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

Wei Wang¹, Shibo Zhou², Jingxi Li², Xiaohua Li², Junsong Yuan¹, Zhanpeng Jin¹

¹ University at Buffalo, ² Binghamton University
OUTLINE

I. Introduction
II. Motivation
III. Rationale & Methods
IV. Experiment & Evaluation
V. Discussion & Conclusion
Introduction

Objective

- Explore a new paradigm for object recognition in autonomous driving:
  - Excellent accuracy
  - Ultra low power consumption and time delay

Contribution

- An event driven model
  Spiking neural networks that can directly process raw LiDAR pulses (without point-cloud or voxelization).

- Sim LiDAR dataset
  A comprehensive temporal pulses dataset that simulates LiDAR reflection of different road conditions and target objects in diverse noise environments.

- Extraordinary time- and energy-efficiency on real-world data
Motivation

- Why Spiking Neural Network (SNN) (rather than conventional neural network)
  - **Data representation**: Spikes have inherent temporal information.
  - **Info. processing**: Event driven, asynchronism
  - **Hardware friendly**: energy efficient

- Why Raw Temporal Pulses (rather than point-clouds/voxels)
  - Eliminates the restrictions of frames.
  - Can achieve better time efficiency
  - Have less computational overhead.
Rationale & Methods

☐ Neuronal Model

**Input / Output**
- **ANN**: continuous real value / non-sparse
- **SNN**: discrete spikes / sparse

**Neuronal Model**
- **ANN**: linear combination + activation (weighted sum)
- **SNN**: accumulation + thresholding (integrate-and-fire)

*Biological Plausible*
Rationale & Methods

Neuronal Model

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

Integrate:
- Synaptic current $I$: when there is a spike propagating
- Membrane voltage $V$: accumulates from the current

Fire:
- generate a spike when $V >$ threshold

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

\[
\kappa(t) = u(t)e^{-\frac{t}{\tau}}
\]

\[
u(t) = \begin{cases} 
1, & \text{if } t \geq 0 \\
0, & \text{otherwise}
\end{cases}
\]

Rationale & Methods

- **Encoding Techniques of Spike Train**
  - Spike count / rate
  - Temporal coding (arrival/interval time)
- **System Flow & Network structure**

\[
e^{t_j} = \sum_{i \in C} e^{t_i} \sum_{\ell \in C} w_{ji} - 1
\]
Experiment & Evaluation

Datasets

- **Sim LiDAR**
  - Arrival time
  - 32 categories
  - Noise injection

- **KITTI**
  - Truncated
  - 8 categories
  - 32,456/8,000 samples (train/test)

- **Dynamic visual sensor (DVS) dataset [2]**
  - event-based camera (asynchronous)
  - 36 categories (a ~ z, 0 ~ 9)
  - 3,453/3,000 samples (train/test)

Objects:
- Pedestrian (a)
- Car (b), Van (c)
- Truck (d), Cyclist (e)
- Tram (f)
- Person sitting (g)
- Misc (h, i)

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

Diverse noise levels

Converted to gray-scale images ONLY for visualization

DVS barrel dataset [2]

## Experiment & Evaluation

### Results

- **DVS Dataset**  
  (compare with existing models)

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2]</td>
<td>HFirst Temporal</td>
<td>84.9%</td>
</tr>
<tr>
<td>[3]</td>
<td>CNN Spike-based</td>
<td>91.6%</td>
</tr>
<tr>
<td>[3]</td>
<td>CNN Frame-based</td>
<td>95.2%</td>
</tr>
<tr>
<td>Our model</td>
<td>Spiking MLP</td>
<td>99.5%</td>
</tr>
</tbody>
</table>

- **KITTI**  
  (compare with conventional CNN)

### Sim LiDAR  
(robustness against noise)

<table>
<thead>
<tr>
<th>Noise Range</th>
<th>0 - 0.10</th>
<th>0 - 0.20</th>
<th>0 - 0.33</th>
<th>0 - 0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>99.83%</td>
<td>96.16%</td>
<td>82.66%</td>
<td>68.16%</td>
</tr>
</tbody>
</table>

- **Efficiency of proposed model**

The proposed SNN only needs part of the input pulses!
Discussion & Conclusion

Conclusion
- The proposed SNN has remarkable accuracy
- Extraordinary time and energy efficiency
- Great potential in resource- and/or time-constrained applications

Future Work
- Event-driven effective and efficient 3D object detection
- Combination with neuromorphic hardware
- End to end event driven solution
  - Event-driven sensor
  - Spiking hardware with LIF unit
  - SNN model
Thank You!