Robust Learning and Recognition of Visual Patterns in Neuromorphic Electronic Agents

Dongchen Liang, Raphaela Kreiser, Carsten Nielsen, Ning Qiao, Yulia Sandamirskaya, and Giacomo Indiveri
Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland
dongchen@ini.uzh.ch

***Abstract*—Mixed-signal analog/digital neuromorphic circuits are characterized by ultra-low power consumption, real-time processing abilities, and low-latency response times. These features make them promising for robotic applications that require fast and power-efficient computing. However, the unavoidable variance inherently existing in the analog circuits makes it challenging to develop neural processing architectures able to perform complex computations robustly. In this paper, we present a spiking neural network architecture with spike-based learning that enables robust learning and recognition of visual patterns in noisy silicon neural substrate and noisy environments. The architecture is used to perform pattern recognition and inference after a training phase with computers and neuromorphic hardware in the loop. We validate the proposed system in a closed-loop hardware setup composed of neuromorphic vision sensors and processors, and we present experimental results that quantify its real-time and robust perception and action behavior.**

***Index Terms*—Neuromorphic computing, noisy spiking neural networks, robust object recognition, unsupervised learning**

I. Introduction

Neuromorphic engineering is concerned with emulating the dynamics of biological neurons and synapses in silicon [1], as well as the organizing and computing principles of real neural processing systems. A primary goal of these studies is to take advantage of the unique features of the brain, such as low-power consumption, massive parallelism, and low-latency processing, in order to perform efficient cognitive computations. In contrast to the classical von Neumann architecture of state-of-the-art digital computers, in neuromorphic hardware, memory and processing are co-localized in the synapses and neurons present in such devices. Previously developed neuromorphic computing hardware devices are using either asynchronous mixed-signal analog/digital [2]–[4] or purely digital circuits. Asynchronous neuromorphic systems process data and transmit signals only if and when they receive and produce events (spikes). Unlike purely digital approaches, the mixed-signal analog/digital approach has recently produced promising technologies for implementing computing architectures based on silicon neurons and synapses which exhibit dynamics that are similar to their biological counterparts [5].

Similar to the widely observed variance across biological neural networks [6], variability exists in all analog spiking neurons due to the unavoidable circuit noise and manufacturing mismatch, which has a strong effect on the behavior of neurons when the circuits are working in the sub-threshold domain [7], [8]. It is very challenging to implement desired behaviors and computations on such hardware [9]. However, it has been observed in the brain that despite the noisy neural substrate and environment, the behavior of biological spiking neural networks remains robust and reliable. It raises the question: can proper network connectivity enable mixed-signal spiking neural systems to perform computations with high robustness despite their noise and mismatch?

To address this question, we present a spiking neural architecture that comprises three biologically plausible mechanisms geared toward variability reduction and robust computation. Namely, we use a dis-inhibition mechanism to reduce the effect of noise and enable robust feature detection, an up- and down-scaling mechanism of connections which leads to the spatial invariance during recognition, and a group of Neural State Machine (NSM) structures [10], [11] to perform unsupervised learning. Besides, we present an on-line spike-based learning rule for both excitatory and inhibitory plastic synapses that enables the network to memorize trained visual patterns and perform visual pattern recognition. Along with the inherent ultra-low-power and event-driven computing paradigm of the mixed-signal analog/digital devices, the high robustness achieved on such hardware makes it possible to build an always-on and massively distributed system. Such a system reports an event or takes action only if and when it recognizes a visual stimulus as a pre-trained pattern; otherwise, the network keeps running with ultra-low power consumption. In Section II, we describe the used mixed-signal neuromorphic hardware in this work. In Section III, we present the network architecture and the learning rule. In Section IV, we validate the network in a real-time real-world task.

II. Methods

The setup used to train and configure the neuromorphic electronic system proposed is illustrated in Fig. 1. It is composed of a neuromorphic vision sensor - the Dynamic Vision Sensor (DVS) (DAVIS240C) [12] and a group of Dynamic Neuromorphic Asychronous Processor (DYNAP) chips [3]. The DVS emulates the dynamics of biological retina cells in silicon using mixed-signal analog/digital technologies. There are 240×180 pixels integrated on each chip. Each of them independently detects the illumination intensity change in a small area of the visual scene. It captures fast moving objects from the environment in a wide range of lighting

This work is supported by the China Scholarship Council (CSC) and by the Institute of Neuroinformatics, University of Zurich and ETH Zurich.

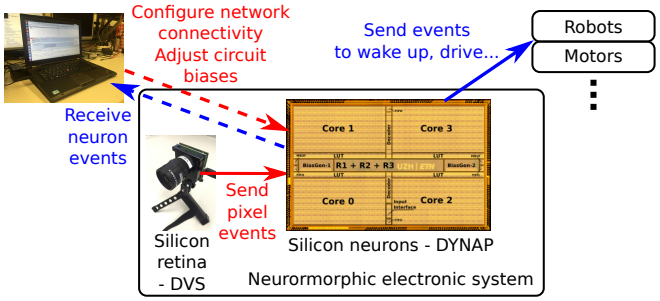

Fig. 1. The neuromorphic electronic system consists of DVS sensors and DYNAP chips. A computer is used to configure the sensors and chips, monitor their activities, or perform learning algorithms. Once a network is set up, or the learning stops, the computer can be disconnected. The perception and processing in neuromorphic sensors and processors are ultra-low-power.

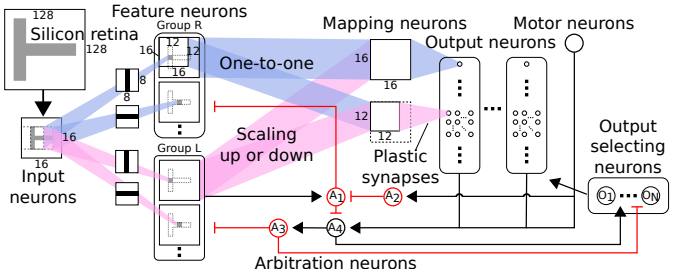

Fig. 2. The network architecture. The network is composed of an input layer, a feature layer, and an output layer. The neurons of these layers form a learning pathway and a recognition pathway. At the bottom of the figure, there is an arbitration mechanism implemented by neuron populations A_1, A_2, A_3, and A_4 that control which pathway will become dominant. In this paper, each of the populations A_1, A_2, A_3, and A_4 is implemented by 4 silicon neurons to make their behaviors robust.

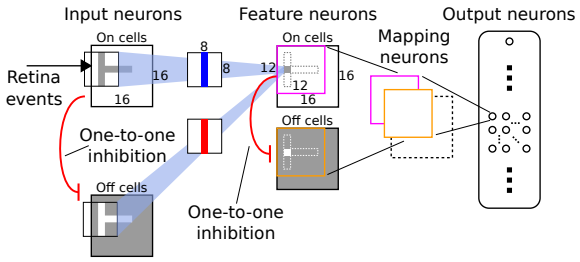

Fig. 3. The detailed structure of the feed-forward pathways from input neurons to output neurons. To simplify the illustration, here we only show the neurons of the recognition pathway while the learning pathway has the same on and off cells and the same connectivity principle. The on cells receive spikes from the previous layer, whereas the off cells tend to fire spontaneously due to a continuously supplied stimuli or a constant current. However, they receive a strong inhibition from the on cells in a one-to-one manner.

conditions. The DYNAP integrates 1024 silicon neurons on each chip. Every neuron features 64 programmable synapses and can stimulate 4096 destination synapses. The neurons realize adaptive exponential integrate-and-fire dynamics with biologically realistic time constants [13]. The synapses are non-plastic but can be trained with a computer in the loop. The parameters of the neurons and their connectivity are configurable via on-chip analog bias generators and digital latches. The events generated by the DVS silicon retina are sent to the silicon neurons on-chip using the Address-Event Representation (AER) protocol.

In this setup, spike-based learning can be run on a computer in real time with the neuromorphic chips in the loop. Whenever a neuron emits a spike, it will send its neuron ID and the timing of when it emits the spike to the computer. A learning algorithm on the computer will write the received information into a ring buffer, for example, at position S. After that, the algorithm will read the ring buffer from the position $S-1$ to $S-N$ to go through the history of the received spikes. The parameter N is chosen by the algorithm to make sure the maximum timing difference between the read-out spikes and the newly arrived spike is within a time window ($50\,ms$ in this implementation). For each pair of the neurons that emit the read-out spike and the newly arrived spike, the algorithm checks a look-up table. The table stores the information of whether there is a plastic synapse between two neurons. If there is one between this checked neuron pair, the algorithm will calculate a new synaptic weight according to the current weight and the timing difference between the two spikes.

As for biological retinas, illumination intensity change is critical for the silicon retina to perceive the environment. To ensure that there is retinal activity also for static scenes biological vision systems resort to two types of eye movements: saccades [14] and microsaccades [15]. In our setup, we simulate these eye movements by fixing the DVS and moving the objects in front of it. The mean rate of the events generated by DVS indicates the speed of the relative movement between the silicon retina and the objects. The online learning algorithm monitors the generated events and simulates the firing of the motor neurons which activate the saccades.

III. Spiking Neural Network Architecture

The architecture we propose is illustrated in Fig. 2. It has a feed-forward structure consisting of three layers of neurons. It starts from a 16×16 input layer. We select the center 128×128 pixels on the silicon retina and down-sample them to match the 16×16 input neurons. A feature layer follows the input layer. Due to the limited number of neurons on the prototype DYNAP chip, we choose to configure the feature layer to detect only horizontal and vertical bars. The feature layer neurons receive spikes from the input layer in a convolutional manner with kernel size 8×8 and stride 1. The output layer consists of multiple groups of neurons, each of which learns to recognize a different visual pattern. Output selecting neurons excite them to start the learning.

There are two pathways in the network. One pathway is for learning, and the other one is for recognition and inference, denoted by blue and pink colors respectively in Fig. 2. Each pathway has its group of feature neurons: neuron groups L and R for the learning and recognition pathways respectively. For recognition, we connect neuron group L to the output neurons in a convolutional manner with different kernel sizes. For learning, to the contrary, there is an up- and down-scaling mechanism through which we connect all the neurons in R to each output neuron and the connection maintains the spatial

information of the feature neurons. However, the connections are not direct. There is a group of mapping neurons that relay the spikes from the feature neurons to the output neurons. The learning takes place at the plastic synapses between the mapping neurons and the output neurons. Here we present the detailed structures and the learning rule as follows:

a) Dis-inhibition mechanism: The input layer and the feature layer consist of on and off cells. The on and off cells are excitatory and inhibitory neurons respectively. They cooperatively perform a dis-inhibition mechanism that makes the recognition robust to the noise from inputs and the inherent mismatch of analog circuits. The connectivity of the on and off cells is illustrated in Fig. 3. The on and off cells have the same connectivity manner to the next layer like the one shown in Fig. 2. However, only the on cells receive spikes from the previous layer whereas the off cells are not connected to the previous layer but receive a one-to-one inhibition from the on cells. Also, the off cells receive excitation from a continuously supplied stimuli or a constant current. In this way, the on and off cells always exhibit opposite behaviors.

b) Arbitration mechanism: The two pathways alternate to perform the learning and recognition of visual patterns. They are controlled by the arbitration neuron groups A_1, A_2, A_3, and A_4 at the bottom of Fig. 2. The learning pathway gains and maintains the opportunity to start and finish the learning as follows. Since the Post-Synaptic Potential (PSP) of inhibition is longer and stronger than that of excitation for the feature neuron group L, given a visual pattern, they will start to fire only if both neuron groups A_3 and A_4 are silent. They are silent when the following condition is satisfied: the output layer does not recognize the visual pattern, and the simulated eye-movement motor neurons stop to fire. If the neuron groups A_3 and A_4 are firing, they will reset the output selecting neurons, which send spikes to the output neurons to start the learning. Thus, it ensures that if the visual pattern is already learned or it is not stable at the input layer, the learning phase will not start. Once the neuron group L starts to fire, they will excite the neuron group A_1 to ensure that the neuron groups R and A_4 are silent during the learning process so that they will not interfere with the learning. The arbitration neuron group A_2 inhibit A_1 when the motor neurons fire. This ensures that after learning when the silicon retina or the visual pattern starts to move, the neuron group A_3 will inhibit L and the recognition phase starts.

c) Spike-based learning rule: The learning happens on the plastic synapses between the mapping neurons and the output neurons. Every spike of the post-synaptic neuron triggers the learning of a plastic synapse. The learning rule is Hebbian-like for excitatory synapses and anti-Hebbian-like for inhibitory synapses. It can be described as:

$$w_{ij}(t) = max((w_{ij}(t) + \alpha \Delta w_{ij}(t) \cdot S_j(t), 0) \quad (1)$$

$$\Delta w_{ij}(t) = \beta \sum_{t-\Delta t<t'<t} S_i(t') - 1 \quad (2)$$

where i and j denote the indexes of the pre-synaptic and post-synaptic neurons respectively. $\alpha > 0$ if the pre-synaptic neuron i is an excitatory neuron, and $\alpha < 0$ if neuron i is an inhibitory neuron. β denotes a learning rate. $w_{ij}(t_0) = 0$, where t_0 denotes when the learning starts. $S_i(t) = \sum_k \delta(t - t_k^i)$ denotes whether there is a spike generated by neuron i at time t, where t_k^i represents the timing of spikes and $\delta(x) = 1$ when $x = 0$ otherwise $\delta(x) = 0$. Once the weight change of a synapse accumulated to be larger than a predefined threshold, the learning algorithm will update the new synaptic weight onto the chips. Otherwise, the synapse will keep the current weight. When a neuron is activated longer than a threshold period ($80\,ms$ in this implementation), the learning of its incoming synapses will stop. After learning, the synapses can memorize the pre-synaptic neural activities. This learning rule ensures that the learning phase will only affect the firing rate but not the firing-or-not activity of the post-synaptic neurons. It makes learning more robust and predictable.

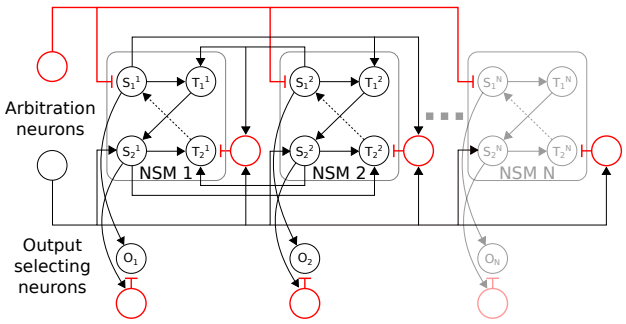

Fig. 4. Network structure that implements the automatic selection of output neurons during learning. The NSM structure is modified from [10]. Here within each NSM, the connections between neuron groups S_1 and T_1, and S_2 and T_2, as well as the winner-take-all structure built upon the neuron groups S_1 and S_2 are simplified for visualization. It only shows the connectivity between NSM 1 and NSM 2 as an example. Every pair of the NSMs has the same connectivity. The dotted line represents plastic synapses. Each circle represents a neuron population of 4 neurons that share the same connectivity.

d) Unsupervised learning: The output selecting neurons and a plastic winner-take-all mechanism are illustrated in Fig. 4. Multiple NSM structures of the type described in [10] implement the plastic winner-take-all mechanism. Each NSM has two states: S_1 and S_2. The network connectivity ensures that anytime only one NSM can stay at state S_1 as the winner and all the other NSMs are at state S_2. Once a NSM wins the competition, it will self-maintain its activity. Meanwhile, the synapses that are coming into the neuron group S_1 of the winner NSM will decrease their weights. Therefore, next time, this NSM will not be selected as the winner again. In detail, each NSM could stay at either state S_1 or S_2, meaning that the neuron groups S_1 or S_2 of the NSM are firing respectively. The neuron groups T_1 and T_2 will fire only if and when its two groups of in-coming synapses both receive spikes [10]. The inter-connected NSMs carry out a winner-take-all behavior because they always push each other to alter the states. Anytime only one NSM can stay at state S_1 and the others are at state S_2. After the competition, the neuron group S_1 of the winner NSM will decrease their in-coming synaptic weights according to the same learning rule as

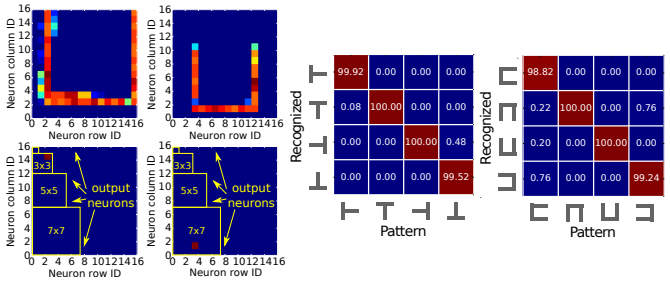

Fig. 5. The recognition of visual patterns after learning. Left: recorded neural activities of the input and output layers in a duration of 20 ms. Right: performance of the network in two simple tasks. Each element in the confusion matrix represents the percentage of spikes that belong to each of the four groups of output neurons compared to the total number of spikes when we show a visual pattern to the neuromorphic system.

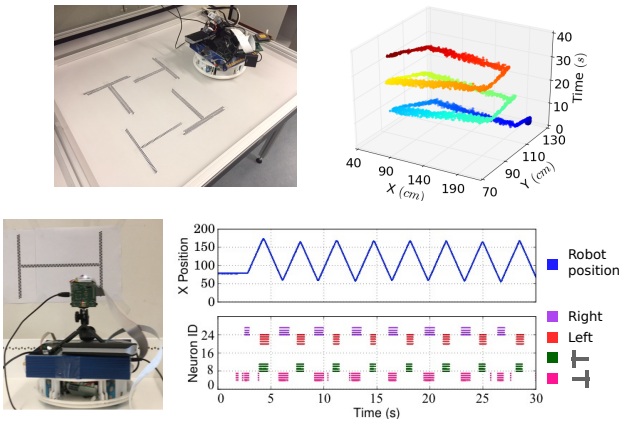

Fig. 6. Real-time real-world experimental results on an omnidirectional robot. We mount the silicon retina and neuromorphic processors on the robot. The output neurons on-chip send out events to drive the robot's motors to go either forward, backward, left, or right. Up: Robot navigates in the arena by recognizing the surface markings on the ground. Down: The robot keeps changing between left and right movements.

discussed above, except that initially $w_{ij}(t_0) > 0$. Next time this NSM will not become the winner again. After learning a visual pattern, with the firing of the motor neurons, the arbitration neuron groups A_3 and A_4 send spikes to reset all the NSMs to state S_2. This reset is necessary to prepare the learning of the next pattern. Because each group of the neurons S_1 excites a different group of the output neurons, the output neurons will take turns to learn different visual patterns. This mechanism enables learning in an unsupervised manner.

IV. Experimental Results

We present experiments in which objects either move with speed in the range of [60, 1.2k] pps (pixels per second) or continuously shake in front of the silicon retina. Here the pps represents how fast the perceived visual pattern moves on the silicon retina. 60 pps and 1.2k pps are measured values. 60 pps is the minimum speed, below which the generated events are too few for the network to recognize. 1.2k pps is the maximum speed, above which the response of the neurons and synapses on-chip are not fast enough to distinguish the visual patterns.

a) Performance and robustness: Because of the limited number of neurons on the prototype chips, every time we train the network with four different visual patterns composed of horizontal and vertical bars, namely a 'T' shape or a '⊔' shape symbol with different directions. Due to the up- and down-scaling mechanism within the connections, the output neurons reserve and represent the position and distance information of the input pattern during recognition. To test the recognition performance of the network after training, we show each pattern to the silicon retina for 1000 s. Fig. 5 shows that the network is robust to the noise and mismatch of neuromorphic devices and the environment. The dis-inhibition mechanism is necessary to achieve high recognition performance. For the same feed-forward network in Fig. 2, the measured recognition performance on-chip is 99.69% and 56.91% with and without dis-inhibition respectively. Although we choose to use these simple patterns due to the limited number of available neurons, the same principle can be scaled up for more complex patterns.

b) Power consumption: Without incoming events, the spiking neural network architecture implemented on neuromorphic devices only consumes static power. The static power dissipation is 945 μW for the DYNAP chip [16]. The primary source of the dynamic power consumption is due to neurons firing and spikes generation. The average mean firing rate of the neurons is 41.76 Hz when there is no visual pattern given to the input layer, and 55.73 Hz when there is one. The silicon retina chip consumes less than 5 mW when the input activity is not very intensive [12]. Thus the total power consumption of this neuromorphic hardware system is estimated to be less than a few mW.

c) Real-time real-world tasks: The time cost for recognizing each pattern depends on the properties of the neurons and synapses (e.g., the time constant, firing threshold, and the length of PSPs) on the feed-forward pathway. Each neuron takes time to integrate evidence from the asynchronous input. This delayed time, however, enables a temporal invariance in the range from tens of to several ms. It is essential for real-world tasks where the robot or object is not moving smoothly. We tested two scenarios in which the robot moved in an arena, continuously detecting and recognizing patterns that are associated with actions accordingly. Fig. 6 illustrates the neural activity and the robot trajectory. It proves that our spiking neural network architecture performs well in the noisy real-world environment.

V. Conclusion

We presented a spiking neural network architecture and an on-line off-chip training method that enable robust learning and recognition of visual patterns in noisy spiking neural networks and noisy environments. We demonstrated pattern recognition tasks in a closed-loop system composed of asynchronous neuromorphic sensors, processors, and robotic agents. In addition to solving practical engineering problems, the proposed network architecture might shed light on how the neurons are organized to carry out robust computation in biological networks.

REFERENCES

[1] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–36, 1990.

[2] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri, "A re-configurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses," *Frontiers in Neuroscience*, vol. 9, no. 141, pp. 1–17, 2015.

[3] S. Moradi, N. Qiao, F. Stefanini, and G. Indiveri, "A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs)," *Biomedical Circuits and Systems, IEEE Transactions on*, pp. 1–17, 2017.

[4] "Brain-inspired multiscale computation in neuromorphic hybrid systems (BrainScaleS)," FP7 269921 EU Grant, 2011–2015.

[5] E. Chicca, F. Stefanini, C. Bartolozzi, and G. Indiveri, "Neuromorphic electronic circuits for building autonomous cognitive systems," *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1367–1388, 9 2014.

[6] A. A. Faisal, L. P. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nature reviews neuroscience*, vol. 9, no. 4, pp. 292–303, 2008.

[7] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbruck, and R. Douglas, *Analog VLSI:Circuits and Principles*. MIT Press, 2002.

[8] J. Binas, G. Indiveri, and M. Pfeiffer, "Spiking analog VLSI neuron assemblies as constraint satisfaction problem solvers," in *International Symposium on Circuits and Systems, (ISCAS), 2016*. IEEE, 2016, pp. 2094–2097.

[9] T. Pfeil, A. Grübl, S. Jeltsch, E. Müller, P. Müller, M. Petrovici, M. Schmuker, D. Brüderle, J. Schemmel, and K.Meier, "Six networks on a universal neuromorphic computing substrate," *Frontiers in Neuroscience*, vol. 7, 2013.

[10] D. Liang and G. Indiveri, "Robust state-dependent computation in neuromorphic electronic systems," in *Biomedical Circuits and Systems Conference, (BioCAS), 2017*. IEEE, Oct. 2017, pp. 108–111.

[11] E. Neftci, J. Binas, U. Rutishauser, E. Chicca, G. Indiveri, and R. Douglas, "Synthesizing cognition in neuromorphic electronic systems," *Proceedings of the National Academy of Sciences*, vol. 110, no. 37, pp. E3468–E3476, 2013.

[12] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "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.

[13] G. Indiveri, E. Chicca, and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike–timing dependent plasticity," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 211–221, Jan 2006.

[14] H. Deubel and W. X. Schneider, "Saccade target selection and object recognition: Evidence for a common attentional mechanism," *Vision research*, vol. 36, no. 12, pp. 1827–1837, 1996.

[15] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," *Nature Reviews Neuroscience*, vol. 5, no. 3, p. 229, 2004.

[16] G. Indiveri, F. Corradi, and N. Qiao, "Neuromorphic architectures for spiking deep neural networks," in *Electron Devices Meeting (IEDM), 2015 IEEE International*. IEEE, Dec. 2015, pp. 4.2.1–4.2.14.