Deep Spiking Neural Network for object tracking

F. M. Quintana[1] M.P. Guerrero-Lebrero[1], F. Perez-Peña[1], P. L. Galindo[1], and E. Guerrero[1]
fernando.quintana@uca.es
maria.guerrero@uca.es
fernandoperez.pena@uca.es
pedro.galindo@uca.es
elisa.guerrero@uca.es

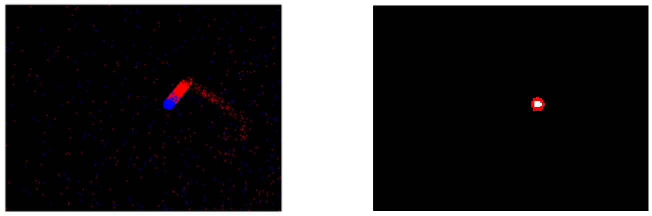
Universidad de Cádiz, Spain

1 Introduction

Recent developments in neuromorphic computing systems have focused on the creation of new hardware based on biological characteristics. An example of such hardware are event cameras, in which pixels of the Dynamic Vision Sensor (DVS) work independently (i.e., asynchronously). Each pixel announces when it discloses a relative change in the illumination intensity that is above or below a defined intensity threshold. Event cameras are robust and accurate motion detectors and automatically filter out any temporarily redundant information [1]. This makes them extremely useful for scenes with motion like high-speed counting, or driving safety systems. Different spiking neuron models have been introduced, such as Izhikevich model [2], Leaky Integrate and Fire (LIF) model [3] and Spike Response Model (SRM) [4]. Despite the success of gradient methods in training traditional Artificial Neural Networks (ANNs), their implementation on Spiking Neural Networks (SNNs) is still problematic when extracting gradient information from output spike times, and biological inspired algorithms have been used [5]. Some kind of metaheuristics can be applied for training SNN. For example, in [4] the connectivity and the sign of the connectivity (the neuron is excitatory if the sign is positive and inhibitory if negative) of a SRM are evolved for vision-based robot control using a Genetic Algorithm (GA). In [6], an adaptive genetic algorithm is adopted to evolve the weights of the SRM model for robot navigation, while in [7] a parallel differential evolution has been employed to evolve the weights of the SRM. Some other works mixed topology and learning, as in [8], where a Pareto-based multi-objective genetic algorithm is used to evolve the connectivity, weights and delays of the SRM or in [9], where they propose a population-based neuroevolution strategy for a recurrent network learning and evolution of the network topology. As an alternative approach surrogate gradient overcomes the limitations related to the discontinuous nonlinearity, as well as can help to control the algorithmic complexity associated with the SNN learning. Different methods have been proposed for training SNNs. Especially surrogated gradient are being used, since they can be very suitable for custom low-power neuromorphic devices [10, 11]. The task that is intended to be addressed here corresponds to the Visual Object Tracking (VOT) problem in which, given the initialization of a specific objective (position of the object x,y), the trajectory of the objective will be followed in the recording set obtained by the event camera. VOT have been developed very fast in the past few years, most of these works have been developed using Deep Learning (DL) (especially Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)), however, few of them have used SNN to solve this type of problem [12]. This work aims to show the behavior of SNN in prediction problems where different tasks are carrying out to develop a learning system for object detection at high speed: 1) Creation of an event-based dataset 2) Development of a neuromorphic system based on surrogate gradient optimization for the position prediction of moving objects at high speed. 3) Validation and test.

2 Methodology

In order to perform supervised learning with BackPropagation Through Time (BPTT), it is necessary to have a labeled dataset. In our case, as we are dealing with a regression problem using data obtained with a DVS for object tracking, we need a dataset containing the position of that object at each time instant we specify. Our dataset consists of a set of DVS [1] recordings of an object (in our case a ball) and its position each microsecond. Due to the nature of the data, they are asynchronous events with a temporal resolution of $1\mu s$, so we obtained the position of the object at that resolution. To do this, we used the E2VID [13] software to generate a series of frames from the event flow (Fig. 1a), in which the centers of the ball are calculated applying computer vision techniques (Fig. 1b). Once the position in each frame is obtained, a cubic interpolation is performed to calculate the position in each desired instant of time.



(a) Frame reconstruction of 640x480 pixels using E2VID software.

(b) Frame processing of 1a to detect the center of the object.

2.1 Deep SNN

For the network used in this project, the neuron model used is a LIF with a soft-reset mechanism (Eq. 2)

$$v_j(t) = \alpha v_j(t-1) + \sum_i W_{ij} x_i(t) - z_j(t-1) v_{th} \tag{1}$$

$$z_j(t) = H(v_j(t) - v_{th}) \tag{2}$$

Where v_j is the neuron voltage, α the decay constant, W_{ij} the synaptic weight between pre-synaptic neuron i and post-synaptic neuron j, v_{th} the neuron threshold, z_j the output spike and H the activation function (step function). Since we are working with spatio-temporal data, the topology of the network consists of a set of convolutional layers with LIF neurons, to record the spatial and temporal information from the DVS, following a fully-connected layer and an output layer that predicts the position of the object detected based on the voltage value of the neuron (Fig. 2). Since we are obtaining the absolute position of the object, the neuron reset mechanism on the last layer is not applied, and a sigmoid activation function over the voltage is applied instead of the step function. Due to the non-differentiable nature of the activation function, surrogate gradient

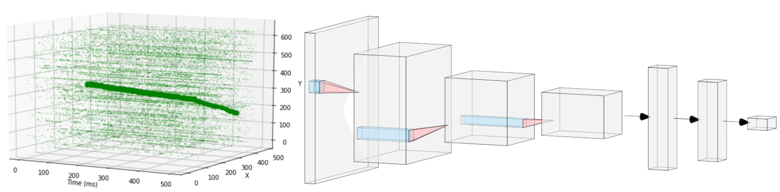


Fig. 2: Neural Network achitecture used. The spatio-temporal pattern obtained from the DVS is fed into the network, composed of 4 convolutional layers, 1 fully-connected layer and the output corresponding the the (X, Y) object position prediction.

[10, 14] is a method that is being widely used to train this type of neural network using known techniques such as BPTT. This method approximates the gradient of the non-differentiable step function.

3 Conclusions and Future Work

We are investigating surrogate gradient as optimization methods in Deep SNN for regression problems. A SNN able to detect a ball at high speed is being developed in which the voltage potential of the output neurons correspond, in real time, with its position, making possible its application in robotic systems that require fast object tracking. As a future work, training and validation over the network and dataset design would be performed using PyTorch framework, as well as the deployment of the system into a robotic platform, for object identification and tracking.

Acknowledgements

Fernando M. Quintana would like to acknowledge the Spanish *Ministerio de Ciencia, Innovación y Universidades* for the support through FPU grant (FPU18/04321). This work was also supported by the project NEMOVISION from the *Ministerio de Ciencia e Innovación y Agencia Estatal de Investigación* , PID2019-109465RB-I00/ MICIU/AEI/10.13039/501100011033.

References

1. IniVation, "Understanding the performance of neuromorphic event-based vision sensors," tech. rep., 2020.
2. E. Izhikevich, "Solving the distal reward problem through linkage of STDP and dopamine signaling," *Cereb. Cortex*, Jan. 2007. In press.
3. A. Burkitt, "A review of the integrate-and-fire neuron model: II. inhomogeneous synaptic input and network properties," *Biological Cybernetics*, 2006.
4. W. Gerstner and W. Kistler, *Spiking neuron models: single neurons, populations, plasticity*. Cambridge University Press, 2002.
5. F. M. Quintana, F. Perez-Peña, and P. L. Galindo, "Bio-plausible digital implementation of a reward modulated stdp synapse," *Neural Computing and Applications*, vol. 34, pp. 15649–15660, 9 2022.
6. H. Hagras, A. Pounds-Cornish, M. Colley, V. Callaghan, and G. Clarke, "Evolving spiking neural network controllers for autonomous robots," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2004, pp. 4620–4626, 2004.
7. N. G. Pavlidis, D. K. Tasoulis, V. P. Plagianakos, G. Nikiforidis, and M. N. Vrahatis, "Spiking neural network training using evolutionary algorithms," *Proceedings of the International Joint Conference on Neural Networks*, vol. 4, pp. 2190–2194, 2005.
8. Y. Jin, R. Wen, and B. Sendhoff, "Evolutionary multi-objective optimization of spiking neural networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4668 LNCS, pp. 370–379, 2007.
9. H. Qiu, M. Garratt, D. Howard, and S. Anavatti, "Evolving spiking neural networks for nonlinear control problems," *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI 2018*, pp. 1367–1373, 1 2019.
10. E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 51–63, 2019.
11. F. M. Quintana, F. Perez-Peña, P. L. Galindo, E. O. Neftci, E. Chicca, and L. Khacef, "Etlp: Event-based three-factor local plasticity for online learning with neuromorphic hardware," 1 2023. [arXiv preprint http://arxiv.org/abs/2301.08281].
12. Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *International Journal of Computer Vision*, vol. 113, p. 54–66, 2015.
13. H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3852–3861, 2019.
14. F. Zenke and T. P. Vogels, "The Remarkable Robustness of Surrogate Gradient Learning for Instilling Complex Function in Spiking Neural Networks," *Neural Computation*, vol. 33, pp. 899–925, 03 2021.