Appendix 1 to
ESA AO/1-1-10873/21/NL/GLC/my

Investigation of low energy Spiking Neural Networks based on temporal coding for scene classification

Study reference number: 21-8601
Type of activity: Standard study

Project Summary

Objective:

Study the applicability of Spiking Neural Networks (SNNs) based on temporal coding for onboard Artificial Intelligence applications, focusing on scene classification case studies. To this aim, we will investigate the trade-offs between accuracy and complexity of different SNN models based on temporal coding. In addition, assuming the availability of dedicated chips, the project aims to establish an upper-bound for the energy per inference of one of the investigated SNN models.

Target university partner competences

Artificial Intelligence, Spiking Neural Networks

ACT provided competences.

Artificial Intelligence, embedded computing, onboard AI

Keywords

Spiking Neural Networks, Scene classification, temporal coding, energy efficiency, embedded computing

Study Objective

This study aims to perform a preliminary investigation of the potential benefits of Spiking Neural Networks (SNNs) based on temporal coding for onboard Artificial Intelligence (AI) applications, considering the case study of scene classification. To achieve this goal, state of the art SNN models are to be compared in terms of accuracy and complexity (here considered as the number of synaptic operations, number of spikes per neuron, and others) on the EuroSAT RGB and EuroSAT multispectral datasets. To this aim, proper training algorithms for the SNN models shall be also evaluated and selected. Eventually, we aim to establish an upper-bound for the energy per inference of selected SNN models, assuming the availability of a dedicated neuromorphic chip.
The results of the analysis will highlight the possible advantages and drawbacks of SNN models compared to Artificial Neural Networks (ANNs), which represent the state of the art for scene classification.

Background and Study Motivation

Interest in AI, and in particular ANNs, on board satellites is growing. The use of onboard AI could be exploited to mitigate the bandwidth requirements of Earth Observation satellites by avoiding the download of corrupted or unmeaningful data, to perform early-detection of potential catastrophic events [1,2] and other applications. However, small Earth observation satellites, such as CubeSats, tend to have small memory and power budgets [2]. Therefore, the use of energy-efficient algorithms could increase the number of potential AI algorithms executed on board, and thus the degree of 'intelligence' of the satellite, given a fixed energy budget.

Previous work carried out by ESA's Advanced Concepts Team (ACT), aimed to study the use of a specific neuromorphic sensor for space landing scenarios [3-5]. This project aims to investigate the usability of SNNs and neuromorphic computing for onboard scene classification problems.

SNNs have attracted the interest of researchers due to their low-power [6,7] and energy-efficient computing properties [8-14]. These characteristics are due to the brain-inspired nature of such networks, which are based on layers of neurons that communicate through spikes. Although there are different models of spiking neurons [8-16], the latter generally accumulate incoming spike currents in different timesteps, increasing their membrane voltage. When the latter exceeds a fixed threshold, the neuron spikes, and its membrane voltage is reset [10,16]. Compared to standard ANNs, whose inference requires updating all neuron activations and synapses, synaptic operations for spiking neurons are performed when an input event occurs, leading to a sparse computing paradigm [6,8].

By implementing SNNs on event-based neuromorphic hardware, it is possible to benefit from their sparse computation, leading to solutions that generally outperform those based on ANNs in terms of power consumption [6,7] and, depending on the model of neurons [7,8,15], information encoding [7,8,13], number of timesteps used, input data [6,7,13] and hardware implementation [7], in terms of energy efficiency.

Despite the potential energy efficiency of SNNs, the state-of-the-art of these models cannot compete with deep ANNs in terms of accuracy for many applications due to the lack of training algorithms capable of scaling to deep models [10,11]. Some researchers have studied approaches that allow conversion between deep ANNs and SNNs [11,13,17]. In many of these cases, the conversion process is based on the close correlation between the activation rate of spiking neurons and the activation of Rectified Linear Units (ReLUs) in standard ANNs. Thanks to these methods, rate-based conversion between ANNs and SNNs can be done with minimal loss of accuracy [6,13,17]. However, the use of rate-based SNNs usually requires high fire-rates and a high number of timesteps to provide acceptable accuracy [11,16]. In this respect, this approach seems to bring real benefits in terms of energy efficiency only for event-based datasets [7]. On the contrary, for standard static images, which represent the data of interest for many remote sensing applications, the gain in terms of energy savings seems to be reduced for complex datasets due to the higher number of timesteps required, which also leads to higher processing latencies [6,7].

In that respect, methods based on *temporal-coding* might offer more promising trade-offs [8,10-12, 14,15]. Such approaches encode information in the fire time of neurons. For example, according to the *time-to-first-spike* coding, neurons are generally forced to fire once at most during an inference; moreover, the more a neuron is activated, the shorter is its firing delay [8,10,11,12]. In view of the reduced firing-rate activations, SNN models based on temporal coding are attractive for energy-constrained applications.

Some energy efficiency benefits of time-coding based SNNs for static data are shown in [10] for the MNIST dataset on the BrainScaleS-2 neuromorphic processor. However, the applications of SNNs for space applications are still limited [18,19,20,21], and the ability of these models to cope with complex features such as those included in scene classification datasets still awaits a convincing demonstration.

Proposed Methodology

To assess the benefits of using SNNs for on-board scene classification tasks, a comparison in terms of energy efficiency/inference time/accuracy [1,2,20] is to be performed on scene classification datasets. One possibility is to deploy spiking models on different hardware solutions (analogue, digital, hybrid) and measure directly their performances. However, this approach would limit the results to the specific hardware platforms chosen and would require the simultaneous availability of the platforms.

This study will follow a different approach and perform, instead, a theoretical analysis comparing simulations of different SNN models in terms of accuracy and complexity. The latter will be related to parameters such as the number of spikes, the number of synaptic operations, the number of timesteps, the number of neuron updates, and others. Although a reliable estimation of latency and energy per inference through a theoretical approach is difficult due to the strong dependence of these parameters on the hardware implementation, proxies such as those mentioned can be used to form a fair rank of different models, helping to identify which approach would eventually lead to lower latency and energy usage [6,13,22].

After having proposed an overall methodology to perform the study, the applicants should propose various SNN models to be compared in terms of complexity/accuracy on the EuroSAT RGB and EuroSAT multispectral datasets [23] as well as propose training approaches for SNN models, which could include ANN/SNN conversion [11] or spike domain training [8,10,12,15].

EuroSAT is a reference dataset for scene classification that includes multispectral imagery provided by the Sentinel-2A satellite, divided into 13 bands over 10 different land use and land cover classes. An RGB version of the dataset is also provided.

As a final goal, the study aims also at establishing an upper-bound for the energy per inference of the SNN model that is found to offer the best trade-offs in terms of accuracy and energy efficiency. To find such an upper bound, the availability of a dedicated hardware chip developed for the selected SNN model is to be assumed.

For this purpose and to validate and improve the energy estimation proxies, the characterisation of one or more models on a specific neuromorphic device can also be proposed and discussed.

ACT Contribution

The project will be conducted in close scientific collaboration with ESA/ACT researchers. ESA/ACT researchers will provide technical expertise in embedded systems and onboard AI to identify the best metrics to use in the evaluation of the different SNN models.

ESA/ACT will also implement one of the identified/agreed models using a common, agreed, framework.

Bibliography

[1] Giuffrida, Gianluca, et al. "CloudScout: A deep neural network for on-board cloud detection on hyperspectral images." *Remote Sensing* 12.14 (2020): 2205.
[2] Furano, Gianluca, et al. "Towards the Use of Artificial Intelligence on the Edge in Space Systems: Challenges and Opportunities." *IEEE Aerospace and Electronic Systems Magazine* 35.12 (2020): 44-56.
[3] Neuromorphic Computation of Optic Flow Data (08-6303a/6303b) – Initial study description: http://www.esa.int/gsp/ACT/doc/ARI/ARI%20Call%20doc/Study%20descriptions/ARI_study_08-6303.pdf
[4] Neuromorphic Computation of Optic Flow Data (08-6303a) – Report: http://www.esa.int/gsp/ACT/doc/ARI/ARI%20Study%20Report/ACT-RPT-BIO-ARI-08-6303-Neuromorphic-Computation-Optic-Flow-Zurich.pdf
[5] Neuromorphic Computation of Optic Flow Data (6303b) – Report: http://www.esa.int/gsp/ACT/doc/ARI/ARI%20Study%20Report/ACT-RPT-BIO-ARI-08-6303b-NeuromorphicOpticFlow-Marseille.pdf
[6] Han, Bing, et al. "Cross-layer design exploration for energy-quality tradeoffs in spiking and non-spiking deep artificial neural networks." *IEEE Transactions on Multi-Scale Computing Systems* 4.4 (2017): 613-623.
[7] Bouvier, Maxence, et al. "Spiking neural networks hardware implementations and challenges: A survey." *ACM Journal on Emerging Technologies in Computing Systems (JETC)* 15.2 (2019): 1-35.
[8] Kheradpisheh, Saeed Reza, and Timothée Masquelier. "S4NN: temporal backpropagation for spiking neural networks with one spike per neuron." *arXiv preprint arXiv:1910.09495* (2019).
[9] Voelker, Aaron R., Daniel Rasmussen, and Chris Eliasmith. "A spike in performance: Training hybrid-spiking neural networks with quantized activation functions." *arXiv preprint arXiv:2002.03553* (2020).
[10] Göltz, Julian, et al. "Fast and deep: energy-efficient neuromorphic learning with first-spike times." *arXiv preprint arXiv:1912.11443* (2019).
[11] Stöckl, Christoph, and Wolfgang Maass. "Recognizing Images with at most one Spike per Neuron." *arXiv preprint arXiv:2001.01682* (2019).
[12] Kheradpisheh, Saeed Reza, Maryam Mirsadeghi, and Timothée Masquelier. "BS4NN: Binarized Spiking Neural Networks with Temporal Coding and Learning." *arXiv preprint arXiv:2007.04039* (2020).
[13] Hunsberger, Eric, and Chris Eliasmith. "Training spiking deep networks for neuromorphic hardware." *arXiv preprint arXiv:1611.05141* (2016).
[14] Park, Seongsik, et al. "T2FSNN: Deep Spiking Neural Networks with Time-to-first-spike Coding." *arXiv preprint arXiv:2003.11741* (2020).
[15] Comsa, Iulia M., et al. "Temporal coding in spiking neural networks with alpha synaptic function." *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020.
[16] Neftci, Emre O., Hesham Mostafa, and Friedemann Zenke. "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks." *IEEE Signal Processing Magazine* 36.6 (2019): 51-63.
[17] Diehl, Peter U., et al. "Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing." *2015 International Joint Conference on Neural Networks (IJCNN)*. ieee, 2015.
[18] Bose, Pritam, et al. "Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series." *IEEE Transactions on Geoscience and Remote Sensing* 54.11 (2016): 6563-6573.
[19] Tao, Xiaoli, and Howard E. Michel. "Novel artificial neural networks for remote-sensing data classification." *Optics and Photonics in Global Homeland Security*. Vol. 5781. International Society for Optics and Photonics, 2005.
[20] Lemaire, Edgar, et al. "An FPGA-based Hybrid Neural Network accelerator for embedded satellite image classification." *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2020.
[21] Suresh, Shilpa, Devikalyan Das, and Shyam Lal. "A Framework for Quality Enhancement of Multispectral Remote Sensing Images." *2017 Ninth International Conference on Advanced Computing (ICoAC)*. IEEE, 2017.
[22] Martin, Erwann, et al. "Eqspike: spike-driven equilibrium propagation for neuromorphic implementations." *iScience* (2021): 102222.
[23] Helber, Patrick, et al. "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.7 (2019): 2217-2226.