Temporal Pulses Driven Spiking Neural Network for Time and Power Efficient Object Recognition in Autonomous Driving

Wei Wang [1], Shibo Zhou [2], Jingxi Li [2], Xiaohua Li [2], Junsong Yuan [1], Zhanpeng Jin [1]
wwang49@buffalo.edu, {szhou19, xli}@binghamton.edu, {jsyuan, zjin}@buffalo.edu
[1] University at Buffalo (SUNY), [2] Binghamton University (SUNY)

Abstract

- We proposed an **SNN model with temporal coding** to directly **process LiDAR pulses** for object recognition.
- We created a comprehensive **temporal pulses dataset**, **"Sim LiDAR"**, which simulated LiDAR reflection of different road conditions and target objects in diverse noise environments.
- Demonstrates novel contribution in **time and computational efficiency** for real-time deep learning applications.

Motivation & Rationale

Why SNN: [See Figure 1]

A. Spikes have **inherent temporal information**.
B. Integrate and fire, more **biological plausible**.
C. Event driven, **asynchronism**, **energy efficient**.

Why raw temporal pulses (rather than point clouds)**:**

A. Eliminates the **restrictions of frames**.
B. Can achieve **better time efficiency** with **less computational overhead**.

Methods

Neuronal model:

non-leaky integrate and fire (n-LIF) neuron [1]

$$\frac{dv_j(t)}{dt} = \sum_i w_{ji}\kappa(t-t_i)$$

$$\kappa(t) = u(t)e^{-\frac{t}{\tau}} \qquad u(t) = \begin{cases} 1, & \text{if } t \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Temporal coding: $$e^{t_j} = \sum_{i \in C} e^{t_i} \frac{w_{ji}}{\sum_{\ell \in C} w_{j\ell} - 1}$$

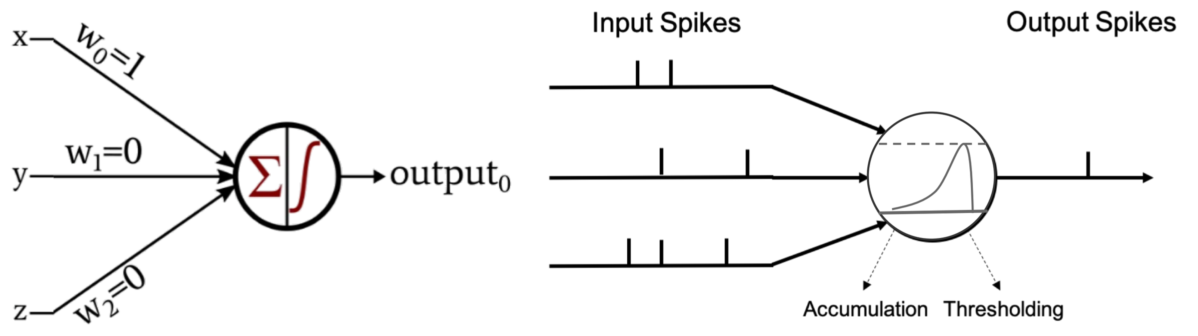

Figure 1. Conventional Neuron vs. Spiking Neuron

Network structure: *Spiking CNN*

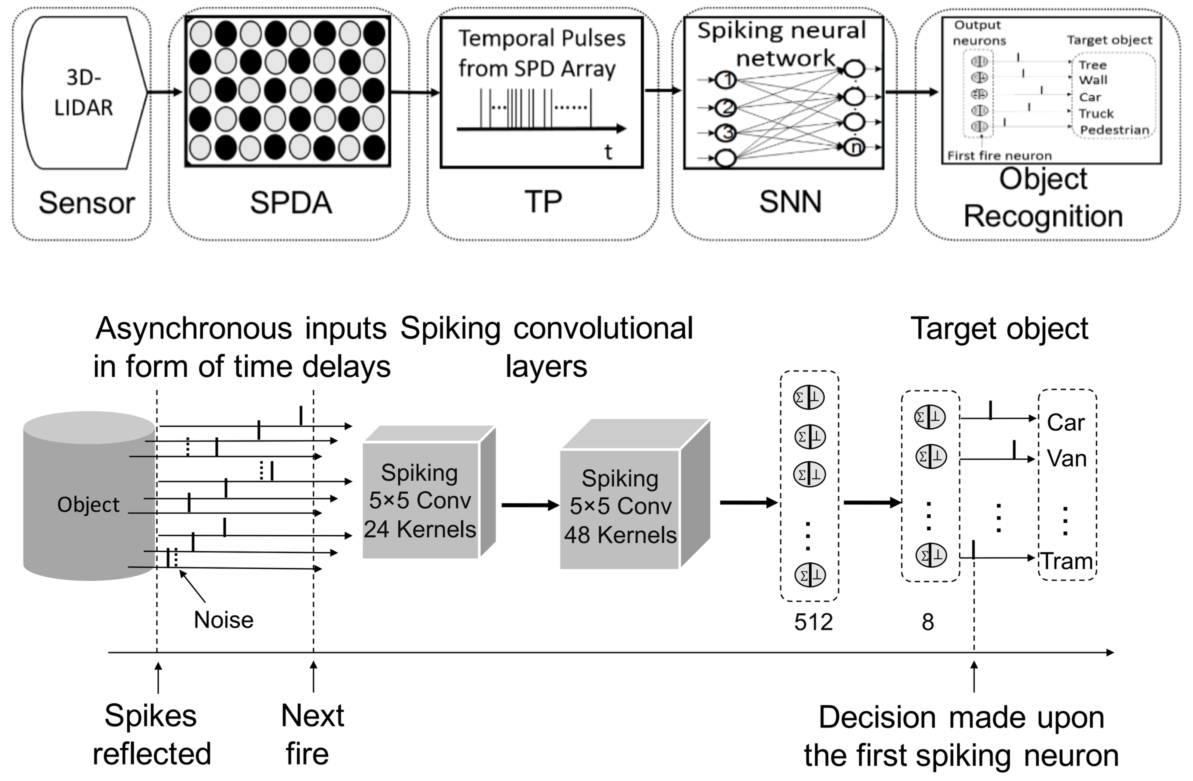

Figure 2. Flow diagram (top) and spiking CNN network (bottom) for object recognition

Datasets

Sim LiDAR (32 categories)

Objects:
car, pedestrian, truck

Road conditions:
tunnel, open road
lower/upper bridge
road (walls on one/two side)
road (lamps on one/two side)

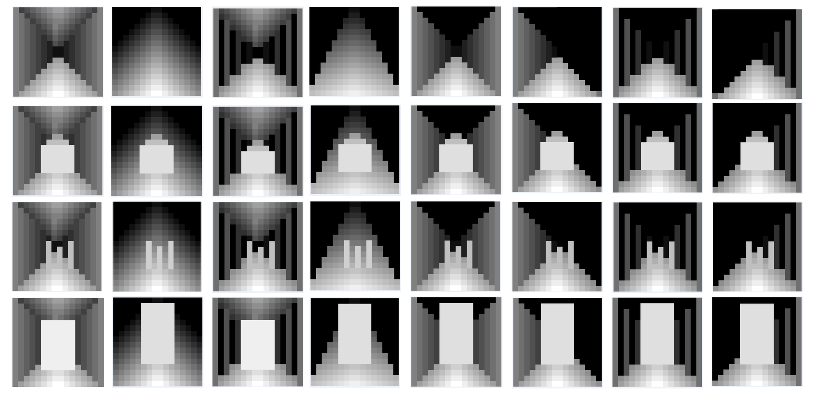

Figure 3. Sim LiDAR covers different objects and road conditions

KITTI (truncated, 8 categories)

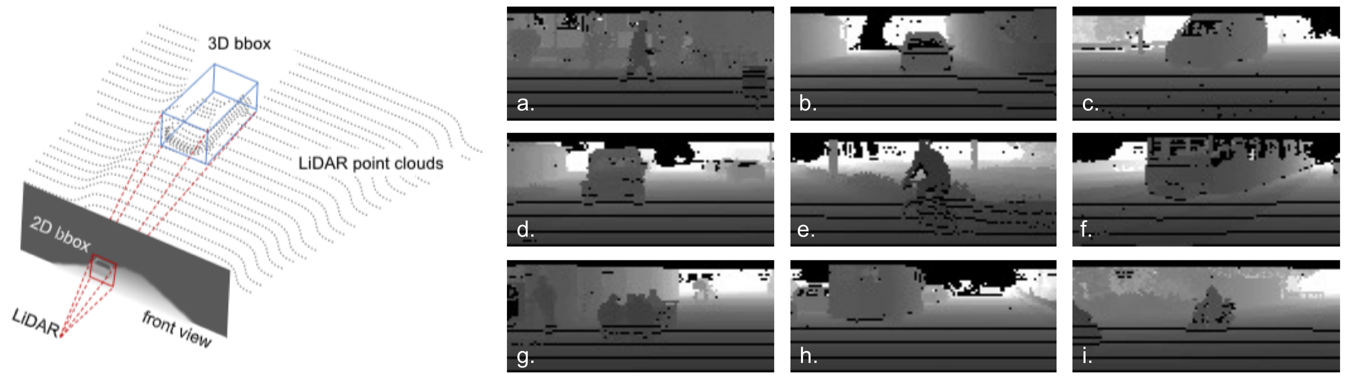

Figure 4. Truncation of KITTI with examples

Dynamic vision sensor (DVS) dataset (36 categories)

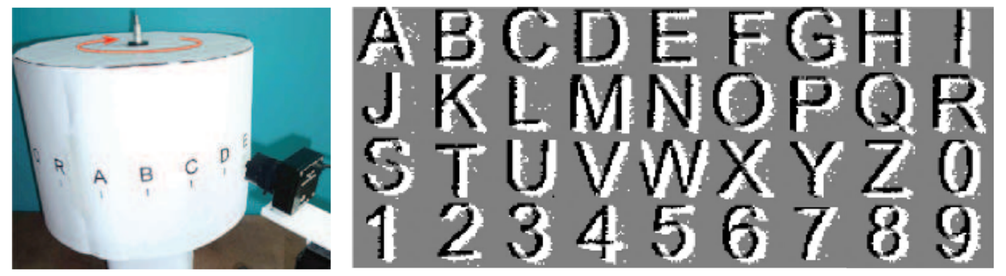

Figure 5. DVS barrel dataset [2]

Results

DVS Dataset (compare with existing models)

Model	Method	Accuracy
[2]	HFirst Temporal	84.9%
[3]	CNN Spike-based	91.6%
[3]	CNN Frame-based	95.2%
Our model	Spiking MLP	**99.5%**

KITTI (compare with conventional CNN)

Model	Acc. (%)	T_{rec} (ms)	R_{data} (%)	Power Consumption (α=0.37 pJ / α=45 pJ)
SCNN	96.62	2.02	76	29.83 nJ/3.63 μJ
CNN	88.22	2.58	100	0.67 J
VGG-16	92.72	11.34	100	2.95 J
ResNet-50	92.84	71.30	100	18.54 J

$$\text{data ratio: } R_{data} = \frac{N_{\text{contributing\_pulses}}}{N_{\text{all\_pulses}}}$$

The proposed SNN only needs part of the input pulses!

Sim LiDAR (robustness against noise)

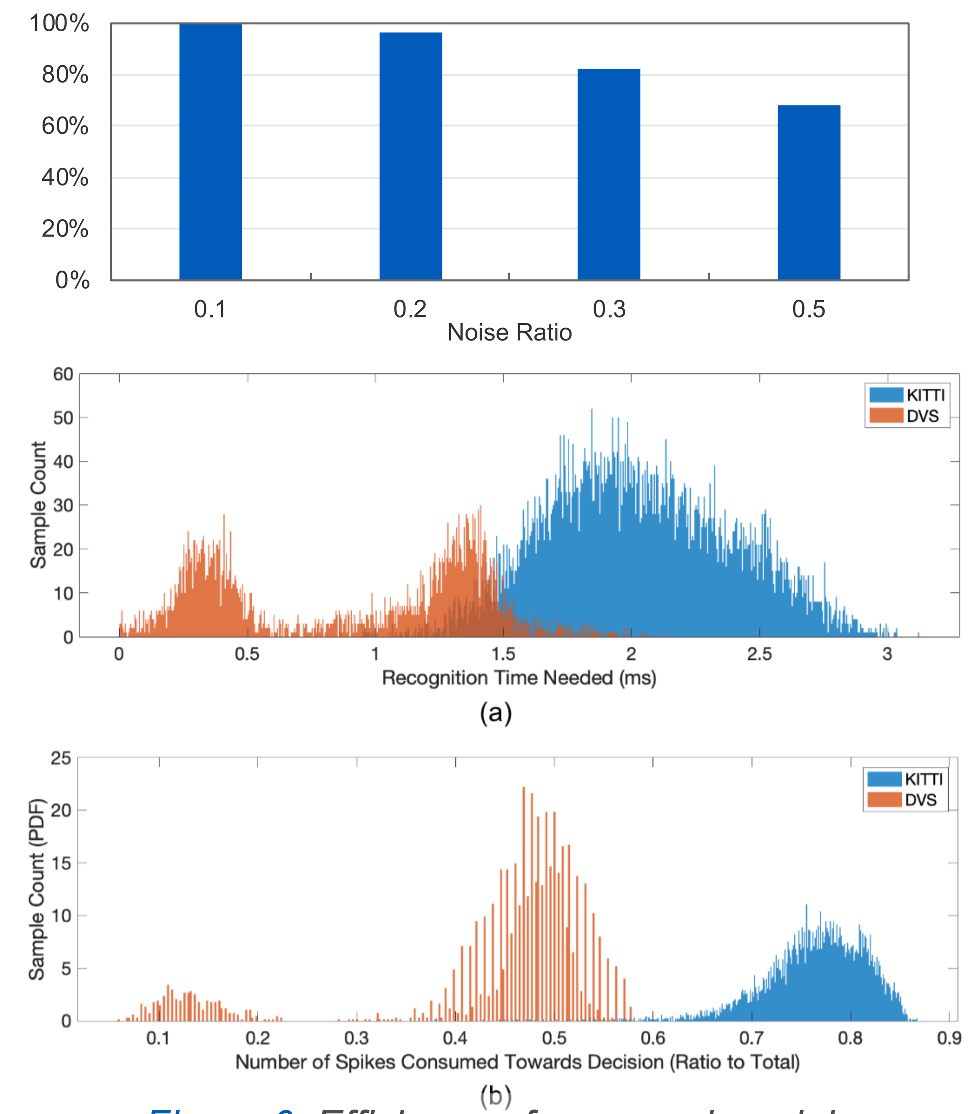

Figure 6. Efficiency of proposed model

Conclusion

- The proposed SNN has remarkable accuracy
- Extraordinary time and energy efficiency
- Great potential in resource- and/or time-constrained applications.
- Calls for the combination with neuromorphic hardware.

References

[1] H. Mostafa, "Supervised learning based on temporal coding in spiking neural networks", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 3227–3235, 2018.
[2] G. Orchard, C. Meyer, R. Etienne-Cummings, C. Posch, N. Thakor, and R. Benosman, "HFirst: a temporal approach to object recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37,no. 10, pp. 2028–2040, 2015.
[3] J. A. P´erez-Carrasco, B. Zhao, C. Serrano, B. Acha, T. Serrano-Gotarredona, S. Chen, and B. Linares-Barranco, "Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing–application to feedforward ConvNets", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 11, pp. 2706–2719, 2013