

PAPER • OPEN ACCESS

VCSEL-based photonic spiking neural networks for ultrafast detection and tracking

To cite this article: Joshua Robertson *et al* 2024 *Neuromorph. Comput. Eng.* **4** 014010

View the [article online](#) for updates and enhancements.

You may also like

- [Brain-inspired nanophotonic spike computing: challenges and prospects](#)
Bruno Romeira, Ricardo Adão, Jana B Nieder et al.
- [A spiking central pattern generator for the control of a simulated lamprey robot running on SpiNNaker and Loihi neuromorphic boards](#)
Emmanouil Angelidis, Emanuel Buchholz, Jonathan Arreguit et al.
- [Exploiting deep learning accelerators for neuromorphic workloads](#)
Pao-Sheng Vincent Sun, Alexander Titterton, Anjlee Gopiani et al.



PAPER

VCSEL-based photonic spiking neural networks for ultrafast detection and tracking

OPEN ACCESS

RECEIVED

3 November 2023

REVISED

2 February 2024

ACCEPTED FOR PUBLICATION

27 February 2024

PUBLISHED

13 March 2024

Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Joshua Robertson^{1,*} , Paul Kirkland², Gaetano Di Caterina² and Antonio Hurtado¹¹ Institute of Photonics, SUPA Department of Physics, University of Strathclyde, Glasgow, United Kingdom² Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, United Kingdom

* Author to whom any correspondence should be addressed.

E-mail: joshua.robertson@strath.ac.uk**Keywords:** neuromorphic computing, semiconductor laser, spiking neural networks, image processing, detection and trackingSupplementary material for this article is available [online](#)**Abstract**

Inspired by efficient biological spike-based neural networks, we demonstrate for the first time the detection and tracking of target patterns in image and video inputs at high-speed rates with networks of multiple artificial spiking optical neurons. Using photonic systems of in-parallel spiking vertical cavity surface emitting lasers (VCSELs), we demonstrate the implementation of multiple convolutional kernel operators which, in combination with optical spike signalling, enable the detection and tracking of target features in images/video feeds at an ultrafast photonic operation speed of 1 ns per pixel. Alongside a single layer optical spiking neural network (SNN) demonstration, a multi-layer network of photonic (GHz-rate) spike-firing neurons is reported where the photonic system successfully tracks a large complex feature (Handwritten Digit 3). The consecutive photonic layers perform spike-enabled image reduction and convolution operations, and interact with a software-implemented SNN, that learns the feature patterns that best identify the target to provide a high detection efficiency even in the presence of a distractor feature. This work therefore highlights the effectiveness of combining neuromorphic photonic hardware and software SNNs, for efficient learning and ultrafast operation, thanks to the use of spiking light signals, towards tackling complex AI and computer vision problems.

1. Introduction

Photonics, the light-based technology platform, is seeing increasing investigation in the field of neuromorphic (brain-inspired) computing, where optical systems are being proposed as neural network alternatives to traditional electronic computing architectures. Utilizing light instead of electrical signals, photonic systems can operate with several benefits that may help overcome the long-term limitations of electronic processors posed by fundamental effects such as joule heating, parasitic capacitance and crosstalk [1]. With higher bandwidth and additional degrees of freedom (such as polarisation, wavelength) for parallel and scalable processing, as well as higher signal speeds, photonics has the potential to realise faster and more computationally-efficient hardware systems, that are increasingly more suitable for complex neural network problems. An exponential growth in the interest of light-enabled hardware for the acceleration of neuromorphic systems has occurred in recent years (see review articles [2–4]), and overall the literature incorporates a diverse set of techniques and devices. phase-change materials [5], optical modulators [6, 7], and semiconductor lasers [8–11], are a few highlight photonic devices that have achieved neuromorphic or neuromimetic operation. Similarly, electro-optic technologies such quantum resonant tunnelling structures [12, 13], amongst others (see [14] for a review), have now also been proposed as suitable device platforms for brain-inspired computing hardware components. With one of the major benefits of photonics being the highly-reduced energy requirements, the recent demonstrations of photonic artificial neural networks (pANNs) [15–19] and reservoir computing systems [20–23] make an exciting case for photonics being the platform at the forefront of potentially sustainable neuromorphic computing. Importantly, the highly

desirable energy efficiency of pANN systems is believed to be critical to the continued development of the currently booming Artificial Intelligence (AI) revolution. Recent studies show exhaustive amounts of energy are used to train and operate AI models [24], with the large amounts of data currently being processed; therefore, given the inevitable further adoption of AI in everyday applications, the real demand for fast and efficient neuromorphic processors will soon arrive.

One type of artificial neural network structure that has even more promising energy efficient operation are spiking neural networks (SNNs). SNNs, that more closely imitate the networks of biological neurons in the brain, operate with all-or-nothing spiking responses that are digital in amplitude but analog in time. SNNs consequently follow a slight different set of rules to other neural network architectures, granting them efficient low power operation (due to the inactivity of large parts of the network during computation [25]) and alternative learning/training approaches, such as spike-timing dependant plasticity (STDP) [26]. Reports of SNNs have therefore shown that lower operational energy requirements and lower latency are possible with networks of all-or-nothing responses, when compared to larger convolutional neural networks (CNNs). Moreover, like CNNs, image processing tasks such as the recognition, detection and classification of image features and patterns, are well suited to networks of artificial spiking nodes. For example, software-implemented neuromorphic SNNs have been used to develop multiple layer hierarchical convolution operations to extract image features patterns in images for classification purposes [27]. Further, software-implemented SNNs, and the sparsity of their spiking representation, have also been shown to pair especially well with hardware sensors such as single photon detectors [28] and event-based cameras [29], providing image processing functionalities with greater reductions in computational processing. The realisation of SNNs with the photonic platform therefore holds exciting prospects for neuromorphic image processing systems with low computational requirements, light-speed operation, and potentially low power requirements.

In this work, driven by the inherent beneficial attributes of photonic technologies (e.g. high speed, low cross talk, high parallelism), we investigate light-enabled spike processing systems for key image/video processing tasks, namely the detection and tracking of patterns in image/video inputs at ultrafast photonic processing rates (100/1000s of nanoseconds per image/frame). Here we define the photonic operation speed as the time for a photonic layer to process 1 pixel, and the photonic processing rate as the time for the photonic layer to process an entire frame/image. Our neuromorphic photonic system is based on VCSELs, ubiquitous devices commonly found in supermarket barcode scanners, automotive sensors, mobile phones, and telecommunication data centres, that exhibit optical spiking responses analogous to those of biological neurons in the brain, but at up to 9 orders of magnitude faster (1 billion times) [30–32]. These interesting properties make VCSELs ideal candidates for photonic-based SNN processing systems, and have shown exciting neuromorphic information processing capabilities [11], which are of key interest to strategic sectors such as defence, energy and healthcare. In this work, we combine spike processing techniques with artificial photonic VCSEL-neurons to process (using ultrafast optical spikes) a wide range of image inputs, to detect the locations of complex patterns/targets and track their location as they move across consecutive frames. The VCSEL-based systems in this work have previously successfully demonstrated the capability to extract edge features in images at ultrafast rates (down to nanoseconds per pixel operation) [33], and even work in conjunction with software-implemented SNNs to provide image classifications, through their operation as leaky integrate-and-fire neurons [34]. This work goes beyond these previous realisations, by implementing for the first time larger convolutional kernels using multiple artificial VCSEL photonic neurons for processing with a parallel layer of photonic spiking elements. By using larger kernels operators we look to process images with multiple layers of spiking elements to perform consecutive kernel operations (dimensionality reduction and convolution), to form hardware network layers capable of feature extraction, target detection and tracking. Importantly, this demonstration will highlight the high photonic operation speed and processing rate of parallel VCSEL-based spiking layers, the simple hardware friendly implementation of these photonic SNN layers, and the growing potential of VCSELs as future spiking pANNs devices for neuromorphic processing, computer vision and AI systems.

The remainder of this paper is organised as follows. Section 2 describes the hardware and software experimental setup implemented in this work, while section 3 presents the experimental results from this work, accompanied by an extensive discussion of such findings. Finally, section 4 draws final conclusions about the work and suggests future research directions.

2. Experimental setup

2.1. Photonic spiking neural network (PSNN)

Figure 1(a) shows the experimental setup used to realise the layers of spiking VCSELs. The photonic setup uses off-the-shelf fibre-optic telecom components and contains two parallel optical injection lines and two

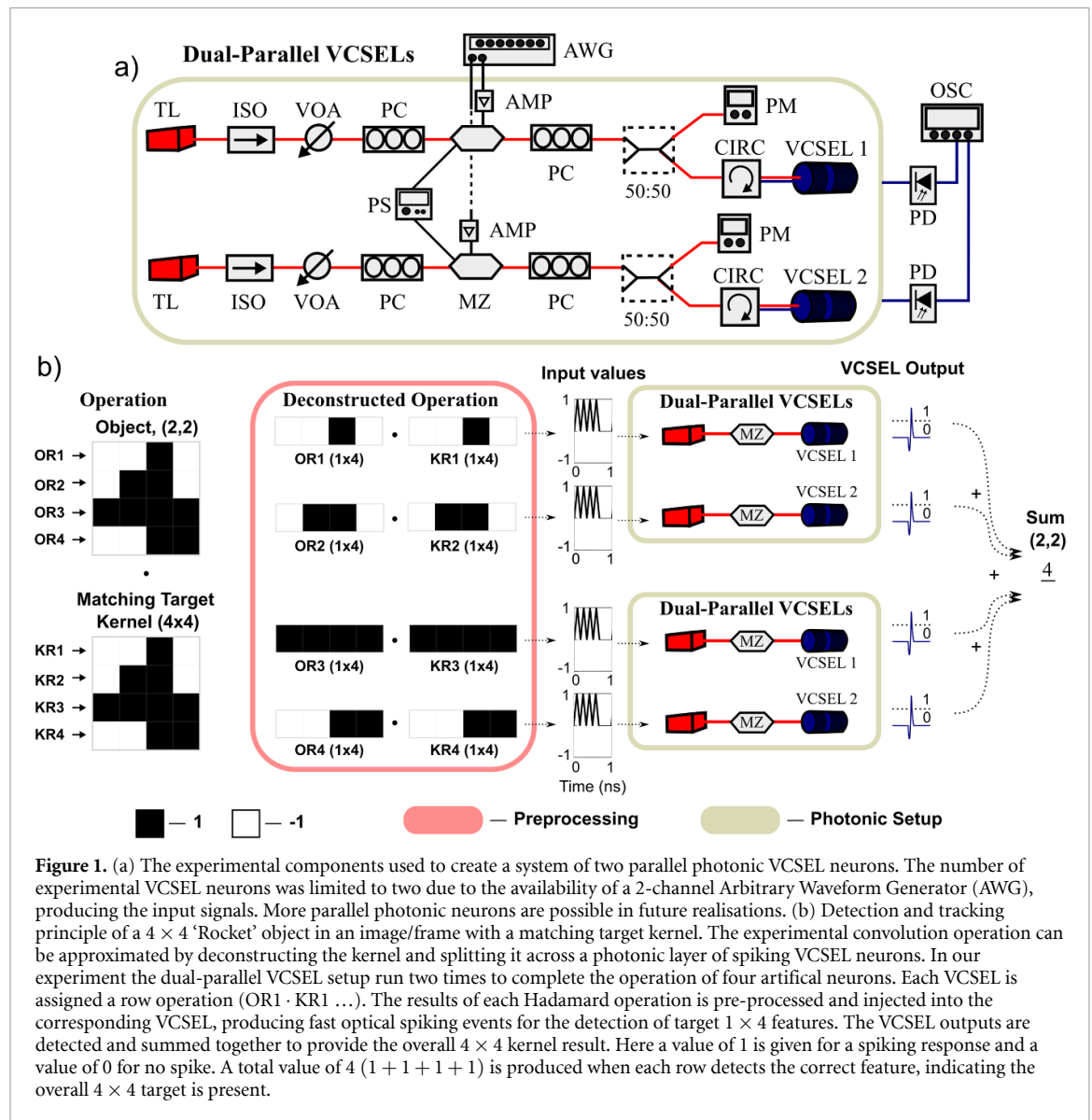


Figure 1. (a) The experimental components used to create a system of two parallel photonic VCSEL neurons. The number of experimental VCSEL neurons was limited to two due to the availability of a 2-channel Arbitrary Waveform Generator (AWG), producing the input signals. More parallel photonic neurons are possible in future realisations. (b) Detection and tracking principle of a 4 × 4 ‘Rocket’ object in an image/frame with a matching target kernel. The experimental convolution operation can be approximated by deconstructing the kernel and splitting it across a photonic layer of spiking VCSEL neurons. In our experiment the dual-parallel VCSEL setup run two times to complete the operation of four artificial neurons. Each VCSEL is assigned a row operation (OR1 · KR1 ...). The results of each Hadamard operation is pre-processed and injected into the corresponding VCSEL, producing fast optical spiking events for the detection of target 1 × 4 features. The VCSEL outputs are detected and summed together to provide the overall 4 × 4 kernel result. Here a value of 1 is given for a spiking response and a value of 0 for no spike. A total value of 4 (1 + 1 + 1 + 1) is produced when each row detects the correct feature, indicating the overall 4 × 4 target is present.

VCSELs, each of which acts as an integrate-and-fire optical spiking neuron as demonstrated in [30]). In each arm of the setup, light from a tuneable laser source (TL) is passed through an optical isolator (ISO—to avoid undesired backward reflections), and a variable optical attenuator to provide the system with a controllable level of optical injection power. A polarisation controller (PC) is then used to maximize the optical input into a 10 GHz Mach Zehnder intensity modulator (MZ), responsible for optically encoding the image input data into the optical injection line. The MZ bias point was controlled using a variable power supply (PS) and was set to produce drops in optical power for positive RF inputs. Image input data is generated by a 2-channel 12 GSa/s, 5 GHz AWG (Keysight M8190a), amplified using an electrical amplifier, and passed to the MZ. A second PC is included after the MZ to match the polarisation of the optically encoded injection to the selected lasing peak of the VCSEL. The encoded light signals are finally injected into the VCSEL cavity via an optical coupler and an optical circulator, and injection power is monitored by a power meter. The optical spiking response from each VCSEL (that contains the resulting detection information) is captured using 9 GHz amplified photodetectors (PDs) and a real-time 16 GHz, 40 GSa/s oscilloscope.

The experimental setup operated using a 2-channel AWG, permitting the realisation and simultaneous control of 2 individual input channels; hence allowing the building of a system with 2 parallel driven VCSEL photonic neurons. Future realisations with a higher number of optical neurons are possible by increasing the number of available independent input arms with additional equipment. Two VCSELs, one with a lasing wavelength of 1550 nm, and one with a lasing wavelength of 1300 nm were used in experiments. Each VCSEL (VCSEL1 and VCSEL2) was stabilized at room temperature (293 K) and driven with a bias current of 3.5 mA (~1.8 times lasing threshold current) and 3.5 mA (~2.5 times lasing threshold current) respectively. Single transverse mode lasing was observed from both VCSELs, with two orthogonal polarisation modes present in

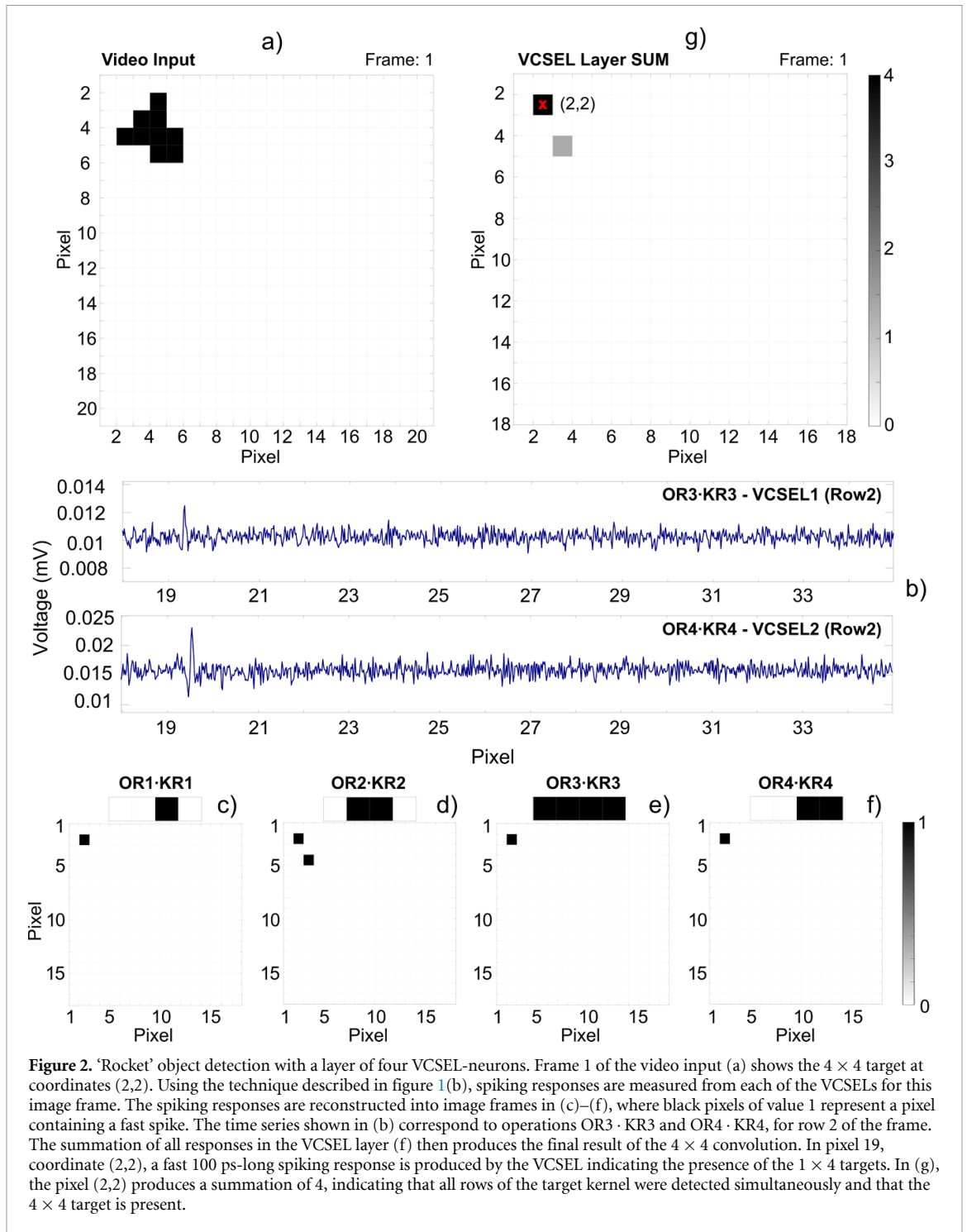
their spectra, as in [31]. Optical injection was made with a negative frequency detuning from the peak of the dominant polarisation mode in VCSEL1 (-3 GHz) and subsidiary polarisation mode in VCSEL2 (-2 GHz). These injection conditions induced an injection locking response in both VCSELs. The optically encoded image inputs were configured to modulate injection amplitude around a mean optical power of $90 \mu\text{W}$ and $111 \mu\text{W}$, respectively. When the appropriate inputs arrived at a VCSEL, within the short (1 ns) time window of the leaky integrate-and-fire behaviour, the inputs induced a transition into a regime of excitable spiking responses in the VCSEL, as demonstrated in [34]. The neuromorphic operation of the VCSEL-neurons, and the triggering of sub-nanosecond (approx. 100 ps-long) spiking events, is therefore used to indicate the presence of target image patterns, directly in the optical domain.

In this work, we deploy for the first time multiple in-parallel VCSELs to achieve operation with larger target kernels, as shown in figure 1(b), to permit the detection of complex patterns. In particular, figure 1(b) reveals the detection and tracking of a 4×4 pixel object with a layer of four spiking VCSELs. Due to the availability of only 2 inputs from our 2-channel AWG, the experimental layer of four VCSELs was achieved by running the Dual-Parallel VCSEL setup twice. However, there are no fundamental limitations (beyond current equipment availability) for 4 (or more) VCSEL neurons to be implemented at once in the future. To achieve the recognition of the target feature the source image/frame is operated on by a 4×4 target kernel, which directly matches the features of the 'Rocket' at coordinates (2,2) (see figure 2(a)). The parallel VCSELs were used to approximate the convolution operation of the 4×4 kernel by reducing it down to four (in-parallel) 1×4 kernel operations. Each VCSEL was fed one deconstructed operation ($\text{OR1} \cdot \text{KR1}$, $\text{OR2} \cdot \text{KR2}$, etc ...) which produced four Hadamard product values. The Hadamard products were calculated in a pre-processing stage outwith the photonic system, before the image inputs were generated and fed into the MZ and VCSEL in the form of short (100 ps-long) input pulses (as detailed in [34]). In this work a configurable pixel duration of 1 nanosecond (ns) per pixel was selected to match the refractory period of the excitable optical spiking VCSEL neuron. A direct match between feature and kernel resulted in a burst of positive input pulses that triggered a fast (sub-nanosecond long) optical spiking event at the output of the VCSEL. The timeseries measured by the PDs are segmented into bins of the appropriate pixel duration (1 ns per pixel) and reconstructed into pixel plots. Bins with spiking responses were counted and assigned a value of 1 (indicated by a black pixel). The result of each 1×4 operation was then combined to provide an approximate convolution output. Final layer summation pixels with a total value of 4 were identified as positions where the entire 4×4 target feature was detected. This operation is an approximate convolution as each VCSEL-neuron produces a binary output using the information of 4 pixels. This differs from traditional convolution where all pixels would contribute to the final pixel value. Instead, in this work the final layer summation is that of 4 binary outputs (total value 4) instead of 16 (one for each pixel). The layer summation of spiking VCSEL responses was again performed offline in a post-processing stage to accommodate the repeated running of the Dual-Parallel VCSEL setup. Given the realisation of a whole VCSEL layer, the analog summation of spiking signals could be performed directly in hardware with a single photodetector.

2.2. Software-based SNN

Building upon the foundational work of the spiking VCSELs, the software-based SNN serves as a mock representation, simulating the potential outcomes of an expanded VCSEL network with more layers. This simulation, run on a digital computer, seamlessly integrates the directly measured outputs from the VCSELs, which function as a first convolution layer of the network, and processes this data towards a classification outcome. The primary objective of this software-based SNN is to provide a glimpse into the capabilities of a scaled-up VCSEL network. By directly taking the output of the VCSEL convolutional layer as its input, the software SNN offers a realistic representation of how the VCSEL network might perform with additional layers. While the approach might seem constrained, especially given the training method and exclusive reliance on convolutional layers, similar systems have demonstrated remarkable efficacy [29]. Notably, this type of simple system has showcased impressive detection capabilities, especially in low signal-to-noise scenarios, and when paired with an event-based vision sensor [35]. Such scenarios, the processing of event-based vision sensor data, are where the speed of a fully realised photonic system could provide a distinct benefit. Given a photonic realisation of the simulated SNN layers, the high speed optical signals (1 ns-per-pixel) of a VCSEL layer could surpass the sampling rate of dynamics vision sensors ($1 \mu\text{s}$) [36], and allow for real-time event-based processing.

The simulated network is trained in an unsupervised fashion, leveraging STDP [26] to hone its feature extraction capabilities. Given the convolutional nature of our network, we opted for a streamlined interpretation of the STDP rule, drawing inspiration from proven methodologies in Spiking CNNs [37]. To enhance the training process, we incorporated a winner take all (WTA) mechanism [38], as evidenced and detailed in previous reports [27, 29]. This approach facilitates the training of simple-shape feature extractors, ensuring that the learned features distinctly represent either an 'on' or 'off' pixel detection. While the



network had the potential to discern tens or even hundreds of features, the WTA mechanism, combined with the set neuron parameters, led it to settle on 16 distinct features. This convergence indicates that the already learned features provided sufficient representation through spiking activation, rendering additional features redundant. Structurally, the network adopts a Conv-Pool-Conv configuration. The terminal convolutional layer serves a dual purpose: it acts as both a convolutional layer and a global pooling mechanism, steering the data towards a classification. This exclusive reliance on convolutional neural layers aligns with our objective of realizing a system predominantly driven by VCSEL-only processing methods. The architecture of the network comprises a 5×5 convolution kernel (with a stride of 1 and a threshold of 5 pixels), followed by a 2×2 max pooling layer (with a stride of 2), and culminating in a 3×3 kernel (with a threshold of 2) acting as a global pool layer with the output feature patterns representing either the Digit 3 or Digit 7.

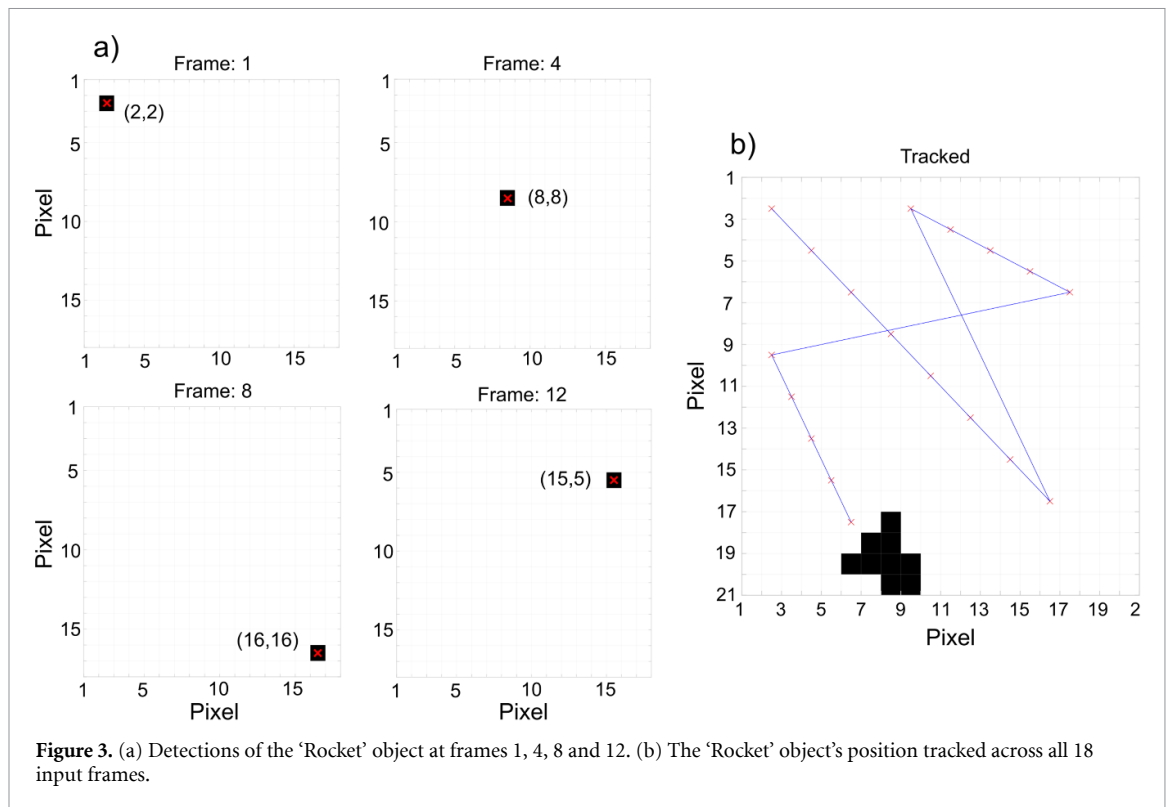


Figure 3. (a) Detections of the 'Rocket' object at frames 1, 4, 8 and 12. (b) The 'Rocket' object's position tracked across all 18 input frames.

3. Detection and tracking results

3.1. Photonic detection and tracking of target patterns

The detection and tracking of a feature pattern (target) in an input image is demonstrated for the first time using the experimental neuromorphic photonic spiking system depicted in figure 1(b). Using the technique described in section 2.1 four parallel VCSEL neurons are tasked with the recognition of a 4×4 pixel 'Rocket' feature in an 18 frame video feed. Supplementary figure 1, provided alongside this manuscript, shows the 18 frame video input. As previously highlighted, the system uses a photonic operation speed of 1ns/pixel, meaning a photonic processing rate of 0.4 micro-seconds per (20×20 pixel) frame is achieved. The position and shape of the 'Rocket' feature (in Frame 1) is shown in figure 2(a). First, the input frame is pre-processed according to figure 1(b), where the image input for each VCSEL is generated following the 1×4 operation. The image input for each optical injection arm is produced by the AWG, fed into each MZ, and shot into each VCSEL. The corresponding optical spiking response of each in-parallel VCSEL spiking neuron (across both experimental runs) is shown in figures 2(c)–(f). The 1×4 feature targeted by each VCSEL is indicated above the plots. The input frames were not padded, hence following the convolution operation the image's pixel dimensions were reduced to 17×17 . In figures 2(c)–(f), black pixels correspond to positions in the scenes where an optical spiking response (output value 1) were obtained in the VCSEL output timeseries. For example, in figure 2(b), part of VCSEL1 and VCSEL2's spiking timeseries is plotted (showing pixels 18–34 of the image, i.e. row 2 of figures 2(e) and (f)). Here, both the 1300 nm and 1550 nm VCSELs produced 100 ps-long spikes in pixel 19, corresponding to position (2,2) in the input frame. Figures 2(c) and (f) show that each of the four parallel VCSELs produced spiking responses at pixel coordinate (2,2). Therefore, in the corresponding VCSEL layer summation (shown in figure 2(g)) the total value of pixel (2,2) is 4, meaning that the 4×4 'Rocket' object was fully detected. An example of a partial detection is also shown in figure 2(g), where only one row of the 4×4 kernel was detected at (3,4). Due to the higher total value in position (2,2), we can identify it as the position of the 'Rocket' feature in the frame. By repeating this process, the rocket feature was detected in each frame of the video. Figure 3 demonstrates that the position of the 'Rocket' object can be determined by following the highest value (4 in this case) in the scene. In figure 3, only pixels with layer summation values of 3 or higher are plotted to filter out partial detections. The trajectory of the 'Rocket' feature can then be identified by following its coordinates in each frame (see figure 3(b) and supplementary figures 2 and 3). In these results the VCSEL layer detected the 'Rocket' feature in all frames for a detection efficiency of 100%. Comparing the trajectory of the original input with the plotted trajectory we also find that in all 18 frames the position the 'Rocket' object was identified correctly. In this case, the 'Rocket' feature was correctly tracked through three diagonal sweeps across the scene.

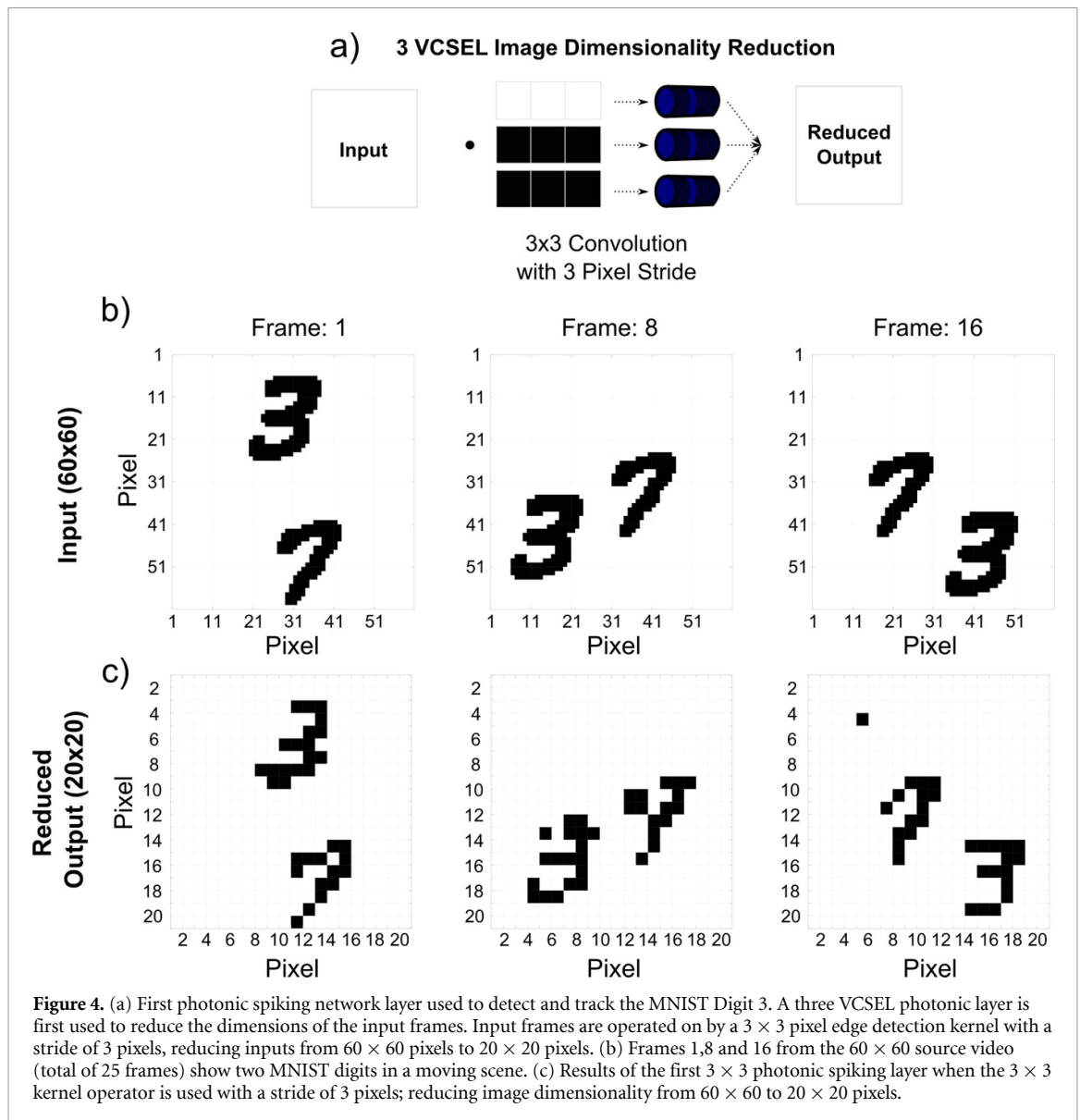
These results show that a layer of four in-parallel VCSELs can perform the detection and tracking of a 16-pixel (4×4) feature in a (18 frame) input video feed, using deconstructed kernel features for an approximate convolution operation. This experimental result was achieved by running two VCSEL-neurons in two experimental runs (effectively acting as a layer of four in-parallel neurons). Due to the parallel running of multiple VCSELs, the 16 value operation was achieved using 4 values per VCSEL (an input burst of four short 100 ps-long pulses). This allowed for a configurable pixel duration of 1 ns per pixel, allowing the photonic integrate-and-fire operation to reach the maximum spike-firing rate of the VCSEL-neurons (1 GHz—refractory period of 1 ns). Therefore, despite the increased (4×4) size of the feature pattern, the processing rate of the photonics was able to operate at the limit of the device [34]. At last, it must be noted that our neuromorphic photonics VCSEL were able to process the optical image inputs of the 20×20 pixel scene in <400 ns (due to the photonic operation speed of 1 ns/pixel). A total of $<7.2 \mu\text{s}$ ($18 \cdot 400$ ns) was therefore required for the VCSEL neurons to process all 18 frames of the input video.

3.2. Photonic detection and tracking of larger, complex and selective targets

A second detection and tracking task of higher complexity was also tested on the photonic system to illustrate the versatility of our platform. This consisted of the detection and tracking of larger, complex patterns, namely 28×28 pixel handwritten digits from the MNIST dataset [39], as they move in a 60×60 pixel scene. The video input (shown in supplementary figure 4) formed of 25 consecutive 60×60 pixel frames, tasked the VCSEL-based neuromorphic optical spike processing system with detecting and tracking the Digit 3 feature only, ignoring the Digit 7 also moving across the scene. Due to the presence of the distractor feature (Digit 7), and the overall larger feature size, this defined a task of considerable higher complexity. In this task, the Digit 3 object required a larger number of pixels to identify its features, while the identifying feature itself would have to be Digit 3 specific (capable of a classification). Therefore, in order to process this image, two network layers of optical spiking VCSEL-neurons were implemented to realise the operation of a multi-layer optical SNN. The first layer of spiking VCSELs is used to reduce the pixel dimensions of the image features, and the second layer is used to identify and track the lower resolution features of the Digit 3.

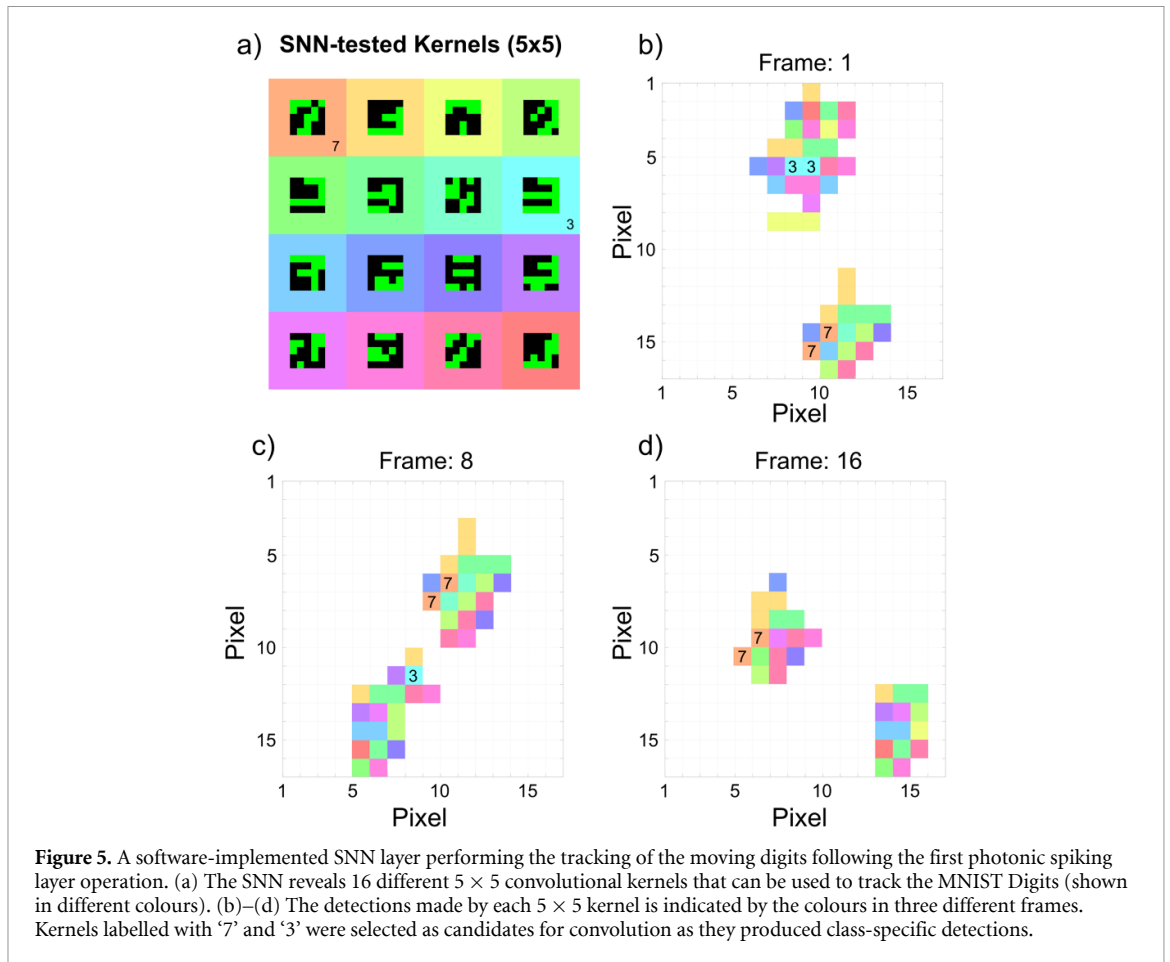
The first layer was implemented experimentally using three in-parallel VCSEL neurons, as shown in figure 4(a). This layer implemented a 3×3 kernel operation according to the technique previously described in section 3. Again, due to the limited number of AWG channels available in our setup, the 3 VCSEL operation was achieved across two experimental runs. A simple kernel operator, which extracted horizontal features, was selected and scanned across the 60×60 pixel frame with a 3 pixel stride. This operation reduced the dimension of the output image by a factor of 3 (from 60×60 to 20×20 pixels), in a technique similar to image pooling. The original 60×60 input video frames (frames 1, 8, and 16) of the moving handwritten digit patterns and their corresponding photonic layer outputs (following the summation of the three spiking VCSELs) are shown in figures 4(b) and (c). Similar to figure 3, only pixels with total values of 2 (spikes) or higher are plotted (black dots) in figure 4(c). These pixels therefore correspond to coordinates that produced spiking responses for multiple rows of the 3×3 edge detection kernel. The images reconstructed from the optical spiking events of the photonic layer show the successful reduction of the image's dimensions. The photonic layer output for each input frame is shown in supplementary figure 5. Now with reduced image dimensions, the Digit 3 object could be more easily described with a 5×5 feature kernels. However, the previously consistent handwritten digit patterns now changed from frame to frame. This result of the 3 pixel stride made the selection of an appropriate convolution kernel (for the specific detection and tracking of the Digit 3) in the subsequent layer increasingly complex. Therefore, in order to realise a suitable 5×5 kernel to describe the Digit 3 features, the output of the first photonic layer was fed into a software-implemented SNN.

The software-implemented SNN layer, discussed in section 2.2 utilised unsupervised learning with STDP to group salient and frequently occurring features in the moving number scene. STDP is a biologically-inspired unsupervised learning mechanism that modifies the strength of synaptic connections between neurons based on the relative timing of their spikes [26]. The binary output of the photonic VCSEL neurons allowed for the efficient computation of spike timing differences, which is a crucial aspect of STDP. The use of STDP also enabled the unsupervised learning of features, which is important for the development of autonomous and adaptive neuromorphic systems. The software-implemented layer consisted of 16 different 5×5 convolutional kernels, each of which learned a different set of features from the input video data. Some of the features extracted were closely related to the spatial features of either the Digit 3 or 7, while others appeared to be a combination of both. The use of multiple kernels allowed the network to learn a diverse set of features that could be combined in subsequent layers to create more representative features for the digits. The software implemented SNN layer served as a blueprint for the extension of the VCSEL-based photonic spike processing system to multiple layers. In a multi-layer architecture, each layer would typically have a reduced spatial resolution, enabling the extraction of increasingly complex and abstract features by combining the learned features from previous layers. This hierarchical learning approach is widely used in



the field of deep learning and is believed to be an effective method for building complex neural networks capable of performing a wide range of tasks. This process therefore produces a number of potential kernel candidates for the subsequent photonic layer.

The results plotted in the colour-coded plots of figures 5(b)–(d) show the detection of the handwritten Digits in frames 1, 8, and 16, using several (16 in total) different kernel operators learned by the software-implemented SNN (shown in figure 5(a)). Supplementary figure 6 shows the software-SNN result for all input frames. In figures 5(b)–(d) each pixel is filled with a colour that corresponds to a kernel operator figure 5(a) successfully detected in the corresponding position in the image. Only the kernel operator with the strongest detection has its colour plotted in each pixel. Again, the dimensions of the resulting outputs are reduced by 4 to 16×16 pixels (from 20×20 pixels) because of the 5×5 kernel operation. The results show that despite the complexity of the moving and changing handwritten Digit 3 and Digit 7 patterns from the first photonic layer, several 5×5 pixel kernels can be used to detect both the Digit 3 and Digit 7 across multiple frames. Importantly, by analysing the detection of each kernel, the software-implemented SNN was able to identify operators that even in one additional layer could work best for each specific handwritten Digit. In figure 8, two examples of kernel operators that act as Digit-specific detectors are highlighted with a ‘3’ and a ‘7’. In frames 1, 8 and 16, figures 5(b)–(d) show that these kernels generate a low number of false-positive identifications. These kernels only appear within their target Digit’s cluster, or not at all, making them suitable for class-specific tracking. Additionally, we note that due to the changing frame-to-frame feature pattern, the selected kernels trigger at different positions within the Digit cluster. Overall, these results demonstrated that the software-implemented SNN layer was able to successfully utilise the outputs of



the photonic spiking hardware VCSEL layer, and that by using a variety of 5×5 kernel operators, the hybrid hardware-software system could detect both digits in all 25 frames, with further potential for Digit-specific detection. Using the kernel identified in figure 5(a), a second software SNN-informed photonic layer of spiking VCSELs was created to implement the detection and tracking of the Digit 3.

The software SNN-informed kernel was implemented experimentally in hardware using a photonic layer of five VCSEL neurons as shown in figure 6(a). This second photonic layer performed the 5×5 kernel operation on the output of the first photonic layer. The second photonic layer was run via consecutive iterations of the dual parallel VCSEL setup (shown in figure 1(a)) to realise the 5×5 operation and deliver the fast (sub-ns) optical spiking responses of five VCSELs. We emphasize again that there are no limitations for future single-run implementations of the photonic layer, once the number of existing independent input channels is increased beyond our current availability. Similar to the first photonic layer (and section 3), the 5×5 kernel operator was deconstructed into five 1×5 operations and run across all five in-parallel VCSELs. Each VCSEL was therefore fed 5 values in the form of 5 short 100 ps-long pulses, enabling the same photonic operation speed of 1 ns/pixel. The output spiking timeseries were collected, summated and reconstructed to produce the frames shown in figures 6(b)–(e). Again, only layer summation values of ≥ 3 are plotted to allow for the partial detection of the changing feature pattern. The complete 25 frame output for the 5 VCSEL layer is shown in supplementary figure 7. In total, the software SNN-informed kernel detected the Digit 3 in 20 of 25 frames. Further, despite the SNN-informed kernel relying on partial detections to identify the changing Digit 3 pattern, the system triggered no false-positive detections of the Digit 7 pattern. The SNN-informed kernel therefore provided a reliable (80% successful) tracking of the Digit 3 while also performing an intrinsic classification functionality. As anticipated, spiking responses were triggered from different parts of the Digit 3 feature pattern, and in some cases multiple triggers were made in the same frame. Where this occurred, the median of the detection coordinate was given as the final position of the Digit 3 in the scene. The trajectory of the tracked Digit 3 (with the SNN-informed kernel) is shown in figure 6(e), and is finally overlaid with the original 60×60 pixel source image in figure 7. Supplementary figure 8 shows the tracking of the Digit 3 feature in all input video frames. Due to the detection of multiple different parts of the Digit 3 feature pattern, the trajectory appears jagged, and not smooth. However, once overlaid on the original source image, it can be seen that the jagged motion is a small-scale effect in comparison to the size of the tracked

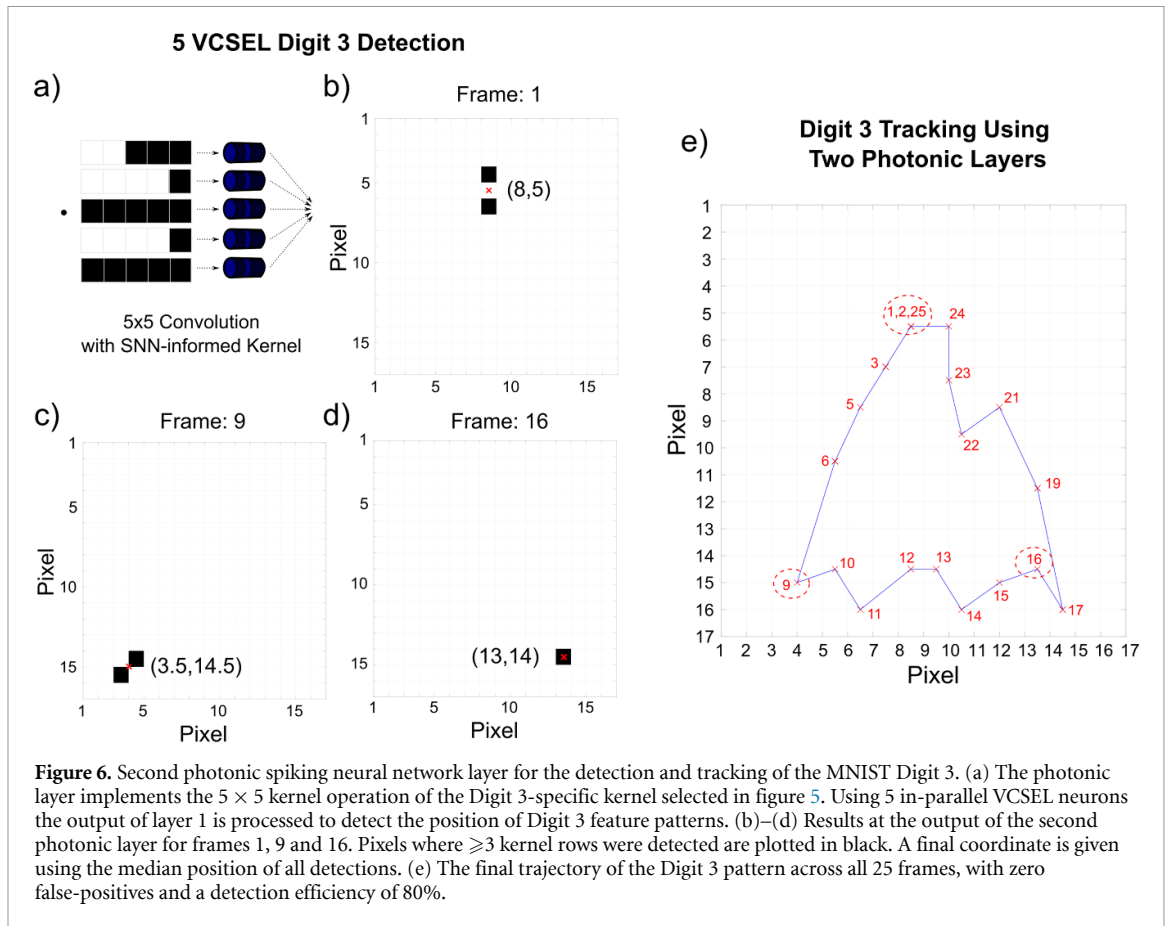


Figure 6. Second photonic spiking neural network layer for the detection and tracking of the MNIST Digit 3. (a) The photonic layer implements the 5×5 kernel operation of the Digit 3-specific kernel selected in figure 5. Using 5 in-parallel VCSEL neurons the output of layer 1 is processed to detect the position of Digit 3 feature patterns. (b)–(d) Results at the output of the second photonic layer for frames 1, 9 and 16. Pixels where ≥ 3 kernel rows were detected are plotted in black. A final coordinate is given using the median position of all detections. (e) The final trajectory of the Digit 3 pattern across all 25 frames, with zero false-positives and a detection efficiency of 80%.

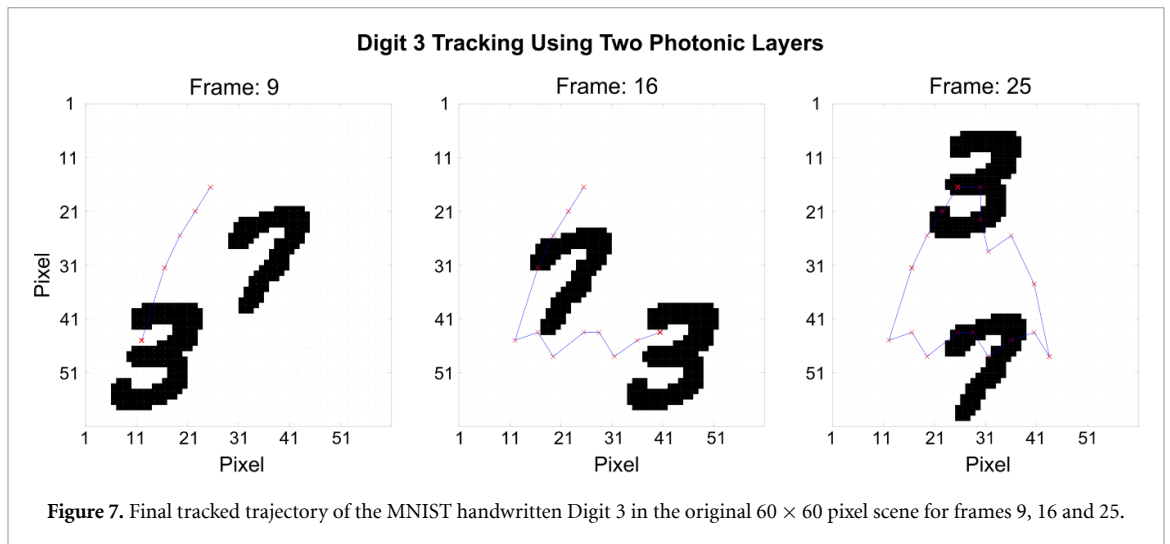


Figure 7. Final tracked trajectory of the MNIST handwritten Digit 3 in the original 60×60 pixel scene for frames 9, 16 and 25.

Digit. Further, due to the increased number of in-parallel spiking photonic VCSEL-neurons in the second layer, the optical image inputs of the larger kernel operation were successfully processed at the same photonic operation speed (1 ns/pixel), meeting the maximum spike-firing rate of the photonic devices. The photonic processing rate of this second task can be calculated by adding the photonic processing rate of each layer, given the preselection of kernels (no SNN training). The photonic processing rate for the first dimensionality reduction layer is < 3600 ns/frame (60×60 pixels at 1 ns/pixel) and the processing rate of the second photonic detection layer was < 400 ns/frame (20×20 pixels at 1 ns/pixel). This means that while maintaining a 1 ns/pixel photonic operation speed, the two layer photonic system achieved detection and tracking of a larger 28×28 pixel target (in a 60×60 pixel scene) at a photonic processing rate of < 4000 ns/frame. Therefore, using two photonic layers of spiking VCSELs, and an intermediate software-implemented SNN layer, we were able to classify, recognise and track a moving handwritten digit target with significantly more

pixels than previous VCSEL-neuron based image processing demonstrations, with a fast photonic operation speed and processing rate.

4. Conclusions

In this work, we successfully implemented and experimentally demonstrated, for the first time, a neuromorphic photonic spike processing system enabling the detection and tracking of target patterns in image and video inputs at ultrafast rates. This is achieved by implementing PSNN layers built with in-parallel VCSEL-based spiking optical neurons. By deconstructing larger kernel operators into smaller components, convolution operations were shared across a number of VCSEL-neurons, enabling the detection of larger, more complex, and individually-selected target patterns. Further, by operating with a layer of VCSEL-neurons a photonic operation speed of 1.0 ns/pixel was achieved, determined by the maximum spike-firing rate of the VCSEL neurons, despite the size of kernel operations. In the first experimental demonstration, we showed that a 4×4 feature pattern ('rocket' object) could be detected and tracked using a layer of 4 optical spiking VCSEL-neurons. By directly matching the feature pattern with a 4×4 kernel operator, we successfully detected the moving target in all frames of the scene without misidentification, and tracked the targets trajectory. In the second experimental demonstration we revealed the successful operation of the neuromorphic spiking system at a more difficult task, namely the detection and tracking of larger, more complex, class-specific moving targets (i.e. hand-written digits) in scenes. To achieve this functionality, we implemented multiple photonic spiking network layers. We applied a photonic layer of 3 VCSELs to first reduce the dimensionality of the input frames with a strided edge detection kernel. The output of this first photonic layer was then sent to a software-implemented SNN, where suitable kernels were calculated for the class-specific tracking of the target feature pattern (handwritten Digit 3). Finally, using the information provided by the software-implemented SNN, a second photonic layer of 5 VCSEL-based spiking neurons was implemented in hardware to perform a second (spike-enabled) convolution operation. This approach permitted the successful detection and tracking of the selected complex pattern (handwritten Digit 3) using the identified 5×5 kernel operator. The photonic neuromorphic system achieved a recognition efficiency of 80% , despite the continuously varying feature pattern of the Digit 3. Further, this successful detection and tracking was achieved without any false-positive identifications of the distractor feature also in the scene (handwritten Digit 7). Moreover, by expanding the trajectory result of the Digit 3 feature pattern, the target was successfully tracked though the original scene at the original image resolution. The class-specific detection and tracking of the moving complex pattern (Digit 3) was therefore achieved using two layers of photonic spiking VCSEL neurons and a software-implemented SNN. The hybrid photonic-software SNN interaction demonstrates the key advantage of unsupervised learning in such systems, and provides the hardware layers with critical knowledge on how to most appropriately operate to achieve the desired performance. Overall, software SNN-informed layers of photonic spiking VCSEL neurons have demonstrated the ability to process images with large target kernels, while maintaining GHz-rate photonic operation speeds and fast (100/1000s nanosecond per image/frame) processing rates. We believe that such photonic systems have exciting prospects in neuromorphic spike-based implementation of hardware for numerous information processing tasks for AI, including computer vision, image processing, target detection and tracking.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.15129/0b2eece5-daf4-4142-a2fc-5d648d1bd48b>.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors acknowledge this work was supported by the UK Defense and Security Accelerator, DASA (ACC2022261) and the UKRI Turing AI Acceleration Fellowships Programme (EP/V025198/1). GD acknowledge support from Leonardo UK as part of Strathclyde's Leonardo Lectureship programme.

ORCID iD

Joshua Robertson  <https://orcid.org/0000-0001-6316-5265>

References

- [1] Miller. D A B 2017 Attojoule optoelectronics for low-energy information processing and communications *J. Lightwave Technol.* **35** 346–96
- [2] Christensen D V et al 2022 2022 roadmap on neuromorphic computing and engineering *Neuromorph. Comput. Eng.* **2** 022501
- [3] Bai Y et al 2023 Photonic multiplexing techniques for neuromorphic computing *Nanophotonics* **12** 795–817
- [4] Shastri B J, Tait A N, de Lima T F, Pernice W H P, Bhaskaran H, Wright C D and Prucnal P R 2021 Photonics for artificial intelligence and neuromorphic computing *Nat. Photon.* **15** 102–14
- [5] Feldmann J, Youngblood N, Wright C D, Bhaskaran H and Pernice W H P 2019 All-optical spiking neurosynaptic networks with self-learning capabilities *Nature* **569** 208–14
- [6] George J K, Mehrabian A, Amin R, Meng J, de Lima T F, Tait A N, Shastri B J, El-Ghazawi T, Prucnal P R and Sorger V J 2019 Neuromorphic photonics with electro-absorption modulators *Opt. Express* **27** 5181
- [7] de Lima T F, Tait A N, Saeidi H, Nahmias M A, Peng H-T, Abbaslou S, Shastri B J and Prucnal P R 2020 Noise analysis of photonic modulator neurons *IEEE J. Sel. Top. Quantum Electron.* **26** 1–9
- [8] Anirudh Pammi V, Alfaro-Bittner K, Clerc M G and Barbay S 2020 Photonic computing with single and coupled spiking micropillar lasers *IEEE J. Sel. Top. Quantum Electron.* **26** 1–7
- [9] Selmi F, Braive R, Beaudoin G, Sagnes I, Kuszelewicz R, Erneux T and Barbay S 2016 Spike latency and response properties of an excitable micropillar laser *Phys. Rev. E* **94** 42219
- [10] Dillane M, Lingnau B, Viktorov E A, Dubinkin I, Fedorov N and Kelleher B 2021 Asymmetric excitable phase triggering in an optically injected semiconductor laser *Opt. Lett.* **46** 440
- [11] Skalli A, Robertson J, Owen-Newns D, Hejda M, Porte X, Reitzenstein S, Hurtado A and Brunner D 2022 Photonic neuromorphic computing using vertical cavity semiconductor lasers *Opt. Mater. Express* **12** 2395
- [12] Hejda M, Arturo Alanis J, Ortega-Piwonka I, Lourenço J ao, Figueiredo J, Javaloyes J, Romeira B and Hurtado A 2022 Resonant tunneling diode nano-optoelectronic excitable nodes for neuromorphic spike-based information processing *Phys. Rev. Appl.* **17** 024072
- [13] Zhang W, Hejda M, Malysheva E, Raghib Ali Al-Taai Q, Javaloyes J, Wasige E, Figueiredo J M L, Dolores-Calzadilla V, Romeira B and Hurtado A 2023 Tunable presynaptic weighting in optoelectronic spiking neurons built with laser-coupled resonant tunneling diodes *J. Appl. Phys.* **56** 084001
- [14] Xu B, Huang Y, Fang Y, Wang Z, Yu S and Xu. R 2022 Recent progress of neuromorphic computing based on silicon photonics: electronic-photonic co-design, device and architecture *Photonics* **9** 698
- [15] Tait A N, de Lima T F, Zhou E, Wu A X, Nahmias M A, Shastri B J and Prucnal P R 2017 Neuromorphic photonic networks using silicon photonic weight banks *Sci. Rep.* **7** 7430
- [16] Schneider M, Toomey E, Rowlands G, Shainline J, Tschirhart P and Segall K 2022 SuperMind: a survey of the potential of superconducting electronics for neuromorphic computing *Supercond. Sci. Technol.* **35** 053001
- [17] Tait A N, Nahmias M A, Shastri B J and Prucnal P R 2014 Broadcast and weight: an integrated network for scalable photonic spike processing *J. Lightwave Technol.* **32** 4029–41
- [18] Feldmann J et al 2021 Parallel convolutional processing using an integrated photonic tensor core *Nature* **589** 52–58
- [19] Chen Z et al 2023 Deep learning with coherent VCSEL neural networks *Nat. Photon.* **17** 723–30
- [20] Sackesyn S, Ma C, Dambre J and Bienstman P 2021 Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation *Opt. Express* **29** 30991
- [21] Borghi M, Biasi S and Pavesi L 2021 Reservoir computing based on a silicon microring and time multiplexing for binary and analog operations *Sci. Rep.* **11** 15642
- [22] Bueno J, Robertson J, Hejda M and Hurtado A 2021 Comprehensive performance analysis of a VCSEL-based photonic reservoir computer *IEEE Photonics Technol. Lett.* **33** 920–3
- [23] Tsurugaya T, Hiraki T, Nakajima M, Aihara T, Diamantopoulos N-P, Fujii T, Segawa T and Matsuo S 2022 Cross-gain modulation-based photonic reservoir computing using low-power-consumption membrane SOA on Si *Opt. Express* **30** 22871
- [24] Strubell E, Ganesh A and McCallum A 2019 Energy and policy considerations for deep learning in NLP *Proc. 57th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics) pp 3645–50
- [25] Park J, Sohmyung H, Theodore Y, Neftci E and Cauwenberghs G 2014 A 65k-neuron 73-Mevents/s 22-pJ/event asynchronous micro-pipelined integrate-and-fire array transceiver *2014 IEEE Biomedical Circuits and Systems Conf. (BioCAS) Proc. (IEEE)* pp 675–8
- [26] Guo-qiang B and Poo M-ming 2001 Synaptic modification by correlated activity: Hebb's postulate revisited *Annu. Rev. Neurosci.* **24** 139–66
- [27] Kirkland P, Di Caterina G, Soraghan J and Matich G 2020 Perception understanding action: adding understanding to the perception action cycle with spiking segmentation *Front. Neurobot.* **14** 568319
- [28] Kirkland P, Kapitany V, Lyons A, Soraghan J, Turpin A, Faccio D, Di Caterina G 2020 Imaging from temporal data via spiking convolutional neural networks *Emerging Imaging and Sensing Technologies for Security and Defence V; and Advanced Manufacturing Technologies for Micro- and Nanosystems in Security and Defence III (SPIE)* Farsari M et al ed p 15
- [29] Kirkland P, Di Caterina G, Soraghan J, Andreopoulos Y and Matich G 2019 UAV Detection: A STDP Trained Deep Convolutional Spiking Neural Network *Retina-Neuromorphic Approach* pp 724–36
- [30] Robertson J, Hejda M, Bueno J and Hurtado A 2020 Ultrafast optical integration and pattern classification for neuromorphic photonics based on spiking VCSEL neurons *Sci. Rep.* **10** 1–8
- [31] Robertson J, Wade E, Kopp Y, Bueno J and Hurtado A 2020 Towards neuromorphic photonic networks of ultrafast spiking laser neurons *IEEE J. Sel. Top. Quantum Electron.* **26** 7700715
- [32] Hejda M, Robertson J, Bueno J and Hurtado A 2020 Spike-based information encoding in vertical cavity surface emitting lasers for neuromorphic photonic systems *J. Phys. Photonics* **2** 44001
- [33] Robertson J, Zhang Y, Hejda M, Bueno J, Xiang S and Hurtado A 2020 Image edge detection with a photonic spiking VCSEL-neuron *Opt. Express* **28** 37526–37

- [34] Robertson J, Kirkland P, Arturo Alanis J, Hejda M, Bueno J, Di Caterina G and Hurtado A 2022 Ultrafast neuromorphic photonic image processing with a VCSEL neuron *Sci. Rep.* **12** 4874
- [35] Kirkland P, Clemente C, Macdonald M, Di Caterina G and Meoni G 2023 Neuromorphic sensing and processing for space domain awareness *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium (Pasadena CA, USA)* pp 4738–41
- [36] Gallego G *et al* 2022 Event-based vision: a survey *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 154–80
- [37] Reza Kheradpisheh S, Ganjtabesh M, Thorpe S J and Masquelier T 2018 STDP-based spiking deep convolutional neural networks for object recognition *Neural Netw.* **99** 56–67
- [38] Maass W 2000 On the computational power of winner-take-all *Neural Comput.* **12** 2519–35
- [39] LeCun Y, Cortes C and Burges C J 2010 Mnist handwritten digit database *ATT Labs* (available at: <http://yann.lecun.com/exdb/mnist>)