Hindawi
Complexity
Volume 2021, Article ID 6689337, 19 pages
https://doi.org/10.1155/2021/6689337

Review Article

Data-Driven Technology in Event-Based Vision

Ruolin Sun,[1] **Dianxi Shi**,[2,3] **Yongjun Zhang**,[2] **Ruihao Li**,[2,3] **and Ruoxiang Li**[1]

[1]*College of Computer, National University of Defense Technology, Changsha, China*
[2]*Artificial Intelligence Research Center, National Innovation Institute of Defense Technology, Beijing, China*
[3]*Tianjin Artificial Intelligence Innovation Center, Tianjin, China*

Correspondence should be addressed to Dianxi Shi; dxshi@nudt.edu.cn

Received 28 November 2020; Revised 9 March 2021; Accepted 17 March 2021; Published 28 March 2021

Academic Editor: Dimitri Volchenkov

Event cameras which transmit per-pixel intensity changes have emerged as a promising candidate in applications such as consumer electronics, industrial automation, and autonomous vehicles, owing to their efficiency and robustness. To maintain these inherent advantages, the trade-off between efficiency and accuracy stands as a priority in event-based algorithms. Thanks to the preponderance of deep learning techniques and the compatibility between bio-inspired spiking neural networks and event-based sensors, data-driven approaches have become a hot spot, which along with the dedicated hardware and datasets constitute an emerging field named event-based data-driven technology. Focusing on data-driven technology in event-based vision, this paper first explicates the operating principle, advantages, and intrinsic nature of event cameras, as well as background knowledge in event-based vision, presenting an overview of this research field. Then, we explain why event-based data-driven technology becomes a research focus, including reasons for the rise of event-based vision and the superiority of data-driven approaches over other event-based algorithms. Current status and future trends of event-based data-driven technology are presented successively in terms of hardware, datasets, and algorithms, providing guidance for future research. Generally, this paper reveals the great prospects of event-based data-driven technology and presents a comprehensive overview of this field, aiming at a more efficient and bio-inspired visual system to extract visual features from the external environment.

1. Introduction

In event-based visual systems comprising perception and processing, algorithms in the processing part are designed to maintain the intrinsic advantages of event cameras, among which data-driven approaches are basically the most prevalent and promising algorithms. Since the development of algorithms is inseparable from relevant hardware and datasets, data-driven approaches along with the required hardware and datasets collectively constitute an emerging research field named data-driven technology. Focusing on data-driven technology in event-based vision, this paper successively presents the background knowledge, the reasons why data-driven technology becomes a research focus, current status of data-driven technology, and future trends, as illustrated in Figure 1. Both current development and future trends of data-driven technology in event-based vision are discussed in terms of algorithms, hardware, and datasets' field.

For numerous smart devices, a visual system that can perceive the external environment and extract visual features of interest from it acts as a crucial prerequisite for performing specific tasks. New requirements such as low power consumption, stability in challenging environments, and real-time response have emerged as well. Though frame-based visual sensors have taken over the market for more than a century, recording instant and rich information about the whole viewed scene, their operating principle leads to natural failures in case of fast motions, or difficult lighting scenarios. Moreover, transmission of massive redundant information also increases the power consumption and latency, limiting application in scenarios that demand immediate reaction or lacking computation power. Inspired by traditional frame-based vision and biological mechanisms, a new type of visual sensor named event camera [1, 2] is on the rise, aiming at applications where traditional cameras fail. Equipped with merits such as high dynamic range, high temporal resolution, low latency, and low

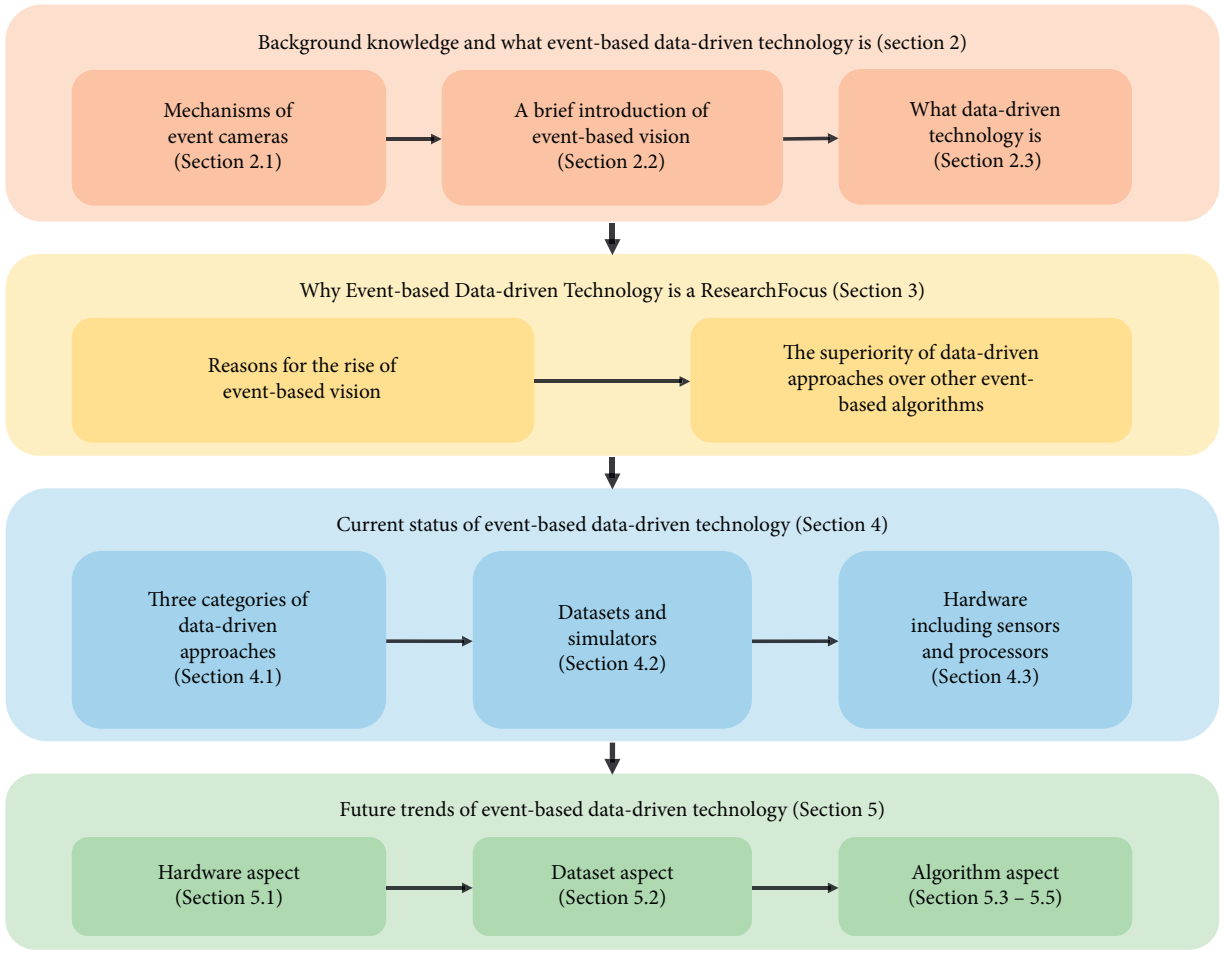

Figure 1: The general framework of this paper, which successively presents the background knowledge, reasons for the rise of data-driven technology, current status of data-driven technology, and future trends. Submodules in the figure correspond to each section of this paper and arrows therein indicate the order of each section.

power consumption, event cameras own a bright future in various practical applications owing to their efficiency and robustness.

Outline: the comprehensive knowledge embodied in this paper is illustrated in Figure 2 and the rest of the paper is organized as follows. Section 2 gives a brief introduction to the relevant background and raises the concept of data-driven technology. In Section 3, two levels of reasons for the predominance of data-driven technology are elaborated according to practical application requirements and existing technologies. Section 4 reviews the current situation of data-driven technology. Concretely, data-driven algorithms are introduced in categories, as well as the relevant hardware and datasets. Section 5 proposes several promising orientations spanning algorithms, datasets, and hardware, providing guidance for future research. This paper ends with a conclusion in Section 6.

2. Background Knowledge and Data-Driven Technology

Compared with traditional cameras that capture whole frames at a fixed rate, event cameras generate event data at a different data form. The novel working principle naturally leads to unique advantages and inherent properties of event data. Underlying the superiority of event cameras exists a subfield named event-based vision [3], which drives a paradigm shiftäin visual features' acquisition and comprises multiple mutually promoting modules such as startups, applications, algorithms, hardware, and datasets. Event-based visual systems are designed to extract interested visual features from the external environment. According to the practical application requirements, maintaining the advantages of robustness and efficiency of event cameras is of utmost urgency in event-based visual systems. Therefore, event-based algorithms should balance the trade-off between accuracy and efficiency, among which data-driven approaches stand out on the basis of existing theories and technologies. The relationship among various parts in the relevant background is shown in Figure 3.

2.1. Mechanisms of Event Cameras. A growing understanding of biological vision is inspiring multiple efficient means of perceiving the environment. Inside an eye ball, the network of various linking cells provides a spatio-temporal filtering mechanism, emphasizes edges and temporal

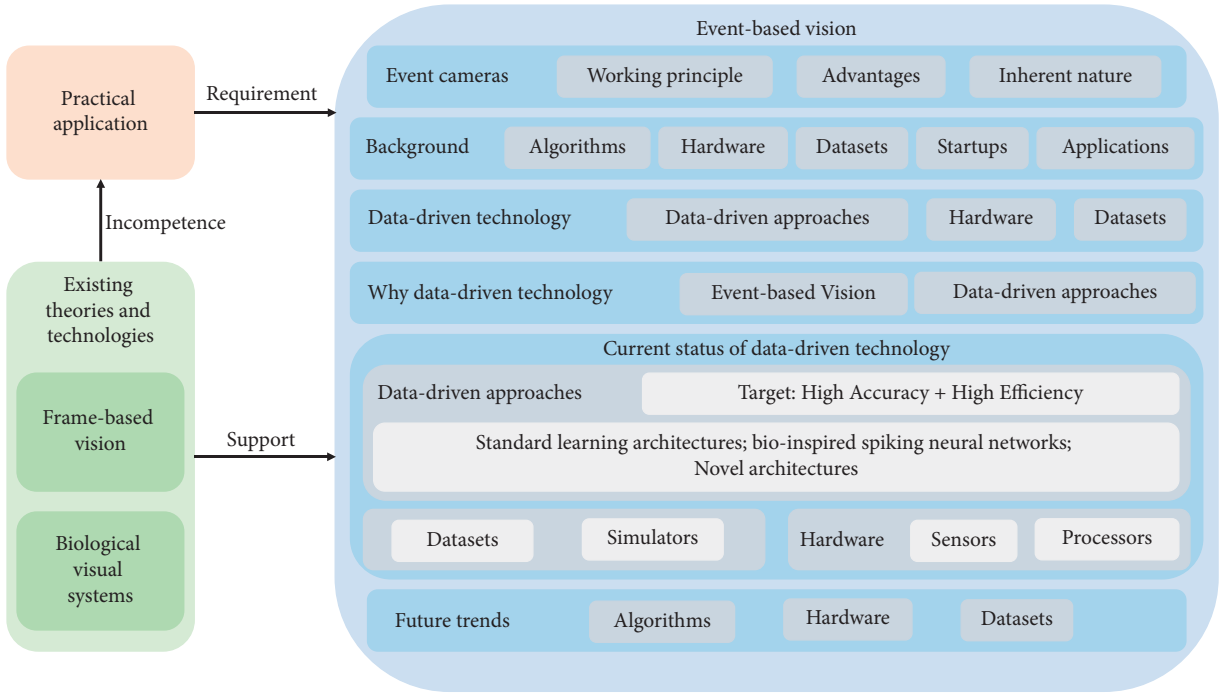

Figure 2: The comprehensive knowledge system embodied in this paper, spanning areas of practical application, existing theories and technologies, and event-based vision, with an emphasis on data-driven technology in event-based vision.

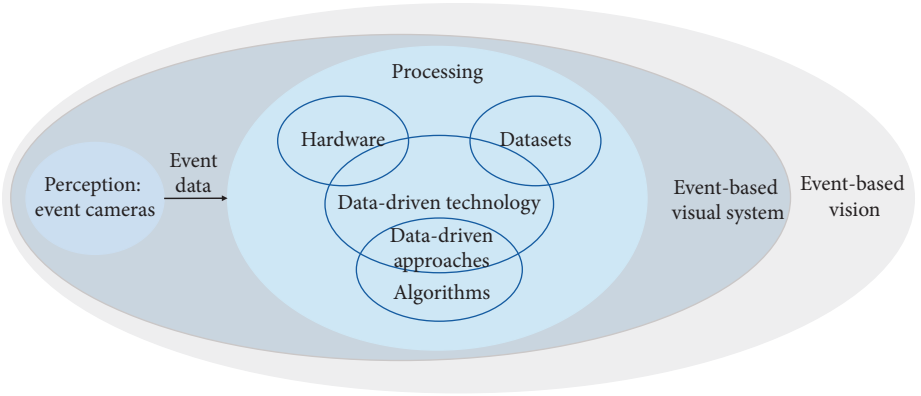

Figure 3: The relationship among event-based vision, event-based visual system, data-driven approaches, and data-driven technology. In event-based vision, event-based visual systems which are widely used in robots, autonomous vehicles, and other applications have become an important research field. Event-based visual systems comprise two parts, namely, perception and processing. In the perception part, event cameras collect information from the external environment and pass down event data to the postprocessing part. The processing part focuses on extracting visual features from the raw event data, comprising hardware, datasets, and algorithms' aspect. Because data-driven approaches are gaining increasing popularity in event-based algorithms, this paper mainly focuses on event-based data-driven technology, covering hardware, datasets, and algorithms' aspect.

changes in the viewed scene, and discards redundant information. Event-based visual systems are designed by mimicking the biological retina and the subsequent processing in the brain. Event cameras are bio-inspired visual sensors that respond to scene dynamics and record log brightness changes rather than full images as frame-based visual sensors do. With pixel-wise logic, the response is defined by the viewed scene, quite like the human eye. Event cameras output a sequence of events represented with four different components: two coordinates (x, y) for the location, a timestamp (t), and a polarity (p) for the change (i.e., brightness increase 1 or decrease −1), denoted as

$$\log\left(I_{x,y,t+\Delta t}\right) - \log\left(I_{x,y,t}\right) \geq pC, \tag{1}$$

where C is the predefined threshold, Δt is the time interval, and I is the intensity. Events are generated with various data rates depending on the magnitude of brightness changes in the scene. Since illumination in the scenario is usually constant, events are mainly triggered by object movements.

Specifically, event cameras own microsecond time resolution and submillisecond transmission latency, offering fast reaction time and outstanding efficiency.

The novel working principle that each pixel reports log-scale intensity changes independently and asynchronously brings vision to the very edge and focus more on the scene dynamics, thus mitigating latency and data redundancy dramatically. In summary, event cameras deliver numerous advantages over frame-based visual sensors, including

(1) High temporal resolution, which enables event cameras to capture fast motions without suffering from motion blur

(2) Low latency, which benefits from the working principle that exempts pixels from global exposure time

(3) High dynamic range and low light sensitivity, suitable for a wider range of lighting conditions, including challenging lighting environments

(4) Low power consumption and high data efficiency, which filters out redundant by recording only pixel-level brightness changes

In a nutshell, its availability in various operating requirements demonstrates a large potential in both machine vision applications and research field.

Event cameras output a stream of events represented as a tuple of location, timestamp, and polarity of the intensity change. Owing to the unique working principle, event data is endowed with the spatial and temporal sparsity nature [4], corresponding to its high efficiency. On the other side, information extraction from event data plays a crucial role for further analysis in event-based vision. Considering the two points mentioned above, an ideal event-based algorithm is supposed to exploit the spatio-temporal sparsity of event data and, at the same time, extract sufficient information from it. In other words, the balance between high accuracy and high efficiency remains a core challenge.

2.2. A Glance over Event-Based Vision. With the emergence and advancement of event cameras, a new field named event-based vision is on the rise, spanning aspects of algorithms, hardware, datasets, manufactures, and applications. During the development of event-based algorithms, existing theories and techniques in biological visual systems and frame-based vision provide considerable support, as illustrated in Figure 4. With the main point of efficiency and accuracy, several rules can be summed up in an event-based vision. First, an event-based vision targets at practical application requirements and takes lessons from existing theories and techniques in biological systems and frame-based computer vision. Moreover, since event-based vision consists of two collaborating parts, namely, perception and processing, comprehensive information should be extracted from event data and passed down to the processing part with the spatio-temporal sparsity well maintained. Last but not the least, since the improvement of algorithms is inseparable from the related hardware and datasets, all three parts should be collectively analyzed in event-based vision.

Event-based vision is still at its early stages compared with frame-based vision. Lessons from frame-based vision include efficient algorithms, dedicated hardware, and large datasets, which also guide the development of event-based vision, as illustrated in Figure 5. Diverse algorithms have been developed to unlock event cameras' potentials ranging from low-level to high-level vision tasks. By adjusting parameters, efficient algorithms extract task-oriented information from event data for optimal accuracy. From model-based methods to data-driven approaches, improvements in event-based algorithms lie primarily in the trade-off between high accuracy and efficiency. Classic architectures include HOTS [5], HATS [13], SNN [14], EST [6], and Ev-flownet [7].

In terms of hardware, development is divided into perception and processing part. As for the perception part, event cameras were first commercialized in 2008 by T. Delbruck under the name of Dynamic Vision Sensor (DVS), followed by several developments such as the Asynchronous Time-Based Image Sensor (ATIS) [1], the Dynamic and Active Pixel Vision Sensor (DAVIS) [2], and the color-DAVIS346 [15]. As the "silicon retina" for event cameras, a corresponding "silicon visual cortex" is required for further disposal. Pairing a neuromorphic processor and an event camera plays a vital role in event-based end-to-end systems, facilitating the development of dedicated hardware for spiking neural networks. As for the processing part, several mature event-based computing platforms have been raised including different neuron model types such as SCAMP, SpiNNaker [8], TrueNorth [9], and Loihi [10], with some of them supporting on-chip learning. Furthermore, recent research studies also aim to design vision systems holistically, that is, co-optimizing hardware with algorithms as a whole vision pipeline [16, 17].

Manufactures for event cameras mainly include Paris-based Prophesee, Zurich-based iniVation, and Shanghai-based CelePixel and Insightness (recently acquired by Sony), cultivating an ecosystem with real products. Recently, with Samsung and Sony collaborating with Swiss-based iniVation and Paris-based Prophesee, respectively, putting their image sensor process technologies on the market, the whole event-based industry has been staging a comeback. Applying a data-efficient and low latency approach to sensing and acquisition by mimicking the human retina, event cameras find utilities in areas such as resource-constrained platforms, highly reactive systems, and limited illumination conditions. The emergence of event cameras also promotes machine vision in applications such as automotive, robotics, ARVR, space, inspection, surveillance, and star tracking. Equipped with outstanding properties, event cameras have unlocked new scenarios previously inaccessible, leaving considerable room for improvement in various aspects.

Sufficient labeled data are of vital importance in the performance evaluation part. By bringing down expenses and providing reliable benchmarks, datasets as well as data simulators are elementary tools for further improvement in event-based vision. Sorted by task, they are generally divided into datasets for regression tasks [18–21] and those for classification tasks [22, 23]. Simulators [24, 25] and

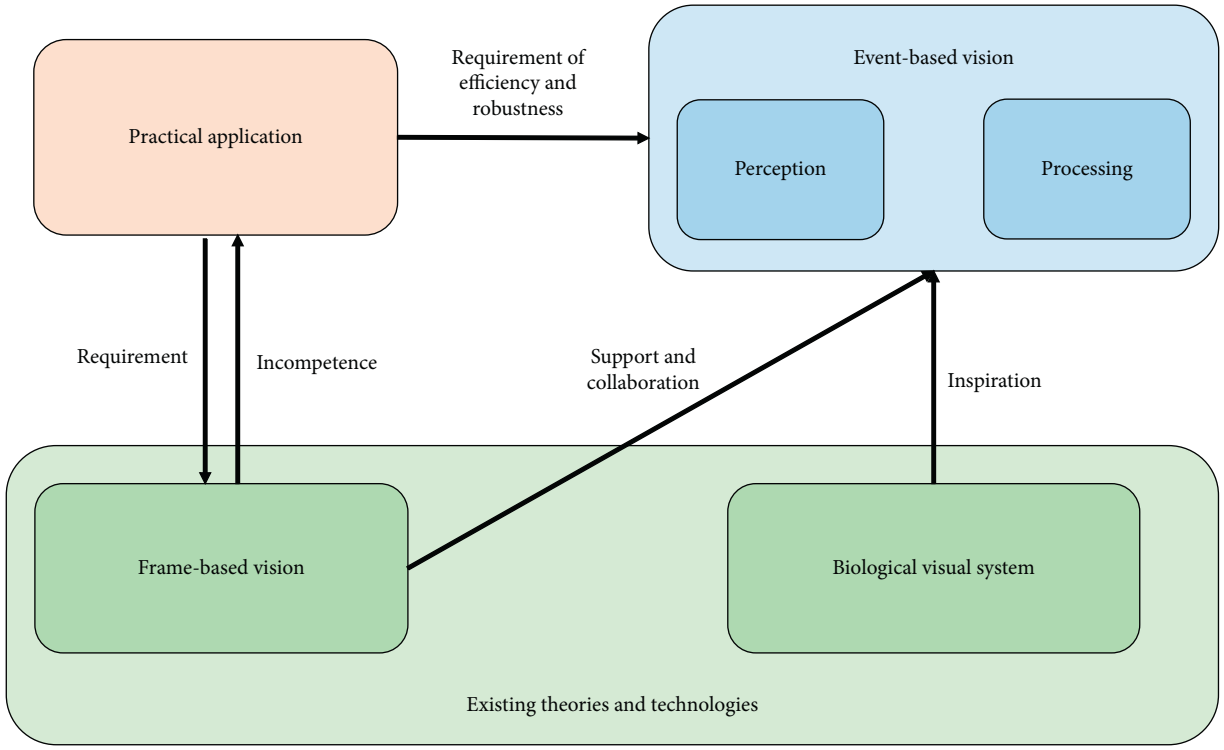

Figure 4: Relationship between practical application requirements, existing theories and technologies (including frame-based vision and biological visual system), and event-based vision.

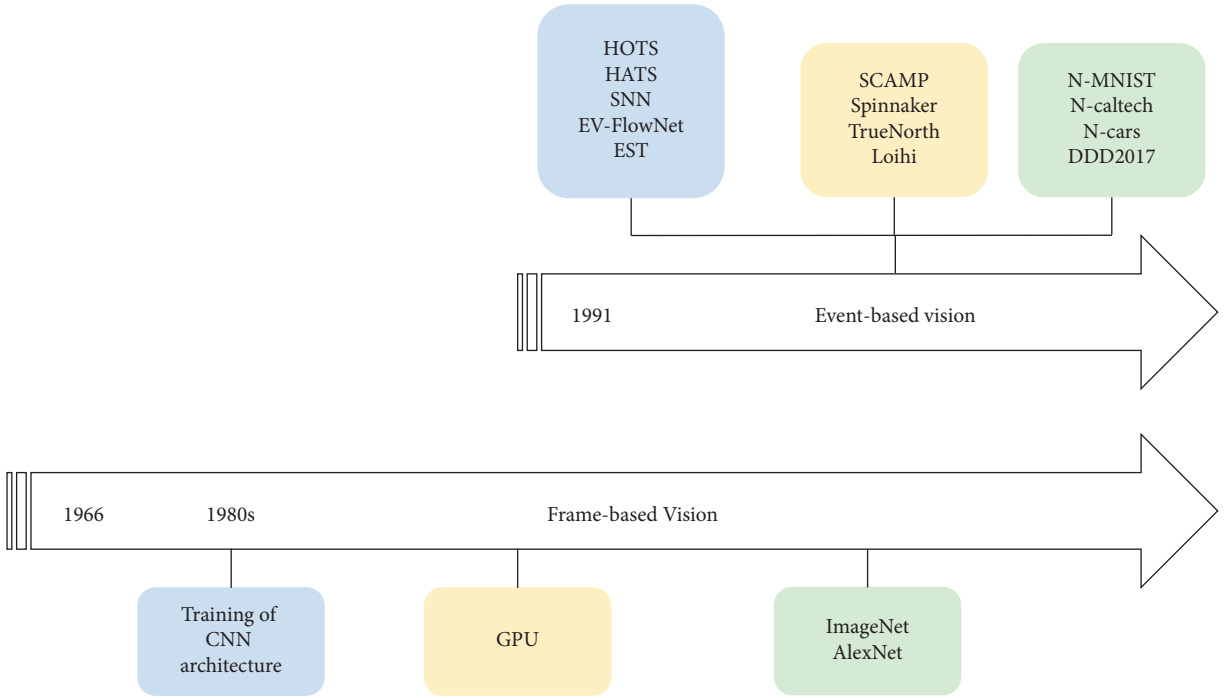

Figure 5: Relationship between event-based vision and frame-based vision, in terms of algorithm (in blue color) [5–7], hardware (in yellow color) [8–10], and dataset (in green color) [11, 12].

emulators [26] imitate the sampling mechanism of event-based sensors, generating data in a low-budget manner, which meets the demand for cheap, high-quality labeled event in algorithm prototyping and algorithm benchmarking. Broadly, efficient algorithms added with proper hardware and sufficient data promise to accelerate progress of the event-based research community, promoting its application in various fields.

2.3. Data-Driven Technology in Event-Based Vision. With the main target of accuracy and efficiency, algorithms in event-based vision are mainly classified as model-based methods and data-driven approaches, and the latter category is gradually occupying the mainstream on the basis of existing theories and technologies. Since the improvement of algorithms is closely related to the development of hardware and datasets, data-driven approaches along with the required hardware and datasets collectively constitute an emerging research field named data-driven technology. A comprehensive analysis on data-driven technology is unfolded in the rest of this paper. Reasons why data-driven technology grows popular are illustrated first, considering the exact need of practical applications and existing technical conditions, and then comes the current status of data-driven technology, among which data-driven approaches as well as the development of relevant hardware and datasets are explained in details. Finally, future trends of data-driven technology in event-based vision are pointed out with respect to algorithms, hardware, and datasets. The holistic structure of data-driven technology is illustrated in Figure 6.

3. Reasons for the Prevalence of Data-Driven Technology

Regardless of what kind of technology it is, one criterion used for assessing its prospects is whether there exist corresponding requirements in practical applications and relevant technical foundation in support of its development. In terms of data-driven technology in event-based vision, reasons for its prosperity can be further divided into two levels considering the criterion mentioned above, as illustrated in Figure 7. Concretely, factors contributing to the rise of event-based vision are presented in the first place, followed by the superiority of data-driven approaches over other kinds of event-based algorithms.

Firstly, as for event-based vision, new areas in commercial markets and industrial fields [27, 28], such as robotics and consumer electronics, give rise to an urgent demand for efficiency and robustness [29] in visual systems, which event cameras can just offer. Despite the lack of experience in event-based vision, mature techniques in frame-based vision and biological visual systems provide valuable guidance in the emerging event-based vision and promote its rapid progress. From another perspective, though frame-based sensors have been adopted by a preponderance of applications, their operating principle that whole frames are recorded at a fixed rate leads to the natural failure in fast motions' scenarios and difficult lighting conditions, which weakens their robustness. Transmission of redundant information also adds to power consumption and latency, resulting in incompetence in highly reactive systems and resource-constrained platforms and weakening their efficiency. Considering the above factors, there are application scenarios that existing techniques cannot satisfy, where event cameras act just as a natural fit. As a result, event-based vision is gaining a slow but steady rise and demonstrates impressive advantages in machine vision applications and consumer electronics.

Secondly, as event-based visual systems comprise perception and processing part, algorithms along with hardware and datasets play a vital role in the latter part, aiming at inheriting the intrinsic advantages of event cameras. Therefore, the trade-off between accuracy and efficiency remains a core challenge for event-based algorithms. An ideal algorithm is supposed to extract sufficient information from events, namely, exploit the fine temporal information of individual events, to ensure accuracy and, at the same time, exploit the sparse and asynchronous nature for low latency and computation, as illustrated in Figure 8.

In terms of model-based methods, common types and their corresponding deficiencies are listed below. A common characteristic of model-based methods is that algorithms are artificial designs based on researchers' knowledge reservation rather than preset models with parameters learned from data. Depending on their relationship with traditional frame-based vision, model-based methods can be further subdivided into two categories: methods that reutilize traditional algorithms [30–33] and methods separate from traditional algorithms [34, 35]. Orthogonally, depending on the degree of information extraction, we can distinguish between methods with information compression [36, 37] and methods that exploit the fine temporal information of each event [31, 33]. A more widely known criterion is to categorize model-based methods by information aggregation manners. One class of research studies [5] uses filtering-based models updated asynchronously with each incoming event. Since an event alone contains little information, each incoming event is usually coupled with extra information from past events for further estimation. Requiring expert knowledge, feature descriptors and measurement update functions [4] require to be handcrafted and task-oriented in these methods, which slows down their widespread adoption in high-level vision such as recognition and segmentation. Despite minimal latency, these methods perform redundant computation owing to frequent system state update, with accuracy sensitive to algorithm parameters. Other works process groups of events simultaneously for high signal-to-noise ratio, integrating events with a fixed number [38] or in a fixed time window [36, 37]. These methods achieve remarkable performance at the cost of losing the inherent low latency property of event data.

Besides, data-driven approaches possess great potentials on the basis of existing theories and technologies, namely, deep learning techniques and biological mechanisms. Recently, growing numbers of event-based research have adopted data-driven approaches [7, 39–41] rather than model-based methods, spanning diverse vision tasks. Inspired by the great success of deep learning methods in traditional frame-based vision, some event-based algorithms aim to reutilize standard learning architectures adopted in frame-based vision after converting groups of event data into frame-like representations. From another perspective, apart from the perception part, biological mechanisms also drive the design of several postprocessing algorithms, such as spiking neural networks (SNN). To some degree, the combination of bio-inspired perception and processing provides an efficient and long-term solution in event-based vision, suiting scenarios where standard cameras fail just well.

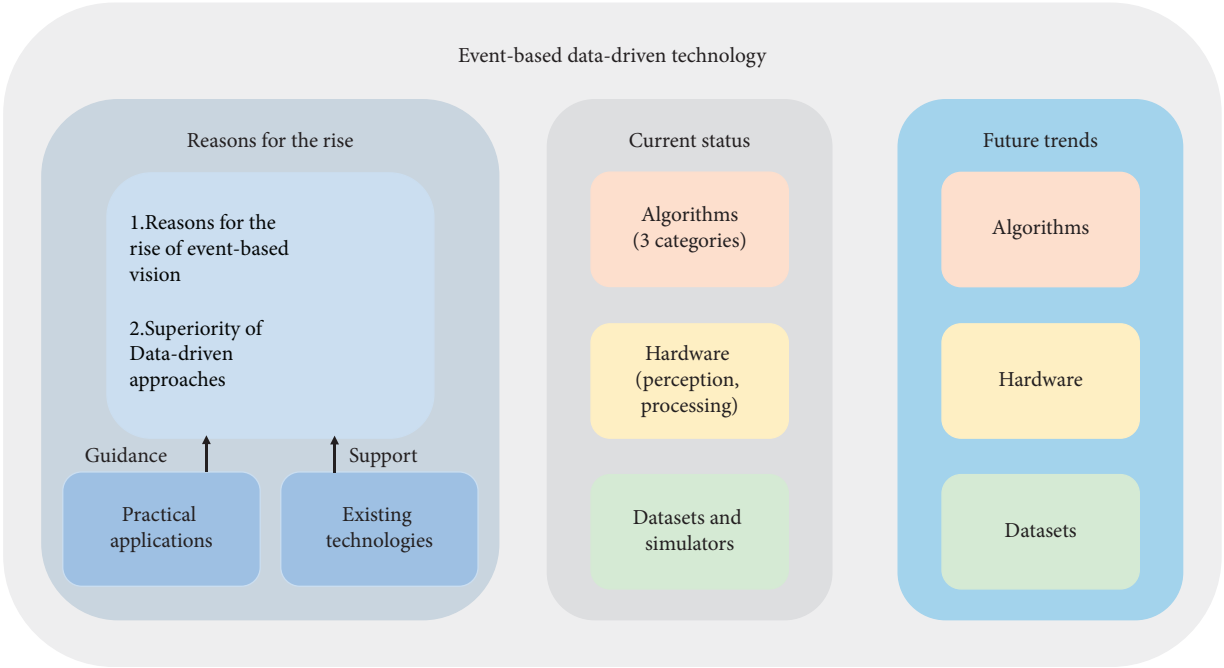

Figure 6: The holistic structure of event-based data-driven technology, including reasons for its rise, current status, and future trends of event-based data-driven technology.

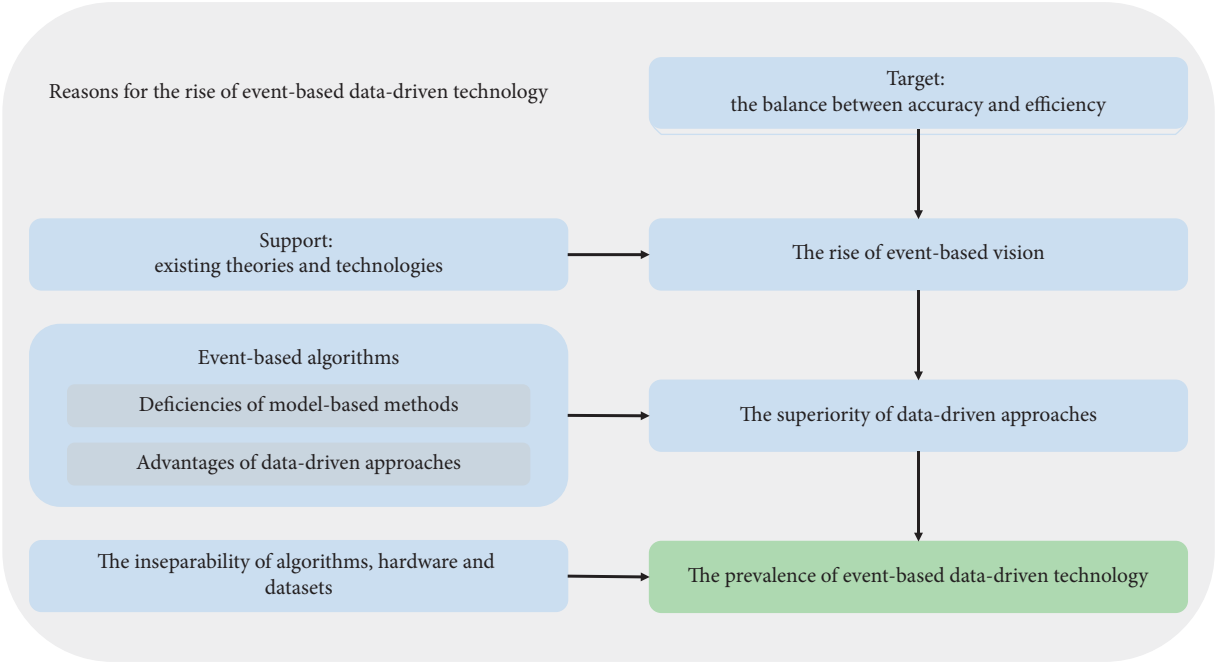

Figure 7: Two reasons for the rise of event-based data-driven technology considering practical application requirements and relevant technical foundation.

In conclusion, the emergence and advancement of event-based vision is to meet the need of efficiency and robustness in several practical applications where traditional frame-based vision shows incompetence. Among event-based algorithms, two main reasons contribute to the prevalence of data-driven approaches: the limitation of model-based methods in complex, high-level vision tasks and the prevalence of deep learning techniques in frame-based vision. Data-driven approaches, dedicated hardware and large datasets remaining core components in data-driven technology, a paradigm shift towards data-driven technology is triggered in event-based vision.

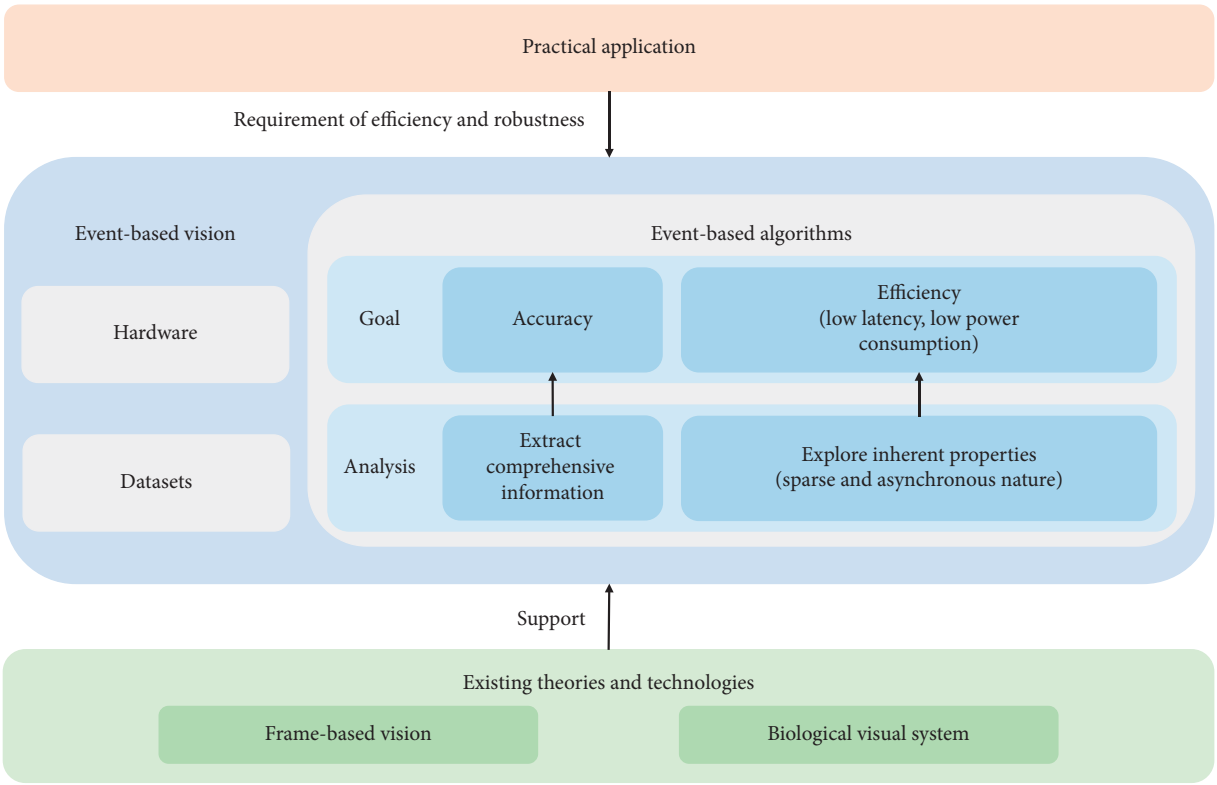

Figure 8: Driven by practical application requirements, the goal of event-based algorithms is to balance the trade-off between high accuracy and high efficiency, namely, extracting comprehensive information and exploring inherent sparse and asynchronous nature. Existing theories and technologies from frame-based vision and biological visual systems provide support for algorithms' design.

4. Current Status of Data-Driven Technology

Since data-driven approaches, dedicated hardware, and relevant datasets collectively constitute the research field of data-driven technology, current status of the three parts will be explained concretely in this section. In terms of algorithms, data-driven approaches are mainly divided into three categories: spiking neural networks, standard learning architectures, and novel architectures. With quantities of event data in urgent need for the implementation of algorithms, large-scale datasets and event camera simulators act as a promising alternative for cost reduction. To improve the holistic performance of event-based visual systems, optimization of hardware ranges from the sensor level to processor level. All three parts influence and promote each other during their development.

4.1. Data-Driven Approaches. With each pixel detecting local luminance changes independently, event cameras pose a paradigm shift in acquisition of scene information, outputting discrete and asynchronous event streams. Due to the unique working principle, deep learning architectures adopted in frame-based vision can not be applied directly to event data, provoking the innovation of several event-based data-driven approaches including spiking neural networks, standard learning architectures, and novel architectures. Spiking neural networks (SNNs) [42] have emerged as a promising candidate since they perform asynchronous inference on specialized neuromorphic hardware [10] with low power consumption, exploiting the inherent spatio-temporal sparsity of event data. However, with the number of spikes drastically vanishing at deeper layers, deep SNNs are notoriously difficult to train, weakening their performance in high-level tasks. Additionally, hardware for SNNs remains expensive and scarce, also limiting their generalization. Some other research [6, 7, 40] aims to reutilize standard learning architectures which are designed for image data by converting groups of events into tensors [5, 6, 13]. These methods have achieved state-of-the-art results in several tasks because of the high-capacity deep neural networks and sufficient signal-to-noise ratio of tensor-like event representations. Correspondingly, the discrete and asynchronous property is sacrificed as a price. Methods above provide solutions with both advantages and disadvantages, and recent research makes optimization by combining complementary advantages that each category has to offer. Current innovation architectures tailored to high-rate, variable-length, and nonuniform event stream have been proposed to balance the trade-off between accuracy and efficiency while protecting the spatio-temporal sparsity [4, 43, 44]. Generally speaking, with high capacity neural networks, features are automatically learned from data by optimizing the corresponding object function in data-driven approaches, exempt from handcrafted feature descriptors. In other words, provided with sufficient high-quality data and suitable

network architecture, parameters are well adjusted according to labeled data through training, regardless of task types.

Inspired by the efficiency and adaptability to changes in biological systems, neuromorphic computing becomes a natural fit to process event data. Though conventional artificial neural networks (ANNs) are predominant tools, the resulting energy requirement hinders their implements on resource-constrained platforms such as embedded devices where event cameras show advantages. Offering asynchronous and sparse computations compatible with event data, spiking neural networks (SNNs) are promising computational partners for event cameras to create an end-to-end event-based system.

Biological neurons work in a separate way from neurons in ANNs, mainly in the way that information propagates between units. Rather than static nonlinearities, neurons in bio-systems are dynamic devices encoding and processing information in the form of discrete spikes, requiring a minute fraction of power. The efficient biological perception and computation principles lead to the realm of SNNs. An SNN is a hierarchical network of dynamic spiking neurons defined by the model parameters, with each unit receiving and outputting spiking signals. Each neuron receives input from its corresponding receptive field, while modifying its membrane potential and emitting an output spike when the state reaches a certain threshold. Neurons connected together form an SNN that processes information with spiking signals rather than numeric values. Universally, input spiking signals are collected from neuromorphic sensors or converted from natural signals. There exist diverse rules for conversion, such as rate encoding, time encoding, and population encoding. Generally, the spiking rate and the temporal pattern contain valuable information about stimulation and computations. Analogously, the output spiking signals are fed to a neuromorphic actuator or converted to natural signals following decoding principles, as illustrated in Figure 9.

Coupling event streams with spiking neural networks exploits the signal's spatio-temporal sparsity and develops event-based end-to-end systems. Without previous conversion from events to image-like tensors, SNNs are more efficient and long-term solutions than conventional frame-based approaches in event-based vision. Despite their asynchronous inference and low power consumption compatible with event cameras, SNNs demand expensive specialized hardware (such as TrueNorth [9] from IBM and Loihi [10] from Intel) and face spikes vanishing problem at deeper layers. Moreover, since the spike-generation mechanism within a neuron is nondifferentiable, feedforward SNNs are difficult to train without effective backpropagation algorithms. Shortcomings adding up together restrict its popularization in complex real-world scenarios. Several event-based research have been working on resolving SNN's shortcomings based on former achievements and utilizing its superiority to the best.

Spiking neural networks (SNNs) [45] have been applied to various event-based fields, including low-level tasks such as optical flow estimation [46–48], high-level tasks such as object recognition [49, 50] and classification [51], and tasks concerning the 3D structure of the scene [52, 53] and robotic visual perception [54]. Benosman et al. [16] used a spiking neural network that is theoretically similar to the classical Lucas–Kanade algorithm to estimate visual motion, exploiting the sparse high temporal resolution event data. Based on [16], the same team [17] demonstrated a fully spiking neural network for optical flow prediction on TrueNorth hardware [9]. Inspired by the visual cortex, a hierarchical feature extraction mechanism has been adopted in SNNs to extract information from the precise timing of the spikes [55]. Authors of [48] estimated event-based optical flow by means of a hierarchical spiking architecture based on Spike-Time-Dependent-Plasticity learning [56]. Researchers in [49] presented a spiking hierarchical model for object recognition utilizing the precise timing information contained in the output data. Tang et al. [50] proposed a hierarchical feedforward spiking neural network for the classification of digital characters recorded by DVS. In addition, Acharya et al. [51] presented a three-layer SNN-based region proposal network operating on event data and applied it to real recordings. Although SNNs have mainly been applied to classification problems [50, 51, 57], a recent research [58] unlocked the potential of SNNs to tackle numeric regression problems in the continuous-time domain for event-based data. Benosman et al. [52, 53] solved the stereo-correspondence problem and estimated depth in a 3D scene in event-based vision with a spiking neural network working on neural processing devices. Moreover, Bing et al. [54] designed an end-to-end SNN based on STDP learning rules for the robotic visual navigation system.

Since spiking neurons' transfer function is naturally nondifferentiable, backpropagation cannot be directly used in training SNNs, limiting the computational potential of SNNs. To breakthrough this bottleneck, some research [59, 60] aims to tailor backpropagation for SNNs and backpropagates error at spike times, reaching limited success. Some other works [61, 62] focus on applying a continuous function as a proxy for spike function derivative, which has been proven effective in deep feedforward SNNs. Since methods above only compute approximate gradients, algorithms for spiking deep convolutional networks [39] remain to be improved in the future study.

Due to the difficult training procedure, limited accuracy in complex tasks remains a challenge for SNN-based methods. Recently, lots of research aims at reutilizing conventional machine learning techniques after converting groups of events into intermediate image-like representations. Mature frame-based learning architectures help accelerate the development of event-based algorithms [5–7, 13].

Additionally, the similarities between tensor-like representations and natural images enable transfer learning with networks pretrained in frame-based vision to some degree [6, 63, 64]. Owing to high capacity neural networks and sufficient signal-to-noise ratio, methods that recur to standard learning architectures have achieved satisfying performance in diverse fields of event-based vision [40, 41, 65]. Broadly, differences between these approaches

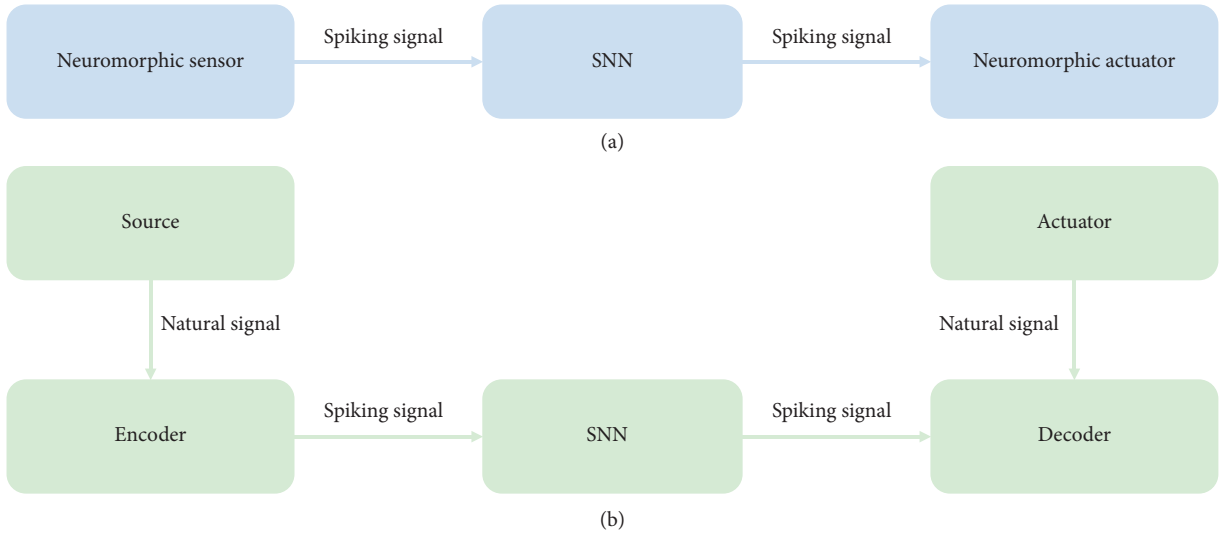

Figure 9: The input and output signal of an SNN (a) presents a direct interface with a neuromorphic sensor and actuator and (b) presents an interface through encoding and decoding.

mainly lie in three aspects: the event representation methods, the network architecture, and the loss functions used for optimization during training.

In order to handle an event stream reutilizing frame-based deep convolutional neural networks (CNN) or recursive neural networks (RNN), events require to be transformed into image-like representations compatible with natural images in advance, as illustrated in Figure 10. The intermediate representations are usually synthesized by batches of events with a fixed number [66] or in a constant temporal window [37]. From another perspective, these representations can be distinguished as updated synchronously [7] or updated asynchronously [13]. Additionally, according to the utilization of temporal information, event data can be represented as the basic dense encoding of event location [40]and dense encoding including temporal information [13, 65]. Several widely adopted representations have emerged during exploration, including event count images [40], maps of most recent timestamps [7, 67], an interpolated voxel grid [64, 65], and the latest end-to-end event representation framework (also referred to as Event Spike Tensor) [6]. Among all these representations, the Event Spike Tensor reserves the maximum amount of information without compression along any dimension.

Maqueda et al. [40] accumulated events of different polarities over a fixed temporal window to predict steering angle. As for optical flow estimation, Zhu et al. [7] proposed a four-dimensional grid encoding the last timestamp and event count of each pixel. Later, the same team made improvements by representing events as a spatio-temporal voxel grid [65], which accumulated events in a linearly interpolated manner using time information as the weight, saving the full spatio-temporal distribution of events. In high-level tasks, Sironi et al. [13] paired event data converted into histograms of averaged time surfaces (HATS) with a support vector machine in object recognition tasks, reutilizing the standard learning pipeline and enabling

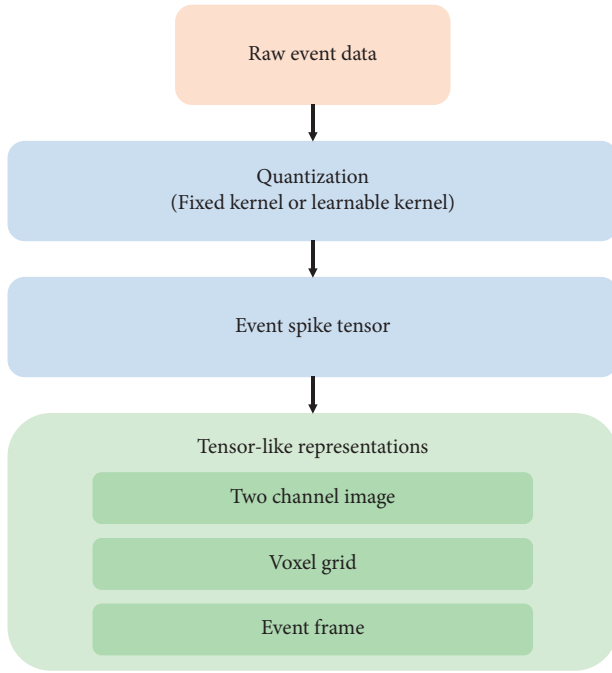

Figure 10: The general framework to convert asynchronous event data into tensor-like representations. These operations are all differentiable [6].

asynchronous update if given sufficient computing power. Alonso et al. [68] contributed a 6-channel image representation in the semantic segmentation task, recording the event count as well as the mean and standard deviation of the timestamps of events. The latest research [6] proposed an end-to-end framework with a learnable kernel, which represented events in a data-driven manner, and evaluated its performance in object recognition and optical flow estimation tasks. Generally, the interpolated voxel grid and Event Spike Tensor are the most popular representations in

recent research [41, 64, 65] for their little information compression in raw event data.

In terms of the network architecture, most research [7, 41, 64, 65, 68, 69] that refers to deep neural networks commonly adopts an encoder-decoder structure inspired by the stacked hourglass [70] and the UNet [71] architecture. The encoder-decoder structure helps ensure the same resolution between the output and the input tensor, satisfying the needs of several tasks, as illustrated in Figure 11. For instance, semantic segmentation is often regarded as a per-pixel classification, with an output at the same resolution of the input. Alonso et al. [68] took an Xception model [72] as the encoder and built a light decoder in event-based semantic segmentation task. Furthermore, fully convolutional networks require less network weight parameters, consequently decreasing the risk of overfitting. Last but not least, the loss can be applied to each intermediate state of the decoder [7], dramatically improving the accuracy of algorithms. Concretely, in optical flow estimation tasks, the EV-FlowNet [7] closely resembled the UNet in terms of the network architecture and designed loss in every intermediate flow from the decoder by downsampling grayscale images. Growing numbers of successful CNN-based or RNN-based approaches have emerged in event-based vision and the encoder-decoder architecture serves as a frequent partner.

As for training methods, supervised learning supported by sufficient labeled data plays a dominant role in classification tasks. For instance, researchers in [6] used the N-Cars dataset [13] and the N-Caltech101 dataset [12] for object recognition evaluation. Authors of [41] compared each reconstruction with the corresponding ground truth grayscale frame pairing a temporal consistency loss with an image reconstruction loss. It is worth mentioning that due to the heavy burden of manually labeling the ground truth and CNNs' robustness to training with approximated labels [73, 74], researchers have successfully used generated labels from grayscale for supervised training in several event-based tasks, such as object detection [75] and semantic segmentation [68].

Since the ground truth for some event-based tasks is difficult to generate and datasets with manually annotated labels remain rare and expensive, self-supervised learning and unsupervised learning have offered an opportunity to learn relevant parameters using event data without corresponding labels, adding to the availability of deep networks in event-based vision. Without enough labeled data [5], a feature extractor trained with unsupervised learning with a classifier is coupled which demands labeled data for training in the recognition task. To predict a vehicle's steering angle [40], it is referred to the pose as a third party ground truth, that is, selecting the frame 1/3 s after the current one as its ground truth. To estimate optical flow [7], a self-supervised loss is applied over the predicted flow using grayscale images generated by the DAVIS camera. Supported by a motion blur based loss function [65], motion information and the structure of the scene in an unsupervised manner are learned using only event data. Generally, unsupervised learning methods make the most of events' coordinates and polarity information and hence do not rely on ground truth or additional information. Gallego et al. [34] provided a unifying framework to handle several estimation problems in event-based vision, obtaining motion parameters that best fit the input data by contrast maximization. The additionally generated motion-corrected edge-like images can also serve several downstream research studies such as feature tracking [38] and object recognition. Researchers carried on this study in later research [35] and analyzed 22 objective functions for unsupervised learning.

An ideal event-based algorithm should extract events' full coordinates and polarity information while exploiting signal's spatio-temporal sparsity to ensure low latency and low power consumption. Data association, namely, establishing inherent correspondences between events, remains a central challenge in event-based vision. With increasing preponderance in event-based algorithms, data-driven approaches are mainly categorized into two classes: asynchronous spiking neural networks [46, 49, 52, 54] and standard learning architectures [7, 40, 65, 68]. As bio-inspired methods, SNNs offer asynchronous inferences at a fraction of power consumption but suffer from the vanishing spike phenomenon [76] in deeper layers. In contrast, the latter methods commonly trade-off efficiency for accuracy and generalize well in complex vision tasks, sacrificing the inherent sparsity of spatio-temporal events. Methods above contribute solutions with both advantages and disadvantages and recent research aims at integrating the possibilities each category has to offer, as illustrated in Table 1. Current innovation architectures tailored to high-rate variable-length event streams have emerged as a promising candidate to balance the trade-off between accuracy and efficiency while protecting the spatio-temporal sparsity [4, 44]. Authors of [44] presented a deep hybrid neural network named Spike-FlowNet for optical flow estimation. The combination of SNNs and ANNs improves its computation efficiency while maintaining performance. Researchers in [4] brought the concept of event-based asynchronous sparse convolutional networks into public sight. To be concrete, a framework that transforms models trained with image-like event representations into asynchronous models is raised to better leverage the asynchronous and sparse nature of events, surmounting the original limitation of deep neural networks in event-based vision.

4.2. Datasets and Simulators. Compared with the mature frame-based computer vision, event-based research is still in its infancy in every aspect, including algorithms, hardware, and datasets. Since event cameras are expensive sensors, only a proportion of research teams can afford this device, severely slowing down the research. In parallel, with the rise of data-driven approaches in the event-based field, a huge amount of data is required for deep neural network training. To circumvent this contradiction, growing numbers of datasets and simulators have been proposed to facilitate the development of event-based algorithms, and they dramatically reduce the research cost and offering quantitative benchmarks for performance evaluation. Available event-based datasets target various applications, including optical

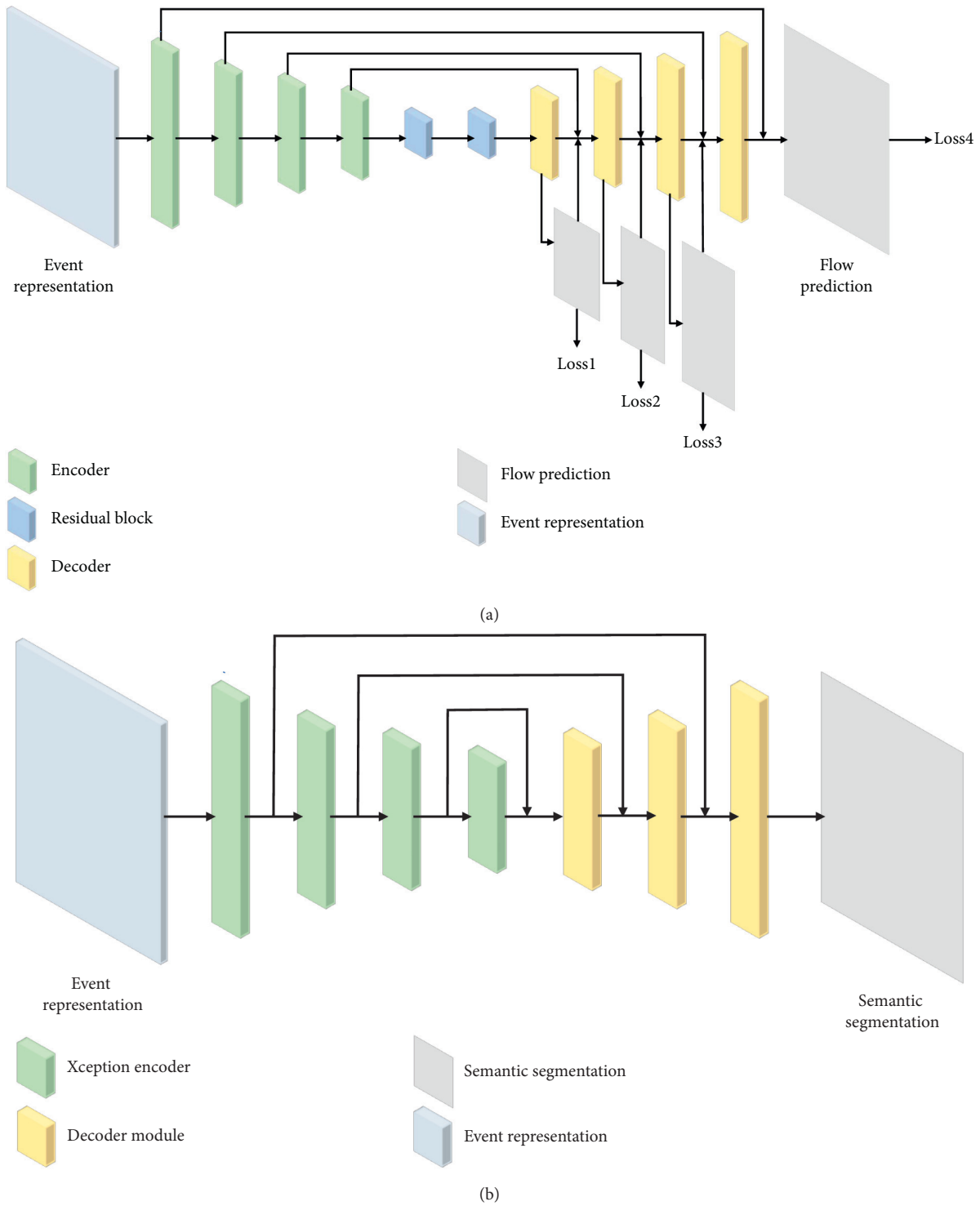

Figure 11: Two examples of the network architectures commonly adopted in event-based algorithms (a) demonstrates the network in optical flow estimation task [7] and (b) demonstrates the network in semantic segmentation task [68].

Table 1: Evaluation of three categories of event-based algorithms.

Category	Superiority	Deficiency
Model-based methods	Low latency	Require prior knowledge
Standard learning architectures	Reutilize frame-based techniques; high accuracy	Redundant computation; high latency
Spiking neural networks	Low latency; low power consumption	Limited accuracy in high-level tasks; require dedicated hardware

flow estimation [20, 77], intensity-image reconstruction from events [78, 79], visual odometry and SLAM [19, 21, 80], segmentation [67, 81], and recognition [22, 23, 82].

For optical flow estimation evaluation, Rueckauer and Delbruck [77] created a dataset recorded with a DAVIS camera and creatively substituted groundtruth optical flow with rate gyro data, since the true optical flow can be computed using gyro data with camera motion restricted to camera rotation. Barranco et al. [20] provided both real-world and simulated datasets to evaluate the performance of visual navigation tasks. Real data was recorded with an RGB-D sensor and a DAVIS sensor on a mobile platform, containing events, images, optical flow, 3D camera motion, and the depth of the scene.

As for image reconstruction task, Binas et al. [11] paired DVS events with APS streams in driving applications and published the DDD17 dataset which contains annotated DAVIS driving recordings in various event-based applications. Researchers in [83] provided the DVS-Intensity dataset recorded with a DAVIS240C camera for event-based image reconstruction. Later, the same team released the CED dataset [78] recorded with a Color-DAVIS346 camera. For the evaluation of event-based video reconstruction, Stoffregen et al. [79] published a High Quality Frames (HQF) dataset with information of events and ground truth frames recorded with a DAVIS240C camera.

Among diverse datasets for SLAM, the most widespread dataset [19] stood out as a benchmark for event-based visual odometry [37, 38, 84] and found great utilities in feature detection [31] and tracking tasks [85] as well. Moreover, for 3D perception tasks, Delmerico et al. [21] released the large MVSEC dataset recorded at different lighting conditions and environments with a DAVIS346 camera. Authors of [86] presented a robotic dataset that contained information about the indoor environment recorded by a mobile robot embedded with two DAVIS240C cameras and an Astra depth camera.

By far, datasets for recognition and classification account for the largest proportion of event-based datasets, including N-MNIST, N-Caltech101 [12], N-CARS [13], and DVS-Gesture [22]. For gesture recognition, the IBM research group collected the DVS-Gesture dataset comprising 11 kinds of hand gestures under 3 illumination levels. Benosman et al. released the first real-world event-based dataset for object classification, namely, the N-CARS dataset which embodied cars at different poses or speeds and various background scenarios. The Dynamic Vision Sensor Human Pose dataset (DHP19) [87] served as a benchmark for human body movements and included a set of 33 movements with 17 subjects from a DAVIS 346 camera.

With the recent preponderance of data-driven approaches in event-based computer vision, a large amount of event data is required to design efficient end-to-end algorithms without intermediate tensor-like representations. To address this issue, apart from large-scale datasets, event camera simulators [19, 24] act as a promising alternative since data and ground truth are easily procurable compared with real data. Simulators in [19] emulated the operation principle of event cameras and generated the corresponding event stream, intensity frames, and depth maps, provided with a virtual scene and a camera trajectory. Simulators in [88] adopted a custom rendering engine to render images from a 3D scene at a very high frame rate, generating the asynchronous output. Different from the fixed-rate sampling approaches mentioned above [24], the event simulator and the rendering engine for a more accurate simulation were tightly coupled. Based on ESIM, the latest research [25, 89] aimed at converting existing video datasets to event datasets, facilitating a variety of event-based applications. Videos of the low frame rate were first transformed into high frame rate ones leveraging an interpolation method [89] and then used for events generation [25]. The v2e toolbox [90] generated events from real or synthetic videos and served as the first candidate to synthesize realistic low light DVS data. Since random noise in neuromorphic sensors remains a challenge for simulation [91], the event probability mask (EPM) to label real-world events with a likelihood was proposed and the DVSNOISE20 dataset for benchmarking denoising was released.

4.3. Sensors and Processors. To improve the holistic performance of event-based visual systems, optimization of hardware mainly consists of improvements from the sensor level and processor level, as illustrated in Figure 12. After the emergence of the first silicon retina, a series of developments have been raised [1, 2, 92]. Among existing event-based sensors, the DVS and its derivatives, namely, DVS128, DAVIS240 [2], and color-DAVIS346 [15], have taken over the academic research, as illustrated in Table 2. Afterward, how to process event data with high efficiency based on brain-inspired chips becomes another focus. Mainstream neuromorphic processors include TrueNorth [9] from IBM, Loihi [10] from Intel, and SpiNNaker [8] from the University of Manchester. Considering the combination, for instance, IBM adopted the DVS128 camera for visual perception before gesture recognition on the TrueNorth chip.

The first commercial camera, the 128×128 pixel DVS128 [93], was published by iniVation and Delbruck et al., with a sampling frequency of $10^6 HZ$ and dynamic range of 120 *db*. It was widely adopted in various tasks including recognition, detection, and tracking. Recently, Samsung has raised a new generation of DVS [94] with higher spatial resolution of 640×480 and smaller pixel size of $9\,\mu m \times 9\,\mu m$.

Recording only brightness changes without absolute intensity information, DVS reduces redundant data and performs well in fast motion scenarios where a ton of data is produced. However, owing to the absence of absolute intensity of the whole scene and the response to only scene dynamics, image reconstruction remained a challenge especially in static scenarios, which stimulated the generation of ATIS [95], DAVIS [2], and CeleX [96]. ATIS [95] reconstructed images referring to an intensity measurement circuit based on a time interval. Cooperating with Posch et al., Prophesee created the 304×240 pixel ATIS [1], with sampling frequency of $10^6 HZ$ and dynamic range of 143 *db*. Financed by Intel, Prophesee also applied ATIS to the visual

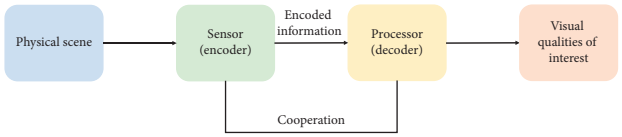

Figure 12: A digital visual system mainly comprises a sensor and a processor. The sensor encodes information from the physical scene and the processor including hardware and algorithms extracts visual qualities of interest. The sensor and the processor cooperate with each other for information acquisition.

Table 2: Comparison between different event cameras.

Type	DVS128	DAVIS240	DAVIS346	DVS-G2
Year	2008	2014	2017	2017
Resolution (pixels)	128×128	180×240	346×260	640×480
Dynamic range (dB)	120	120	120	90
Latency (s)	12	12	20	65–410
Chip size (mm^2)	6.30×6	5×5	8×6	8×5.8
Pixel size (μm^2)	40×40	18.50×18.5	18.50×18.5	9×9
Supply voltage (V)	3.3	1.8 & 3.3	1.8 & 3.3	1.2 & 2.8
Intensity frame	No	Yes	Color	No

system of an autonomous vehicle. However, the mismatch between event data and grayscale reconstruction still exists in fast motion conditions. DAVIS emerged as a combination of DVS and a standard camera by the addition of an extra Active Pixel Sensor, represented as DAVIS240 proposed by iniVation and Delbruck et al. Color-DAVIS346 was later published with a spatial resolution of 346×260. CeleX recorded intensity information by means of expanding event bandwidth to 9*bits*. The recent CeleX-V [97] possesses a competitive spatial resolution of 1280×800, showing great potential in applications such as industrial automation and autonomous vehicles.

Inspired by biological systems, event-based visual systems are designed to pursue high efficiency. Neuromorphic processors unlock the unparalleled computational power of hardware for spiking neural networks, acting as a natural computational partner for event-based sensors. According to neuron types, neuromorphic processors are mainly divided into processors with analog neurons, digital neurons, and software neurons. Several mature architectures such as TrueNorth [9] and Loihi [10] are composed of digital neurons. On-chip learning is also feasible for some processors. Originating from the University of Manchester, the SpiNNaker (Spiking Neural Network Architecture) [8] mimics the biological mechanism in the human brain with constituent neurons as software on the ARM cores, achieving better model flexibility. Except for software neurons adopted in SpiNNaker, IBM's TrueNorth employs digital neurons for real-time computation with networks trained offline, incapable of on-chip learning. IBM has adopted TrueNorth for postprocessing after visual perception with a DVS128 in gesture recognition task [22].The Loihi chip created by Intel also takes digital neurons for inference to support on-chip learning and use programmable rules. Additionally, pretrained nonspiking networks can be transformed into approximate spiking networks with asynchronous inference for Loihi, offering extra possibilities.

5. Future Trends

Inspired by the efficient working principals of biological systems, neuromorphic sensors just stepped into public sight in 1991, in the form of a "silicon retina." The subsequent DVS and DAVIS have significantly promoted event-based computer vision by offering numerous advantages over standard cameras, unlocking scenarios previously unreachable. However, compared with the mature frame-based computer vision, event-based research is still in its infancy considering algorithms, hardware, and datasets. With the coexistence of opportunities and challenges, considerable room is left for future improvement in event-based vision. Several possible directions are pointed out below: spanning hardware, datasets, and algorithms perspectives.

5.1. Hardware. Recently, companies such as Samsung and Sony have built DVS with competitive pixel size, overcoming an initial limit in size and resolution, therefore adding to its fighting chance in practical use. Since the emergence of the first silicon retina by Mahowald and Mead, a series of improvements [1, 2, 92] have been proposed over the past decade, with some of them outputting static along with dynamic information. To accelerate their employment in automotive and consumer applications, further improvement in sensing hardware remains an open challenge. Additionally, since neuromorphic sensors are motivated by the biological retina and subsequent processing principle of the brain, neuromorphic processors act as a natural fit to build an end-to-end event-based system. Several mature processors have emerged, including TrueNorth from IBM [9] and Loihi [10] from Intel. In order to support a branch of event-based algorithms (e.g., SNN), processor-related research is worth carrying on in the future study. Furthermore, inspired by the concept of co-optimizing hardware with postprocessing algorithms holistically [16] and the fact that the contrast threshold of event cameras affect the function of CNN-based algorithms [79], learnable parameters in event-based hardware might be another point of interest.

5.2. Datasets. Since event-based sensors remain scarce and expensive to acquire, large-scale high-quality datasets with ground truth are urgently needed in event-based research community. With the prevalence of data-driven approaches, well-labeled datasets play an increasingly significant role in both algorithm parameters tuning and algorithm benchmarking. Prevailing event-based datasets mainly include real-world datasets [19, 80] recorded with event cameras, simulated datasets [24, 88], and datasets [25, 89] and converted from existing video datasets. Despite those existing

datasets, several problems remain unanswered in this field, leaving considerable room for improvement. For instance, most event-based datasets focus on scenarios where event cameras excel and traditional ones fail, namely, scenarios with fast motion or challenging illumination levels. Only recently an alternative [79] recorded in well-lit conditions with little motion blur was provided. Furthermore, owing to the burden of manually labeling a per-pixel ground truth brought by the novel data form, few large-scale datasets have been published, especially in complex visual tasks such as object detection and segmentation. As a consequence, a large amount of reliable event data is in desperate need to promote event-based research.

5.3. Algorithms. As mentioned in Section 4, event-based data-driven approaches are mainly divided into three categories: spiking neural networks, standard learning architectures, and novel architectures. Inspired by this, improvements of algorithms mainly focus on the three aspects below, namely, event representations, novel architectures, and general frameworks.

Observing algorithms in diverse vision tasks, using deep learning techniques or not, most event-based algorithms comprise two modules: event representation methods and postprocessing inferences. That is, event streams are first summarized using an intermediate representation and then fed to the subsequent algorithms such as a deep neural network or a classic filtering-based structure because of the intrinsic incompatibility between the novel asynchronous event data and traditional frame-based algorithms. Generally, an ideal event representation is supposed to extract comprehensive information from raw event data to ensure accuracy as well as match the input size requirement. In recent research, widely adopted event representations include Surface of Active Events (SAE) [98] and its derivatives, Event frame [66], Event count image [40], Voxel grid [65], and Event Spike Tensor (EST) [6]. However, most of the existing transformations suffer from information compression compared with the raw data. Event representations without information loss are well worth further research.

As novel types of sensors, event cameras transmit per-pixel intensity changes in the form of event streams with each event encoding space-time coordinates and sign information. Basically, existing event-based algorithms either refer to frame-based computer vision techniques or take inspiration from biological systems. However, pure bio-inspired SNN architectures suffer from the vanishing spike phenomenon [76] in deeper layers and the lack of specific hardware. As for standard learning architectures designed for frame-based vision tasks, the reliance of intermediate image-like representation usually leads to the loss of asynchronous property and redundant computation. Efficient long-term solutions specially tailored to event streams are of utmost urgency. To tackle the challenge, recent research studies with novel architectures include a deep hybrid neural network [44] that combines SNNs with ANNs and an efficient processing framework [4] of event-based asynchronous sparse convolutional network, offering valuable reference for the future work.

Apart from local techniques that enhance task-specific performance, event-based processing frameworks available for diverse visual tasks also arouse researchers' enthusiasm owing to their strong generalization and inheritability in relevant studies. Determined by the unique working principle, event data possess several inherent properties: the spatio-temporal sparsity and the observation that events usually encode moving edges of the viewed scene. To some degree, efficient algorithms commonly make the best of events' intrinsic nature based on a concise and essential design concept. For instance, Gehrig et al. [6] presented a general framework that transforms event data into task-specific tensor-like representations in an end-to-end paradigm, suitable for various tasks such as optical flow prediction and object recognition. Gallego et al. [34] also introduced a unifying framework aiming at several estimation problems in event-based vision, obtaining motion parameters that best fit the input data by contrast maximization. Furthermore, the generated motion-corrected edge-like images can serve several downstream tasks such as feature tracking and object recognition. In later research [35], 22 objective functions for unsupervised learning were analyzed to perfect this framework. An efficient processing framework [4] insensitive to event representation, neural network architecture, and task was raised recently, achieving both low latency and high accuracy. Observing the contributions mentioned above, research focusing on general frameworks has a more profound effect over research subject to a specific task, pointing out new opportunities.

6. Conclusion

In challenging scenarios such as resource-constrained platforms, highly reactive systems, or limited illumination conditions, a more efficient and robust way of visual perception is of urgent need. Event-based visual systems including sensing and processing have emerged as promising candidates in applications such as consumer electronics, Internet of Things, industrial automation, and autonomous vehicles. Background knowledge in event-based vision was concisely presented in Section 2, paving the way for further analysis. During the development of event-based vision, the limitation of model-based methods in complex, high-level vision tasks and the great success of deep learning techniques in frame-based vision jointly lead to the predominance of event-based data-driven technology, including data-driven approaches in event-based algorithms, dedicated hardware, and relevant datasets. Reasons for the rise of event-based data-driven technology were concretely analyzed in Section 3. Lessons from frame-based vision include efficient algorithms, dedicated hardware, and large datasets, which also guide the development of event-based vision. In Section 4 of this paper, we focused on the current status of event-based data-driven technology, spanning areas of algorithms (data-driven approaches), datasets, and hardware. During the development of event-based algorithms, existing theories

and techniques from biological systems and frame-based vision provide considerable support. Maintaining the inherent advantages of event-based perception, namely, low power consumption and low latency, stands as a priority on a par with high accuracy in event-based algorithms. Among various event-based algorithms, data-driven approaches is gaining growing prevalence in prospect, considering the preponderance of deep learning techniques and the compatibility of bio-inspired spiking neural networks with event-based sensors. With two main categories of existing algorithms both suffering from intrinsic limitations, efficient long-term solutions specially tailored to event streams to balance the trade-off between efficiency and accuracy are of utmost urgency. Several recent research studies with novel architectures have been proposed. To circumvent the contradiction between the lack of event-based sensors and the huge demand for event data driven by learning algorithms, growing numbers of datasets and simulators have been proposed to facilitate the development of event-based algorithms, offering quantitative benchmarks for performance evaluation. To advance the holistic performance of an event-based visual system, optimization of hardware in terms of sensors and processors also plays a vital role. Last but not least, event-based research is still in its infancy compared with frame-based computer vision, leaving considerable room for future improvement. Several opportunities were pointed out in Section 5, spanning hardware, datasets' and algorithms' perspectives.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] 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 Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 259–275, 2011.

[2] C. Brandli, R. Berner, M. Minhao Yang, S. Shih-Chii Liu, and T. Delbrück, "A 240×180 130 dB 3 μs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014.

[3] G. Gallego, T. Delbrück, G. Orchard et al., "Event-based vision: a survey," *CoRR*, https://arxiv.org/abs/1904.08405, 2019.

[4] N. Messikommer, D. Gehrig, A. Loquercio, and D. Scaramuzza, "Event-based asynchronous sparse convolutional networks," in *Proceedings of the Computer Vision-ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part VIII, Vol. 12353 of Lecture Notes in Computer Science*, A. Vedaldi, H. Bischof, T. Brox, and J. Frahm, Eds., pp. 415–431, Springer, 2020.

[5] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, R. B. Benosman, and HOTS "HOTS: a hierarchy of event-based time-surfaces for pattern recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 7, pp. 1346–1359, 2017.

[6] D. Gehrig, A. Loquercio, K. G. Derpanis, and D. Scaramuzza, "End-to-end learning of representations for asynchronous event-based data," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019*, pp. 5632–5642, IEEE, Seoul, Korea (South), October-November 2019.

[7] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Ev-flownet: self-supervised optical flow estimation for event-based cameras," in *Proceedings of the Robotics: Science and Systems XIV*, H. Kress-Gazit, S. S. Srinivasa, T. Howard, and N. Atanasov, Eds., Carnegie Mellon University, Pittsburgh, PA, USA, June 2018.

[8] S. B. Furber, D. R. Lester, L. A. Plana et al., "Overview of the spinnaker system architecture," *IEEE Transactions on Computers*, vol. 62, no. 12, pp. 2454–2467, 2013.

[9] F. Akopyan, J. Sawada, A. Cassidy 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*, vol. 34, no. 10, pp. 1537–1557, 2015.

[10] M. Davies, N. Srinivasa, T.-H. Lin et al., "Loihi: a neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.

[11] J. Binas, D. Neil, S. Liu, and T. Delbrück, "DDD17: end-to-end DAVIS driving dataset," *CoRR*, https://arxiv.org/abs/1711.01458, 2017.

[12] G. Orchard, A. Jayawant, G. Cohen, and N. V. Thakor, "Converting static image datasets to spiking neuromorphic datasets using saccades," *CoRR*, https://arxiv.org/abs/1507.07629, 2015.

[13] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "HATS: histograms of averaged time surfaces for robust event-based object classification," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pp. 1731–1740, IEEE Computer Society, Salt Lake City, UT, USA, June 2018.

[14] S. K. Esser, P. A. Merolla, J. V. Arthur et al., "Convolutional networks for fast, energy-efficient neuromorphic computing," *Proceedings of the National Academy of Sciences*, vol. 113, no. 41, pp. 11441–11446, 2016.

[15] D. P. Moeys, F. Corradi, C. Li et al., "A sensitive dynamic and active pixel vision sensor for color or neural imaging applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 1, pp. 123–136, 2018.

[16] V. Sitzmann, S. Diamond, Y. Peng et al., "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," *ACM Trans. Graph.*vol. 37, no. 4, pp. 1–114, 2018.

[17] J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," in *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019*, pp. 10192–10201, IEEE, Seoul, Korea (South), October-November 2019.

[18] D. Weikersdorfer, D. B. Adrian, D. Cremers, and J. Conradt, "Event-based 3d SLAM with a depth-augmented dynamic vision sensor," in *2014 IEEE International Conference on Robotics and Automation, ICRA 2014*, pp. 359–364, IEEE, Hong Kong, China, May-June 2014.

[19] E. Mueggler, H. Rebecq, G. Gallego, T. Delbrück, and D. Scaramuzza, "The event-camera dataset and simulator: event-based data for pose estimation, visual odometry, and SLAM," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.

[20] F. Barranco, C. Fermuller, Y. Aloimonos, and T. Delbruck, "A dataset for visual navigation with neuromorphic methods," *Frontiers in Neuroscience*, vol. 10, p. 49, 2016.

[21] J. A. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are we ready for autonomous drone racing? the UZH-FPV drone racing dataset," in *Proceedings of the International Conference on Robotics and Automation, ICRA 2019*, pp. 6713–6719, IEEE, Montreal, QC, Canada, May 2019.

[22] A. Amir, B. Taba, D. J. Berg et al., "A low power, fully event-based gesture recognition system," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 7388–7397, IEEE Computer Society, Honolulu, HI, USA, July 2017.

[23] P. de Tournemire, D. Nitti, E. Perot, D. Migliore, and A. Sironi, "A large scale event-based detection dataset for automotive," *CoRR*, https://arxiv.org/abs/2001.08499, 2020.

[24] H. Rebecq, D. Gehrig, and D. Scaramuzza, "ESIM: an open event camera simulator," in *Proceedings of the 2nd Annual Conference on Robot Learning, CoRL 2018, Vol. 87 of Proceedings of Machine Learning Research*, pp. 969–982, PMLR, Zürich, Switzerland, October 2018.

[25] D. Gehrig, M. Gehrig, J. Hidalgo-Carrió, and D. Scaramuzza, "Video to events: recycling video datasets for event cameras," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pp. 3583–3592, IEEE, Seattle, WA, USA, June 2020.

[26] M. L. Katz, K. Nikolic, and T. Delbrück, "Live demonstration: behavioural emulation of event-based vision sensors," in *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems, ISCAS 2012*, pp. 736–740, IEEE, Seoul, Korea (South), May 2012.

[27] Y. Jiang, S. Yin, and O. Kaynak, "Data-driven monitoring and safety control of industrial cyber-physical systems: basics and beyond," *IEEE Access*, vol. 6, pp. 47374–47384, 2018.

[28] Y. Jiang, S. Yin, J. Dong, and O. Kaynak, "A review on soft sensors for monitoring, control and optimization of industrial processes," *IEEE Sensors Journal*, 2020.

[29] Y. Jiang, S. Yin, and O. Kaynak, "Performance supervised plant-wide process monitoring in industry 4.0: a roadmap," *IEEE Open Journal of the Industrial Electronics Society*, vol. 2, pp. 21–35, 2021.

[30] V. Vasco, A. Glover, and C. Bartolozzi, "Fast event-based harris corner detection exploiting the advantages of event-driven cameras," in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4144–4149, IEEE, 2016.

[31] E. Mueggler, C. Bartolozzi, and D. Scaramuzza, "Fast event-based corner detection," in *Proceedings of the British Machine Vision Conference 2017, BMVC 2017*, BMVA Press, London, UK, September 2017.

[32] M. Liu and T. Delbrück, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors," in *Proceedings of the British Machine Vision Conference 2018, BMVC 2018*, p. 88, BMVA Press, Newcastle, UK, September 2018.

[33] R. Li, D. Shi, Y. Zhang, K. Li, and R. Li, "Fa-harris: a fast and asynchronous corner detector for event cameras," in *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2019*, pp. 6223–6229, IEEE, Macau, SAR, China, November 2019.

[34] G. Gallego, H. Rebecq, and D. Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pp. 3867–3876, IEEE Computer Society, Salt Lake City, UT, USA, June 2018.

[35] G. Gallego, M. Gehrig, and D. Scaramuzza, "Focus is all you need: loss functions for event-based vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pp. 12280–12289, Computer Vision Foundation / IEEE, Long Beach, CA, USA, June 2019.

[36] A. Z. Zhu, N. Atanasov, and K. Daniilidis, "Event-based feature tracking with probabilistic data association," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation, ICRA 2017*, pp. 4465–4470, IEEE, Singapore, Singapore, May-June 2017.

[37] A. Z. Zhu, N. Atanasov, and K. Daniilidis, "Event-based visual inertial odometry," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 5816–5824, IEEE Computer Society, Honolulu, HI, USA, July 2017.

[38] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Ultimate slam? combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 994–1001, 2018.

[39] J. Lee, T. Delbrück, and M. Pfeiffer, "Training deep spiking neural networks using backpropagation," *CoRR*, https://arxiv.org/abs/1608.08782, 2016.

[40] A. I. Maqueda, A. Loquercio, G. Gallego, N. García, and D. Scaramuzza, "Event-based vision meets deep learning on steering prediction for self-driving cars," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pp. 5419–5427, IEEE Computer Society, Salt Lake City, UT, USA, June 2018.

[41] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *CoRR*, https://arxiv.org/abs/1906.07165, 2019.

[42] J. A. Pérez-Carrasco, B. Bo Zhao, C. Serrano et al., "Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing--application to feedforward ConvNets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2706–2719, 2013.

[43] Y. Sekikawa, K. Hara, and H. Saito, "Eventnet: asynchronous recursive event processing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pp. 3887–3896, Computer Vision Foundation / IEEE, Long Beach, CA, USA, June 2019.

[44] C. Lee, A. Kosta, A. Z. Zhu, K. Chaney, K. Daniilidis, and K. Roy, "Spike-flownet: event-based optical flow estimation with energy-efficient hybrid neural networks," in *Proceedings of the Computer Vision-ECCV 2020—16th European Conference, Part XXIX, Vol. 12374 of Lecture Notes in Computer Science*, A. Vedaldi, H. Bischof, T. Brox, and J. Frahm, Eds., Springer, Glasgow, UK, pp. 366–382, August 2020.

[45] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997.

[46] G. Orchard, R. Benosman, R. Etienne-Cummings, and N. V. Thakor, "A spiking neural network architecture for visual motion estimation," in *Proceedings of the 2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 298–301, IEEE, Rotterdam, The Netherlands, October-November 2013.

[47] G. Haessig, A. Cassidy, R. Alvarez, R. Benosman, and G. Orchard, "Spiking optical flow for event-based sensors using IBM's TrueNorth neurosynaptic system," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 4, pp. 860–870, 2018.

[48] F. Paredes-Vallés, K. Y. W. Scheper, and G. C. H. E. de Croon, "Unsupervised learning of a hierarchical spiking neural

network for optical flow estimation: from events to global motion perception," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2051–2064, 2020.

[49] G. Orchard, C. Meyer, R. Etienne-Cummings et al., "HFirst: a temporal approach to object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2028–2040, 2015.

[50] 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 Trans. Neural Networks Learn. Syst.*vol. 26, no. 9, p. 1963, 1978.

[51] J. Acharya, V. Padala, and A. Basu, "Spiking neural network based region proposal networks for neuromorphic vision sensors," in *Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS 2019*, pp. 1–5, IEEE, Sapporo, Japan, May 2019.

[52] M. Osswald, S.-H. Ieng, R. Benosman, and G. Indiveri, "A spiking neural network model of 3d perception for event-based neuromorphic stereo vision systems," *Scientific Reports*, vol. 7, p. 40703, 2017.

[53] G. Haessig, X. Berthelon, S.-H. Ieng, and R. Benosman, "A spiking neural network model of depth from defocus for event-based neuromorphic vision," *Scientific Reports*, vol. 9, no. 1, pp. 1–11, 2019.

[54] Z. Bing, C. Meschede, K. Huang et al., "End to end learning of spiking neural network based on R-STDP for a lane keeping vehicle," in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation, ICRA 2018*, pp. 1–8, IEEE, Brisbane, Australia, May 2018.

[55] H. Akolkar, C. Meyer, X. Clady et al., "What can neuromorphic event-driven precise timing add to spike-based pattern recognition?," *Neural Computation* vol. 27, no. 3, pp. 561–593, 2015.

[56] P. U. Diehl and M. Cook, "Unsupervised learning of digit recognition using spike-timing-dependent plasticity," *Frontiers Comput. Neurosci.*vol. 9, p. 99, 2015.

[57] P. U. Diehl, D. Neil, J. Binas, M. Cook, S. Liu, and M. Pfeiffer, "Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing," in *Proceedings of the 2015 International Joint Conference on Neural Networks, IJCNN 2015*, pp. 1–8, IEEE, Killarney, Ireland, July 2015.

[58] M. Gehrig, S. B. Shrestha, D. Mouritzen, and D. Scaramuzza, "Event-based angular velocity regression with spiking networks," in *Proceedings of the 2020 IEEE International Conference on Robotics and Automation, ICRA 2020*, pp. 4195–4202, IEEE, Paris, France, May-August 2020.

[59] B. Schrauwen and J. Van Campenhout, "Improving spikeprop: enhancements to an error-backpropagation rule for spiking neural networks," in *Proceedings of the 15th ProRISC workshop*, vol. 11, pp. 301–305, 2004.

[60] S. B. Shrestha and Q. Song, "Event based weight update for learning infinite spike train," in *Proceedings of the 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, pp. 333–338, IEEE Computer Society, Anaheim, CA, USA, December 2016.

[61] F. Zenke and S. Ganguli, "Superspike: supervised learning in multilayer spiking neural networks," *Neural Comput*, vol. 30, no. 6, 2018.

[62] S. B. Shrestha and G. Orchard, "SLAYER: spike layer error reassignment in time," in *Proceedings of the Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018*, S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., pp. 1419–1428pp. 1419–, Montréal, Canada, December 2018.

[63] H. Rebecq, T. Horstschaefer, G. Gallego, and D. Scaramuzza, "EVO: a geometric approach to event-based 6-DOF parallel tracking and mapping in real time," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 593–600, 2017.

[64] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: bringing modern computer vision to event cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pp. 3857–3866, Computer Vision Foundation/IEEE, Long Beach, CA, USA, June 2019.

[65] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pp. 989–997, Computer Vision Foundation/IEEE, Long Beach, CA, USA, June 2019.

[66] H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization," in *Proceedings of the British Machine Vision Conference 2017, BMVC 2017*, BMVA Press, London, UK, September 2017.

[67] A. Mitrokhin, C. Ye, C. Fermüller, Y. Aloimonos, and T. Delbrück, "EV-IMO: motion segmentation dataset and learning pipeline for event cameras," in *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2019*, pp. 6105–6112, IEEE, Macau, SAR, China, November 2019.

[68] I. Alonso and A. C. Murillo, "Ev-segnet: semantic segmentation for event-based cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019*, pp. 1624–1633, Computer Vision Foundation/IEEE, Long Beach, CA, USA, June 2019.

[69] C. Scheerlinck, H. Rebecq, D. Gehrig, N. Barnes, R. E. Mahony, and D. Scaramuzza, "Fast image reconstruction with an event camera," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, pp. 156–163, IEEE, Snowmass Village, CO, USA, March 2020.

[70] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proceedings of the Computer Vision-ECCV 2016—14th European Conference, Proceedings, Part VIII, Vol. 9912 of Lecture Notes in Computer Science*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Springer, Amsterdam, The Netherlands, pp. 483–499, October 2016.

[71] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference, Part III, vol. 9351 of Lecture Notes in Computer Science*, N. Navab, J. Hornegger, and A. F. Frangi, Eds., Springer, Munich, Germany, pp. 234–241, October 2015.

[72] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 1800–1807, IEEE Computer Society, Honolulu, HI, USA, July 2017.

[73] I. Alonso, A. B. Cambra, A. Muñoz, T. Treibitz, and A. C. Murillo, "Coral-segmentation: training dense labeling models with sparse ground truth," in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2017*, pp. 2874–2882, IEEE Computer Society, Venice, Italy, October 2017.

[74] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proceedings of the IEEE International Conference on Computer Vision, ICCV 2017*, pp. 843–852, IEEE Computer Society, Venice, Italy, October 2017.

[75] N. F. Y. Chen, "Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018*, pp. 644–653, IEEE Computer Society, Salt Lake City, UT, USA, June 2018.

[76] P. Panda, S. A. Aketi, and K. Roy, "Towards scalable, efficient and accurate deep spiking neural networks with backward residual connections, stochastic softmax and hybridization," *CoRR*, https://arxiv.org/abs/1910.13931, 2019.

[77] B. Rueckauer and T. Delbruck, "Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor," *Frontiers in Neuroscience*, vol. 10, p. 176, 2016.

[78] C. Scheerlinck, H. Rebecq, T. Stoffregen, N. Barnes, R. E. Mahony, and D. Scaramuzza, "CED: color event camera dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019*, pp. 1684–1693, Computer Vision Foundation / IEEE, Long Beach, CA, USA, June 2019.

[79] T. Stoffregen, C. Scheerlinck, D. Scaramuzza et al., "How to train your event camera neural network," *CoRR*, https://arxiv.org/abs/2003.09078, 2020.

[80] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: an event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018.

[81] A. Mitrokhin, C. Fermüller, C. Parameshwara, and Y. Aloimonos, "Event-based moving object detection and tracking," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2018*, pp. 1–9, IEEE, Madrid, Spain, October 2018.

[82] S. Miao, G. Chen, X. Ning et al., "Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection," *Frontiers Neurorobotics*, vol. 13, p. 38, 2019.

[83] C. Scheerlinck, N. Barnes, and R. E. Mahony, "Continuous-time intensity estimation using event cameras," in *Proceedings of the Computer Vision-ACCV 2018—14th Asian Conference on Computer Vision, Revised Selected Papers, Part V. Vol. 11365 of Lecture Notes in Computer Science*, C. V. Jawahar, H. Li, G. Mori, and K. Schindler, Eds., Springer, Perth, Australia, pp. 308–324, December 2018.

[84] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1425–1440, 2018.

[85] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, "Asynchronous, photometric feature tracking using events and frames," in *Proceedings of the Computer Vision-ECCV 2018—15th European Conference, Part XII, Vol. 11216 of Lecture Notes in Computer Science*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Springer, Munich, Germany, pp. 766–781, September 2018.

[86] S. Leung, E. J. Shamwell, C. Maxey, and W. D. Nothwang, "Toward a large-scale multimodal event-based dataset for neuromorphic deep learning applications," in *Micro-and Nanotechnology Sensors, Systems, and Applications* Xvol. 10639, p. 106391T, International Society for Optics and Photonics, 2018.

[87] E. Calabrese, G. Taverni, C. A. Easthope et al., "DHP19: dynamic vision sensor 3d human pose dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019*, pp. 1695–1704, Computer Vision Foundation/IEEE, Long Beach, CA, USA, June 2019.

[88] W. Li, S. Saeedi, J. McCormac et al., "Interiornet: mega-scale multi-sensor photo-realistic indoor scenes dataset," in *Proceedings of the British Machine Vision Conference 2018, BMVC 2018*, p. 77, BMVA Press, Newcastle, UK, September 2018.

[89] T. Delbrück, Y. Hu, and Z. He, "V2E: from video frames to realistic DVS event camera streams," *CoRR*, https://arxiv.org/abs/2006.07722, 2020.

[90] H. Jiang, D. Sun, V. Jampani, M. Yang, E. G. Learned-Miller, and J. Kautz, "Super slomo: high quality estimation of multiple intermediate frames for video interpolation," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pp. 9000–9008, IEEE Computer Society, Salt Lake City, UT, USA, June 2018.

[91] R. W. Baldwin, M. Almatrafi, V. K. Asari, and K. Hirakawa, "Event probability mask (EPM) and event denoising convolutional neural network (edncnn) for neuromorphic cameras," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pp. 1698–1707, IEEE, Seattle, WA, USA, June 2020.

[92] M. Guo, J. Huang, and S. Chen, "Live demonstration: a 768×640 pixels 200 meps dynamic vision sensor," in *Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS 2017*, p. 1, IEEE, Baltimore, MD, USA, May 2017.

[93] P. Lichtsteiner, C. Posch, and T. Delbrück, "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[94] B. Son, Y. Suh, S. Kim et al., "4.1 A 640×480 dynamic vision sensor with a 9 μm pixel and 300meps address-event representation," in *Proceedings of the 2017 IEEE International Solid-State Circuits Conference, ISSCC 2017*, pp. 66-67, IEEE, San Francisco, CA, USA, February 2017.

[95] C. Posch, D. Matolin, and R. Wohlgenannt, "An asynchronous time-based image sensor," in *Proceedings of the International Symposium on Circuits and Systems (ISCAS 2008)*, pp. 2130–2133, IEEE, Seattle, WA, USA, May 2008.

[96] J. Huang, M. Guo, and S. Chen, "A dynamic vision sensor with direct logarithmic output and full-frame picture-on-demand," in *Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS 2017*, pp. 1–4, IEEE, Baltimore, MD, USA, May 2017.

[97] S. Chen and M. Guo, "Live demonstration: celex-v: a 1m pixel multi-mode event-based sensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019*, pp. 1682-1683, Computer Vision Foundation / IEEE, Long Beach, CA, USA, June 2019.

[98] R. Benosman, C. Clercq, X. Lagorce, S. Sio-Hoi Ieng, and C. Bartolozzi, "Event-based visual flow," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 407–417, 2014.