Sequence Approximation Using FeedForward Spiking-Neural Network For Spatiotemporal Learning

서울시립대학교
UNIVERSITY OF SEOUL

Key Contributions

- We prove that any spike-sequence-to-spike-sequence mapping functions on a compact domain can be approximated by feedforward SNN with one neuron per layer using skip-layer connections, which cannot be achieved if no skip-layer connection is used.
- We prove that using heterogeneous neurons having different dynamics and skip-layer connection increases the number of memory pathways a feedforward SNN can achieve and hence, improves SNN's capability to represent arbitrary sequences.
- We develop complex SNN architectures using the preceding theoretical observations and experimentally demonstrate that they can be trained with supervised BPTT and unsupervised STDP for spatiotemporal data classification.
- We design a dual-search-space option for Bayesian optimization process to sequentially optimize network architectures and neuron dynamics of a feedforward SNN considering heterogeneity and skip-layer connection to improve learning and classification of spatiotemporal patterns.

Lemma 1

Lemma 1 *For any input spike sequence with period t_{in} in range $[T_{min}, T_{max}]$, there exists a spiking neuron n with fixed parameters v_{th}, v_{reset}, a, R_m and τ_m, such that by changing synaptic conductance G, it is possible to set the neuron response rate γ_n to be any positive integer.*

Definition 1

Lemma 1 *For any input spike sequence with period t_{in} in range $[T_{min}, T_{max}]$, there exists a spiking neuron n with fixed parameters v_{th}, v_{reset}, a, R_m and τ_m, such that by changing synaptic conductance G, it is possible to set the neuron response rate γ_n to be any positive integer.*

Definition 1 *Neuron Response Rate* γ For a spiking neuron n with membrane potential at v_{reset} and input spike sequence with period t_{in}, γ is the number of input spike n needs to reach v_{th}.

Neuron Response Rate: t_{in}이라는 시간 동안 최소 γ의 스파이크가 들어와야 세포막 전위(potential)가 역치(threshold) v_{th}에 닿아서, 출력 스파이크를 내뱉는다.

Lemma 1 from definition 1

Lemma 1 *For any input spike sequence with period t_{in} in range $[T_{min}, T_{max}]$, there exists a spiking neuron n with fixed parameters $v_{th}, v_{reset}, a, R_m$ and τ_m, such that by changing synaptic conductance G, it is possible to set the neuron response rate γ_n to be any positive integer.*

Neuron Response Rate: t_{in}이라는 시간 동안 최소 γ의 스파이크가 들어와야 세포막 전위(potential)가 역치(threshold) v_{th}에 닿아서, 출력 스파이크를 내뱉는다.

Lemma 1: 우리는 뉴런의 Hyperparameters를 그대로 놔두고 입력 전류 비례상수 (synaptic conductance) G 만을 변경해서, 세포가 작동하기 위한 스파이크 밀도를 지정할 수 있다!

Meaning of Lemma 1

Lemma 1 *For any input spike sequence with period t_{in} in range $[T_{min}, T_{max}]$, there exists a spiking neuron n with fixed parameters $v_{th}, v_{reset}, a, R_m$ and τ_m, such that by changing synaptic conductance G, it is possible to set the neuron response rate γ_n to be any positive integer.*

Lemma 1: 우리는 뉴런의 Hyperparameters를 그대로 놔두고 입력 전류 비례상수 (synaptic conductance) G 만을 변경해서, 세포가 작동하기 위한 스파이크 밀도를 지정할 수 있다!

Idea: G matrix를 만들고 학습시킨다면, 우리는 많은 뉴런이 각각 다른 스파이크 밀도에 작동하게 할 수 있다!

Theorem 1

Theorem 1 *For any input and target output spike sequence pair with periods* $(t_{in}, t_{out}) \in [T_{min}, T_{max}] \times [T_{min}, T_{max}]$, *there exists a minimal-layer-size network with skip-layer connections that has memory pathway with output spike period function* $P(t)$ *such that* $|P(t_{in}) - t_{out}| < \epsilon$.

Theorem 1: 우리는 어떤 입력-출력 쌍에 대해서 memory pathway 를 가지고 입력과 출력 시간(period) 오차가 ϵ 이하인 SNN을 만들 수 있다! (Universal Approximation Thm. of SNN)

But what is memory pathway?
And what is Universal Approximation Theorem? 🤔

Theorem 1 – Memory Pathway

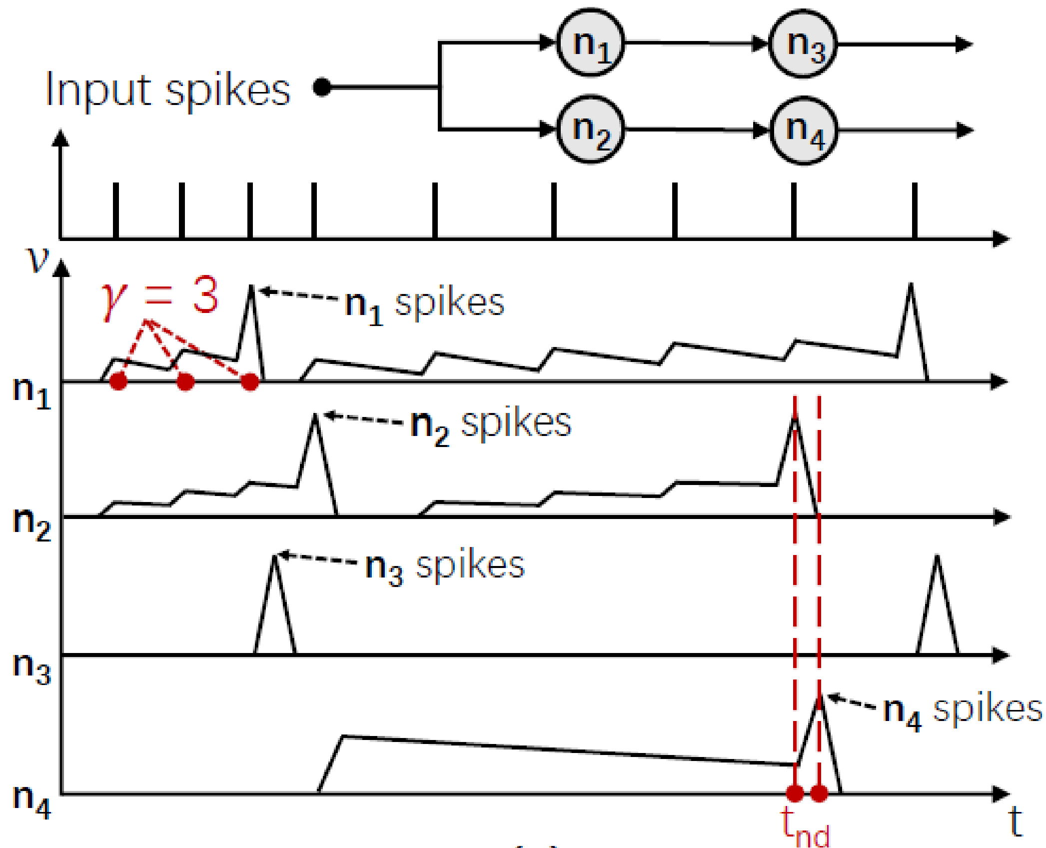
Theorem 1 *For any input and target output spike sequence pair with periods* $(t_{in}, t_{out}) \in [T_{min}, T_{max}] \times [T_{min}, T_{max}]$, *there exists a minimal-layer-size network with skip-layer connections that has memory pathway with output spike period function* $P(t)$ *such that* $|P(t_{in}) - t_{out}| < \epsilon$.

Theorem 1: 우리는 어떤 입력-출력 쌍에 대해서 memory pathway를 가지고 입력과 출력 시간(period) 오차가 ϵ 이하인 SNN을 만들 수 있다!

Definition 2 *Memory Pathways* For a feedforward SNN with m layers, a memory pathway is defined as a spike propagation path from input to the output layer. Two memory pathways are considered distinct if the set of neurons contained in them is different.

Memory Pathways: SNN에서 가능한 모든 spike 전파 경로!

Theorem 1 – Memory Pathway



(a)

Theorem 1 – Memory Pathway

Theorem 1 *For any input and target output spike sequence pair with periods* $(t_{in}, t_{out}) \in [T_{min}, T_{max}] \times [T_{min}, T_{max}]$*, there exists a minimal-layer-size network with skip-layer connections that has memory pathway with output spike period function* $P(t)$ *such that* $|P(t_{in}) - t_{out}| < \epsilon$.

Theorem 1: 우리는 어떤 입력-출력 쌍에 대해서 memory pathway를 가지고 입력과 출력 시간(period) 오차가 ϵ 이하인 SNN을 만들 수 있다!

Definition 2 *Memory Pathways* For a feedforward SNN with m layers, a memory pathway is defined as a spike propagation path from input to the output layer. Two memory pathways are considered distinct if the set of neurons contained in them is different.

"For any input … , there exists a minimal-layer-size network … that has memory pathway"
라는 말은, 어떤 입력-출력 쌍이 주어지면 입력->출력을 유도하는 Spiking FFNN를 **무조건** 만들 수 있다는 뜻!

Theorem 1

Theorem 1 *For any input and target output spike sequence pair with periods* $(t_{in}, t_{out}) \in [T_{min}, T_{max}] \times [T_{min}, T_{max}]$, *there exists a minimal-layer-size network with skip-layer connections that has memory pathway with output spike period function* $P(t)$ *such that* $|P(t_{in}) - t_{out}| < \epsilon$.

Theorem 1: 우리는 어떤 입력-출력 쌍에 대해서 memory pathway를 가지고 입력과 출력 시간(period) 오차가 ϵ 이하인 SNN을 만들 수 있다!

"For any input ... , there exists a minimal-layer-size network ... that has memory pathway"
라는 말은, 어떤 입력-출력 쌍이 주어지면 입력->출력을 유도하는 Spiking FFNN를 **무조건** 만들 수 있다는 뜻!

"... with output spike period function ... $|P(t_{in}) - t_{out}| < \epsilon$" 라는 말은, 근사 오차를 임의로 줄일 수 있다는 의미!

Theorem 1

Theorem 1 *For any input and target output spike sequence pair with periods* $(t_{in}, t_{out}) \in [T_{min}, T_{max}] \times [T_{min}, T_{max}]$, *there exists a minimal-layer-size network with skip-layer connections that has memory pathway with output spike period function* $P(t)$ *such that* $|P(t_{in}) - t_{out}| < \epsilon$.

Theorem 1: 우리는 어떤 입력-출력 쌍에 대해서 memory pathway를 가지고 입력과 출력 시간(period) 오차가 ϵ 이하인 SNN을 만들 수 있다!

The meaning of Theorem 1: We can approximate any spike sequence with SNN with skip-layer!

But what is Universal Approximation Thm.?

보편 근사 정리(Universal approximation theorem)는 하나의 은닉층을 갖는 인공신경망은 임의의 연속인 다변수 함수를 원하는 정도의 정확도로 근사할 수 있다는 정리이다. 모든 인공신경망과 모든 활성화 함수에 대해 증명된 것은 아니다.

사례 [편집]

1989년 조지 시벤코(Cybenko)가 발표한 **시벤코 정리**(Cybenko's theorem)는 다음과 같다.

> φ를 시그모이드 함수 형식의 연속 함수라 하자(예, $\varphi(\xi) = 1/(1+e^{-\xi})$). $[0,1]^n$ 또는 R^n의 부분집합에서 실수의 연속 함수 f와 $\epsilon > 0$가 주어지면, 다음을 만족하는 벡터 $\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_N}, \alpha$, θ와 매개 함수 $G(\cdot, \mathbf{w}, \alpha, \theta) : [0,1]^n \to R$이 존재한다.
>
> $|G(\mathbf{x}, \mathbf{w}, \alpha, \theta) - f(x)| < |\epsilon|$ for all $\mathbf{x} \in [0,1]^n$
>
> 이때,
>
> $$G(\mathbf{x}, \mathbf{w}, \alpha, \theta) = \sum_{i=1}^{N} \alpha_j \varphi(\mathbf{w}_j^T \mathbf{x} + \theta_j)$$
>
> 이고, $\mathbf{w}_j \in R^n, \alpha_j, \theta_j \in R, \mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_N), \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_N)$이다.

이 정리는 하나의 은닉층을 갖는 인공신경망은 임의의 연속인 다변수 함수를 원하는 정도의 정확도로 근사할 수 있음을 말한다. 단, $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N, \alpha$와 θ를 잘못 선택하거나 은닉층의 뉴런 수가 부족할 경우 충분한 정확도로 근사하는데 실패할 수 있다.

Lemma 2

Lemma 2 *With no skip-layer connection, there does not exist a minimal-layer-size network that has output spike period function* $P(t)$ *such that for any input and target output spike sequence pair with periods* $(t_{in}, t_{out}) \in [T_{min}, T_{max}] \times [T_{min}, T_{max}]$, $|P(t_{in}) - t_{out}| < \epsilon$.

The meaning of Lemma 2: We cannot approximate spike sequence without skip-layer!

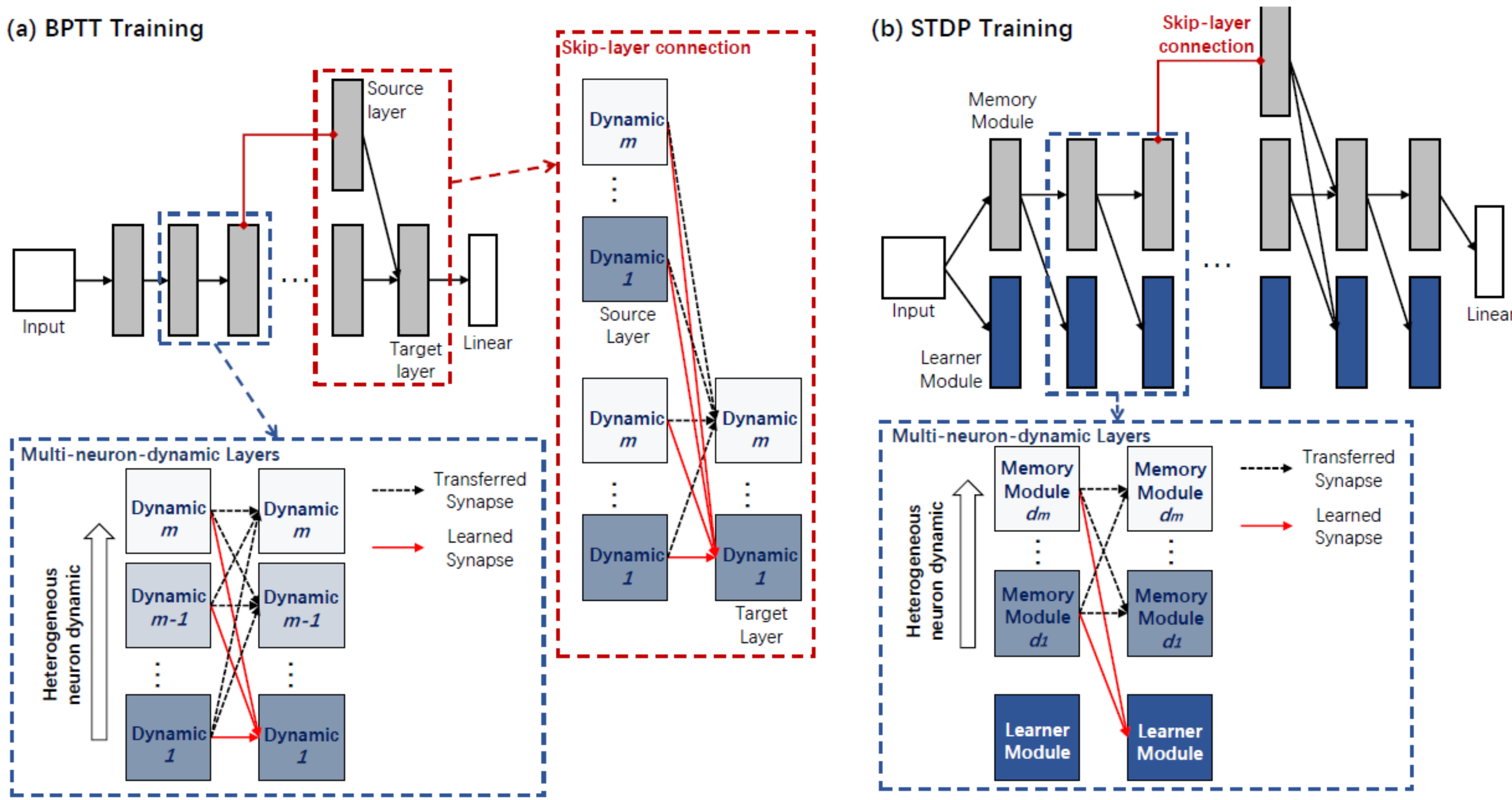
Lemma 3, 4, 5

Lemma 3 *A spiking neuron has cutoff period* $\omega_c = \tau_m \ln(\frac{v_{reset}-a}{v_{reset}-a+\frac{R_m}{\tau_m}G})$ *above which input spike sequence cannot cause the spiking neuron to spike.*

Lemma 4 *For an mMND network with* m *layers and* $\{\lambda_1, \lambda_2, ...\lambda_m\}$ *number of different neuron dynamics in each layer, the least upper bound of the number of distinct memory pathways is* $\prod_{i=1}^{m} \lambda_i$.

Lemma 5 *For an mMND network with* m *layers and* $\{\lambda_1, \lambda_2, ...\lambda_m\}$ *different neuron dynamics in each layer and a skip-layer connection made between layer* l_a *and* l_b, *s.t.* $a, b \in \{1, 2, ...m\}$ *and* $(b-a) > 1$, *the least upper bound of the number of memory pathways is* $\prod_{i=1}^{m} \lambda_i + (\prod_{i=1}^{a} \lambda_i \cdot \prod_{i=b}^{m} \lambda_i)$

Proposed Model



Dual-search-space Baysian Optimization

Dual-search-space Bayesian Optimization Bayesian optimization uses Gaussian process to model the distribution of an objective function, and an acquisition function to decide points to evaluate. For data points in a target dataset $x \in X$ and the corresponding label $y \in Y$, an SNN with network structure $\mathcal{V}$ and neuron parameters $\mathcal{W}$ acts as a function $f_{\mathcal{V},\mathcal{W}}(x)$ that maps input data x to predicted label $\tilde{y}$. The optimization problem in this work is defined as

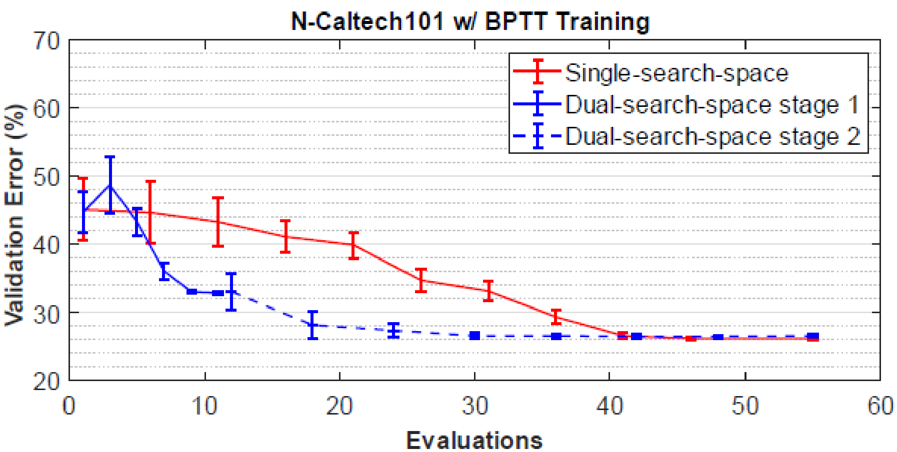
$$\min_{\mathcal{V},\mathcal{W}} \mathcal{P} \quad \text{where } \mathcal{P} = \sum_{x \in X, y \in Y} \mathcal{L}(y, f_{\mathcal{V},\mathcal{W}}(x)) \tag{3}$$

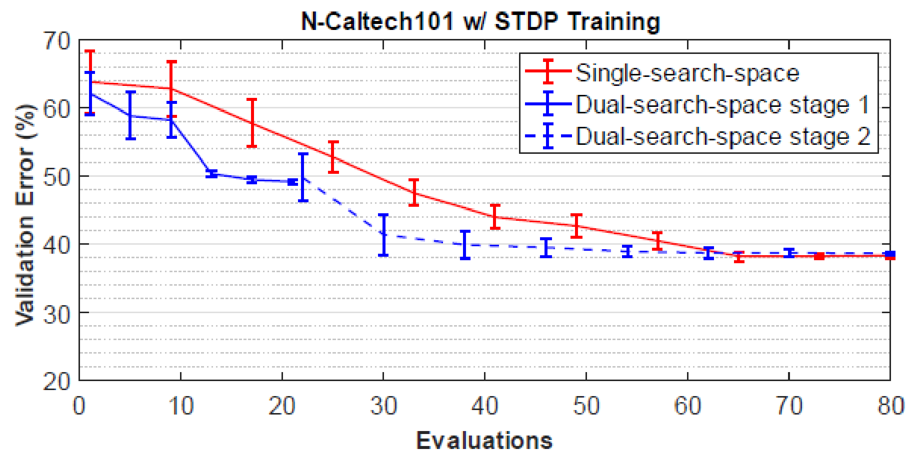
$\mathcal{V}$ contains the number of layers N_{layers}, the number of memory dynamics $N_{dynmaic}$ and skip-layer connection configuration variables N_{skip}, L_{start} and L_{end}, each controlling the number of skip-layer connections, the first layer and last layer to implement skip-layer connections. All of the values are discrete. $\mathcal{W}$ contains the values for a, τ_m and R_m in (1), which are continuous. We separate the discrete and continuous search spaces by implementing a dual-search-space optimization process, where $\mathcal{V}$ is first optimized with fixed, manually tuned neuron parameters. After an optimal structure is found, $\mathcal{W}$ are optimized for the selected $\mathcal{V}$. Details on the configurations of the optimization process are listed in the appendix. To achieve Bayesian optimization with constraints, we implement a modified expected improvement (EI) acquisition function similar to the one shown by Gardner (Gardner et al., 2014), which uses a Gaussian process to model the feasibility indicator due to its high evaluation cost. In this work, since the constraint function can be explicitly defined, we use a feasibility indicator that is directly evaluated. The modified EI function is defined as:

$$I_c(\mathbb{W}) = \Delta(\mathbb{W}) \cdot \max\{0, \mathcal{P}(\mathbb{W}) - \mathcal{P}(\mathbb{W}^+)\} \tag{4}$$

where $\mathbb{W}$ is the network configuration containing $\mathcal{W}$ and $\mathcal{V}$. $\mathbb{W}^+$ is the test point that provided the best result. $\Delta(\mathbb{W})$ is the explicitly defined indicator function that takes the value of 1 when all constraints are satisfied and 0 otherwise.

Dual-search-space Baysian Optimization





Dual-search-space Baysian Optimization

Table 3: Accuracy (%) for DVS Gesture (top) and N-Caltech101 (bottom)

Model	Labeled Data % In Training				Parameter Number
	100%	50%	30%	10%	
ConvLSNN (Salaj et al., 2020)	97.1	95.3	92.0	84.3	2.9M
DECOLLE (Kaiser et al., 2020)	97.5	95.0	91.2	83.9	1.3M
(Fang et al., 2021)	97.8	-	-	-	-
HATS (Sironi et al., 2018)	95.2	94.1	91.6	83.7	-
H-SNN (She et al., 2021)	96.2	95.8	93.7	88.2	0.74M
This Work-STDP Training	96.6	**96.0**	**94.1**	**91.2**	0.81M
This Work-BPTT Training	**98.0**	95.3	91.1	82.4	1.1M

Model	Labeled Data % In Training				Parameter Number
	100%	70%	50%	30%	
ConvLSNN (Salaj et al., 2020)	63.1	58.7	51.3	45.4	3.0M
DECOLLE (Kaiser et al., 2020)	66.9	61.9	56.2	50.6	2.0M
HATS (Sironi et al., 2018)	64.2	61.0	54.3	48.8	-
H-SNN (She et al., 2021)	42.8	41.9	37.0	34.6	1.7M
This Work-STDP Training	58.1	57.8	**57.2**	**54.6**	1.4M
This Work-BPTT Training	**71.2**	**65.4**	56.0	52.5	1.7M