SpikeConverter: An Efficient Conversion Framework Zipping the Gap between Artificial Neural Networks and Spiking Neural Networks

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

先进计算机体系结构实验室 Advanced Computer Architecture Laboratory

Fangxin Liu[1], Wenbo Zhao[1], Yongbiao Chen[1], Zongwu Wang[1] and Li Jiang[1]

Shanghai Jiao Tong University

36TH AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE
A VIRTUAL CONFERENCE
FEBRUARY 22 - MARCH 1, 2022
AAAI-22

ANN-Converted SNN

- **ANN:** High Precision, Dense Computation.
- **SNN:** Low Computation cost.
  - **Directly Trained SNN:** Low Precision, Moderate Delay.
  - **ANN-Converted SNN:** High Precision, Large Delay.
- **Highlight, this work:**
  - **High Precision:** As High Precision as ANN.
  - **Small Delay:** Log Scale of that of ANN-Converted SNN.

Comparison of Current Methods

	ANN	Directly Trained SNN	ANN-Converted SNN	This Work, SpikeConverter
Network Type	ANN	SNN	SNN	***SNN***
Training Method	BP	STBP	Not Needed	***Not Needed***
Computation Cost	High	Low	Medium	***Low***
Precision	High	Low	High	***High***
Delay	Small	Medium	Large	***Small***

Challenges

- **Low Precision of Directly Trained SNN**

  Unlike ANN training that have mature and developing algorithm based on gradient descent, they can not be deployed on SNN. Other training alternative also leads to inferior accuracy to ANN.
- **Large Delay of ANN-Converted SNN**

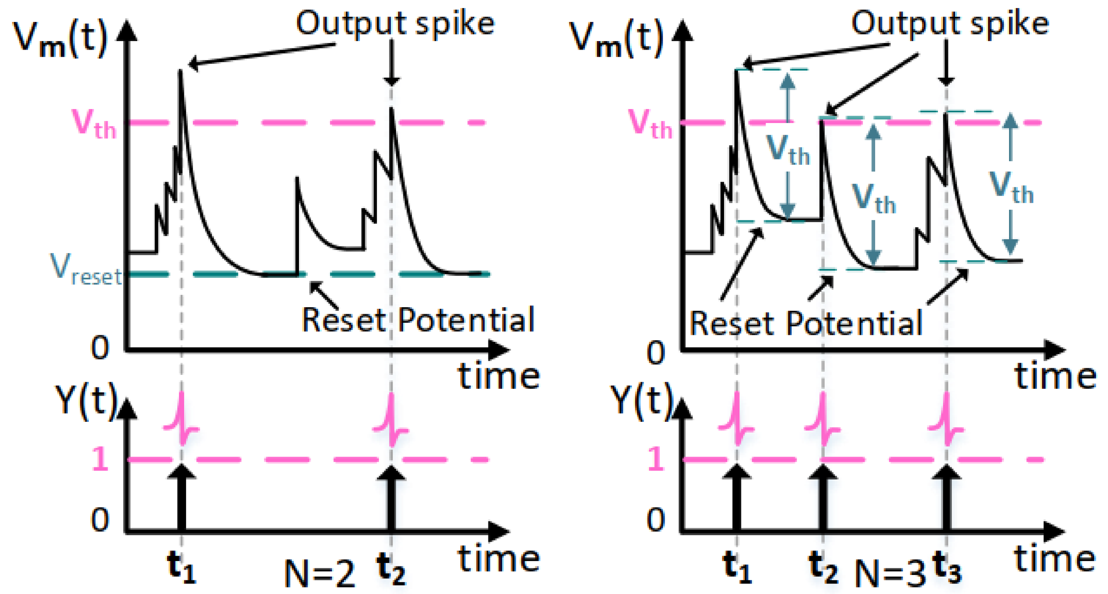
  Although ANN-converted SNN can match the precision of ANN, their spike trains are very long to represent adaquate information comparable with ANN.

SpikeConverter

Soft Reset

- Reduce threshold from membrane voltage when fires, instead of resetting to zero in hard reset mechanism.
- Enables the accurate construction of **ideal conversion identical relation**:

$$\sum_i w_i \cdot \left(\sum_{t=1}^{T} X_i[t] \cdot k^{T-t} \right) = V_{th} \cdot \sum_{t=1}^{T} Y[t] \cdot k^{T-t}$$



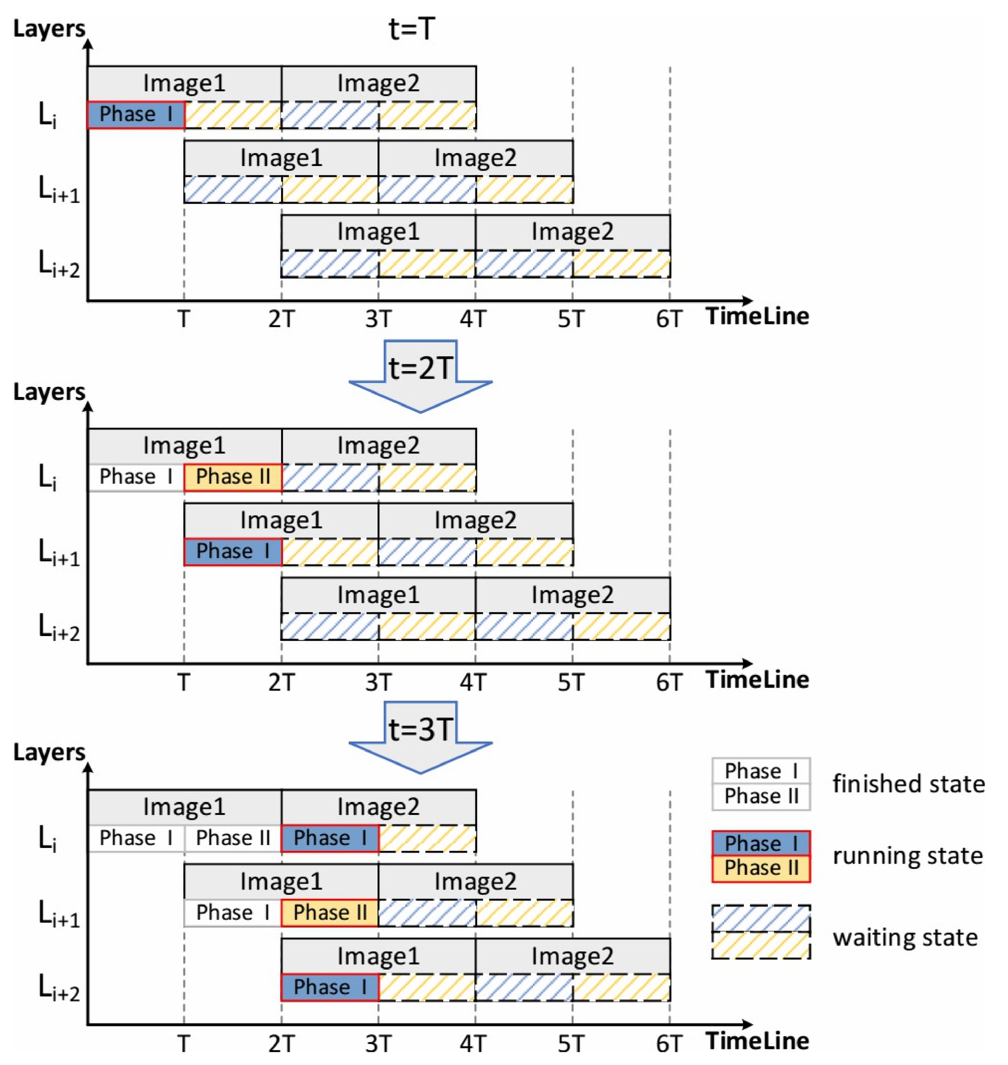
(a) LIF Neuron with hard reset mechanism
(b) LIF Neuron with soft reset mechanism

Ideal Conversion Identical Relation

- The identical relation between input and output spike trains.
- Holds under ideal condition: soft reset mechanism, **zero** membrane voltage after the last time step.

Temporal Seperation

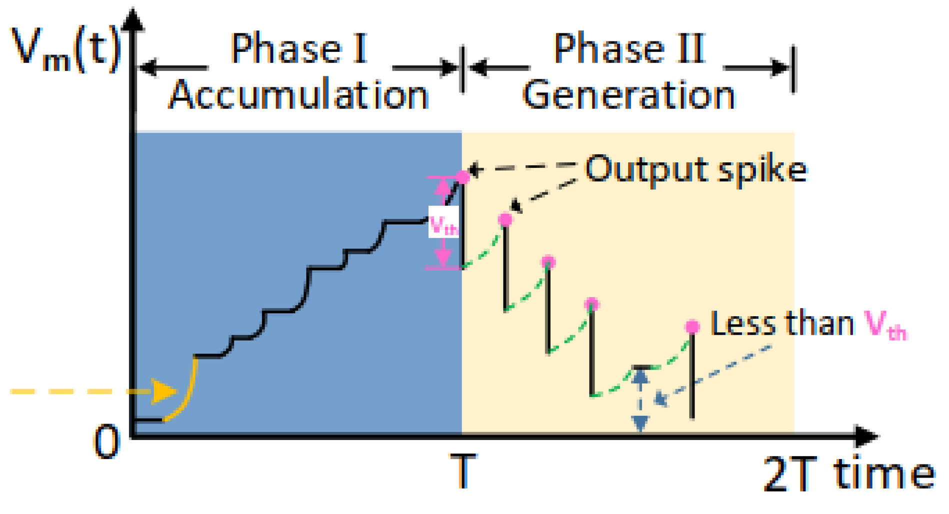
- SNN can not respond to accumulated negative membrane voltage.
- Lead to failing to follow the identical relation and influence the accuracy.
- Solution: temporal seperation, fire spikes after full accumulation.
- Pipeline multiple inputs to make full ues of resource.



Pipeline with Temporal Seperation

Inverse-LIF Neuron

- Membrane voltage inflates each time step.
- Necessary for the implementation of temporal seperation.
- Optional additiional time step to increase precision.

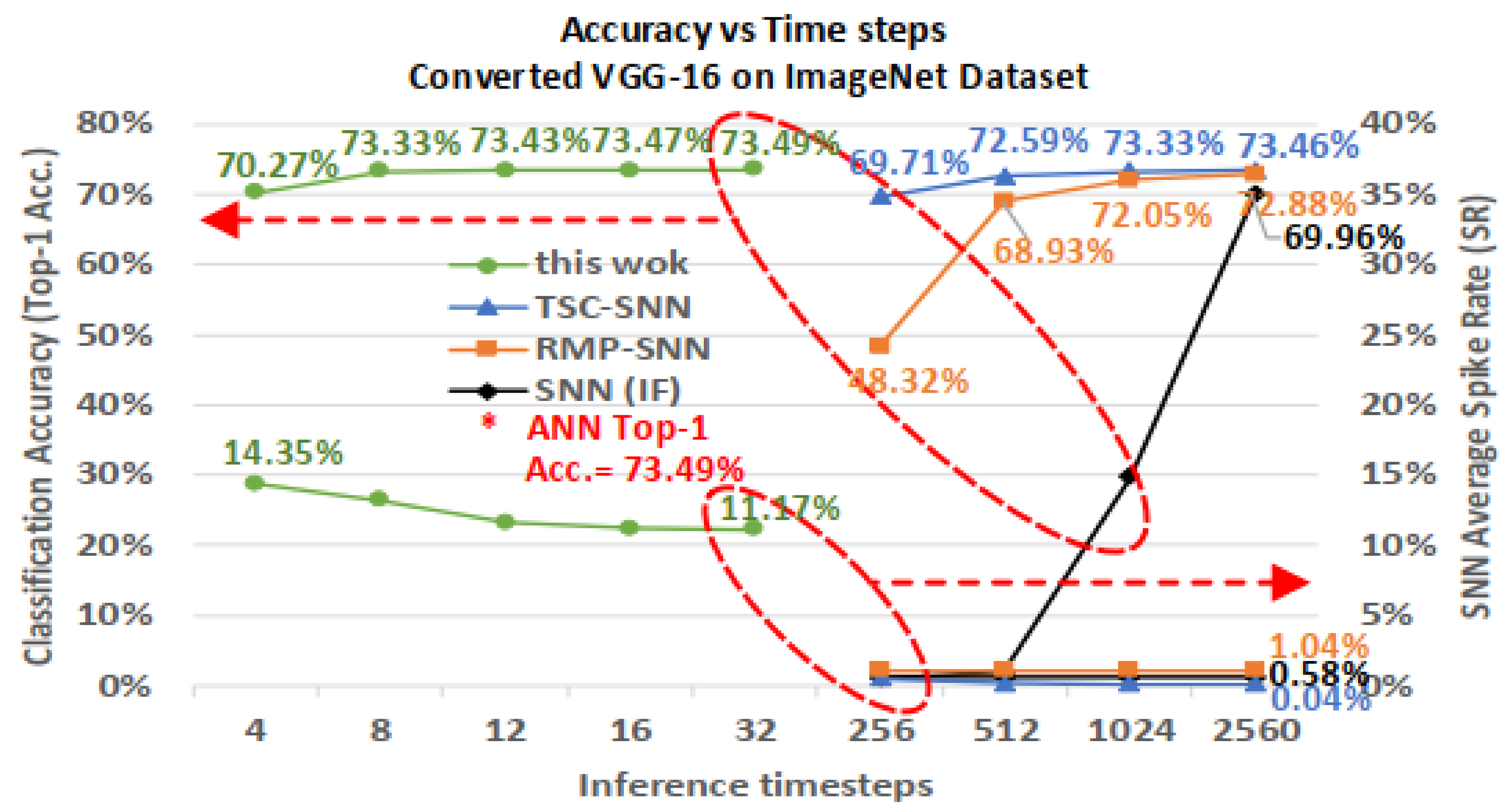


Results

Results on ImageNet

Network Architecture	Spiking Neuron Model	ANN (Top-1)	Time Steps	SNN (Top-1)	Accuracy Loss
ResNet-34 (Sengupta et al. 2019)	IF (hard-reset)	70.69%	4096	65.47%	5.22%
ResNet-34 (Han, Srinivasan, and Roy 2020)	RMP (soft-reset)	70.64%	4096	69.89%	0.75%
ResNet-34 (Han and Roy 2020)	TSC (soft-reset)	70.64%	4096	69.93%	0.71%
ResNet-34 [This work]	**SpikeConverter (soft-reset)**	**70.64%**	**16**	**70.57%**	**0.07%**
VGG-16 (Rueckauer et al. 2017)	Converted-SNN (hard-reset)	63.89%	400	49.61%	14.28%
VGG-16 (Sengupta et al. 2019)	IF (hard-reset)	70.52%	2560	69.96%	0.56%
VGG-16 (Han, Srinivasan, and Roy 2020)	RMP (soft-reset)	73.49%	2560	73.09%	0.4%
VGG-16 (Han and Roy 2020)	TSC (soft-reset)	73.49%	2560	73.46%	0.03%
VGG-16 (Deng and Gu 2021)	ReLU+threshold (soft-reset)	73.47%	128	71.06%	2.41%
VGG-16 [This work]	**SpikeConverter (soft-reset)**	**73.49%**	**16**	**73.47%**	**0.02%**
MobileNet-v2 [This work]	**SpikeConverter (soft-reset)**	**71.88%**	**16**	**71.71%**	**0.17%**

Inference Performance



Inference Computation Cost