Movement Classification and Segmentation Using Event-Based Sensing and Spiking Neural Networks

Paul Kirkland and Gaetano Di Caterina

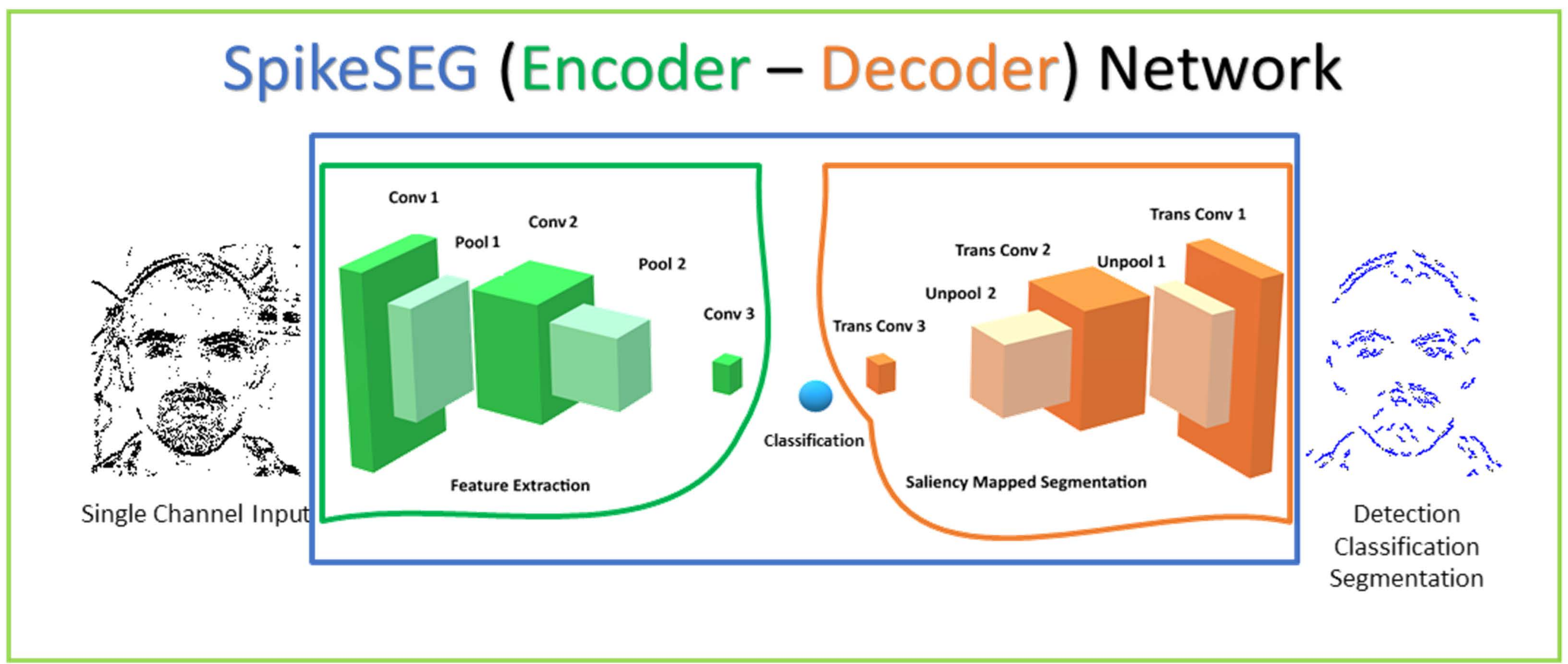
Neuromorphic Sensor Signal Processing Lab,
Department of Electronic and Electrical Engineering, University of Strathclyde, Scotland

Neuromorphic Sensor Signal Processing Lab

University of Strathclyde Engineering

Introduction

The development of Spiking Neural Networks (SNN) and the discipline of Neuromorphic Engineering has resulted in a paradigm shift in how Machine Learning (ML) and Computer Vision (CV) problems are approached. At the heart of this shift is the adoption of event-based sensing and processing methods. The production of sparse and asynchronous events that are dynamically connected to the scene is possible with an event- based vision sensor, allowing for the acquisition of not just spatial data but also high-fidelity temporal data. In this work, we describe a novel method for performing instance segmentation of objects, only using their spatio-temporal movement patterns, by utilising the weights of an unsupervised Spiking Convolutional Neural Network that was originally trained for object recognition and extending it to instance segmentation. This takes advantage of the network's spatial and temporal characteristics encoded within its internal feature representation, to offer this additional discriminative ability. Exploiting this new paradigm allows for an efficient method of segmenting track paths from moving objects. It does so by using the continual data input to accumulate information over time, and as soon as it matches a known feature the passes the information forward in the system. This system could be utilised in two ways, either to segment known movement patterns for identification, or to mask these known movements to help declutter the scene for unknown movement patterns to be recognised easier.

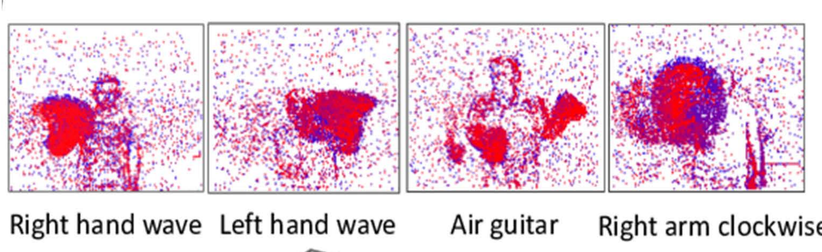


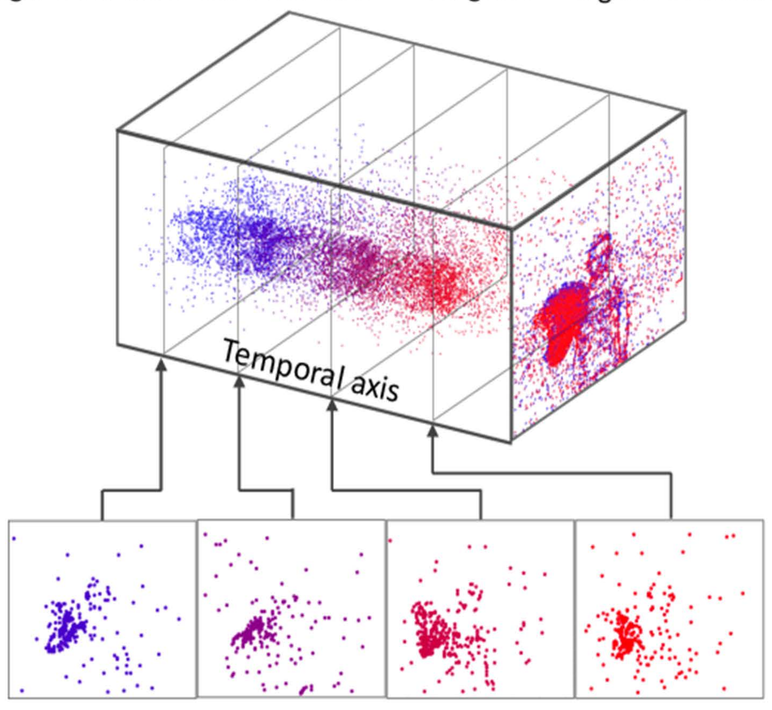
What Is Neuromorphic / Event-Based Vision?

An event camera, also known as a neuromorphic camera, silicon retina or dynamic vision sensor, is an imaging sensor that responds to local changes in brightness. Event cameras do not capture images using a shutter as conventional (frame) cameras do. Instead, each pixel inside an event camera operates independently and asynchronously, reporting changes in brightness as they occur, and staying silent otherwise. This is depicted within the three images showing normal motion and the frame output compared to the event output. Then fast motion with motion blur in the frames and the events capture the motion smoothly. Lastly the no motion example shows how if no changes are detected then no there is no output from the camera.

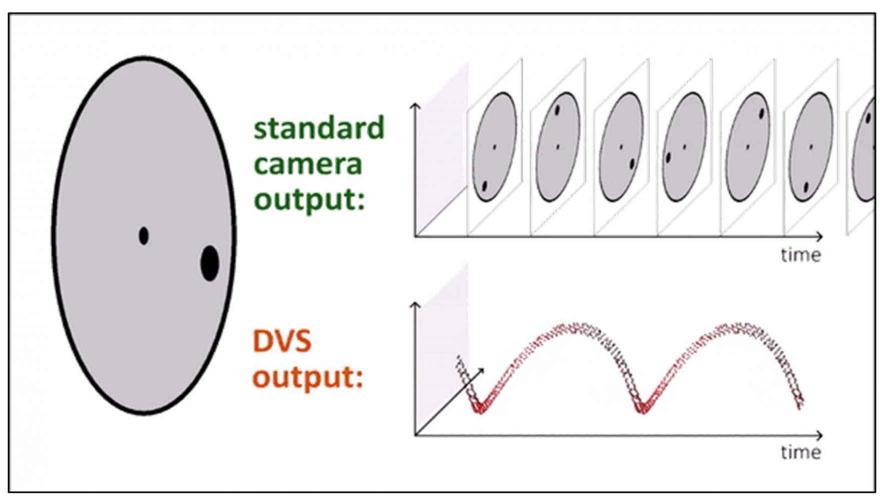
Event cameras offer attractive properties compared to traditional cameras: high temporal resolution (in the order of µs), very high dynamic range (140 dB), low power consumption, and high pixel bandwidth (on the order of kHz) resulting in reduced motion blur. Hence, event cameras have a large potential for robotics and computer vision in challenging scenarios for traditional cameras, such as low-latency, high speed, and high dynamic range.

The sensor has the innate ability to be able to capture temporal information in a more useful manner. However with the shift in sensing ability also comes the requirement to change the processing approach in order to take full advantage of the benefits provided by this sensor.



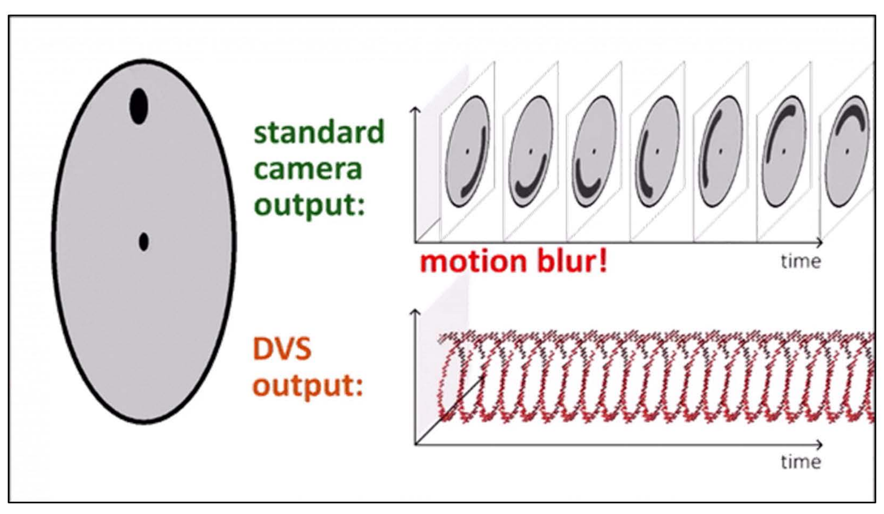


Normal Motion



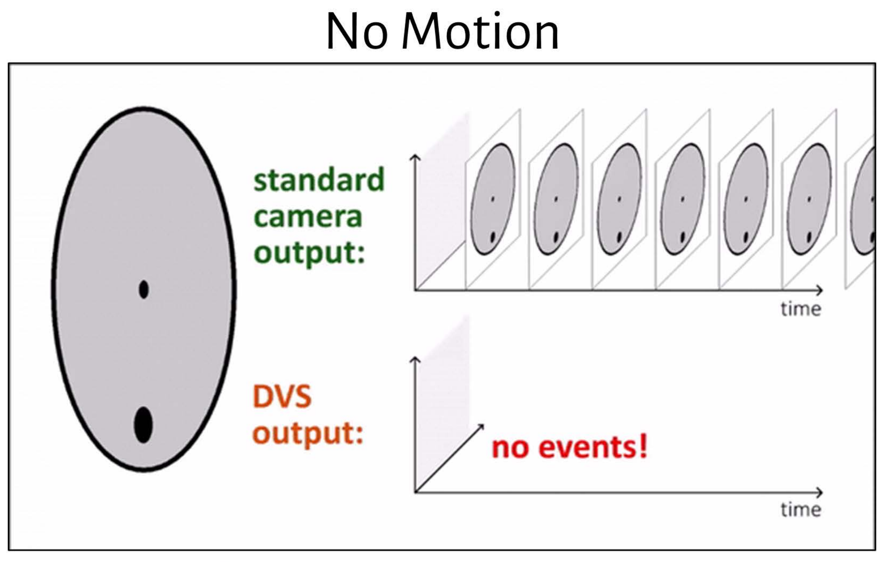
Sparse, high temporal resolution output of elements that have changed

Fast Motion



Dynamic relationship with the scene, more change = more output events, no fixed sampling rate

No Motion



No change, no output, no processing

What Are Spiking Neural Networks?

Artificial neural networks that closely mimic natural neural networks are known as Spiking Neural Networks (SNNs) and are seen as the third generation of neural networks. In addition to neuronal and synaptic status, SNNs incorporate time into their working model. The idea is that neurons in the SNN do not transmit information at the end of each propagation cycle (as they do in traditional ANNs), but only when a membrane potential – a neuron's intrinsic quality related to its membrane potential – reaches a certain value, known as the threshold. The neuron fires when the membrane potential hits the threshold, sending a signal to neighbouring neurons, which increases or decreases their potential in response to the signal. A spiking neuron model is a neuron model that fires at the moment of threshold crossing. What distinguishes a traditional ANN from an SNN is the information propagation approach. SNN aspires to be as close to a biological neural network as feasible. As a result, rather than working with continually changing time values as ANN does, SNN works with discrete events that happen at defined times. SNN takes a set of spikes as input and produces a set of spikes as output (a series of spikes is usually referred to as spike trains).

Advantages:

SNNs are a dynamical system. Ideal for dynamic processing tasks like speech and video.

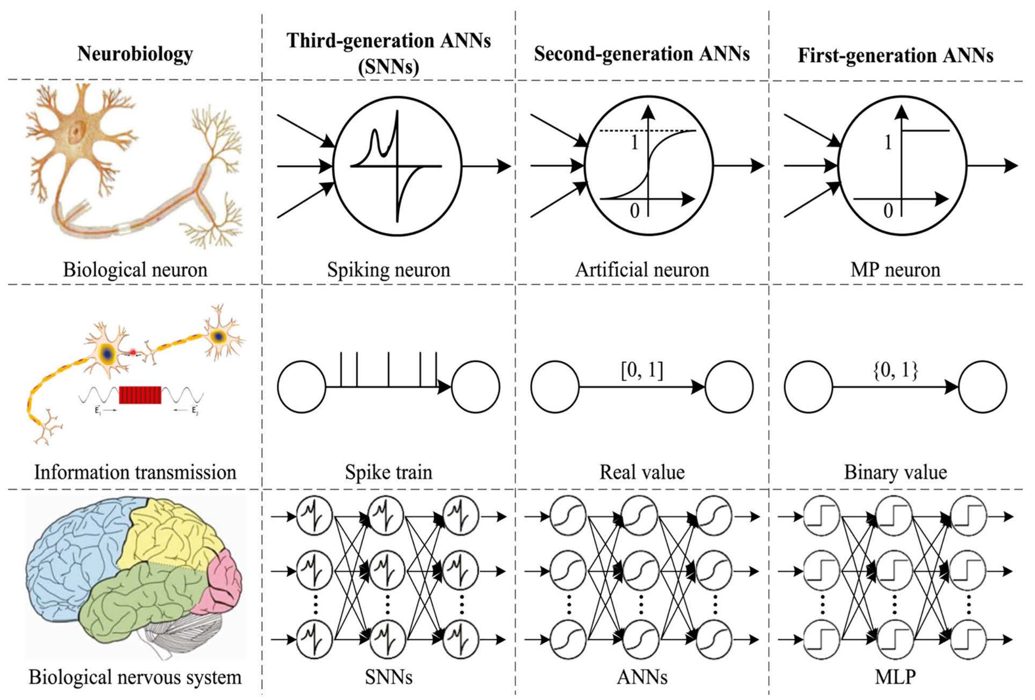
An SNN can perform online/continual learning due to unsupervised learning.

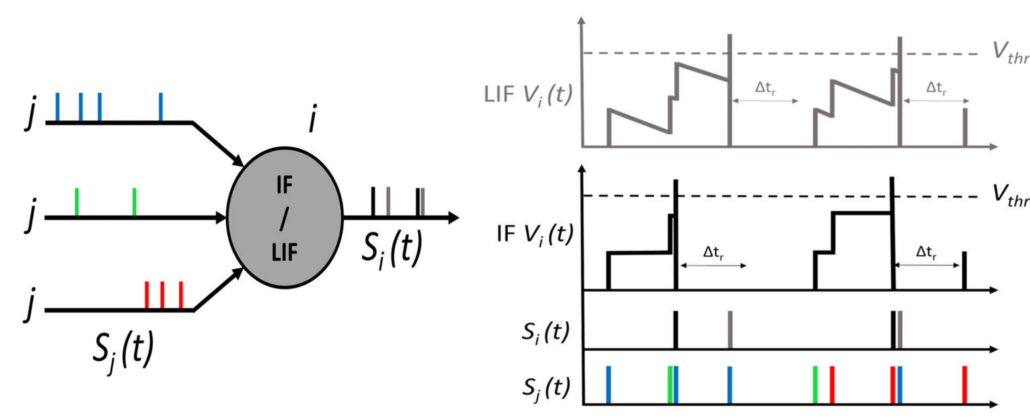
SNNs have a lower throughput latency due to the binary communication and asynchronous nature

Because they leverage the temporal presentation of information, SNNs have boosted information processing productivity and noise immunity.

SNNs naturally accumulate information over time without the need for recurrence, due to the integrating nature of the neurons.

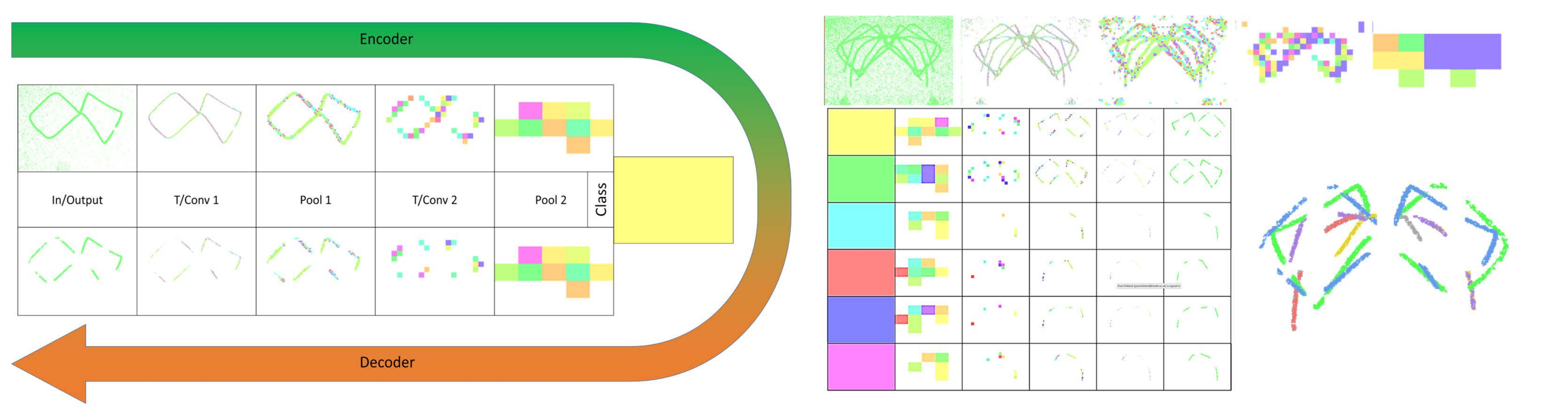
SNNs can leverage much of the research of ANNs from training, architectures, and hyperparameter tuning techniques.



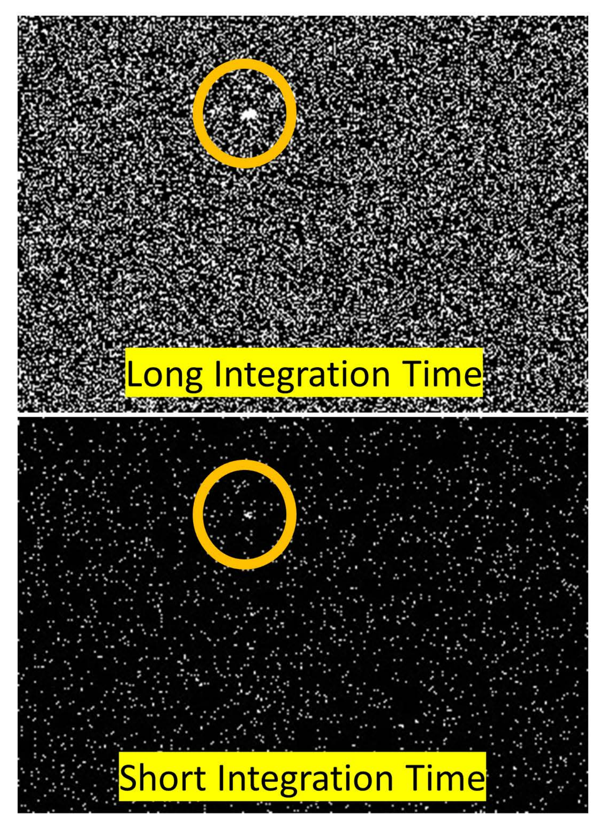


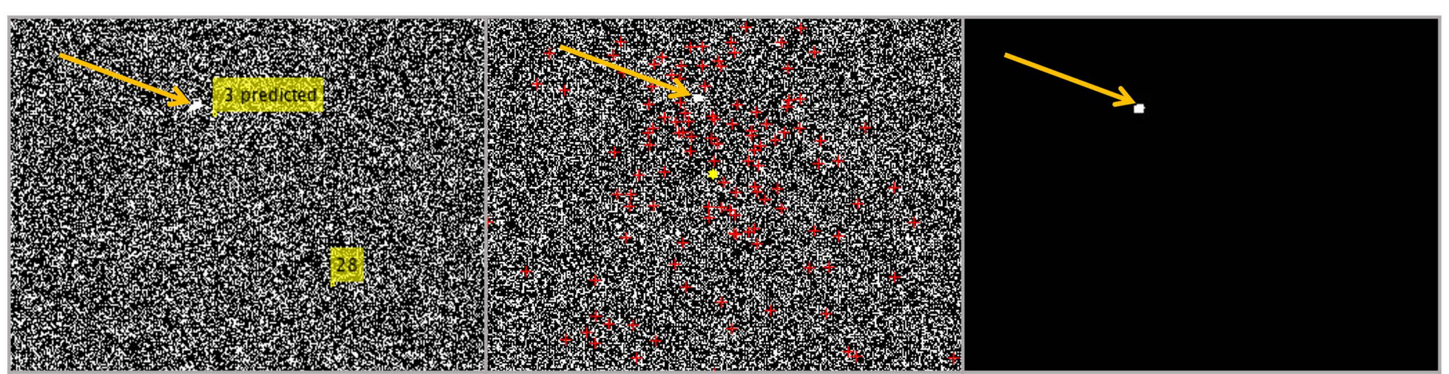
Movement Segmentation and Application in Track before Detect

Once it was established that featural-temporal information could be extracted from the spatial features of the spiking event data, the next step was to test the feature extraction ability on spatio-temporal information. As such, the spatial information alone is not representative of anything meaningful, so a longer integration period is required to ascertain if there. This is a temporal component of the spatial information presented. This was tested under the assumption of an unknown object (small dot) completing a set number of movement patterns



In the situation of a low signal-to-noise ratio (SNR). This problem is highly related to the principle behind Track before Detect (TBD), as detection is based on tracking or accumulating information on any objects of interest within a scene. However, the time scales required for movement detection are far shorter than that required in the previous classification task. Regardless, it became clear that neuromorphic sensing and processing could be utilised to great effect in the more challenging TBD domain. The neuromorphic event-based sensor allows for the accumulation of spatio-temporal information on higher fidelity and variable/incoherent scale, due to its high temporal resolution and asynchronous readout. This means the sensor can accumulate small amounts of time to detect pixel motion while mitigating the effects of the sensor noise and clutter. When a crude version of the SNN is compared with equally crude Kalman Filters and Particle Filters. The SNN provides a compelling argument for further development, as it can exploit this non-linear relationship between noise and signal to its advantage as shown in image below.





Kalman Filter	Particle Filter	Spiking Neural Network

Augmented Reality Poster Scan to Unlock