Brain-Inspired Spatiotemporal Processing Algorithms for Efficient Event-Based Perception

Biswadeep Chakraborty*
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia
biswadeep@gatech.edu

Uday Kamal*
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia
uday.kamal@gatech.edu

Xueyuan She
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia
xshe6@gatech.edu

Saurabh Dash
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia
saurabhdash@gatech.edu

Saibal Mukhopadhyay
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia
saibal.mukhopadhyay@ece.gatech.edu

***Abstract*—Neuromorphic event-based cameras can unlock the true potential of bio-plausible sensing systems that mimic our human perception. However, efficient spatiotemporal processing algorithms must enable their low-power, low-latency, real-world application. In this talk, we highlight our recent efforts in this direction. Specifically, we talk about how brain-inspired algorithms such as spiking neural networks (SNNs) can approximate spatiotemporal sequences efficiently without requiring complex recurrent structures. Next, we discuss their event-driven formulation for training and inference that can achieve real-time throughput on existing commercial hardware. We also show how a brain-inspired recurrent SNN can be modeled to perform on event-camera data. Finally, we will talk about the potential application of associative memory structures to efficiently build representation for event-based perception.**

***Index Terms*—event-based camera, brain-inspired, efficient, spiking neural network**

I. Introduction

Event-based cameras can strongly benefit the future of real-time machine vision applications such as robotics, and autonomous driving [1], [2] owing to their ultra-low power, high dynamic range, high temporal resolution, and low latency. The current state-of-the-art convolutional and recurrent neural network-based methods, originally developed for frame-based cameras, have demonstrated good perception accuracy on event cameras [3]. However, these methods rely on temporal aggregation of the events to create a frame-like dense representation as input, thereby discarding the inherent sparsity of event data and resulting in high computational costs. Recent works have explored event-based processing methods for object recognition to exploit data sparsity. These methods adopt event-based processing to lower computational costs but do not achieve similar performance compared to the dense representation-based methods. This necessitates computationally efficient algorithms that exploit sparsity and achieve high accuracy. To solve these problems, we discuss the feasibility of using biologically plausible brain-inspired SNNs, which are highly energy efficient compared to standard DNNs and take the input as discrete spikes, making them ideal for event-based data processing. Instead of capturing static frames and transmitting them discretized in time, spiking-based models use the continuous spike streams produced by the event sensors at each pixel location.

This work briefly discusses the recent developments in brain-inspired processing methods for event-based perception. Specifically, we discuss the learning capabilities of a heterogeneous feedforward [4] and heterogeneous recurrent SNN models [5] and their superior performance on event-based datasets. We also discuss the improvements in model robustness, performance, and efficiency due to the heterogeneity in the hyperparameters. To meet the high-speed processing requirement of the event-camera data, we discuss an event-driven SNN learning and inference framework [6] that enables near-real-time processing performance on commercial GPU hardware. Furthermore, we discuss the recent developments in brain-inspired event-camera data processing methods alternative to the SNNs. Specifically, we explore a novel memory-augmented representation learning framework for asynchronous and efficient event-based perception–EventFormer [7], which learns to store, retrieve and update its memory representation in the latent form of higher-order spatiotemporal dynamics of the events. The rest of the paper is structured as follows: Section II discusses the existing works on SNN-based event-data processing. Section III discusses the potential application of heterogeneous neurons in diverse SNN architectures, followed by their near-real-time processing in section IV. Finally, we discuss the brain-inspired associative memory-based event-processing in section V and conclude the study with a brief discussion in section VI.

* equal contribution

978-3-9819263-7-8/DATE23/© 2023 EDAA

II. Spiking Neural Networks for Event-Based Perception

The use of Neuromorphic hardware offers energy-efficient and low-latency processing for event-based signals. Event cameras capture rapid brightness changes, and neuromorphic processors consume much less energy and have lower latency than traditional von Neumann architectures. To fully utilize the potential of these systems, we need to adopt an event-based processing approach where events are directly passed between the event-based sensor and the neuromorphic processor running an event-based algorithm without any intermediate processing or accumulation. With the increasing importance of low-power devices in a real-world environment, such as medical robots [8], self-driving cars [9], and drones [10], more energy-efficient neural networks are required. However, conventional Artificial Neural Networks (ANNs) incur a substantial computational cost. Whereas SNNs can emulate the functionality of a biological neuron by processing visual information with binary events, resulting in highly low-power implementation on neuromorphic hardware [11], [12]. Therefore, learning in SNNs is an emerging research topic since the asynchronous and binary event has different characteristics than the conventional float activation value of ANNs. An event-based camera is a suitable visual input sensor for SNNs to build efficient neuromorphic systems for real-world applications. The individual pixel of the event camera generates asynchronous events following the change of luminance. This bio-plausible visual sensor has the advantage of high processing speed, low energy consumption, and less blurring effect than a conventional frame-based camera [13]. However, training deep SNNs with spike events from an event camera remains challenging. The main reason is that Leaky-Integrate-and-Fire (LIF) neurons in SNNs diminish spike activation in deeper layers. The sparsity in event data limits the depth of SNNs. Most of the previous training algorithms for SNNs focus on frame-based static images [14], [15] where they convert a static image into a binary spike train across multiple time steps. Therefore SNNs learn temporal dynamics from the given spike data. Hence, She et al. [16] focus on improving the spike-based backpropagation method in which SNNs are trained from asynchronous event data directly. SNNs can operate directly on the event data instead of aggregating them, and recent works use the concept of time surfaces [17]. Escobar et al. [18] proposed a feed-forward SNN for action recognition using the mean firing rate of every neuron and synchrony between neuronal firing. Other recent works [19], [20] have used shallow SNNs to learn human action recognition using a gradient descent-based learning mechanism by encoding the trajectories of the joints as spike trains. Recent research learned video activities with limited examples using this idea of reservoir computing [21], [5]. Again, recurrent networks of spiking neurons can be trained to achieve competitive performance compared to DNN-based recurrent neural networks. Demin et al. [22] showed that using recurrence could reduce the number of layers in SNN models and potentially form various functional network structures. Zhang et al. [23] proposed a spike-train level recurrent SNN backpropagation method to train the deep RSNNs, which achieves excellent performance in image and speech classification tasks. On the other hand, Wang et al. [24] used the recurrent LIF neuron model with the dynamic pre-synaptic currents and trained by the BP based on surrogate gradient. However, such supervised learning algorithms require a massive amount of labeled training data for good performance. Unsupervised learning methods, such as STDP, have shown great generalization and trainability properties [15]. Previous works have used STDP for training the recurrent spiking networks [25]. Several other works have used a reservoir-based computing strategy, as described above. Liquid State Machines [26], equipped with unsupervised learning models like STDP [27], and BCM [28] have also shown promising results.

III. Heterogeneity in Spiking Neural Networks

Inspired by the biological observations, recent empirical studies showed potential for improving SNN performance with heterogeneous neuron dynamics [29], [4]. Current literature primarily looks into how heterogeneity in neuronal timescales improves the model performance. They do not study how heterogeneity can be leveraged to design sparse neural networks. All these models use a uniform parameter distribution for spiking neuron parameters and their learning dynamics. There has been little research leveraging heterogeneity in the model parameters and their effect on performance and generalization. Recently, Perez-Nieves et al. [30] introduced heterogeneity in the neuron time constants and showed this improves the model's performance in the classification task and makes the model robust to hyperparameter tuning. She et al. [31] also used a similar heterogeneity in the model parameters of a feedforward spiking network and showed it could classify temporal sequences. Again, modeling heterogeneity in the brain cortical networks, Zeldenrust et al. derived a class of RSNNs that tracks a continuously varying input online [32]. Chakraborty et al. [33] showed that heterogeneity among the neuronal dynamics improves the memory capacity of the model while heterogeneity in the synaptic dynamics reduces the spiking activation of neurons and maintains memory capacity. Hence, a heterogeneous model has a lesser firing rate than its homogeneous counterpart.

A. *Feedforward Network with Heterogeneous Neurons*

A dynamical system of spiking neurons with only feedforward connections can classify spatiotemporal patterns without recurrent connections. The theoretical construct of a feedforward SNN for approximating a temporal sequence is not well explored, making optimizing SNN architectures for learning complex spatiotemporal patterns challenging. She et al. [4] presented a theoretical framework for analyzing and improving the ability of feedforward SNNs to approximate complex spatiotemporal patterns. The authors demonstrated that a feedforward SNN with one neuron per layer and skip-layer connections could approximate the mapping function

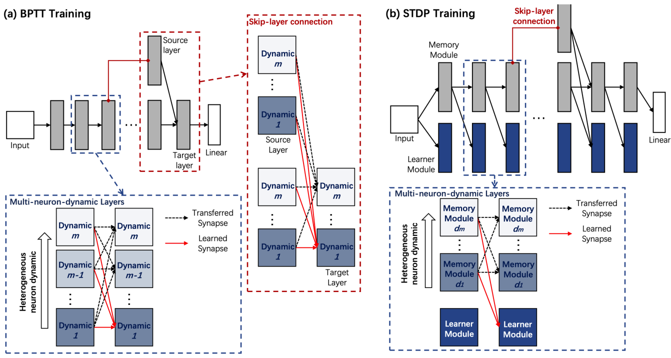

Fig. 1: (a) The network with BPTT training, each multi-neuron-dynamic layer contains a set of neuron dynamics from d1 to dm. (b) The network with STDP training. Figure sourced from [31] with the author's permission.

between input and output spike trains on a compact domain. Heterogeneous neurons with different dynamics and skip-layer connections have been found to increase the number of memory pathways in a feedforward SNN and, thus, improve the SNN's ability to represent arbitrary sequences. The authors proposed a dual-search-space Bayesian optimization process to optimize the architecture and parameters of the SNN, considering heterogeneity and skip-layer connections, to improve the learning and classification of spatiotemporal patterns. The experimental studies on feedforward SNN for spatiotemporal classifications showed that the basic design principles for improving sequence approximation could be adopted to optimize SNN architectures and improve performance for spatiotemporal classification tasks. The complete flowchart of the method is illustrated in Fig. 1.

B. RSNNs with Heterogeneous Neurons and Synapses

Although recent SNNs trained with supervised backpropagation show classification accuracy comparable to deep networks, the performance of unsupervised learning-based SNNs still needs to improve. Chakraborty et al. [5] developed a heterogeneous recurrent SNN (HRSNN) for spatiotemporal classification of video activity recognition tasks on event-based datasets (DVS128 Gesture). HRSNN has heterogeneity in both the LIF neuron parameters and STDP dynamics, which improves the performance and enables the development of smaller models with sparse connections and less training data. The authors adopted unsupervised STDP learning to train the network, which resulted in a similar performance

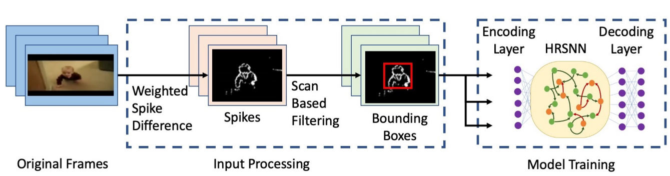

Fig. 2: Flowchart for the input processing and model training. Figure sourced from [5] with the author's permission

	Models	Accuracy
DNN	RG-CNN [34]	97.2
	PointNet [35]	95.3
	I3D [36]	96.5
Homogeneous SNN (Supervised)	Liu et al [37]	92.7
Homogeneous SNN (Supervised)	ConvLSNN [38]	97.1
	DECOLLE [39]	97.5
	BPTT-Homogeneous Recurrent SNN [4]	97.1
Heterogeneous SNN (Supervised)	Perez et al. [30]	82.9
	BPTT- Feedforward Heterogeneous SNN [4]	98.0
	BPTT-Recurrent Heterogeneous SNN [5]	98.1
Homogeneous SNN (Unsupervised)	GRN-BCM [40]	77.2
	CMA-ES [41]	89.3
Heterogeneous SNN (Unsupervised)	STDP-Feedforward Heterogeneous SNN [4]	96.6
	Recurrent Homogeneous SNN [5]	90.3
	Recurrent Heterogeneous SNN [5]	96.5

TABLE I: Table shows the performance and model complexity for DNN and Supervised and Unsupervised SNN models. Table sourced from [4], [5] with the authors' permission.

to state-of-the-art supervised SNNs but with fewer neurons and connections and less training data. The study contributes to the understanding of why heterogeneity in both neuronal and synaptic parameters can lead to improved performance and smaller models with less training data. It is noteworthy that heterogeneity is a property of the human brain, thereby making it biologically plausible.

C. Results

Incorporating heterogeneity is beneficial for designing high-performance feedforward SNNs (with heterogeneity in only neuronal parameters) and recurrent SNNs (with heterogeneity in both neuronal and synaptic parameters) for classifying complex spatiotemporal datasets for action recognition tasks. The input processing and model training flowchart is shown in Fig. 2. The performance of these models on event-based datasets (DVS-Gesture 128) is shown in Tab. I. The Table shows that the heterogeneous HRSNN model has a much lesser average neuronal activation than the homogeneous RSNN and the other unsupervised SNN models. Thus, we conclude that while the model with heterogeneity in only neuronal dynamics improves the model performance, heterogeneity in both neuronal and synaptic dynamics induces sparse activation and sparse coding of information.

IV. Near-Real-Time Inference with Heterogeneous Neurons

Our discussion so far primarily focused on the recent developments in sequence modeling capability, superior unsupervised and spike-efficient learning capacity of the SNNs with heterogeneous neuronal dynamics [5], [31], [33]. One major challenge to their practical application is the high-speed processing requirement of the SNNs to match the high temporal resolution of the event cameras. On modern GPU hardware, existing specialized software tools such as ParallelSpikeSim [16] and BindsNet [42] have 10^3 frames per second processing capability, whereas typical event-camera can generate 10^5 events per second [43]. These existing frameworks leverage discrete-time formulation of the SNN processing, where a global update to all the neurons of the network occurs at each time step. Such formulation results in lower throughput for highly sparse event-camera data since the processing speed is constrained by the time-step resolution, not the event rate. SPEED is an alternate and effective solution to this challenge, event-driven learning, and inference framework for event-camera data [6]. Unlike discrete-time formulation, which suffers from inefficiency due to the lower neuronal activation inside the network, the event-driven neuron simulation updates only those that receive an input signal. This is achieved by recording the exact timestamp of the past inactive, update, and spike occurrence and the corresponding membrane potential of each neuron. Only one neuron in layer $n-1$ spikes, and the corresponding three post-synaptic neurons in layer n are updated. One of these three neurons' potential exceeds the threshold voltage and initiates the update of the next three neurons in layer $n+1$. The rest of the inactive neurons in the network does not concur any additional computation. Event-driven unsupervised learning is achieved by introducing a checker function for all the pre and post-synaptic neurons connected to the spiked neurons to keep track of the time difference between the current time and the last spike time, which is a crucial component of the STDP-based learning. The number of such checker functions is typically small, thanks to the convolution structure of the popular networks. We use a layer-wise learning mechanism that allows the framework to update the neurons layer by layer. This enables initiation of either LTP or LTD depending on the synaptic type (pre or post) of the source neuron. Fig. 3 shows the event-driven SNN processing described in the paper, and Table II shows the quantitative efficacy of the proposed method. On high-performance hardware (AMD), a 6× and 167× throughput improvement in learning and inference can be observed, respectively, which is almost similar to the low-power hardware (Intel) benchmark (6.3× in learning and 170× in inference). For the GPU benchmark, although the improvement is relatively smaller (3.5× in learning and 8.2× in inference), we achieve near real-time processing performance with an inference speed of 35320 events/second.

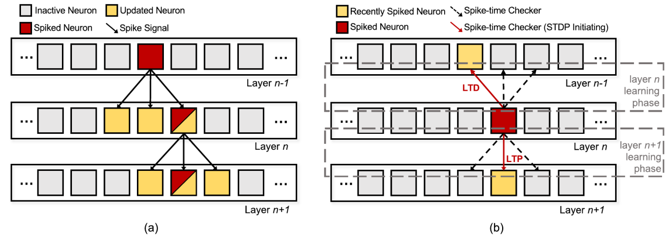

Fig. 3: Event-driven SNN processing where each neuron is connected to three neurons in the next layer. (a) The update occurs only to those neurons receiving input. (b) Event-driven STDP learning only for those neurons connected to the spiked neuron. Figure sourced from [7] with the author's permission.

Single Threaded (Intel i5-4278U)	**Discrete-time Simulation**	**Event-driven Simulation**
Learning Throughput	26 events/s	164 events/s
Inference Throughput	32 events/s	5424 events/s
Single Threaded (AMD Ryzen 5600X)		
Learning Throughput	46 events/s	274 events/s
Inference Throughput	55 events/s	9216 events/s
GPU Parallel (NVIDIA RTX 2080Ti)		
Learning Throughput	3224 events/s	11351 events/s
Inference Throughput	4325 events/s	35320 events/s

TABLE II: Processing speed comparison of discrete-time and event-driven SNN simulation across diverse hardware platforms. Table sourced from [4] with the authors permission.

V. Associative Memory: Brain-inspired alternative paradigm for event-based perception

While SNNs have shown remarkable efficiency in event-camera data processing, they still lag behind the modern dense architectures (CNN and Graph Networks) in terms of algorithmic performance (accuracy). This is mostly due to the lack of exact and accurate differentiability of the spiking operations coupled with the complex and highly sensitive hyperparameter space. On the other hand, modern network architectures require either temporal aggregation or redundant computation (storing and processing past events) to process event-camera data. More specifically, methods that employ convolution-based operations require temporal aggregation of the events to create a frame-like representation and adopt existing CNN-architectures (originally designed for frame-based data) for processing, which suffers from low-throughput, and synchronous operation [3]. Recent works have explored sparse alternatives to convolution, such as a graph neural network-based approach that represents incoming events in a spatiotemporal graph and point-cloud-based method [35] that leverage PointNet-like architecture by treating events as a spatiotemporal point cloud. Despite their computational efficiency, they still suffer from redundant computation resulting from storing and re-processing past events every time a new event occurs. This is illustrated in Fig. 4. Unlike these methods, the human brains can much more efficiently perceive by combining the immediate sensory inputs, and the past stored patterns in the memory [44], [45]. Associative memory in the human brain can efficiently store and retrieve such patterns that highly correlate with sensory input. Efficient processing for Event-based perception also requires a similar computing paradigm that can leverage the benefits of artificial associative memory to compute and store relevant representations from

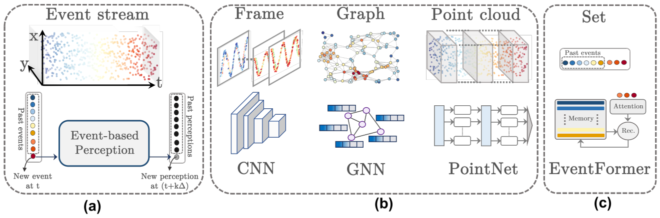

Fig. 4: (a) Event-based perception algorithms. (b) Existing methods for processing event-camera data. (c) EventFormer with an associative memory to correlate with past events. Figure sourced from [7] with the authors' permission.

the past. An event can be triggered at any pixel location depending on the object's motion and scene dynamics. The key algorithmic challenge is that a single event independently does not represent useful perception unless properly correlated with past events across space and time. As events are triggered asynchronously, an event-based processing algorithm must generate and maintain a higher-order representation from the events and efficiently update that representation for each new event to properly correlate a new event with the past events across both space and time (Fig. 4 (a)). This necessitates recursive processing with internal memory to store useful information from past events. To address this challenge, we discuss the recently proposed associative memory-augmented event-representation learning framework [7], inspired by the human-brain memory model. We maintain a spatiotemporal representation associated with past events, occurring at various pixels as the hidden states of an Associative Memory. This enables shifting the spatiotemporal correlation from input to the compact latent space, thereby reducing computation by an order of magnitude. EventFormer encodes the positional coordinates of the incoming event streams into a higher-dimensional embedding space using a positional encoder, followed by their higher-order interaction by computing self-attention among the positional embeddings. Next, it retrieves the past representation stored in the memory using a query-key-based association. A recurrent module takes these past states and combines them with the present states to produce a refined representation and writes it back to the memory for future reference. A task-specific head operates directly on this memory representation and learns to perform the target task (classification). Table III shows the competitive performance of EventFormer on the NCaltech101 dataset. Thanks to its latent-space spatiotemporal processing capability, EventFormer achieves higher performance (accuracy and compute efficiency) than the existing frame-based and sparse methods.

Methods	Representation	Acc.	MFlops/ev
H-First [46]	Spike	0.054	-
Gabor-SNN [47]	Spike	0.284	-
HOTS [48]	Time-Surface	0.210	54.0
HATS [49]	Time-Surface	0.642	4.3
DART [50]	Time-Surface	0.664	-
EST [51]	Event-Histogram	0.817	4150
Matrix-LSTM [52]	Event-Histogram	0.843	1580
YOLE [53]	Voxel-Grid	0.702	3659
AsyNet [54]	Voxel-Grid	0.745	202
EvS-S [55]	Graph	0.761	11.5
AEGNN [56]	Graph	0.668	0.369
EventFormer [7]	Set	0.848	0.048

TABLE III: Performance comparison on NCaltech101 dataset. Table sourced from [7] with the authors' permission.

VI. Conclusion

This work briefly overviews our recent efforts in developing brain-inspired spatiotemporal processing paradigms for efficient event-based perception. We discussed how a simple feedforward SNN could approximate complex spatiotemporal sequences without using any recurrent structure. Our work on event-driven processing makes it possible to achieve near-real-time performance on event-camera data processing using SNNs. We show that heterogeneous neuronal dynamics can enhance the processing capability of the SNNs to enable them to achieve better performance. Finally, we provide an alternate formulation for event-based processing leveraging an associative memory structure by efficiently correlating with past events. These advancements require further research to develop more compute-efficient processing algorithms for event-based perception.

Acknowledgement

This work is supported in part by the Army Research Office and Defense Advanced Research Projects Agency (DARPA) under Grant Numbers W911NF-19-1-0447 and GR00019153, respectively. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Department of Defence, DARPA, or the U.S. Government.

References

[1] Davide Falanga et al. Dynamic obstacle avoidance for quadrotors with event cameras. *Science Robotics*, 5(40):eaaz9712, 2020.
[2] Sihao Sun et al. Autonomous quadrotor flight despite rotor failure with onboard vision sensors: Frames vs. events. *IEEE Robotics and Automation Letters*, 6(2):580–587, 2021.
[3] Daniel Gehrig et al. End-to-end learning of representations for asynchronous event-based data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5633–5643, 2019.
[4] Xueyuan She et al. A heterogeneous spiking neural network for unsupervised learning of spatiotemporal patterns. *Frontiers in Neuroscience*, 14:1406, 2021.
[5] Biswadeep Chakraborty et al. Heterogeneous recurrent spiking neural network for spatio-temporal classification. *arXiv preprint arXiv:2211.04297*, 2022.
[6] Xueyuan She et al. Speed: Spiking neural network with event-driven unsupervised learning and near-real-time inference for event-based vision. *IEEE Sensors Journal*, 21(18):20578–20588, 2021.
[7] Uday Kamal et al. Associative memory augmented asynchronous spatiotemporal representation learning for event-based perception. In *International Conference on Learning Representations*, 2023.
[8] Claus Bing et al. A Survey of Robotics Control Based on Learning-Inspired Spiking Neural Networks. *Frontiers in neurorobotics*, 12(NA):35–35, 2018.

[9] Jacob Hwu et al. IJCNN - A self-driving robot using deep convolutional neural networks on neuromorphic hardware. *2017 International Joint Conference on Neural Networks (IJCNN)*, NA(NA):635–641, 2017.

[10] David Salt et al. Parameter Optimization and Learning in a Spiking Neural Network for UAV Obstacle Avoidance Targeting Neuromorphic Processors. *IEEE transactions on neural networks and learning systems*, 31(9):3305–3318, 2019.

[11] Jun Akopyan et al. TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(10):1537–1557, 2015.

[12] Narayan Davies et al. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning. *IEEE Micro*, 38(1):82–99, 2018.

[13] Elias Gallego, Jon E. A.; Mueggler et al. Event-Based, 6-DOF Camera Tracking from Photometric Depth Maps. *IEEE transactions on pattern analysis and machine intelligence*, 40(10):2402–2412, 2017.

[14] Biswadeep Chakraborty et al. A fully spiking hybrid neural network for energy-efficient object detection. *IEEE Transactions on Image Processing*, 30:9014–9029, 2021.

[15] Biswadeep Chakraborty et al. Characterization of generalizability of spike timing dependent plasticity trained spiking neural networks. *Frontiers in Neuroscience*, 15:695357, 2021.

[16] Xueyuan She et al. Fast and low-precision learning in gpu-accelerated spiking neural network. In *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 450–455. IEEE, 2019.

[17] Jean-Matthieu Maro et al. Event-based gesture recognition with dynamic background suppression using smartphone computational capabilities. *Frontiers in neuroscience*, 14:275, 2020.

[18] Maria-Jose Escobar et al. Action recognition using a bio-inspired feedforward spiking network. *International journal of computer vision*, 82(3):284–301, 2009.

[19] Jing Yang et al. Real time human motion recognition via spiking neural network. In *IOP Conference Series: Materials Science and Engineering*, volume 366, page 012042. IOP Publishing, 2018.

[20] Wei Wang et al. Temporal spiking recurrent neural network for action recognition. *IEEE Access*, 7:117165–117175, 2019.

[21] Arun M George et al. A reservoir-based convolutional spiking neural network for gesture recognition from dvs input. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2020.

[22] Vyacheslav Demin et al. Recurrent spiking neural network learning based on a competitive maximization of neuronal activity. *Frontiers in neuroinformatics*, page 79, 2018.

[23] Wenrui Zhang et al. Spike-train level backpropagation for training deep recurrent spiking neural networks. *Advances in neural information processing systems*, 32, 2019.

[24] Zijian Wang et al. Recurrent spiking neural network with dynamic presynaptic currents based on backpropagation. *International Journal of Intelligent Systems*, 2021.

[25] Matthieu Gilson et al. Stdp in recurrent neuronal networks. *Frontiers in computational neuroscience*, 4:23, 2010.

[26] Wolfgang Maass. Liquid state machines: motivation, theory, and applications. *Computability in context: computation and logic in the real world*, pages 275–296, 2011.

[27] Fangzheng Xue et al. Computational capability of liquid state machines with spike-timing-dependent plasticity. *Neurocomputing*, 122:324–329, 2013.

[28] Vladimir Ivanov et al. Increasing liquid state machine performance with edge-of-chaos dynamics organized by astrocyte-modulated plasticity. *Advances in Neural Information Processing Systems*, 34, 2021.

[29] Xueyuan She et al. Sequence approximation using feedforward spiking neural network for spatiotemporal learning: Theory and optimization methods. In *International Conference on Learning Representations*, 2022.

[30] Nicolas Perez-Nieves et al. Neural heterogeneity promotes robust learning. *bioRxiv*, 12(1):2020–12, 2021.

[31] Xueyuan She et al. Sequence approximation using feedforward spiking neural network for spatiotemporal learning: Theory and optimization methods. In *International Conference on Learning Representations*, 2021.

[32] Fleur Zeldenrust et al. Efficient and robust coding in heterogeneous recurrent networks. *PLoS computational biology*, 17(4):e1008673, 2021.

[33] Biswadeep Chakraborty et al. Heterogeneous neuronal and synaptic dynamics for spike-efficient unsupervised learning: Theory and design principles. In *International Conference on Learning Representations*, 2023.

[34] Yin Bi et al. Graph-based spatio-temporal feature learning for neuromorphic vision sensing. *IEEE Transactions on Image Processing*, 29:9084–9098, 2020.

[35] Qinyi Wang et al. Space-time event clouds for gesture recognition: From rgb cameras to event cameras. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1826–1835. IEEE, 2019.

[36] Joao Carreira et al. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017.

[37] Qianhui Liu et al. Event-based action recognition using motion information and spiking neural networks. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, Z.-H. Zhou, Ed. International Joint Conferences on Artificial Intelligence Organization*, volume 8, pages 1743–1749, 2021.

[38] Darjan Salaj et al. Spike-frequency adaptation provides a long short-term memory to networks of spiking neurons. *bioRxiv*, 5(number), 2020.

[39] Jacques Kaiser et al. Synaptic plasticity dynamics for deep continuous local learning (decolle). *Frontiers in Neuroscience*, 14:424, 2020.

[40] Yan Meng et al. Modeling activity-dependent plasticity in bcm spiking neural networks with application to human behavior recognition. *IEEE transactions on neural networks*, 22(12):1952–1966, 2011.

[41] Yan Zhou et al. Surrogate-assisted evolutionary search of spiking neural architectures in liquid state machines. *Neurocomputing*, 406:12–23, 2020.

[42] Hananel Hazan et al. Bindsnet: A machine learning-oriented spiking neural networks library in python. *Frontiers in neuroinformatics*, 12:89, 2018.

[43] Arnon Amir et al. A low power, fully event-based gesture recognition system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7243–7252, 2017.

[44] Thomas D Albright. On the perception of probable things: neural substrates of associative memory, imagery, and perception. *Neuron*, 74(2):227–245, 2012.

[45] Jin-Hui Wang et al. Associative memory cells and their working principle in the brain. *F1000Research*, 7, 2018.

[46] Cedric Orchard et al. HFirst: A Temporal Approach to Object Recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(10):2028–2040, 2015.

[47] Alan C. Bovik et al. Multichannel texture analysis using localized spatial filters. *IEEE transactions on pattern analysis and machine intelligence*, 12(1):55–73, 1990.

[48] Garrick Lagorce et al. HOTS: A Hierarchy of Event-Based Time-Surfaces for Pattern Recognition. *IEEE transactions on pattern analysis and machine intelligence*, 39(7):1346–1359, 2017.

[49] Amos Sironi et al. Hats: Histograms of averaged time surfaces for robust event-based object classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1731–1740, 2018.

[50] Bharath Ramesh et al. Dart: distribution aware retinal transform for event-based cameras. *IEEE transactions on pattern analysis and machine intelligence*, 42(11):2767–2780, 2019.

[51] Antonio Gehrig et al. ICCV - End-to-End Learning of Representations for Asynchronous Event-Based Data. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, NA(NA):5632–5642, 2019.

[52] Marco Cannici et al. Matrix-lstm: a differentiable recurrent surface for asynchronous event-based data. In *Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK*, pages 23–28, 2020.

[53] Marco Cannici et al. Asynchronous convolutional networks for object detection in neuromorphic cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[54] Daniel Messikommer et al. Event-based Asynchronous Sparse Convolutional Networks. *arXiv: Computer Vision and Pattern Recognition*, NA(NA):NA–NA, 2020.

[55] Yijin Li et al. Graph-based asynchronous event processing for rapid object recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 934–943, 2021.

[56] Simon Schaefer et al. Aegnn: Asynchronous event-based graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12371–12381, 2022.