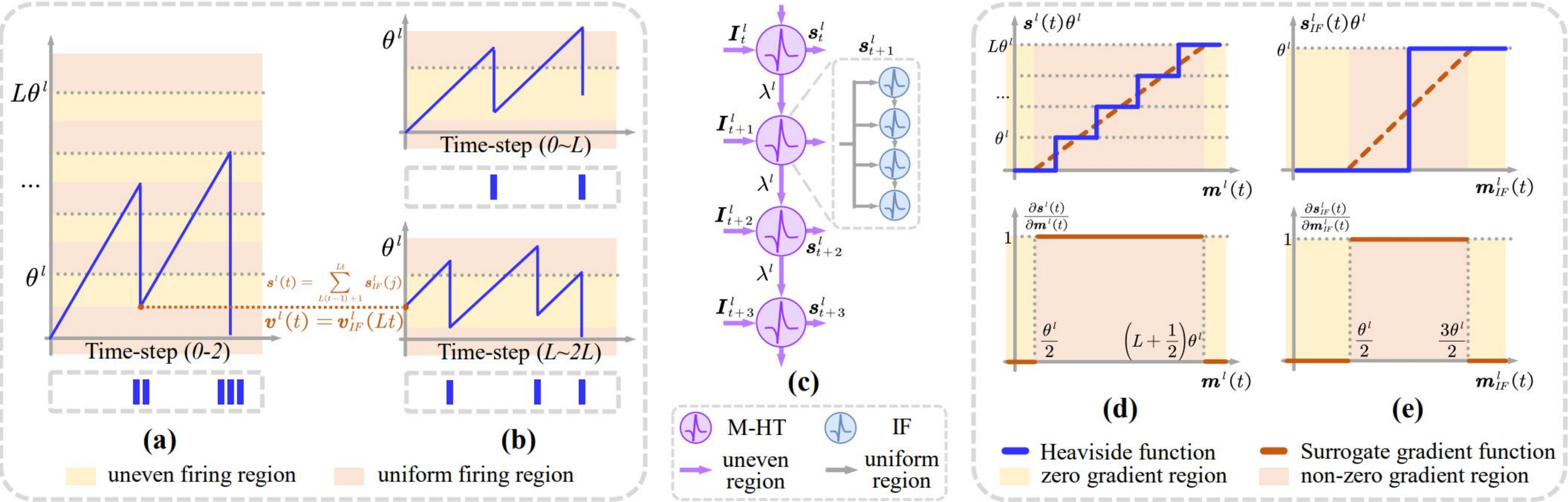
LM-HT SNN: Enhancing the Performance of SNN to ANN Counterpart through Learnable Multi-hierarchical Threshold Model

Zecheng Hao[1], Xinyu Shi[1,2], Yujia Liu[1], Zhaofei Yu[1,2]* & Tiejun Huang[1,2]
[1] School of Computer Science, Peking University
[2] Institute for Artificial Intelligence, Peking University

From Single-threshold to Multi-threshold Model

- We point out that the essence of the **equidistant multi-threshold** model is to simulate the spike firing situation of the vanilla spiking model within **specific time windows**. Specially, when the input current follows a completely uniform distribution on the time dimension, its spike firing rate is mathematically equivalent to the activation output of **quantized ANNs**.



Learnable Multi-hierarchical Threshold Model

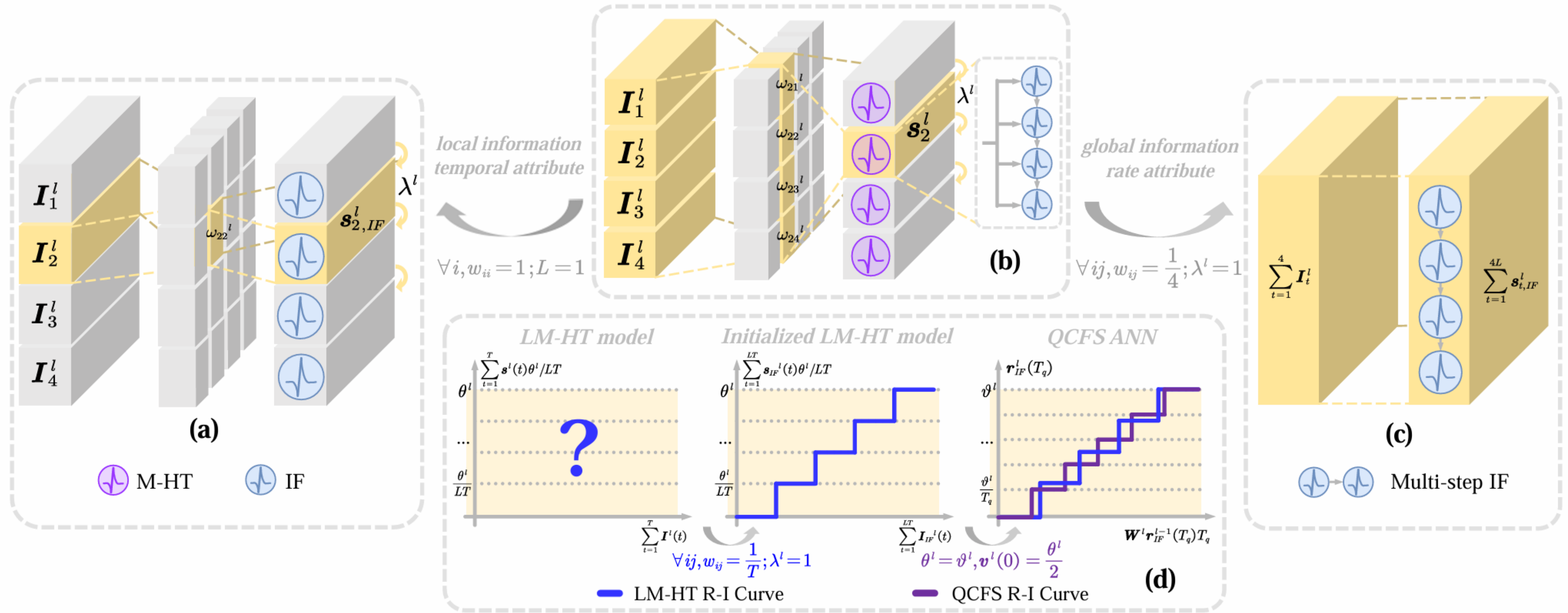
- Under the condition that the input current completely follows a uniform distribution, the spike firing rate of the L-level multi-threshold model within T time-steps is equivalent to the LT-level quantized ANN.
- In the initial state of hybrid training, from the perspective of mathematical expectation, the spike firing rate corresponding to STBP fine-tuning training with any number of time-steps is equivalent to the spike firing rate corresponding to ANN with any quantization level.

Theorem 4.2. *When $\lambda^l = 1, \boldsymbol{v}^l(0) \in [0, \theta^l)$, for a M-HT model with L-level threshold, after T time-steps, we will derive the following conclusions:*
(i) If we further assume $\forall t \in [1,T], \boldsymbol{I}^l(t) \in [0, L\theta^l)$, we will have: $\forall t \in [1,T], \boldsymbol{s}^l(t) = \sum_{j=L(t-1)+1}^{Lt} \boldsymbol{s}_{IF}^l(j), \boldsymbol{v}^l(t) = \boldsymbol{v}_{IF}^l(Lt), \sum_{t=1}^{T} \boldsymbol{s}^l(t) = \sum_{j=1}^{LT} \boldsymbol{s}_{IF}^l(j)$.
(ii) If we further assume $\boldsymbol{I}^l(1) = \ldots = \boldsymbol{I}^l(T)$, we will have: $\sum_{t=1}^{T} \boldsymbol{s}^l(t) = \mathrm{clip}\left(\left\lfloor \frac{\boldsymbol{v}^l(0)+\sum_{t=1}^{T}\boldsymbol{I}^l(t)}{\theta^l} \right\rfloor, 0, LT\right)$.
Here the IF model has uniform input currents $\boldsymbol{I}^l(1)/L, \ldots, \boldsymbol{I}^l(T)/L$ respectively within every L steps and satisfies $\boldsymbol{v}_{IF}^l(0) = \boldsymbol{v}^l(0)$.

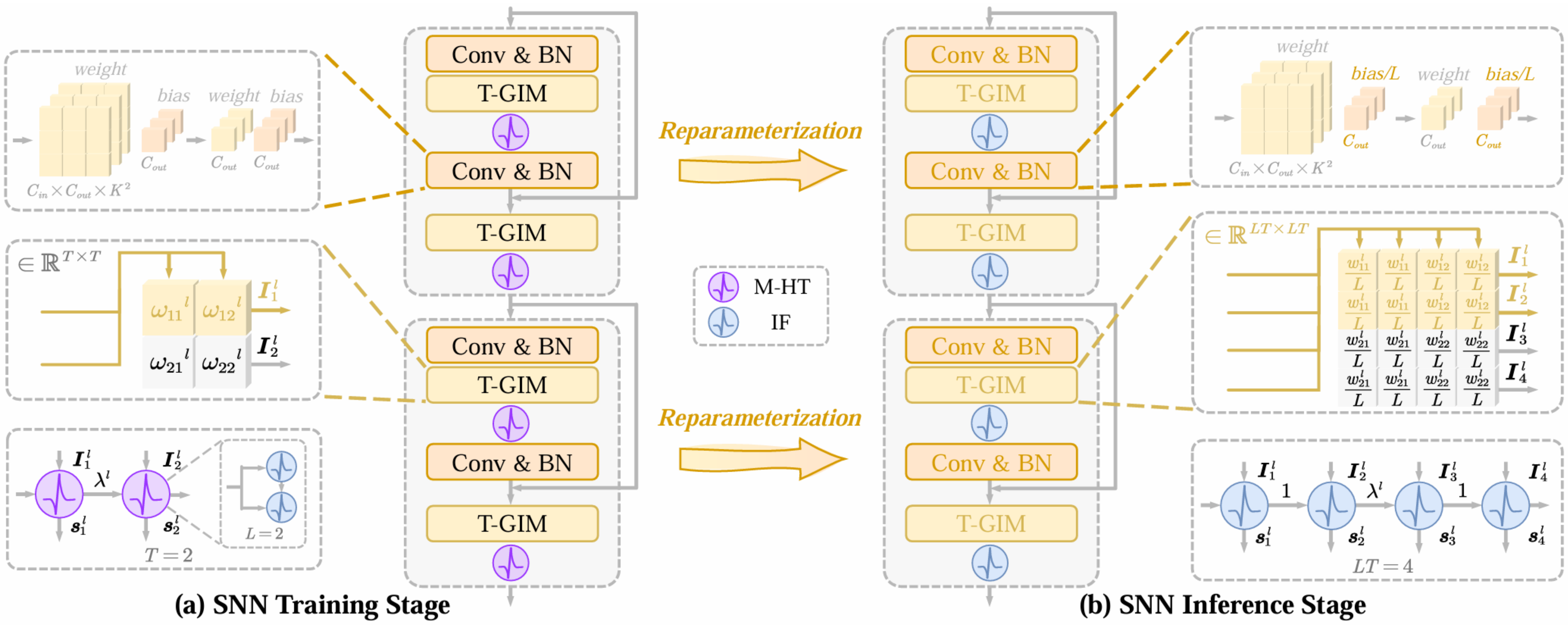
Theorem 4.4. *When $\sum_{t=1}^{T} \boldsymbol{I}^l(t)/LT = \boldsymbol{W}^l \boldsymbol{r}_{IF}^{l-1}(T_q)$ and $\sum_{t=1}^{T} \boldsymbol{I}^l(t) \in [0, LT\theta^l]$, if $\forall i,j \in [1,T], \omega_{ij}^l = \frac{1}{T}$ and $\lambda^l = 1, \theta^l = \vartheta^l, \boldsymbol{v}^l(0) = \frac{\theta^l}{2}$, for L, T, T_q with arbitrary values, we have:*
$$\mathbb{E}\left(\frac{\sum_{t=1}^{T}\boldsymbol{s}^l(t)\theta^l}{LT} - \frac{\vartheta^l}{T_q}\mathrm{clip}\left(\left\lfloor \frac{\boldsymbol{W}^l \boldsymbol{r}_{IF}^{l-1}(T_q)T_q}{\vartheta^l} + \frac{1}{2}\right\rfloor, 0, T_q\right)\right) = 0.$$

Learnable Multi-hierarchical Threshold Model

- By adopting different parameter initialization schemes, the LM-HT model can further establish a bridge between the vanilla STBP and quantized ANNs training.
- A brand-new hybrid training framework: ANN-SNN Conversion + STBP training based on LM-HT model (no more than 30 epochs)



Reparameterization Process for LM-HT Model



Experiments

Table 1: Ablation study for the LM-HT model on a subset of ImageNet-1k.

Model	T-GIM	Arch.	Acc.(%)	SOPs(G)	E.(mJ)
L=1,T=4	w/o	ResN-18	76.22	1.60	1.44
L=2,T=2	w/o		80.52	1.08	0.97
L=2,T=2	w/		80.56	0.73	0.66
L=1,T=4	w/o	ResN-34	65.62	3.20	2.88
L=2,T=2	w/o		82.18	2.42	2.18
L=2,T=2	w/		82.72	1.72	1.55

Table 2: Validation for the reparameterization procedure.

Arch.	Acc.(%)	SOPs(G)	E.(mJ)
Before reparameterization (L=2, T=2)			
VGG-13	61.64	0.26	0.23
ResN-18	64.44	0.52	0.46
After reparameterization (L=1, T=4)			
VGG-13	61.66	0.26	0.23
ResN-18	64.50	0.52	0.46

Table 4: The performance of hybrid training based on the LM-HT model for CIFAR-100 dataset.

Method	Time-steps	VGG-16		ResNet-20	
		ANN Acc.(%)	SNN Acc.(%)	ANN Acc.(%)	SNN Acc.(%)
RMP [16]	32, 64, 128	71.22	- , - , 63.76	68.72	27.64, 46.91, 57.69
SNM [43]	32, 64, 128	74.13	71.80, 73.69, 73.95	-	-
SRP [17]	5, 6, 8	76.28	71.52, 74.31, 75.42	69.94	46.48, 53.96, 59.34
QCFS (T_q=4) [2]	2, 4, 8	76.11	63.33, 69.70, 74.12	63.90	38.04, 52.28, 61.77
LM-HT (L=2)	**2**	-	**75.97 (+6.27)**	-	**63.55 (+11.27)**
	4	-	**76.49 (+2.37)**	-	**64.87 (+3.10)**
LM-HT (L=4)	**2**	-	**76.38 (+2.26)**	-	**63.43 (+1.66)**
QCFS (T_q=8) [2]	2, 4, 8	77.31	64.85, 70.50, 74.63	69.56	19.76, 34.17, 55.50
LM-HT (L=2)	**2**	-	**76.31 (+5.81)**	-	**67.08 (+32.91)**
	4	-	**76.79 (+2.16)**	-	**69.00 (+13.50)**
LM-HT (L=4)	**2**	-	**76.08 (+1.45)**	-	**67.21 (+11.71)**

Experiments

北京大学 PEKING UNIVERSITY

Table 3: Comparison with previous state-of-the-art works.

Dataset	Method	Type	Architecture	Time-steps	Accuracy(%)
CIFAR-10	STBP-tdBN [54]	Direct Training	ResNet-19	4	92.92
	Dspike [29]	Direct Training	ResNet-18	4	93.66
	TET [8]	Direct Training	ResNet-19	4	94.44
	SLTT [34]	Online Training	ResNet-18	6	94.44
	GLIF [51]	Direct Training	ResNet-18	2, 4, 6	94.15, 94.67, 94.88
			ResNet-19	2, 4, 6	94.44, 94.85, 95.03
	LM-HT (L=2)	**Direct Training**	**ResNet-18**	**2**	**96.25**
			ResNet-19	**2**	**96.89**
CIFAR-100	Dspike [29]	Direct Training	ResNet-18	4	73.35
	TET [8]	Direct Training	ResNet-19	4	74.47
	SLTT [34]	Online Training	ResNet-18	6	74.38
	GLIF [51]	Direct Training	ResNet-18	2, 4, 6	74.60, 76.42, 77.28
			ResNet-19	2, 4, 6	75.48, 77.05, 77.35
	RMP-Loss [14]	Direct Training	ResNet-19	2, 4, 6	74.66, 78.28, 78.98
	LM-HT (L=2)	**Direct Training**	**ResNet-18**	**2**	**79.33**
			ResNet-19	**2**	**81.76**
ImageNet-200	DCT [13]	Hybrid Training	VGG-13	125	56.90
	Online-LTL [48]	Hybrid Training	VGG-13	16	54.82
	Offline-LTL [48]			16	55.37
	ASGL [44]	Direct Training	VGG-13	4, 8	56.57, 56.81
	LM-HT (L=2)	**Direct Training**	**VGG-13**	**2, 4**	**61.09, 61.75**
	LM-HT (L=4)			**2**	**62.05**
ImageNet-1k	STBP-tdBN [54]	Direct Training	ResNet-34	6	63.72
	TET [8]	Direct Training	ResNet-34	6	64.79
	MBPN [15]	Direct Training	ResNet-34	4	64.71
	RMP-Loss [14]	Direct Training	ResNet-34	4	65.17
	SEW ResNet [11]	Direct Training	ResNet-34	4	67.04
	GLIF [51]	Direct Training	ResNet-34	4	67.52
	LM-HT (L=2)	**Direct Training**	**ResNet-34**	**2**	**70.90**
CIFAR10-DVS	STBP-tdBN [54]	Direct Training	ResNet-19	10	67.80
	Dspike [29]	Direct Training	ResNet-18	10	75.40
	MBPN [15]	Direct Training	ResNet-19	10	74.40
	RMP-Loss [14]	Direct Training	ResNet-19	10	76.20
	LM-HT (L=2)	**Direct Training**	**ResNet-18**	**2, 4**	**80.70, 81.00**
	LM-HT (L=4)			**2**	**81.90**

Thanks for Listening!

Zecheng Hao[1]**, Xinyu Shi**[1,2]**, Yujia Liu**[1]**, Zhaofei Yu**[1,2]* **& Tiejun Huang**[1,2]

[1] School of Computer Science, Peking University

[2] Institute for Artificial Intelligence, Peking University