# EventF2S: Asynchronous and Sparse Spiking AER Framework using Neuromorphic-Friendly Algorithm

Lakshmi Annamalai
*Centre for Artificial Intelligence and Robotics*
*Defence Research and Development Organization*
Bangalore, India
lakshmia@alum.iisc.ac.in

Chetan Singh Thakur
*Department of Electronics Systems Engineering*
*Indian Institute of Science*
Bangalore, India
csthakur@iisc.ac.in

***Abstract*—Bio-inspired Address Event Representation (AER) sensors have attracted significant popularity owing to their low power consumption, high sparsity, and high temporal resolution. Spiking Neural Network (SNN) has become the inherent choice for AER data processing. However, the integration of the AER-SNN paradigm has not adequately explored asynchronous processing, neuromorphic compatibility, and sparse spiking, which are the key requirements of resource-constrained applications. To address this gap, we introduce a brain-inspired AER-SNN object recognition solution, which includes a data encoder integrated with a First-To-Spike recognition network. Being fascinated by the functionality of neurons in the visual cortex, we designed the solution to be asynchronous and compatible with neuromorphic hardware. Furthermore, we have adapted the principle of denoising and First-To-Spike coding to achieve optimal spike signaling, significantly reducing computation costs. Experimental evaluation has demonstrated that the proposed method incurs significantly less computation cost to achieve state-of-the-art competitive accuracy. Overall, the proposed solution offers an asynchronous and cost-effective AER recognition system that harnesses the full potential of AER sensors.**

***Index Terms*—Road segmentation, Semantic segmentation, flickering**

## I. Introduction

ADdress Event Representation (AER) sensors [1] [2] [3] [4] [5] [6] [7], which aims to mimic biological eyes, has gained a lot of attention from different communities owing to its unique features. The inherent working principle of AER sensors results in the benefit of data sparsity, thus providing huge energy-saving. In contrast to a conventional camera, each pixel of an event camera individually monitors the dynamics of the scene and records in the form of spiking asynchronous events when the change exceeds a pre-defined threshold.

The nature of AER data poses severe challenges in processing event data efficiently [8] [9] [10] [11] [12]. The asynchronous event-driven paradigm of the AER sensor makes it a natural choice for Spiking Neural Networks (SNN) [13] [14] [15]. SNN can be broadly classified into two approaches: Rate-coded SNN [18] [19] [20] and temporal-encoded SNN [25] [26] [27] [28]. The fundamental problem of rate-coded SNN is that it results in tremendous spike signaling cost and latency. Temporal-encoded SNN in the AER domain involves hand-crafted spatiotemporal feature extraction followed by a spiking recognition layer. The feature extraction phase of these networks makes their implementation challenging. Recently a learnable temporal-coded feature extraction has been proposed in [29]. However, the major limitation of this method comes in the event encoding process, which incurs a memory cost of $O(N^2)$, thus making its implementation non-trivial.

Motivated by the fact that neurons in the visual cortex respond in asynchronous mode with minimal energy cost, we asked ourselves whether a similar hypothesis could be achieved in AER recognition systems. Towards this, we introduce a novel single-spike object recognition system which we refer to here as EventF2S (Event Camera First-To-Spike Spiking Network).

EventF2S consists of two main components: 1) Sparse Spiking Temporal Encoding (SS-TE) layer: This layer provides a neuromorphic implementable algorithm to encode raw events into as few informative early spikes as possible. The SS-TE encoding scheme delivers a drastically reduced number of spikes to computationally heavy downstream applications. 2) Spiking Recognition Network: It consists of a group of differentiable spiking neurons that adapts the First-To-Spike protocol for inter-neuron communication. The learnable spiking neural dynamics enable us to backpropagate the error deep down the network and achieve competitive accuracy with single-spike input encoding. By incorporating the First-To-Spike communication protocol, we aim to achieve further reduction in spike signaling.

To summarize, the significant contributions of EventF2S are as follows: i) **Asynchronous processing:** Asynchronous processing allows real-time responsiveness, low-latency and resource efficiency; ii) **Sparse spiking:** By curtailing the number of spikes, EventF2S offers tremendous energy saving; iii) **Neuromorphic Friendly:** This enables EventF2S to inherit brain's efficiency and computational mechanisms.

## II. Related Work

As the proposed framework is a temporal spiking neural network, we provide a survey specifically in the domain of temporal coding SNN, [25] [26] [27] [28] [32] [33] [34] [35] [36] [37]. HMAX [25] is one of the most popular hierarchical models structured as simple (linear) and complex (non-linear) layers S1, C1, S2, and C2 with SVM as a classifier. Algorithm of [26] extracted position and size invariant features for human

posture recognition. Zhao. et al. [27] proposed an event-driven neuron model in the S1 layer with a forgetting mechanism and replacing max-pooling with MAX operation at the C1 layer. The generated features converted to single-spike encoding have been classified using a Tempotron-based classification layer. Orchard. et al. [28] came up with an HMAX-based hierarchical spiking architecture. However, the learning layers used statistical methods, thus losing the spiking nature.

Liu. et al. proposed [32] proposed MuST (Multi-Scale spatiotemporal) feature representation based on Leaky Integrate and Fire (LIF) neurons followed by an unsupervised recognition approach. The architecture of [33] consists of two modules: Hand-crafted multi-spike encoding followed by tempotron-based SNN learning. Though this method was fully temporally encoded, multiple spikes were required to achieve good accuracy as learning is implemented only in the classification stage.

[34] includes spatiotemporal feature extraction and a novel learning algorithm. The feature extraction phase is similar to that of [33]. As part of spike learning, they have introduced a novel learning algorithm known as the Segmented Probability Maximization (SPA) algorithm, which classifies based on the firing rate of neurons in the classification layer. Though [34] was able to achieve better accuracy than [33], the rate coding of the decision layer discredits its use in the temporal domain. Recently, [29] ventured into a temporal-encoded spiking object recognition system. The spike encoding involves the accumulation of event stream into an image frame, $D(x, y)$, over a period of time from $t_0$ to $t_k$, which incurs a memory cost of $O(N^2)$. This will incur additional memory costs, thus rendering it not suitable for neuromorphic implementation.

## III. Sparse Temporal-Encoded EventF2S Architecture

The proposed spiking architecture is structured as follows: Single Spike Temporal Encoding (SS-TE) layer and a learnable First-To-Spike encoded object recognition spiking network.

### A. Single Spike Temporal Encoding (SS-TE) Layer

SS-TE layer densely covers the pixels of the AER sensor, performing signal-selective single spike per pixel encoding. The new sensing modality of the AER sensor has made it more susceptible to noise. The noisy events of the AER sensor broadly fall under two categories: Type I noise, which is uncorrelated with spatiotemporal neighbor events, and Type II noise, which exhibits self-temporal correlation. To address this issue, we implement event encoding as a collection of parallel asynchronous continuous-time filters whose spatial sensitiveness is defined to respond best to spatiotemporally correlated events while penalizing the noise. The degree of penalization is inversely proportional to the frequency.

*1) SS-TE Encoding Function:* The function of SS-TE encoding can be derived as furnished here. Events are modeled as a summation of impulse signals in continuous time (Eq. 1),

$$E(t) = \sum_{k=1}^{\inf} p_k \delta(t - t_k) \delta_{x_k}(x) \tag{1}$$

Where $\delta(t)$ is the Dirac delta function, $p = \pm 1$ is the polarity depending on the scene brightness change, $x = (u, v)$ is the discrete pixel address, $t_k$ is the timestamp of occurrence of the $k^{th}$ event, relative to the global clock and $\delta_{x_k}(x)$ is the Kronecker delta function which is $= 1$ at $x = x_k$.

Considering the events $E_{x_i^-,\Delta\tau}(t)$ that occurred on a local neighborhood $x_i^- = \{(u, v) : |u - u_i| < \beta, |v - v_i| < \beta, x_i \nexists x$ centered at $i^{th}$ pixel and activated at $[t_i, t_i - \Delta\tau]$, it can be represented as a monodimensional signal in continuous time as follows,

$$E_{x_i^-,\Delta\tau_i}(t) = \int_{t_i-\Delta\tau}^{t_i} \delta(t - \tau) \delta_{x_i^-}(x) d\tau \tag{2}$$

The events $E_{x_i,\Delta\tau}(t)$ that occured at the $i^{th}$ pixel in the time duration $[t_i, t_i - \Delta\tau]$ can be expressed as

$$E_{x_i,\Delta\tau}(t) = \int_{t_i-\Delta\tau}^{t_i} \delta(t - \tau) \delta_{x_i}(x) d\tau \tag{3}$$

Define the following continuous time single dimensional low pass filter $h(t) = e^{-\frac{t}{\tau_c}}$, with cut-off frequency determined by $\tau_c$. In order to remove the Type I and II noise, we need to maximize $h(t) \otimes E_{x_i^-,\Delta\tau}(t)$ and minimize $h(t) \otimes E_{x_i,\Delta\tau}(t)$ respectively. Hence, we need to estimate events such that the following objective function is maximized,

$$h(t) \otimes E_{x_i^-,\Delta\tau}(t) - h(t) \otimes E_{x_i,\Delta\tau}(t) \tag{4}$$

The given objective function can be realized with a set of parallel Leaky Integrate and Fire (LIF) filters whose output is modeled as a convolution between asynchronously sampled continuous time filter $h(t)$ and the events $\left[E_{x_i^-,\Delta\tau}(t), E_{x_i,\Delta\tau}(t)\right]$ weighted with $[+1, -1]$. After firing, the particular neuron shunts all its incoming spikes till a classification decision is reached.

### B. Learning by First-To-Spike Spiking Network

In the learning stage, a single spike per pixel emitted by the LIF neurons of the SS-TE layer is transmitted to the convolution-based feature extraction layers arranged in a feed-forward fashion. The learned multi-channel feature spikes are fed to the classification layer, which consists of $C$ (number of classes) neurons. Once a neuron spikes, temporal inhibition is performed, which will force the active neuron to enter into a dormant stage. As a result, this promotes sparsity in spike representation.

*1) Neuron Dynamics:* This section describes the neural and synaptic dynamics of the neurons. Inspired by [39], we define our recognition network with a set of non-leaky IF neurons with synaptic kernel function $\Theta(t - t_i)e^{-\frac{(t-t_i)}{\tau_m}}$, where $\Theta(t - t_i) = 1$ if $t \geq t_i$ and 0 otherwise. As given in [39], the relation

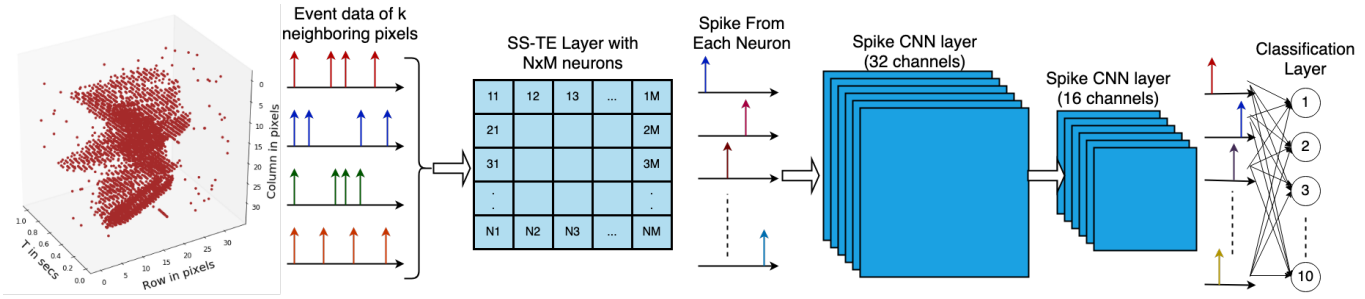


Fig. 1. Architecture of the proposed EventF2S. Events are fed to SS-TE layer which is composed of a grid of $N \times N$ neurons. The neurons will accumulate spikes and emit a single spike when it reaches a threshold. These spikes propagate asynchronously to their corresponding neurons in the convolution layer of recognition network. The recognition network is composed two convolution layers, each with a group of non-leaky IF neurons.

between input and output spike time when the neuron spikes are given as follows,

$$e^{t_{out}} = \frac{\sum_{i \in C} w_i e^{t_i}}{\sum_{i \in C} w_i - 1} \tag{5}$$

Where $C$ is the number of causal spikes which causes the neuron to spike and $w_i$ is the weight of the $i^{th}$ spike. The relation between the neuron's first spike time and input spike time given by Eq. 5 is differentiable with respect to the weights, hence making the network trainable. The network is trained such that the output neuron belonging to the correct class fires early. To achieve the same, the following differentiable cross-entropy loss function is defined, which is minimized using the standard Gradient Descent technique.

$$-\sum_{k=1}^{C} ln \frac{e^{-z_k}}{\sum_i e^{-z_i}} \tag{6}$$

Where $z_k$ is the output of $k_{th}$ output neuron.

### C. Merits of EventF2S

in this section, we justify the key benefits of the proposed algorithm, which are asynchronous processing and neuromorphic compatibility. Sparse spiking has been validated with experiments in the following section.

*1) Neuromorphic Compatibility:* The architecture of proposed SS-TE layer can be realized with set of $N \times N$ Leaky integrate and Fire (LIF) neurons organized in a matrix format of size $N \times N$, where N represents the number of pixels. The SS-TE layer at pixel $i$ operates on spikes occuring at $i$ and its $K$ neighboring pixels. This functionality can be realized by connecting the Time To First Spike (TTFS) encoded event output of neuromorphic camera to these $N \times N$ LIF neurons in convolutional arrangement. The synapse weights are designed such that self-pixel connections have a weight of +1, while connections to neighboring pixels have a weight of -1. Each LIF neuron integrates the incoming spikes into its membrane potential and generates an output spike once its membrane potential exceeds its pre-defined threshold. The neuronal dynamics of LIF spiking neurons helps in memorizing the causal spatio temporal spike occurences without incurring memory complexity.

The recognition network is composed up of non-leaky Integrate and Fire neurons. These neurons are easily realizable in neuromorphic hardware as described in [41].

*2) Asynchronous Processing:* As an illustrative example, lets assume that first event arrives at time $t_1$, followed by a series of events. The neurons of SS-TE layer will start functioning immediately upon the the arrival of first spike at each pixel. When a neuron's membrane potential reaches it threshold, it fires a spike and this spike strats its journey towards recognition network. Note that this process occurs independently at each neuron without waiting for all neurons in the SS-TE layer to finish firing.

A similar kind of operation is observed in each layer of the spiking recognition network. Once a spiking neuron at the classification layer receives enough spikes at time $t_k$ to trigger its threshold, it fires. Importantly, this neuron does not wait for all its pre-synaptic neurons to fire, as is typically the case in synchronous Artificial Neural Network (ANN) processing.

## IV. Experiments and Results

In this section, we conduct two types of experiments: 1) To showcase the capability of the SS-TE encoding layer and 2) To bring out the recognition capability of the overall EventF2S framework. The neuromorphic datasets considered are [22], [21], N-MNIST [30] and MNIST-DVS [40].

### A. SS-TE Experiments

This section highlights the benefits reaped by prefixing the recognition network with the SS-TE encoding layer.

| **SS-TE** | | **Density** [23] | | **BAF** [24] | |
|---|---|---|---|---|---|
| TP | FP | TP | FP | TP | FP |
| **0.98** | **0.37** | 0.99 | 0.99 | 0.99 | 0.99 |
| **0.95** | **0.32** | 0.89 | 0.85 | 0.99 | 0.99 |
| 0.74 | 0.23 | 0.74 | 0.71 | 0.99 | 0.99 |
| 0.39 | 0.14 | 0.60 | 0.55 | 0.99 | 0.99 |
| 0.25 | 0.09 | 0.53 | 0.43 | 0.99 | 0.99 |

TABLE I

Analysis of denoising methods in terms of True Positive (TP) and False Positive (FP) at 0.9 SNR. SS-TE encoding passes most of the signals (high TP) while eliminating all noisy events (low FP)

*1) Denoising Experiments:* Though the primary focus of our research is the object recognition spiking task, we include the performance of our SS-TE encoding layer on the denoising task as a means to validate that the proposed solution retains only the informative single spike for each pixel. Denoising was evaluated on a publicly available noise dataset, recorded by a Davis346 camera under bright and dark conditions [21]. To evaluate the denoising performance, the noise events were mixed with pure event signals. A conventional camera video from [21] was converted to pure events using the v2e algorithm [22].

Table. I shows the comparison of the proposed encoding with state-of-the-art event denoising methods [23] [24] in terms of True Positive (TP) and False Positive (FP) rates at different SNR. The results demonstrate the effectiveness of SS-TE in achieving a high True Positive (TP) rate for a low False Positive (FP) rate.

*2) Sparse Temporal Spiking:* The spikes generated and transmitted in a spiking network should be as low as possible to reap the full potential of the AER sensor. The average number of spikes per pixel ($N_s$) generated by the AER sensor, calculated across test samples, amounts to $49$ and $50$ for NMNIST and MNIST-DVS, respectively. SS-TE encoding layer reduces the spike count to just one per pixel as against $49$ and $50$.

### *B. Recognition Experiments*

For evaluation of our primary object recognition task, the network topology used is as follows: Two convolution layers of $32$ and $16$ channels, each with a kernel size of $5 \times 5$ and stride of $2$, followed by a classifier layer of $10$ neurons. $10$ neurons represent the number of classes. During the training process, the learning rate was set to $1e^{-2}$ for the initial $50$ epochs, and later it was reduced to $1e^{-3}$ for the remaining $50$ epochs.

*1) comparison in terms of Recognition Accuracy:* We have compared our approach with [26], [27], [28], [30] [31] [32] [35] [38]. We have restricted our comparison to non-rate-coded spiking networks as the rate-coded networks provide good accuracy at par with ANN at the cost of huge energy consumptions as compared to temporal-coded networks. Results from the comparison are summarized in Table. II in terms of accuracy. It could be observed that EventF2S achieves good accuracy at par or above the state-of-the-art AER recognition networks.

The performance of EventF2S in terms of accuracy is below [29]. While accuracy is acknowledged as an important metric, the specific goal of the proposed EventF2S is to provide a neuromorphic-friendly solution. This allows us to leverage the benefits of neuromorphic hardware, such as asynchronous processing and low cost. The encoding process mentioned in [29] incurs additional memory cost, which is a major drawback for hardware implementation. Additionally, the processing primitives of the encoding layer of [29] do not align with the device physics of neuromorphic hardware, further limiting the application of [29]. Thus, EventF2S is a promising asynchronous approach for low-cost edge applications.

| **Method** | **NM** | **MD** | **Method** | **NM** | **MD** |
|---|---|---|---|---|---|
| [26] | - | 0.63 | [27] | 0.85 | 0.768 |
| [28] | 0.711 | 0.55 | [30] | 0.83 | - |
| [31] | 0.834 | - | [32] | 0.89 | 0.79 |
| [33](ss) | 0.79 | - | [33](ms) | 0.93 | 0.77 |
| [35] | 0.704 | 0.746 | [38] | 0.80 | 0.80 (full) |
| EventF2S | **0.94** | **0.84** | EventF2S$^{2s}$ | **0.97** | - |

TABLE II

COMPARISON OF RECOGNITION ACCURACY WITH STATE-OF-THE-ART METHODS ON NMNIST (NM) AND MNIST-DVS (MD) DATASETS. ACCURACY OF OTHER METHODS ARE AS REPORTED IN [32] [35] [34] [33] (MD: 100MS). THE PROPOSED METHOD GIVES BETTER ACCURACY THAN OTHER NON-RATE-CODED AER SPIKING NETWORKS.

| **Methods** | **NM** | **MD** | **Methods** | **NM** | **MD** |
|---|---|---|---|---|---|
| [27], [34] | 2.75 | 17.92 | [28] | 2.47 | 16.06 |
| [33] | $N_a \times 2.47$ | $N_a \times 16.06$ | Ours | **0.46** | **6.55** |

TABLE III

COMPUTATION COMPLEXITY ($\times 10^6$) AT FIRST LAYER: $N_a$ IS THE NUMBER OF EVENTS INPUT TO THE FEATURE EXTRACTION LAYER OF [33]. NOTE THE HUGE REDUCTION IN COMPUTATION IN EVENTF2S COMPARED TO OTHER METHODS.

*2) Computation Efficiency via Sparse Spiking:* In this section, we perform experiments to analyze the proposed EventtF2S in terms of computation cost in comparison with [27], [28], [32] [33] and [34]. Initial feature extraction layers contribute to the bulk of the computational complexity as they operate on larger spatial resolution. Hence, we evaluated the computation cost as the number of weight multiplications that takes place in the first layer of the feature extraction stage.

To estimate computation cost, we multiply the number of events input into the first layer ($N_e$) with the number of computations incurred per event ($N_c$). Each event was assumed to be surrounded densely by neighboring events. The values of $N_e$ for the raw event and SS-TE encoding are $(4203, 1156)$ and $(27320, 16384)$ for NMNIST (NM) and MNIST-DVS (MD), respectively. $N_e$ is calculated as the mean of the number of input events that enter the feature extractor across the test samples.

The results are provided in Table. III makes it evident that the computation incurred at the first layer of EventF2S is far lesser as compared to other methods. State-of-the-art methods perform feature extraction on every input event, thus increasing the computational complexity. By making the feature extraction learnable, we were able to reduce the number of computations drastically with single-spike encoding.

## V. CONCLUSION

In this paper, we proposed an asynchronous, sparse-spiking, neuromorphic compatible AER sensor object recognition framework known as EventF2S. The pipeline of EventF2S comprises an event encoding layer followed by an AER recognition network. SS-TE was specifically derived to convert raw and noisy AER sensor data into very few informative early spikes. Experiments conducted demonstrated the denoising capability of the SS-TE layer and how it contributed to sparse

spiking and resultant reduced computation. Sparse spiking can lead to information loss, which has been compensated by utilizing learnable feature extraction spiking neurons. Experiments on neuromorphic datasets revealed the above-mentioned fact by exhibiting competitive recognition accuracy at par with state-of-the-art temporal encoded SNNs. Note being neuromorphic hardware friendly, EventF2S exploits the unique advantages of AER sensors.

## REFERENCES


[1] M. Mahowald, "VLSI analogs of neuronal visual processing: A synthesis of form and function," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, May 1992

[2] G. Gallego, T. Delbruck, G. M. Orchard, et al., "Event-based vision: A survey," IEEE Trans. Pattern Anal. Mach. Intell., pp. 1–1, 2020

[3] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128x128 120dB 15us latency asynchronous temporal contrast vision sensor," IEEE J. Solid-State Circuits, vol. 43, no. 2, pp. 566–576, Feb. 2008

[4] P. Lichtsteiner, C. Posch, and T. Delbruck, A 128×128 120 dB 15$\mu$s latency asynchronous temporal contrast vision sensor, IEEE J. Solid-State Circuits, vol. 43, no. 2, pp. 566–576, 2008

[5] C. Posch, D. Matolin, and R. Wohlgenannt, A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS, IEEE J. Solid-State Circuits, vol. 46, no. 1, pp. 259–275, Jan. 2011

[6] P. Lichtsteiner and T. Delbruck, "64x64 event-driven logarithmic temporal derivative silicon retina," in IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, 2005, pp. 157–160.

[7] P. Lichtsteiner, "An AER temporal contrast vision sensor," Ph. D. Thesis, ETH Zurich, Dept. of Physics (D-PHYS), Zurich, Switzerland, 2006.

[8] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, "Interacting maps for fast visual interpretation," in Int. Joint Conf. Neural Netw. (IJCNN), 2011, pp. 770–776

[9] H. Rebecq, T. Horstschafer, G. Gallego, and D. Scaramuzza, EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real-time, IEEE Robot. Autom. Lett., vol. 2, no. 2, pp. 593–600, 2017.

[10] M. Liu and T. Delbruck, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors," in British Mach. Vis. Conf. (BMVC), 2018.

[11] J. Kogler, C. Sulzbachner, and W. Kubinger, "Bio-inspired stereo vision system with silicon retina imagers," in Int. Conf. Comput. Vis. Syst. (ICVS), 2009, pp. 174–183.

[12] A. I. Maqueda, A. Loquercio, G. Gallego, N. Garcıa, and D. Scaramuzza, "Event-based vision meets deep learning on steering prediction for self-driving cars," in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2018

[13] Chankyu Lee, Adarsh Kumar Kosta, Alex Zihao Zhu, Kenneth Chaney, Kostas Daniilidis, and Kaushik Roy. Spike-flownet: event-based optical flow estimation with energy-efficient hybrid neural networks. In European Conference on Computer Vision, pp. 366–382. Springer, 2020

[14] Nitin Rathi, Gopalakrishnan Srinivasan, Priyadarshini Panda, and Kaushik Roy. Enabling deep spiking neural networks with hybrid conversion and spike timing dependent backpropagation. arXiv preprint arXiv:2005.01807, 2020

[15] Guillaume Bellec, Darjan Salaj, Anand Subramoney, Robert Legenstein, and Wolfgang Maass. Long short-term memory and learning-to-learn in networks of spiking neurons. Advances in neural information processing systems, 31, 2018

[16] Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. In- corporating learnable membrane time constant to enhance learning of spiking neural networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2661–2671, 2021

[17] Yujie Wu, Lei Deng, Guoqi Li, Jun Zhu, Yuan Xie, and Luping Shi. Direct training for spiking neural networks: Faster, larger, better. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pp. 1311–1318, 2019

[18] Ali Samadzadeh, Fatemeh Sadat Tabatabaei Far, Ali Javadi, Ahmad Nickabadi, and Morteza Haghir Chehreghani. Convolutional spiking neural networks for spatio-temporal feature extraction. arXiv preprint arXiv:2003.12346, 2020

[19] Guo, Shasha, and Tobi Delbruck. Low cost and latency event camera background activity denoising. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[20] Y. Hu, S.-C. Liu, and T. Delbruck, "v2e: From video frames to realistic DVS events," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2021

[21] Feng Y and et al. Event density based denoising method for dynamic vision sensor. Applied Sciences,

[22] Liu H and et al. Design of a spatiotemporal correlation filter for event-based sensor. IEEE International Symposium on Circuits and Systems, pages 722–725, 2015

[23] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 3, pp. 411–426, 2007

[24] S. Chen, P. Akselrod, B. Zhao, J. A. P. Carrasco, B. Linares-Barranco, and E. Culurciello, "Efficient feedforward categorization of objects and human postures with address-event image sensors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 2, pp. 302–314, 2012.

[25] B. Zhao, R. Ding, S. Chen, B. Linares-Barranco, and H. Tang, "Feedforward categorization on aer motion events using cortex-like features in a spiking neural network," IEEE transactions on neural networks and learning systems, vol. 26, no. 9, pp. 1963–1978, 2014

[26] Garrick Orchard, Cedric Meyer, Ralph Etienne-Cummings, Christoph Posch, Nitish Thakor, Ryad Benosman, "Hfirst: A temporal approach to object recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.

[27] Shibo Zhou, Wei Wang, Xiaohua Li and Zhanpeng Jin, "A spike learning system for event driven object recognition", 2021

[28] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting static image datasets to spiking neuromorphic datasets using saccades," Frontiers in Neuroscience, vol. 9, p. 437, 2015.

[29] Cohen, G. K, Orchard, G, Leng, S. H, Tapson, J, Benosman, R. B and Van Schaik, A. 2016. Skimming digits: neuromorphic classification of spike-encoded images. Fron- tiers in neuroscience 10:184.

[30] Liu, Q, Pan, G, Ruan, H, Xing, D, Xu, Q, and Tang, H, Unsupervised aer object recognition based on multiscale spatio-temporal features and spiking neurons. arXiv preprint arXiv:1911.08261, 2019

[31] R. Xiao, H. Tang, Y. Ma, R. Yan, and G. Orchard, An event-driven categorization model for aer image sensors using multispike encoding and learning, IEEE Transactions on Neural Networks and Learning Systems, 2020

[32] Qianhui Liu, Haibo Ruan, Dong Xing, Huajin Tang, Gang Pan, Effective AER Object Classification Using Segmented Probability-Maximization Learning in Spiking Neural Networks, AAAI, 2020

[33] X. Peng, B. Zhao, R. Yan, H. Tang, and Z. Yi, "Bag of events: An efficient probability-based feature extraction method for aer image sensors," IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 4, pp. 791–803, 2017

[34] Lagorce, Xavier, et al. "Hots: a hierarchy of event-based time-surfaces for pattern recognition." IEEE transactions on pattern analysis and machine intelligence 39.7 (2016): 1346-1359.

[35] Sironi, Amos, et al. "HATS: Histograms of averaged time surfaces for robust event-based object classification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[36] Lagorce, X.; Orchard, G.; Galluppi, F.; Shi, B. E.; and Benosman, R. B. 2017. Hots: a hierarchy of event-based time-surfaces for pattern recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(7):1346– 1359.
Systems.

[37] Hesham Mostafa, Supervised learning based on temporal coding in spiking neural networks, IEEE Transactions on Neural Network and Learning Systems, 2017.

[38] Serrano-Gotarredona, Teresa, and Bernabé Linares Barranco. "Poker-DVS and MNIST-DVS. Their history, how they were made, and other details." Frontiers in neuroscience 9, 481, 2015.

[39] Zhou, Shibo, et al. "Temporal-coded deep spiking neural network with easy training and robust performance." Proceedings of the AAAI conference on artificial intelligence. Vol. 35. No. 12. 2021.