HAL open science

Unsupervised Learning of Spatio-Temporal Receptive Fields from an Event-Based Vision Sensor

Thomas Barbier, Céline Teulière, Jochen Triesch

To cite this version:

Thomas Barbier, Céline Teulière, Jochen Triesch. Unsupervised Learning of Spatio-Temporal Receptive Fields from an Event-Based Vision Sensor. 29th International Conference on Artificial Neural Networks,, 2021, Bratislava, Slovakia. pp.622-633, 10.1007/978-3-030-61616-8_50 . hal-03049596

HAL Id: hal-03049596

https://hal.science/hal-03049596v1

Submitted on 11 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised Learning of Spatio-Temporal Receptive Fields from an Event-Based Vision Sensor*

Thomas Barbier[1[0000−0003−4711−688X]], Céline Teulière[1[0000−0002−7253−6221]], and Jochen Triesch[1,2[0000−0001−8166−2441]]

[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal, F-63000 Clermont-Ferrand, France
[2] Frankfurt Institute for Advanced Studies, Frankfurt, Germany

Abstract. Neuromorphic vision sensors exhibit several advantages compared to conventional frame-based cameras including low latencies, high dynamic range, and low data rates. However, how efficient visual representations can be learned from the output of such sensors in an unsupervised fashion is still an open problem. Here we present a spiking neural network that learns spatio-temporal receptive fields in an unsupervised way from the output of a neuromorphic event-based vision sensor. Learning relies on the combination of spike timing-dependent plasticity with different synaptic delays, the homeostatic regulations of synaptic weights and firing thresholds, and fast inhibition among neurons to decorrelate their responses. Our network develops biologically plausible spatio-temporal receptive fields when trained on real world input and is suited for implementation on neuromorphic hardware.

Keywords: Event-Based Vision · Unsupervised Learning · Receptive Field · Neuromorphic Engineering · Spiking Neural Network · Spike Timing-Dependent Plasticity.

1 Introduction

Biological vision systems learn to make sense of their environments without much external supervision. Mimicking such unsupervised learning abilities in technical systems may pave the way for vision systems that do not require millions of manually labelled training examples, but that can learn in a much more autonomous fashion. How biological vision systems learn to see without supervision is still poorly understood, however. While fundamental mechanisms of neuronal and synaptic plasticity have been identified, there is still no model that could explain the development of vision in terms of these basic mechanisms. One hurdle

* This work was sponsored by a public grant overseen by the French National Agency as part of the "Investissements d'Avenir" through the IMobS3 Laboratory of Excellence (ANR-10-LABX-0016) and the IDEX-ISITE initiative CAP 20-25 (ANR-16-IDEX-0001). JT acknowledges support from the Johanna Quandt foundation.

to building such models is scale. Simulating detailed large-scale models of, e.g., the primate visual system on conventional (super-)computers is prohibitively expensive. This has led to growing interest in neuromorphic approaches that mimic principles of neural information processing and learning in hardware and can be vastly more energy efficient compared to conventional computing paradigms.

Event based cameras are inspired by the mammalian retina and represent a great starting point for creating biologically inspired artificial vision systems. They offer very high temporal resolution (in the order of μs) and low data rates. Importantly, they operate in a fully asynchronous way, i.e., there are no clearly defined image "frames". This necessitates to drastically change the design of the vision system. Spiking neural networks (SNNs) are the asynchronous analog to conventional neural networks, which makes them ideally suited for unsupervised learning from an event based camera.

Here, we propose a SNN model to learn spatio-temporal receptive fields from the output of a neuromorphic event-based vision sensor in an unsupervised fashion. The network combines three unsupervised learning mechanisms: spike timing-dependent plasticity, a homeostatic regulation of neuronal firing thresholds, and a multiplicative synaptic normalization to prevent unbounded growth of synapses. A simple fast inhibition scheme decorrelates neural responses and effectively prevents multiple neurons from developing identical receptive fields. We show that our network learns motion-sensitive receptive fields in an unsupervised fashion and that the learned receptive fields qualitatively resemble receptive fields observed in visual cortex.

2 Related Work

There have been various previous attempts at solving classification and recognition tasks using SNNs fed with event data. The works fall into two main categories. The first are spiking convolutional neural networks inspired by their frame-based counterparts. Most of the time these are transformations of successful deep learning convolutional neural networks and use a form of supervised training [12][17][14]. The second category uses spike timing-dependent plasticity (STDP) to train an SNN in an unsupervised way [11][4][1][7][6]. For example, Akolkar et al. [1] demonstrate the possibility to learn visual receptive fields (RFS) in an event-driven framework. However, their attempt is limited to learning purely spatial RFs without any time dependence, i.e. they do not consider the encoding of image motion.

The primate visual system uses specific populations of neurons tuned to different motion directions and velocities to estimate object motion. Some work has been done on bio-inspired ways of sensing motion using event-driven data. Specifically, Tschechne et al. [16] have proposed a model using filters with spatio-temporal tuning to compute the optical flow of a scene. Haessig et al. [5] have proposed a spiking neural network implemented on IBM's TrueNorth neurosynaptic system inspired by the Barlow & Levick method for optical flow estimation. Orchard et al. [9] have created a motion sensing SNN using synaptic delays. Each

neuron is designed to detect a specific motion direction and speed. The main limitation is that the delays and orientations for the whole population need to be set by hand. Hopkins et al. [6] also describe motion estimation via synapses with delays, but do not consider learning. Most recently, Paredes-Vallés et al. [10] have proposed a framework for learning motion sensitive receptive fields from an event-based camera via a form of STDP in an unsupervised fashion. Their work is most closely related to ours, but there are a number of differences. First, the depression part of their STDP rule does not require presynaptic input spikes to arrive shortly after a postsynaptic spike, which is an important feature of biological STDP. Second, they use a convolutional network architecture, which enforces the development of identical receptive fields in all parts of the visual field. This prevents the receptive fields at different locations from adapting to systematic differences in the statistics of optical flow signals across the visual field as is common, e.g., during ego motion. Finally, our implementation is fully event-based, which makes it well-suited for implementation on neuromorphic hardware.

3 Methods

3.1 Spiking Neural Network

We designed our spiking neural network to couple efficiently with an event based camera. We use the sensor as our input layer and the second layer comprises a set of spiking neurons, each connected to a specific region of the event based camera via weighted synapses which define the neuron's receptive field. If one of the camera's pixels records an event, the neurons connected to it will receive an excitatory input depending on the strength of the synapse. In order to differentiate ON and OFF events, neurons are connected with at least two synapses to a pixel, one for each event polarity. Furthermore, we allow pixels to connect to the spiking neurons with different synaptic delays (described below), to enable the development of motion tuning.

We chose the well-known Leaky-Integrate and Fire (LIF) point neuron model. $V(t)$ denotes the neuron's membrane potential at time t. Updates of $V(t)$ are only performed upon receiving an event. More precisely, an input event from a pixel i creates an excitatory post-synaptic potential, which increases the membrane potential according to the strength of the corresponding synapse $w_i(t)$. If the neuron's membrane potential exceeds a threshold V_θ, the neuron "spikes" and its membrane potential returns to the resting value, which we define to be zero. Between the arrival of successive synaptic inputs to the neuron at times t and $t + \Delta t$, the membrane potential exponentially decays back to its resting value with the membrane time constant τ_m. Taken together, $V(t)$ evolves according to:

$$\tilde{V}(t + \Delta t) = V(t)e^{\frac{-\Delta t}{\tau_m}} + w_i(t) \tag{1}$$

$$V(t + \Delta t) = \begin{cases} \tilde{V}(t + \Delta t) : \tilde{V}(t + \Delta t) < V_\theta \\ 0 : \tilde{V}(t + \Delta t) \geq V_\theta \ . \end{cases} \tag{2}$$

When the neuron is inhibited (see section 3.4 below), however, its membrane potential is not updated for the duration of the inhibition.

3.2 Synaptic Delays and Spike Timing-Dependent Plasticity

To allow for the learning of motion sensitive receptive fields, each pixel of the event based camera can be coupled to an LIF neuron via D synapses with different delays. Therefore, a neuron can receive, e.g., the same on-event from the same pixel at D different times. Similarly, it can simultaneously receive events having occurred at different pixels at different times and therefore be sensitive to image motion.

Synaptic weights adapt according to a simple model of Spike Timing-Dependent Plasticity (STDP), a well-known unsupervised learning mechanism. In STDP, the sign and magnitude of a synaptic weight change depend on the relative timing of pre- and post-synaptic spikes. In the most common form of STDP, presynaptic spikes arriving shortly before a postsynaptic spike will lead to a long-term potentiation (LTP) of the synapse, while the reverse timing leads to long-term depression (LTD). In our implementation of LTP, as soon as a neuron spikes, its synaptic input connections will increase depending on the timing of the last input they received. This is a symmetric interpretation, where each presynaptic spike is paired with the last postsynaptic spike, and each postsynaptic spike is paired with the last presynaptic spike. Each synapse undergoes an instantaneous change in weight depending on an exponential relationship between the time difference between the timestamp t_i of the last input arriving at synapse i and the time of the postsynaptic spike $t_s > t_i$:

$$\Delta w_i^{\mathrm{LTP}} = A_{\mathrm{LTP}}\, e^{\frac{-|t_i - t_s|}{\tau_{\mathrm{LTP}}}} \quad , \tag{3}$$

with A_{LTP} and τ_{LTP} controlling, respectively, the height and duration of the potentiation window. Any input spike arriving after a postsynaptic spike ($t_i > t_s$) leads to depression of the synaptic weight:

$$\Delta w_i^{\mathrm{LTD}} = -A_{\mathrm{LTD}}\, e^{\frac{-|t_s - t_i|}{\tau_{\mathrm{LTD}}}} \quad , \tag{4}$$

where A_{LTD} and τ_{LTD} control, respectively, the height and duration of the depression window. Note that in this formulation, multiple presynaptic spikes can interact with the last postsynaptic spike to induce depression. Synapses whose weight would become negative due to LTD are set to zero. The STDP rule applies equally to all synapses with different time delays. The relevant time difference $|t_s - t_i|$ is always the one between the *arrival* of the presynaptic spike (occasionally delayed with respect to the time it was generated) and the moment of the postsynaptic spike.

3.3 Threshold Adaptation and Synaptic Normalization

To avoid the necessity to fine-tune the neuron's firing thresholds we use a homeostatic regulation to enforce a certain target firing rate $S^* = 0.75$ spikes.s^{-1}.

The threshold is adapted automatically every second depending on the difference between an estimate of the recent spiking rate $S(t)$ and the desired S^* as:

$$\Delta V_\theta = A_\theta \left(S(t) - S^*\right) , \tag{5}$$

where A_θ is a scalar parameter controlling the rate of change of the spike threshold. To estimate the recent spiking rate $S(t)$, we store the number of spikes that occurred during each of the last 10 seconds in a 10-element ring buffer $S_i(t)$, $i = 1, \ldots, 10$. The spiking rate $S(t)$ is estimated as $S(t) = 0.1 \times \sum_{i=1}^{10} S_i(t)$.

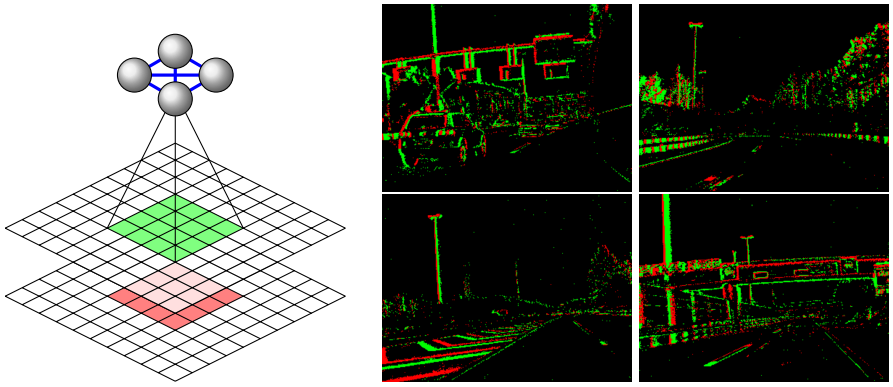
To avoid unbounded growth of synaptic weights we use a simple weight normalization mechanism, which normalizes the weights projecting to a single neuron in a multiplicative fashion. Different groups of synapses onto the same neuron are normalized separately. Each neuron receives inputs from two channels (for on and off events) and D synaptic delays from a pixel array of width W and height H. This amounts to a total number of $2 \times D \times W \times H$ synapses per neuron. The synapses from the same channel and the same synaptic delay form a synapse group, i.e., each synapse group comprises $W \times H$ synapses and there are $2D$ such groups. After every spike of the neuron, all input synapses of all synapse groups are multiplicatively rescaled such that the L_2 norms of the weights of each group equal a parameter λ, chosen empirically. This separate normalization of different synapse groups ensures that on- and off-channels and the different synaptic delays all contribute equally to activating the neuron. Such multiplicative re-scaling of the efficacies of groups of synapses could be the result of a local competition for synaptic building blocks such as neurotransmitter receptors [15].

3.4 Lateral Inhibition

We use a simple lateral inhibition mechanism to facilitate the learning of diverse receptive fields tuned to different orientations and movement directions and speeds. In the network, N neurons are connected to the same set of input pixels as shown in Fig. 1a. Such neurons are linked by inhibitory connections. When one neuron spikes, it immediately inhibits the other $N - 1$ neurons from firing for a fixed duration $T_I = 8$ ms. This prevents neurons which receive input from the same region of the sensor from firing at roughly the same time, which would imply similar synaptic weight updates and lead to similar receptive fields.

3.5 Coupling of the Event Based Camera to the Network

The input to the network comes from the DAVIS346B [3], a DVS sensor of 346×260 pixels. We tile the pixel array into 10×10 pixel squares (ignoring the 6 rightmost columns), giving $\frac{340 \times 260}{10 \times 10} = 884$ tiles. Each input tile projects to $N = 4$ neurons connected by inhibitory synapses (compare Fig. 1a and Sec. 3.4 on lateral inhibition). This implies a network size of 3536 neurons. Each neuron



(a) (b)

Fig. 1: (a) Groups of $N = 4$ neurons are connected to the same patch of input pixels providing On (green) and Off (red) events. The neurons are linked by inhibitory connections (blue). A neuron can be connected to pixels of its receptive field by D synapses with different delays to gain localised motion sensing properties (not shown). (b) Examples of input events from the driving sequence.

receives inputs from $2 \times D \times 100$ synapses corresponding to the 2 event polarities, D time delays, and $W \times H = 100$ input pixels. For $D = 3$, this amounts to 2 121 600 input synapses for the entire network. The event-based spiking neural network simulation was implemented in C++ as a DV-software module. Running on a standard Intel Core i5-8365U CPU @1.60GHz without particular optimization, the network showed real-time performance when using a single synaptic delay ($D = 1$) and near real-time performance when $D = 3$.

4 Results

To demonstrate the learning abilities of our approach we tested different network configurations on both natural and synthetic visual input.

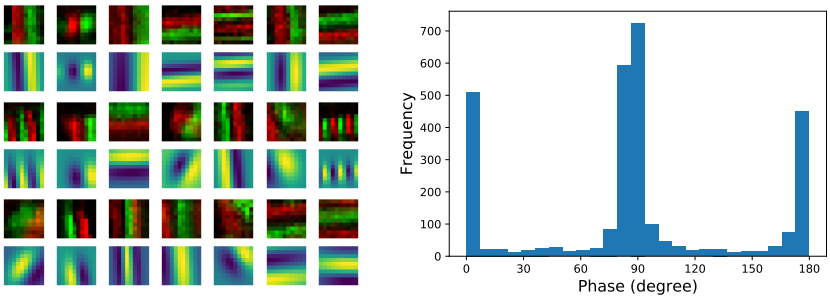
4.1 Development of orientation-tuned receptive fields

To test the network's ability to develop diverse orientation-tuned receptive fields as observed in visual cortex, we used a sequence from the DDD17: DAVIS Driving Dataset [2]. It features a 3 minutes long driving sequence mostly on freeways and highways at a somewhat regular speed. Figure 1b shows 4 examples of short time slices of events. The data set features various types of visual inputs such as cars, traffic signs, poles, trees, buildings, safety barriers, road markings, etc.

Table 1 lists the parameters used in our spiking neural network. Changes in parameters specific to one test will be mentioned in the text. In this first experiment we focus on the spatial structure of the learned receptive fields and

A_{LTP} (mV)	A_{LTD} (mV)	τ_{LTP} (ms)	τ_{LTD} (ms)	τ_m (ms)	V_θ (mV)	A_θ	S^* (spikes.s^{-1})	T_I (ms)	λ
0.077	0.021	7	14	18	30	4	0.75	8	4

Table 1: Parameters used for the learning.



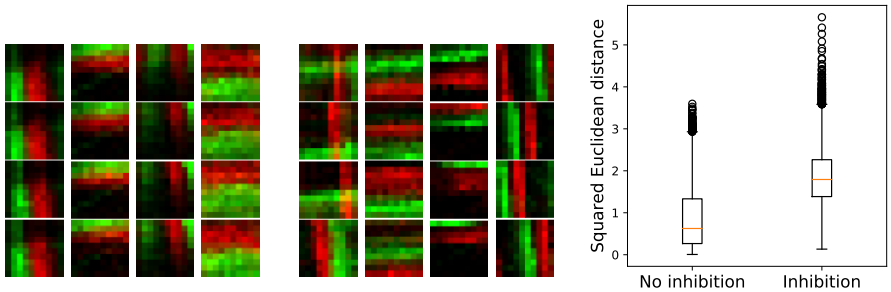
(a) (b)

Fig. 2: (a) Selected examples of learned receptive fields (rows 1, 3 and 5) and corresponding Gabor fits (rows 2, 4 and 6) showing tuning to different orientations and scales. (b) Histogram of the network's receptive field orientations obtained from the fitting of Gabor functions. Horizontal and vertical orientations are strongly over-represented, showing adaptation to the scene statistics in line with the dominance of these orientations in man-made (but also natural) environments.

we thus consider a network without multiple synaptic delays (i.e., $D = 1$). The initial synaptic weights are drawn randomly from a uniform distribution. Figure 2a shows examples of learned receptive fields after 15 minutes of training (5 repetitions of the driving sequence). Green/red pixels represent synapses transmitting On/Off events, respectively. The weight strength is represented by the color intensity. Yellow areas indicate regions where the neuron is sensitive to both On and Off events.

The receptive fields of simple cells in primary visual cortex are well described by Gabor functions. To test the biological plausibility of our learned receptive fields, we fitted Gabor functions to them. 83% of receptive fields obtained a good fit (sum of squared errors ≤ 5). Example fits are shown in Fig. 2a. They exemplify that filters of different orientation and scale are learned. Figure 2b shows the histogram of fitted Gabor orientations. Horizontal and vertical orientations are over-represented, resembling the oblique effect in visual perception [8].

To test the importance of the lateral inhibition mechanism, we studied the effect of disabling it in Fig. 3. Each column in Fig. 3a represents the receptive fields of 4 neurons connected to the same input patch after learning without lateral inhibition. Even though the receptive field initialisation was different, all 4 neurons have learned very similar receptive fields. In contrast, the lateral



(a) (b) (c)

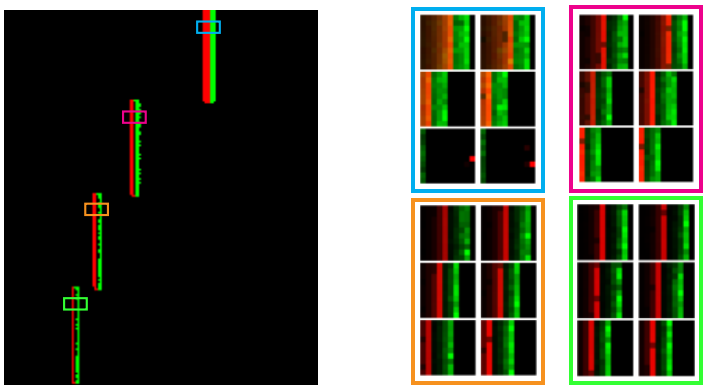
Fig. 3: Lateral inhibition diversifies receptive fields. Examples of 16 neurons' receptive fields learnt without (a) and with (b) lateral inhibition. All neurons in a column receive the exact same inputs from the event sensor, but start with a different random initialisation of their synaptic weights. (c) Boxplot of the squared Euclidean distances between synaptic weights of neurons receiving similar inputs with and without lateral inhibition.

inhibition leads to more diverse receptive fields (Fig. 3b). Figure 3c quantifies this effect by showing box plots of the distributions of all pairwise squared Euclidean distances between the receptive fields learned at the same location. Inhibition greatly improves the diversity of receptive fields.

4.2 Development of motion tuned receptive fields

To test the network's ability to develop motion tuned receptive fields, we introduced synapses with $D = 3$ different time delays of 0, 10, and 20 ms. We used a higher starting threshold $V_\theta = 700$mV to work with the extra synapses, as well as a lower target spiking rate of $S^* = 0.15$ spikes.s^{-1} which is sufficient in this sequence of fewer events. All other parameters are still set to Table 1 values. We first tested the motion learning capacity in a controlled setting using synthetic stimuli. We generated a simple sequence showing 4 vertical bars moving horizontally at predefined speeds of 420, 210, 140 and 105 pixels per second across the pixel array. The video was recorded at a framerate of 1200 frames per second and then converted to events using the Open Event Camera Simulator, ESIM [13]. A snapshot of the events can be seen in Figure 4a.

The speed of the bars influences the number of events produced, which means that neurons should have different threshold values to appropriately fit the input. To accommodate this problem, we first set the neurons' thresholds to a high value to prevent spiking. The threshold adaptation described in Sec. 3.3 then decreases the threshold until the neurons starts spiking and its receptive field will adapt to the input. With the delays between the synapses being fixed at 10 ms, we expected to see a consistent shift in the receptive fields corresponding to the different synaptic delays. For instance, for the fastest bar the expected shift was $10\,\text{ms} \times 420\,\text{px.s}^{-1} = 4.2\,\text{px}$. Indeed, after exposing the network to about



(a) (b)

Fig. 4: Motion sensitive receptive fields develop from STDP with multiple synaptic delays. (a) On and off events recorded by the DV-software visualisation GUI for 4 vertical bars moving at speeds of 420, 210, 140 and 105 px/s (top to bottom). (b) Receptive fields of pairs of neighboring neurons receiving the 4 vertical bars as input (arranged by decreasing speed from top left to bottom right). Each neuron has 3 groups of input synapses with delays of 0, 10, and 20 ms (top to bottom). The expected displacements of the 4 bars during 10 ms are 4.2, 2.1, 1.4 and 1.05 pixels, which matches the horizontal shifts of receptive field structures for different synaptic delays.

50 repetitions of the vertical inputs, we obtained the receptive fields in Fig. 4b matching the expected shifts, even though the expected displacement could go down to a fraction of a pixel.

Next, we investigated the network's ability to develop motion tuned receptive fields for natural input by training it on the driving sequence of Fig. 1b. We were particularly interested in systematic differences in tuning properties across the visual field reflecting typical optic flow patterns occurring during driving. We chose a higher starting membrane potential threshold of $V_\theta = 150$mV. Learned spatio-temporal receptive fields across the entire sensor array are shown in Fig. 5. Only one out of four neurons per location (compare Fig. 3) is shown. The receptive fields of some regions of the sensor are enlarged in Fig. 5b. Overall, a large variety of receptive fields tuned to different orientations, motion directions, and speeds have been learned. Importantly, we observe systematic differences in tuning properties across different parts of the visual field. In particular, the left and right regions of the network have mostly learned vertically tuned receptive fields, whereas the top and bottom parts have developed horizontally tuned receptive fields (compare pink, blue, and orange regions). This is coherent with the statistics of the sensory input. The left and right regions of the data sequence contain many poles, trees, and buildings, which due to the motion of the car will generate vertically aligned events. In contrast, the top and bottom parts con-

tain bridges, highway panels, and road markings, generating mostly horizontally aligned events due to the motion of the car. Overall, the receptive fields show an interesting adaptation to the spatial orientation statistics of the input across the visual field.

The learned receptive fields also capture information about the optic flow statistics of the sequence. The shifts between receptive fields of different synaptic delays approximate the average speed of objects passing through that region of the sensor. We find that bigger shifts occur in the outer region of the sensor, whereas the inner regions exhibit smaller shifts (compare top and bottom part of orange region), corresponding to large optic flow in the periphery and small optic flow in the center. Considering the polarity of shifts across different regions of the pixel array, we find, as expected, that they reflect the dominant optic flow pattern of ego motion corresponding to objects moving radially outward to the periphery.

5 Discussion

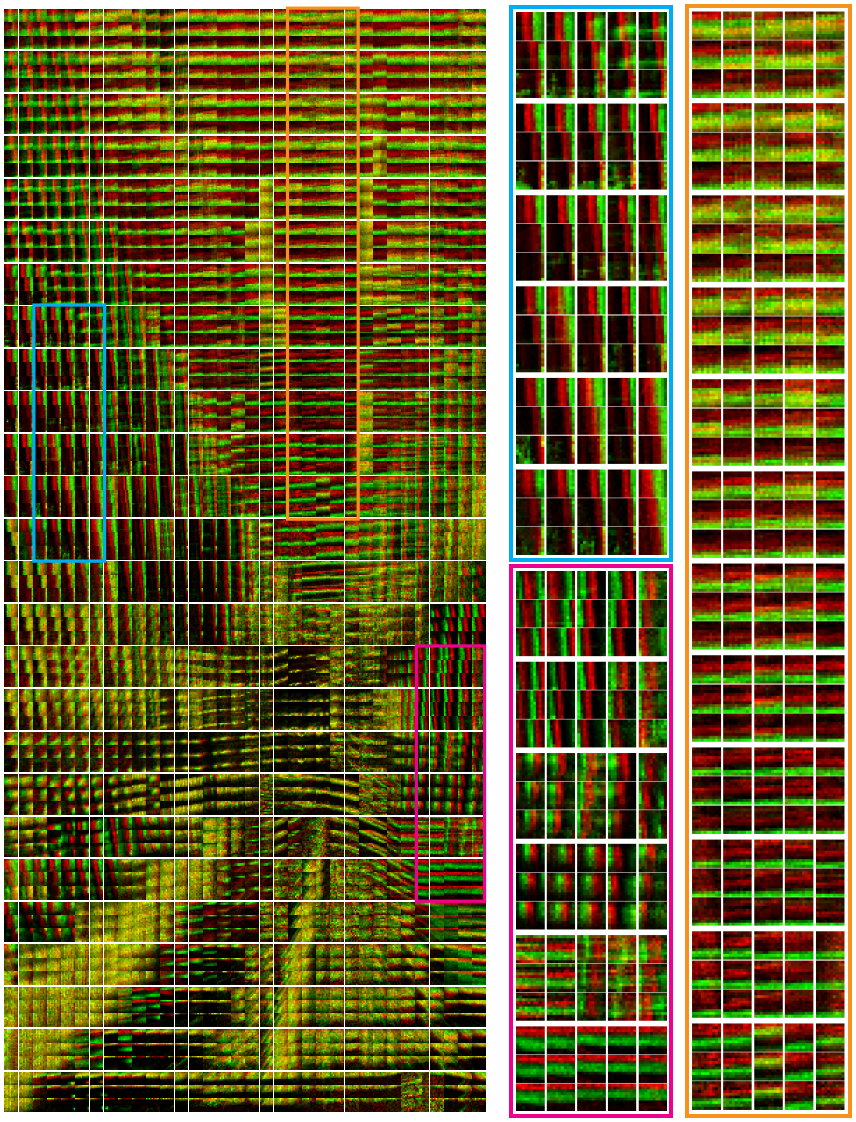
We have presented a spiking neural network that learns motion-sensitive receptive fields from the input of an event based camera in an unsupervised fashion. Motion tuning arises from spike timing-dependent plasticity (STDP) with multiple synaptic delays combined with homeostatic mechanisms and a simple lateral inhibition scheme to diversify tuning properties. The mechanisms used are all biologically inspired, but were not intended as accurate models of biological reality. Among the biggest idealizations are the instantaneous lateral inhibition to decorrelate responses of neurons with overlapping receptive fields and the group-wise normalization of synaptic inputs. The latter could be biologically plausible, however, if the different groups of synapses were considered to reside on separate dendritic branches [15]. This seems plausible for synapse groups with short vs. long delays, which could correspond to inputs to more proximal vs. more distal dendritic branches, respectively. Similarly, On an Off channel inputs could also be sorted to different dendritic branches during development based on their correlations.

Our work suggests a number of avenues for future research. First, we would like to extend our approach to active binocular vision, considering the simultaneous learning of disparity representations and vergence eye movements in a fully spiking implementation. Second, scaling up our approach to more complex spiking neural networks using neuromorphic hardware is an exciting topic for future research.

References

1. Akolkar, H., Panzeri, S., Bartolozzi, C.: Spike time based unsupervised learning of receptive fields for event-driven vision. In: IEEE Int. Conf. on Robotics and Automation. IEEE (May 2015)

2. Binas, J., Neil, D., Liu, S.C., Delbruck, T.: Ddd17: End-to-end davis driving dataset (2017), https://arxiv.org/abs/1711.01458
3. Brandli, C., Berner, R., Yang, M., Liu, S.C., Delbruck, T.: A 240 180 130 dB 3 μs latency global shutter spatiotemporal vision sensor. IEEE Journal of Solid-State Circuits **49**(10), 2333–2341 (Oct 2014)
4. Diehl, P.U., Cook, M.: Unsupervised learning of digit recognition using spike-timing-dependent plasticity. Frontiers in Computational Neuroscience **9** (Aug 2015)
5. Haessig, G., Cassidy, A., Alvarez, R., Benosman, R., Orchard, G.: Spiking optical flow for event-based sensors using IBMs TrueNorth neurosynaptic system. IEEE Transactions on Biomedical Circuits and Systems **12**(4), 860–870 (Aug 2018)
6. Hopkins, M., Pineda-García, G., Bogdan, P.A., Furber, S.B.: Spiking neural networks for computer vision. Interface Focus **8**(4), 20180007 (Jun 2018)
7. Kheradpisheh, S.R., Ganjtabesh, M., Thorpe, S.J., Masquelier, T.: STDP-based spiking deep convolutional neural networks for object recognition. Neural Networks **99**, 56–67 (Mar 2018)
8. Li, B., Peterson, M.R., Freeman, R.D.: Oblique effect: a neural basis in the visual cortex. Journal of neurophysiology **90**(1), 204–217 (2003)
9. Orchard, G., Benosman, R., Etienne-Cummings, R., Thakor, N.V.: A spiking neural network architecture for visual motion estimation. In: IEEE Biomedical Circuits and Systems Conference (BioCAS) (Oct 2013)
10. Paredes-Valles, F., Scheper, K.Y.W., Croon, G.C.H.E.D.: 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 (2019)
11. Paulun, L., Wendt, A., Kasabov, N.: A retinotopic spiking neural network system for accurate recognition of moving objects using NeuCube and dynamic vision sensors. Frontiers in Computational Neuroscience **12** (Jun 2018)
12. Perez-Carrasco, J.A., Serrano, C., Acha, B., Serrano-Gotarredona, T., Linares-Barranco, B.: Spike-based convolutional network for real-time processing. In: IAPR Int. Conf. on Pattern Recognition. IEEE (Aug 2010)
13. Rebecq, H., Gehrig, D., Scaramuzza, D.: Esim: an open event camera simulator. In: Conference on Robot Learning (Oct 2018)
14. Stromatias, E., Soto, M., Serrano-Gotarredona, T., Linares-Barranco, B.: An event-driven classifier for spiking neural networks fed with synthetic or dynamic vision sensor data. Frontiers in Neuroscience **11** (Jun 2017)
15. Triesch, J., Vo, A.D., Hafner, A.S.: Competition for synaptic building blocks shapes synaptic plasticity. Elife **7**, e37836 (2018)
16. Tschechne, S., Sailer, R., Neumann, H.: Bio-inspired optic flow from event-based neuromorphic sensor input. In: Advanced Information Systems Engineering, pp. 171–182. Springer Berlin Heidelberg (2014)
17. Zhao, B., Ding, R., Chen, S., Linares-Barranco, B., Tang, H.: Feedforward categorization on AER motion events using cortex-like features in a spiking neural network. IEEE Transactions on Neural Networks and Learning Systems **26**(9), 1963–1978 (Sep 2015)



(a) (b)

Fig. 5: Motion tuned receptive fields learned at different image locations. (a) Receptive fields learned across the entire field of view. Each receptive field has three sub-fields (arranged vertically) corresponding to different synaptic delays. (b) Enlarged view of marked groups of receptive fields in (a). See text for details.