SPIKEPOINT

香港科技大學(廣州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

ICLR

SpikePoint: An Efficient Point-based Spiking Neural Network for Event Cameras Action Recognition

Action Recognition, SNN, Event Camera

Hongwei Ren, Yue Zhou, Xiaopeng Lin, Yulong Huang, Haotian Fu, Jie Song, Bojun Cheng*

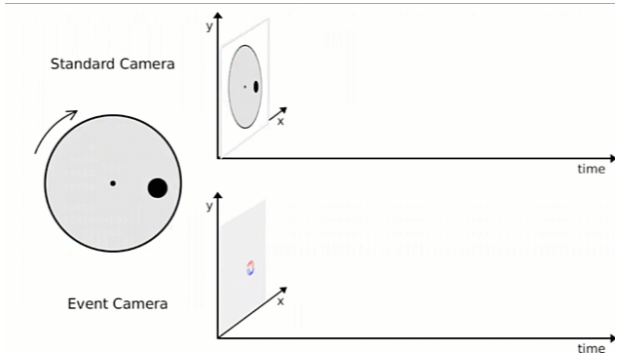
Neuro. Elec & Pho. Lab

Ren H, Zhou Y, Haotian F U, et al. SpikePoint: An Efficient Point-based Spiking Neural Network for Event Cameras Action Recognition[C]//The Twelfth International Conference on Learning Representations. 2023.

Background

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

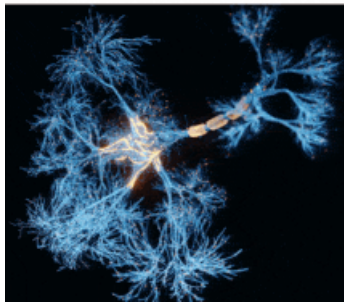
Event Cameras

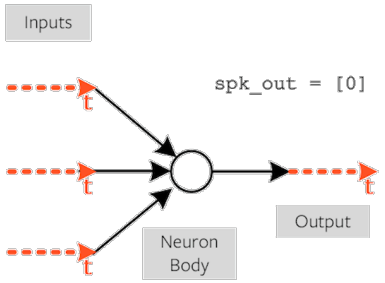


- High dynamic range.
- Low Latency.
- Low Power Consumption.
- Fast Motion Handling.
- Robust to Motion Blur.

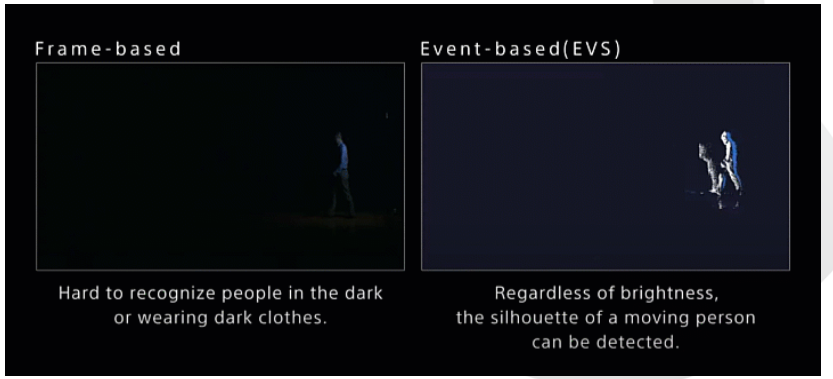
Ultra-low-power Application (action recognition)

Spiking Neural Network





- Binary information.
- Biologically inspired network.
- Low Power Consumption.



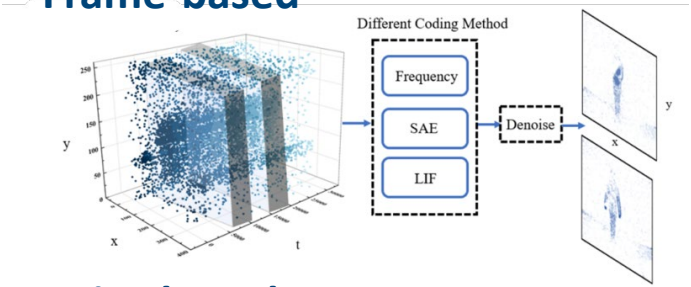
Source link: https://qbi.uq.edu.au/sahgroup, https://pypi.org/project/snntorch/

Background

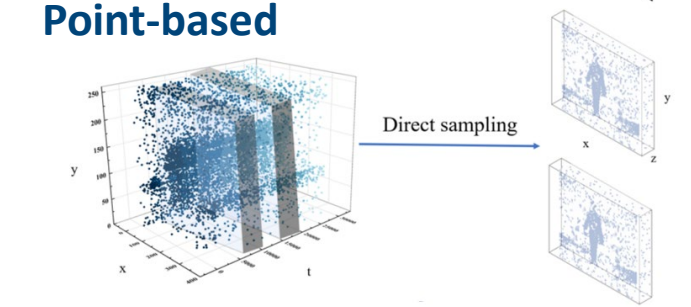
ICLR

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

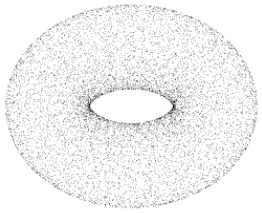
Frame-based



Point-based



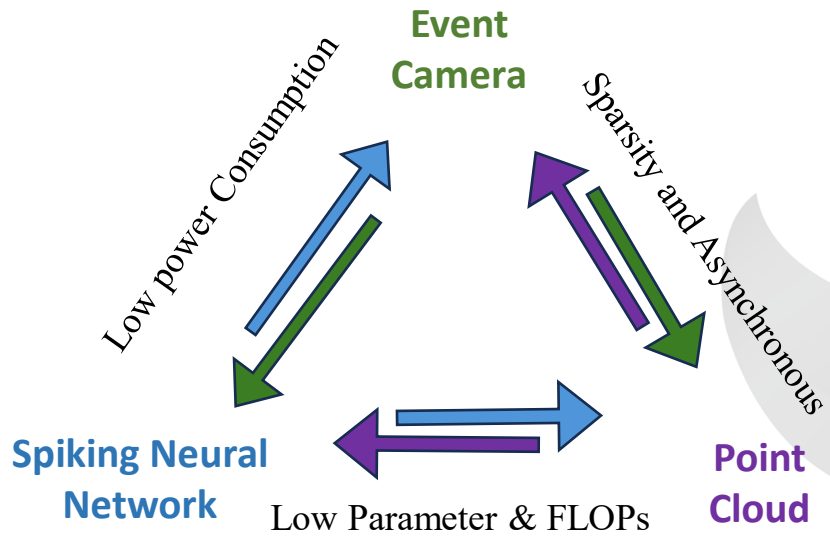
Point Cloud



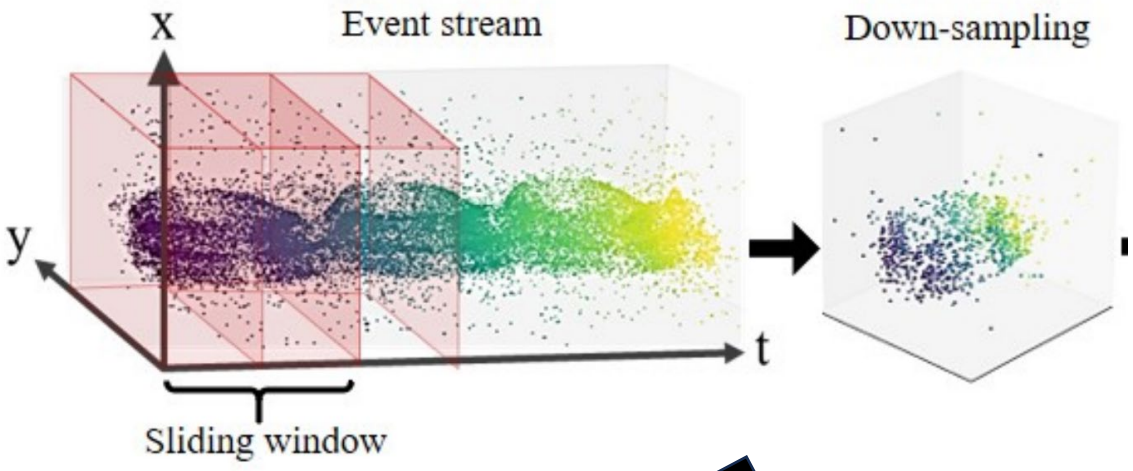
- Sparsity.
- (x,y,z).
- Permutation invariance.

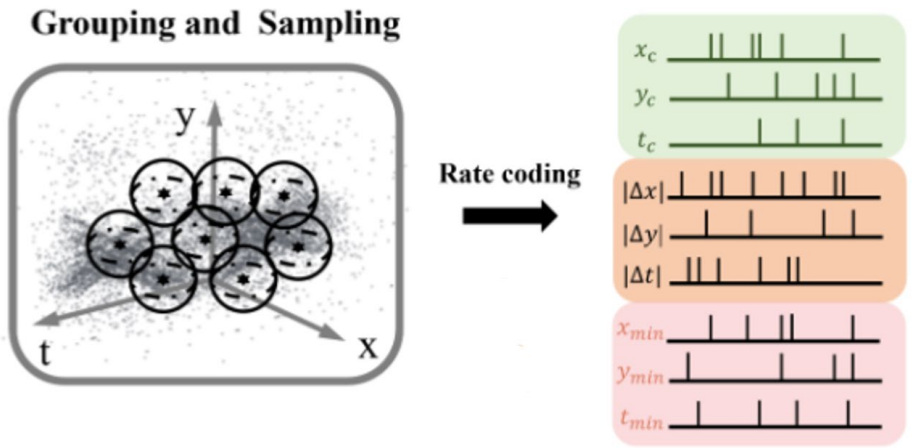
Different Event-based Method

1. Time-consuming Representation conversion.
2. Mature backbone network for Frame (VGG, ResNet, VIT).
3. Dense Computing and Synchronization(not match for DVS).
4. High Accuracy than point-based method.



Preprocessing





1. Ignore events' polarity .

$$\mathcal{E} = \{e_k = (x_k, y_k, t_k, p_k) \mid k = 1, \ldots, n\}$$

2. Divided by sliding window.

$$\mathcal{P}_i = \{e_{j \to l} \mid t_l - t_j = R\} \quad i = 1, \ldots, m$$

3. Down-sampling (1024 events) and normalization.

$$PN_i = (\frac{X_i}{w}, \frac{Y_i}{h}, \frac{T_i - t_j}{t_l - t_j})$$

4. Sampling and Grouping.

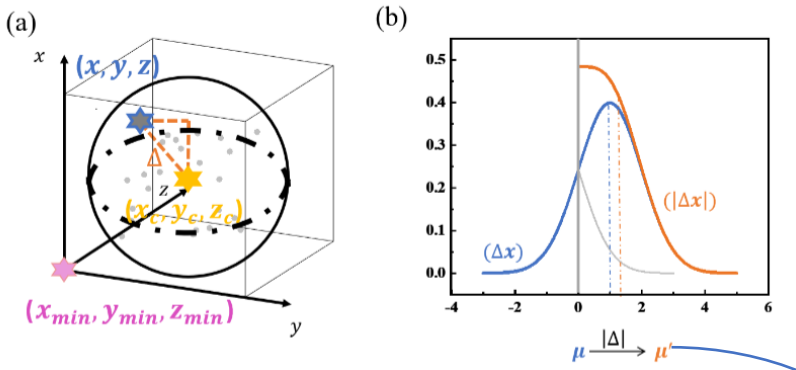
$$PS_i = FPS(PN_i) \quad PG_i = KNN(PN_i, PS_i)$$

5. Standardization.

$$[\Delta x, \Delta y, \Delta z] = \frac{\mathcal{G} - \text{Centroid}}{SD(\mathcal{G})} \sim N(0,1), \quad SD(\mathcal{G}) = \sqrt{\frac{\sum_{i=1}^{n}(g_i - \bar{g})^2}{n-1}} \quad g_i \in \mathcal{G}$$

6. Rate Coding. Challenge!

Challenge: Sampling and Grouping



Visualization the shift of absolute operation. (a) The spatial coordinate of $[x_{min}, y_{min}, z_{min}]$ and $[x_c, y_c, z_c]$. (b) The transformation of the distribution after taking absolute.

$$[\Delta x, \Delta y, \Delta z] = \frac{\mathcal{G} - \text{Centroid}}{SD(\mathcal{G})} \sim N(0,1)$$

There is a negative value that cannot be spike encoded.

$$[\Delta x, \Delta y, \Delta z] = \left| \frac{\mathcal{G} - \text{Centroid}}{SD(\mathcal{G})} \right| \sim N(0,1)$$

Take absolute values, but the data distribution will shift.

$$f(x;\mu,\delta^2) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(x-\mu)^2}{2\delta^2}} \rightarrow \frac{1}{\sqrt{2\pi}\delta}(e^{-\frac{(x-\mu)^2}{2\delta^2}} + e^{-\frac{(x+\mu)^2}{2\delta^2}})(x \geq 0)$$

$$\dot{\mu} = \int_0^\infty x \frac{1}{\sqrt{2\pi}}(e^{-\frac{x^2}{2}} + e^{-\frac{x^2}{2}})dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x e^{-\frac{x^2}{2}} dx = \sqrt{\frac{2}{\pi}} \int_0^\infty e^{-\frac{x^2}{2}} d(\frac{x^2}{2}) = \sqrt{\frac{2}{\pi}}$$

$$\sqrt{\frac{2}{\pi}},$$

By modify the input ($[\Delta x, \Delta y, \Delta z, x_{min}, y_{min}, z_{min}]$), our method results in a 76% reduction in rate encoding error of the coordinates in Daily DVS dataset.

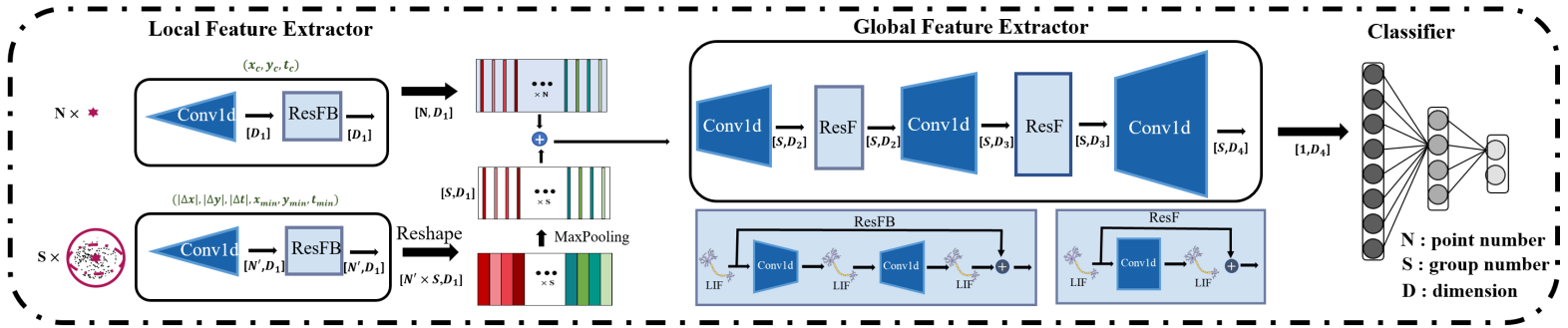
The performance get better.

No.	$[x_{min}, \dots]$	$[x_c, \dots]$	Acc.
1	√	×	97.92%
2	×	√	96.25%

Challenge: Hierarchical Structure

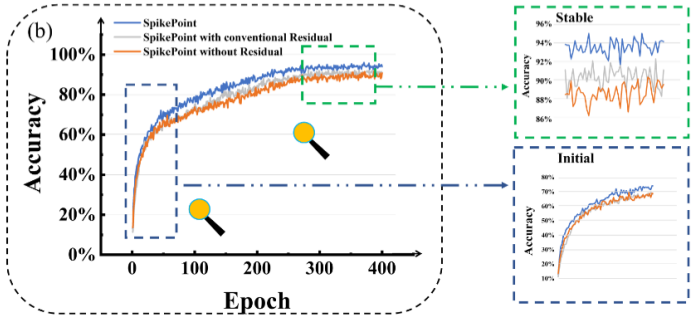
香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

singular stage structure



The overall architecture of SpikePoint.

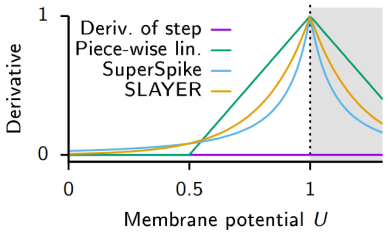
Ablation of different residual connection.



Advantages

1. Global and local structural module. Maintain high precision.

2. Feature extractor designed for SNN, capable of back-propagation training.



$$\triangle W_{ij}^{l} = \frac{\partial L}{\partial W_{ij}^{l}} = \sigma^{'}(I_i^l) S_j^{l-1} (\sum_k \delta_k^{l+1} W_{ik}^{T,l} + \sigma^{'}(I_i^{l+m-1} + S_j^l) \sum_k \delta_k^{l+m} W_{ik}^{T,l+m-1})$$

3. Solve spiking neural network negative coding problem with two-branch architecture.

Experiment

ICLR

香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

1. DVS 128 Gesture.

Name	Method	Param	Acc
TBR+I3D (Innocenti et al., 2021)	ANN	12.25 M	99.6%
PointNet++ (Qi et al., 2017b)	ANN	1.48 M	95.3%
EV-VGCNN (Deng et al., 2021)	ANN	0.82 M	95.7%
VMV-GCN (Xie et al., 2022)	ANN	0.86 M	97.5%
SEW-ResNet (Fang et al., 2021a)	SNN	-	97.9%
Deep SNN(16 layers) (Amir et al., 2017)	SNN	-	91.8%
Deep SNN(8 layers) (Shrestha & Orchard, 2018)	SNN	-	93.6%
Conv-RNN SNN(5 layers)(Xing et al., 2020)	SNN	-	92.0%
Conv+Reservoir SNN (George et al., 2020)	SNN	-	65.0%
HMAX-based SNN (Liu et al., 2020)	SNN	-	70.1%
Motion-based SNN (Liu et al., 2021a)	SNN	-	92.7%
SpikePoint	**SNN**	**0.58 M**	**98.74%**

2. Daily DVS.

Name	Method	Param	Acc
I3D(Carreira & Zisserman, 2017)	ANN	49.19 M	96.2%
TANet(Liu et al., 2021b)	ANN	24.8 M	96.5%
VMV-GCN (Xie et al., 2022)	ANN	0.84 M	94.1%
TimeSformer (Bertasius et al., 2021)	ANN	121.27 M	90.6%
HMAX-based SNN (Xiao et al., 2019)	SNN	-	68.3%
HMAX-based SNN (Liu et al., 2020)	SNN	-	76.9%
Motion-based SNN (Liu et al., 2021a)	SNN	-	90.3%
SpikePoint	**SNN**	**0.16 M**	**97.92%**

Summary

1. Minimal number of parameters.

2. Super high accuracy (results better than most ANNs).

3. Training by back-propagation, not ANN2SNN.

4. Low power consumption Real-time action recognition applications.

3. DVS Action.

Name	Method	Param	Acc
ST-EVNet (Wang et al., 2020)	ANN	1.6 M	88.7%
PointNet (Qi et al., 2017a)	ANN	3.46 M	75.1%
Deep SNN(6 layers) (Gu et al., 2019)	SNN	-	71.2%
HMAX-based SNN (Xiao et al., 2019)	SNN	-	55.0%
Motion-based SNN (Liu et al., 2021a)	SNN	-	78.1%
SpikePoint	**SNN**	**0.16 M**	**90.6%**

4. HMDB51-DVS

Name	Method	Param	Acc
C3D (Tran et al., 2015)	ANN	78.41 M	41.7%
I3D (Carreira & Zisserman, 2017)	ANN	12.37 M	46.6%
ResNet-34 (He et al., 2016)	ANN	63.7 M	43.8%
ResNext-50 (Hara et al., 2018)	ANN	26.5 M	39.4%
P3D-63 (Qiu et al., 2017)	ANN	25.74 M	40.4%
RG-CNN (Bi et al., 2020)	ANN	3.86 M	51.5%
SpikePoint	**SNN**	**0.79 M**	**55.6%**

5. UCF101-DVS

Name	Method	Param	Acc
RG-CNN +Incep.3D (Bi et al., 2020)	ANN	6.95M	63.2%
I3D (Carreira & Zisserman, 2017)	ANN	12.4M	63.5%
ResNext-50 (Hara et al., 2018)	ANN	26.05M	60.2%
ECSNet-SES (Chen et al., 2022)	ANN	-	70.2%
Res-SNN-18 (Fang et al., 2021a)	SNN	-	57.8%
RM-RES-SNN-18 (Yao et al., 2023)	SNN	-	58.5%
SpikePoint	SNN	1.05M	68.46%

Power consumption.

Model	Input	Timestep	Accuracy(%)	OPs(G)	Dynamic(mJ)	Para.(M)	Static(mJ)
SpikePoint	**Point**	**16**	**98.7**	**0.9**	**0.82**	**0.58**	**0.756**
SEW-ResNet(tdBN)[1]	Frame	40	96.9	4.79	4.36	11.7	15.305
Spikingformer[2]	Frame	16	**98.3**	3.72	4.26	**2.6**	**3.401**
Spikformer[3]	Frame	16	97.9	6.33	10.75	2.6	3.401
Deep SNN(16)[4]	Frame	16	91.8	2.74	2.49	1.7	2.223
Deep SNN(8)[5]	Frame	16	93.6	2.13	1.94	1.3	1.7
PLIF [6]	Frame	20	97.6	2.98	2.71	17.4	22.759
TBR+I3D [7]	Frame	ANN	**99.6**	38.82	178.6	**12.25**	**160.23**
Event Frames+I3D [8]	Frame	ANN	96.5	30.11	138.5	12.37	16.18
RG-CNN [9]	voxel	ANN	96.1	0.79	3.63	19.46	25.45
ACE-BET [10]	voxel	ANN	98.8	2.27	10.44	11.2	14.65
VMV-GCN [11]	voxel	ANN	97.5	0.33	1.52	0.84	1.098
PoinNet++ [12]	point	ANN	95.3	0.872	4.01	1.48	1.936

香港科技大學(廣州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Thank You!

Hongwei Ren