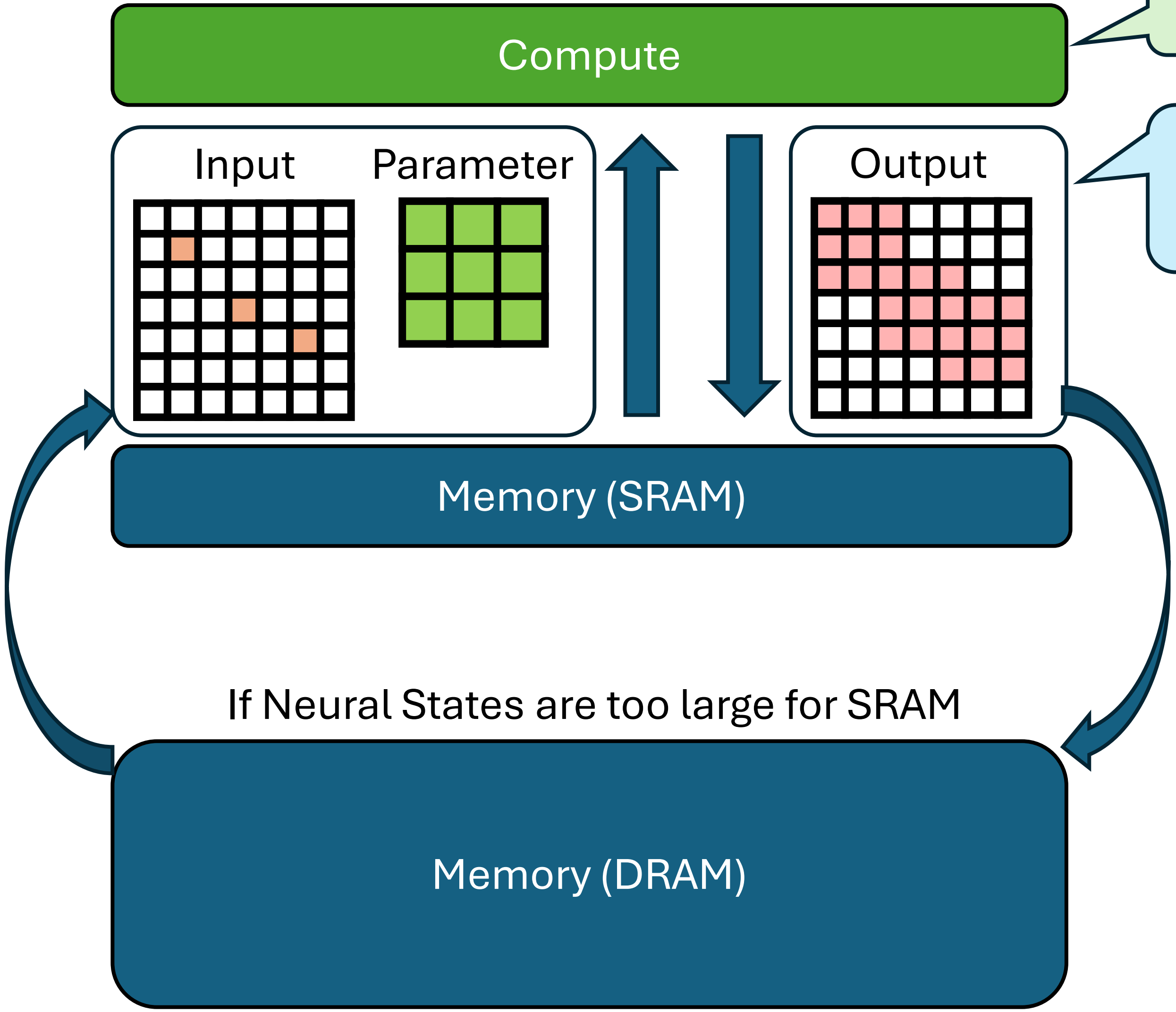
Sparse Convolutional Recurrent Learning for Efficient Event-based Object Detection

<u>Guangzhi Tang</u>
Assistant Professor
Department of Advanced Computing Sciences
Maastricht University

Making CNN more efficient on Digital Neuromorphic Processor

Layer-by-layer CNN processing on GPU

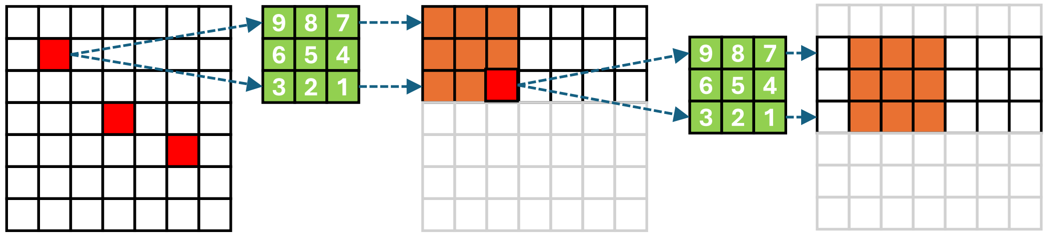


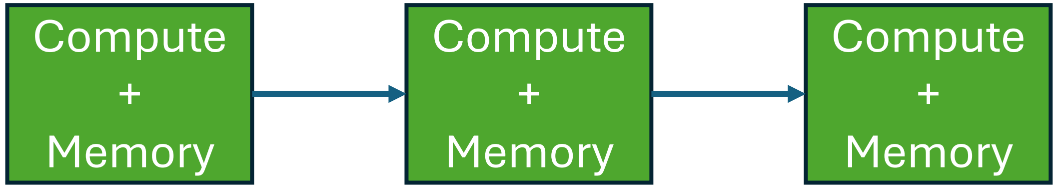
Can we exploit **sparse activation** inputs?

Can we avoid **intensive data movement and memory cost** of neural states and activations?

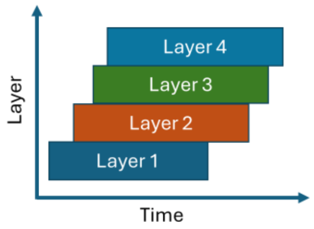
Event-driven

Depth-first Convolution





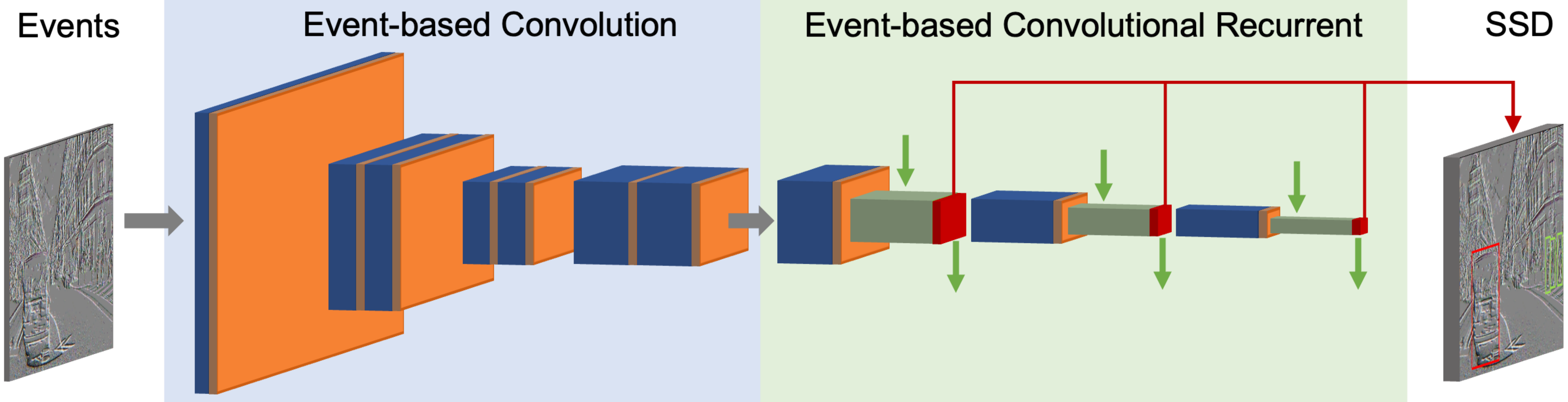
Reduce memory cost
Reuse 3 rows state memory



Reduce latency
Layer starts as soon as possible

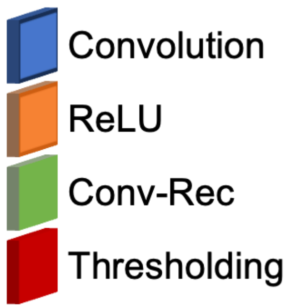
Xu, Yingfu, et al. "Optimizing event-based neural networks on digital neuromorphic architecture: a comprehensive design space exploration.“ 2024

Using Fully Convolutional Network for Event-based Object Detection



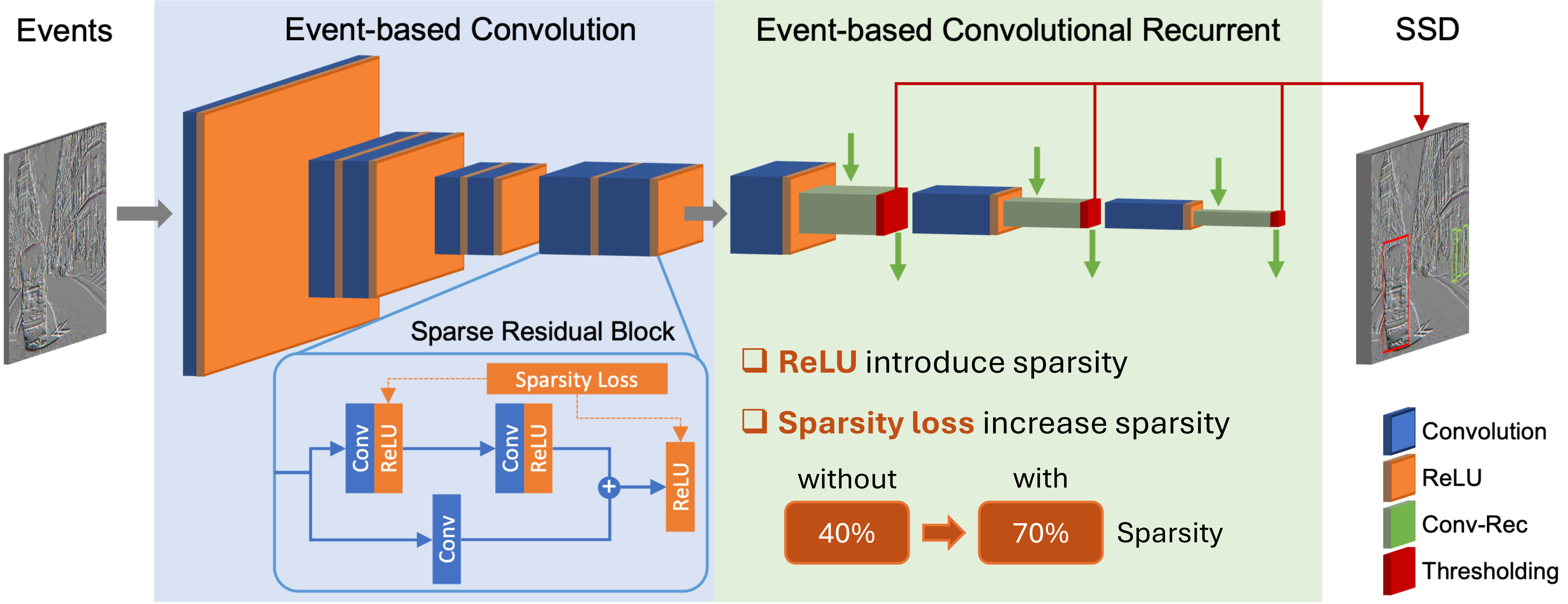
We can deploy a fully convolutional network with event-driven depth-first convolution on neuromorphic processor

But how to make the activation events sparse?



Wang, Shenqi, et al. "Sparse convolutional recurrent learning for efficient event-based neuromorphic object detection." In review, 2025.

Using Fully Convolutional Network for Event-based Object Detection

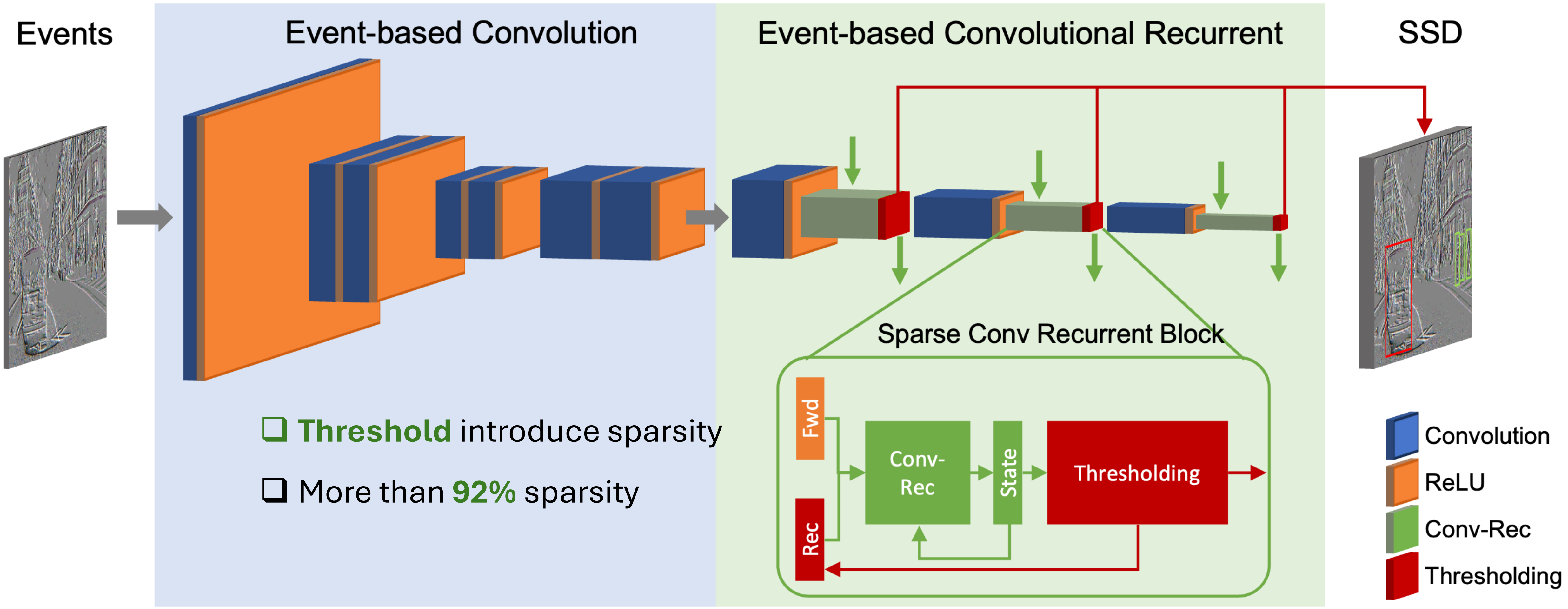


- **ReLU** introduce sparsity
- **Sparsity loss** increase sparsity

	without		with	
	40%	→	70%	Sparsity

Wang, Shenqi, et al. "Sparse convolutional recurrent learning for efficient event-based neuromorphic object detection." In review, 2025.

Using Fully Convolutional Network for Event-based Object Detection

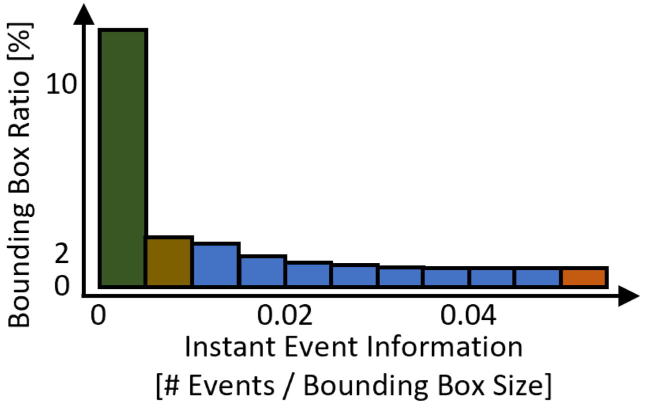


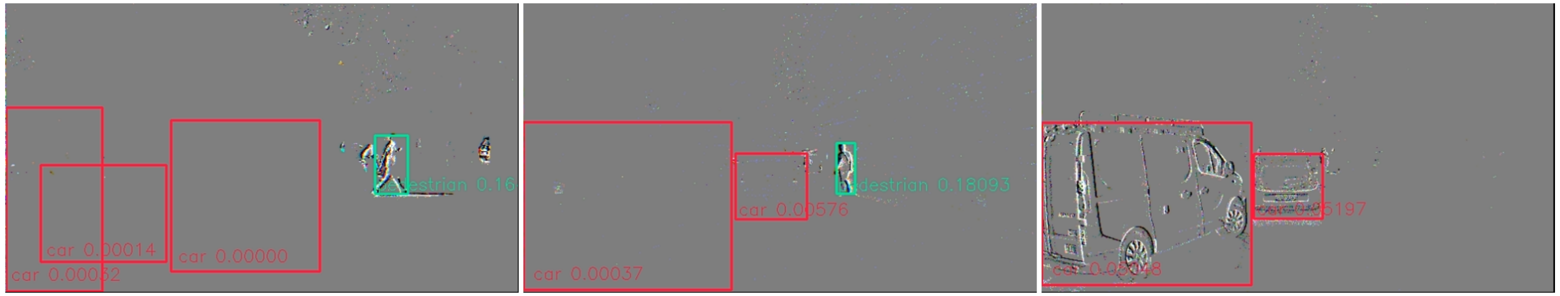
- ❑ **Threshold** introduce sparsity
- ❑ More than **92%** sparsity

Is there still temporal learning with less than 10% of activation pass to next step?

Wang, Shenqi, et al. "Sparse convolutional recurrent learning for efficient event-based neuromorphic object detection." In review, 2025.

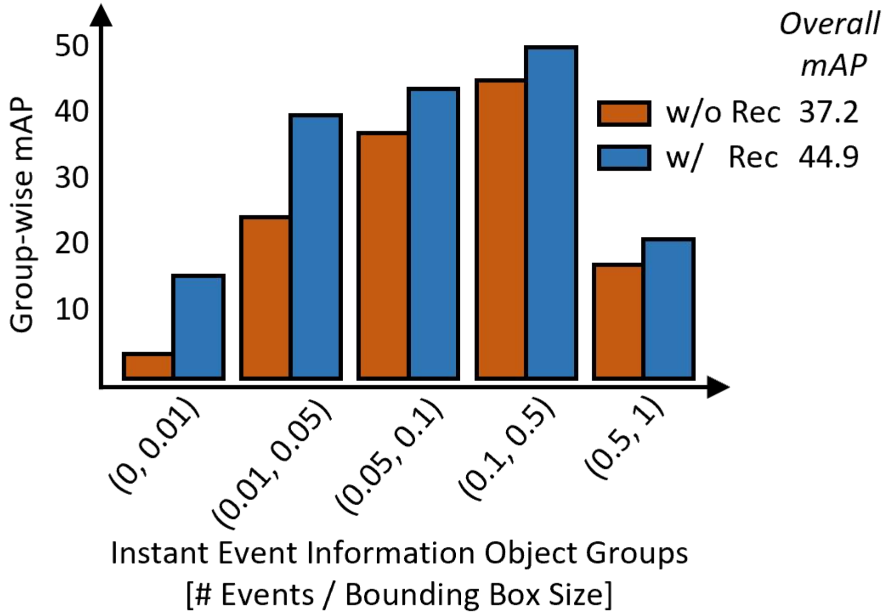
Effectiveness of Sparse Convolutional Recurrent Learning

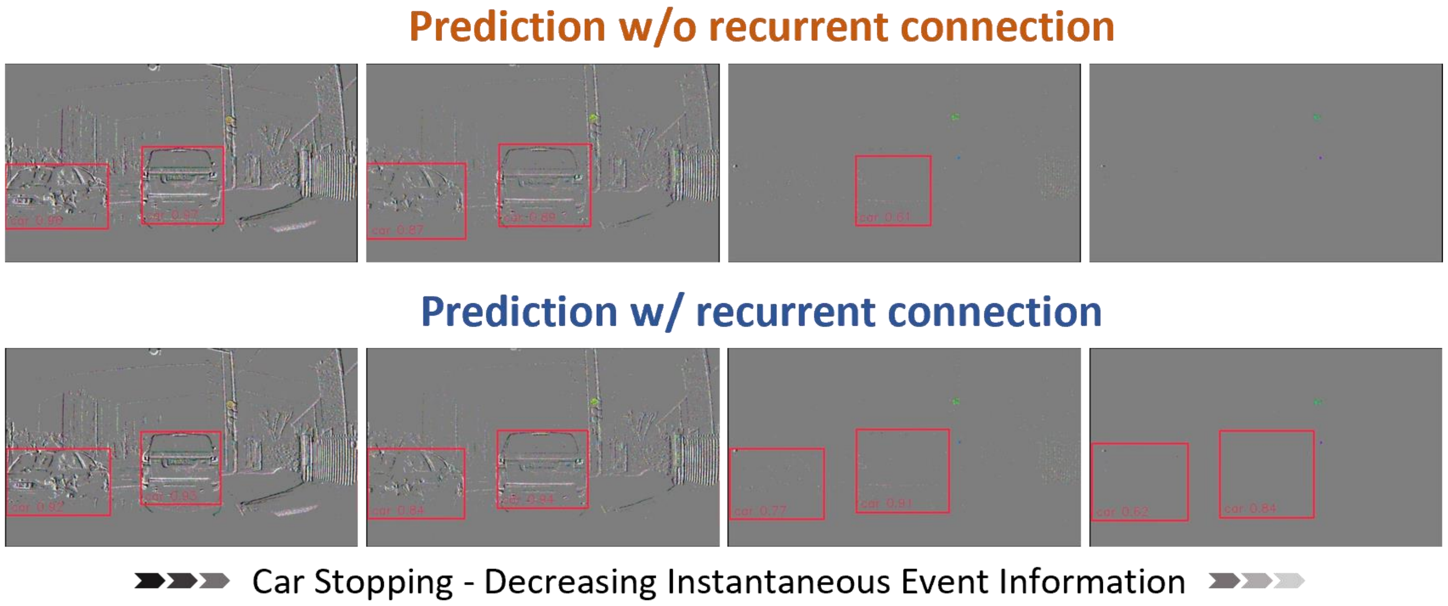




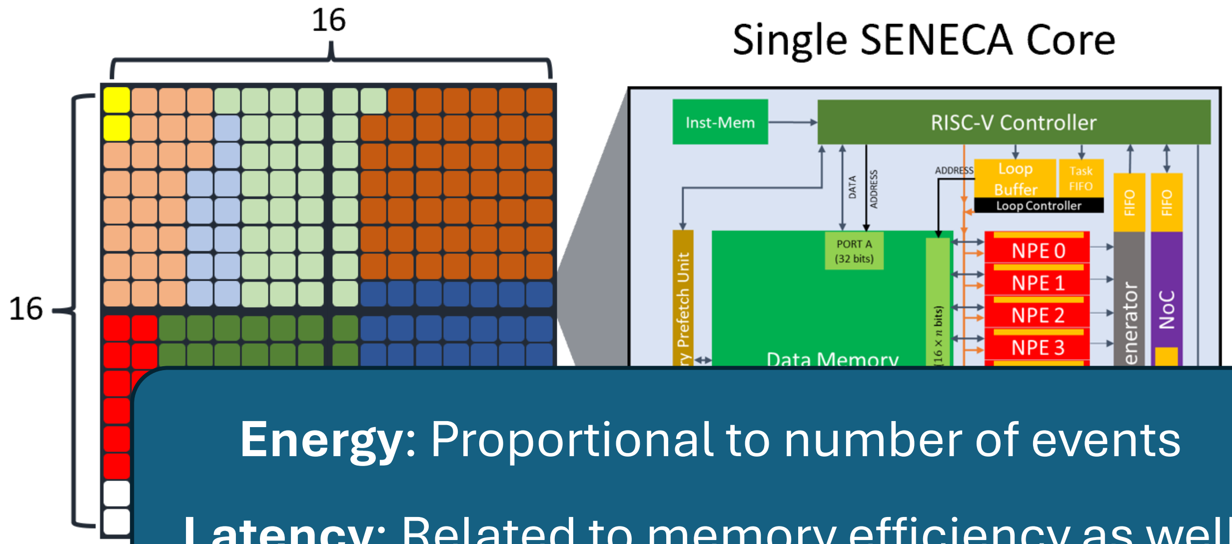
Car Stop (0 – 0.0003)　　Car Stop (0.005)　　Car Start Moving (0.05)

Less or no events when the relative speed of camera and objects reduces





Hardware Simulation Study on the SENECA Neuromorphic Processor



Energy: Proportional to number of events

Latency: Related to memory efficiency as well

Event Integration

MLD // Load State	3.7pJ
MLD // Load Weight	3.7pJ
MUL	1.4pJ
ADD	1.4pJ
MST // Store State	3.9pJ
Total	*14.1pJ*

Event Generation

MLD // Load State	3.7pJ
MLD // Load Bias	3.7pJ
MAX	1.2pJ
EVC (+1.1pJ/Event)	0.5pJ
MST // Store State	3.9pJ
Total	*13.0pJ*

Method	Recurrent Dimension	Sparse Recurrent	Sparsity Loss	Recurrent Unit	Energy (mJ)	Latency (ms)	Memory (Mb)	Cores	GSOp
SEED	128	Yes	Yes	3× GRU	39.3	21.6	90.8	236	2.75
	256	Yes	Yes	3× GRU	54.7	44.9	245.1	254	3.83
SEED (w/ ablations)	128	Yes	No	3× GRU	68.1	44.0	90.8	256	4.79
	256	Yes	No	3× GRU	80.9	81.43	245.1	254	5.69
	256	No	No	3× GRU	284.3	240.1	226.1	234	20.1
RED (w/o SE)	256	No	No	5× LSTM	368.6	1077.3	404.1	343	26.1
	256	No	No	5× LSTM	368.6	40.7	404.1	1446	26.1

Tang, Guangzhi, et al. "Open the box of digital neuromorphic processor: Towards effective algorithm-hardware co-design." 2023.

Shenqi Wang

Yingfu Xu

Amirreza Yousefzadeh

Sherif Eissa

Henk Corporaal

Federico Corradi

imec

Maastricht University

KDT JU Key Digital Technologies Joint Undertaking

TEMPO

DAIS

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

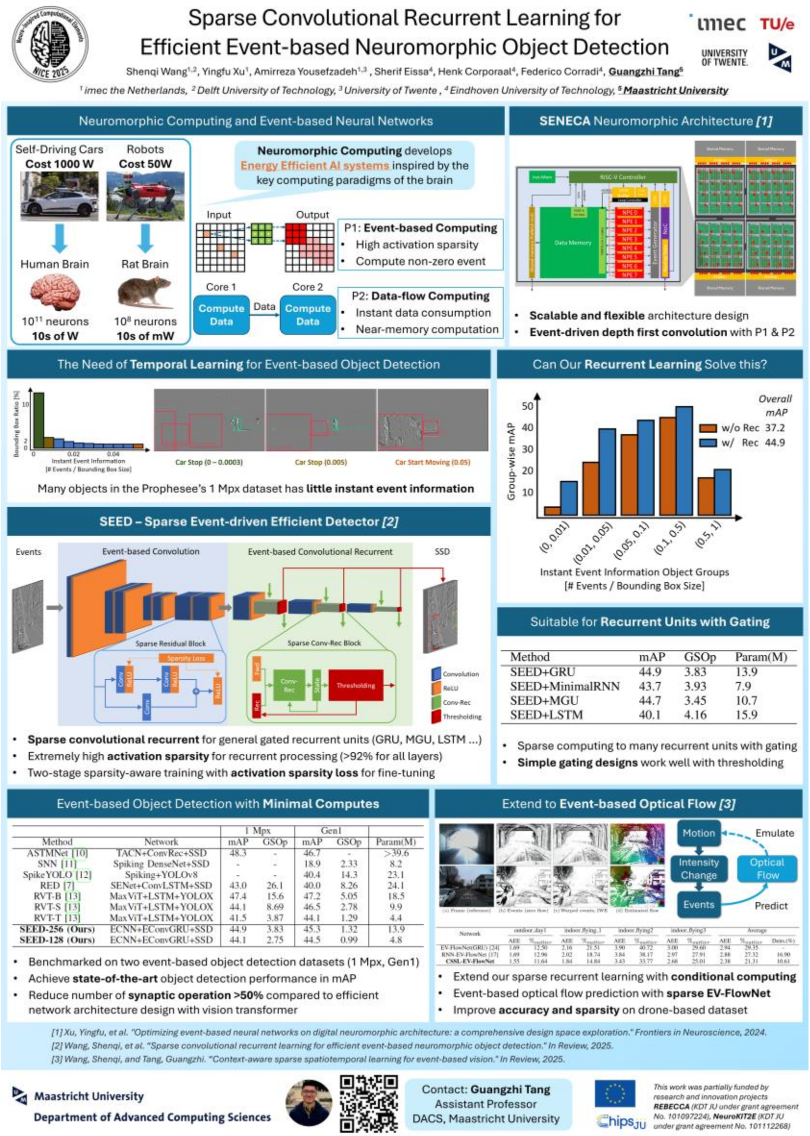
UNIVERSITY OF TWENTE.

ECSEL Joint Undertaking

ANDANTE

REBECCA

ChipsJU



Please visit our poster for more discussions

Maastricht University

Thank You