RESCIENCE C

Reproducibility report - Wearable-based Human Activity Recognition with Spatio-Temporal Spiking Neural Networks, Li et al.

Umar Faruk Abdullahi, Mira Budenova, Oluseyi Samuel Berom[1]
[1]Åbo Akademi University

Written by
Li et al.

Published
14 November 2022

DOI
https://arxiv.org/abs/2212.02233v1

Reproducibility Summary

Scope of Reproducibility — According to the paper, the objective of human activity recognition (HAR) includes sports injury detection, well-being management, medical diagnosis, smart building solutions and elderly care. Using Artificial Neural Networks (ANNs) require a big amount of energy consumption and computation resources. Proposed solution is implementing Spiking Neural Networks (SNNs).

Methodology — Replacing non-linear activation in deep learning networks with Leaky-Integrate-and-Fire (LIF) neurons model for spiking neurons, which constantly receive inputs and outputs spikes through time. Authors use convolutional neural network (FCN) and deep convolutional long-short memory (DCL) architectures for ANNs, and for SNN they added spike neurons to the same archotectures: SFCN, and SDCL.

Results — Leveraging the identical architectures and parameters meticulously outlined by the original authors, we crafted four distinct machine models on the same datasets. Within this ensemble, we deployed two fundamental models: CNN and DCL (DeepConvLSTM), both employing the framework of Artificial Neural Networks (ANN). We then trained the adaptations of these models using Spiking Neural Networks (SNN). Remarkably, our models demonstrated accuracy scores that align closely with those reported by the esteemed authors, validating the robustness and fidelity of our implementations.

What was easy — The authors cleary specified the scope and methodology of the experiment which made the replication process lucid and well-defined. Also, the accompanying code repository provided detailed documentation on the required software and datasets needed to perform the experiment.

What was difficult — A major challenge emerged in the process of configuring the existing codebase to operate on a distinct hardware setup, divergent from that of the original authors. Notably, in the context of the SDCL models, the prolonged duration of training prompted a strategic reduction in the number of training epochs to expedite results.

Communication with original authors — An email inquiry was initiated with the authors to solicit insights into the comparative analysis of energy and computation efficiency between the SNN architecture models during both training and inference. Regrettably, as of the present report, no response has been received from the authors.

Code is available at Github

1 Introduction

With the rapid development of smart devices such as phones and fitness trackers, sensing user activities or behavioral insights becomes more important for healthcare purposes. The objective of human activity recognition includes sports injury detection, well-being management, medical diagnosis, smart building solutions and elderly care.

2 Scope of reproducibility

They use convolutional layers in the Artificial Neural Networks (ANNs) and optimize the model with gradient descent, but ANNs use full precision (i.e. 32-bit floating-point operations) computation and incur low sparsity, bringing huge computation complexity and energy consumption to wearable devices. To overcome the above limitations, authors utilize Spiking Neural Networks (SNNs) combined with convolutional layers for dealing with time series data in HAR. 2 advantages using SNN:

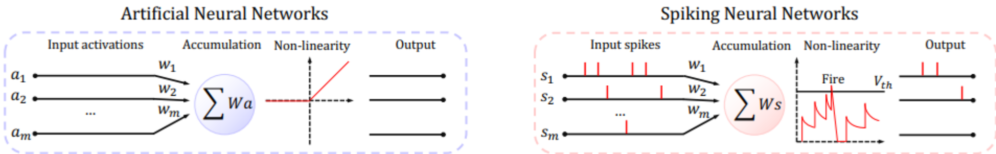

Figure 1. The ANNs and SNNs

- SNNs take advantage of binary spikes (either 0 or 1) and thus enjoy multiplication-free and highly sparse computation that lowers energy consumption on time-series data;
- SNNs can inherently model the temporal dynamics in time series data. The spiking neurons from SNNs maintain a variable called the membrane potential through time. As long as the membrane potential exceeds a pre-defined threshold, the neuron will a spike in the current time step.

The main claims throughout the paper is following:

- SNNs provide higher accuracy score, e.g., the validation accuracy score for SDCL is 98.4% while the ANN's DCL accuracy score is 96%.
- It can be seen that SNNs consume up to 94% less energy than ANNs, which could largely promote the battery life in smart devices. However, we could not test this claimIt can be seen that SNNs consume up to 94% less energy than ANNs, which could largely promote the battery life in smart devices.

3 Methodology

Implementing Leaky-Integrate-and-Fire (LIF) neurons gradually decrease energy consumption and the amount of computation resources for wearable devices. Authors integrate spiking neurons into deep neural networks by replacing their non-linear activation with LIF. Specifically, since the time series data naturally has a time dimension, authors also integrate the pre-synaptic potential charge along this time dimension. The aim of our reproducibility is to to re-implement the approach from the description in the paper.

3.1 Model descriptions

Two seminal Artificial Neural Network (ANN) models were subjected to training, namely the Convolutional Neural Network (CNN) and the Deep Convolutional LSTM (DCL). The CNN model comprised three fully convolutional blocks, with Rectified Linear Unit (ReLU) serving as the activation function. In the case of the DCL model, a distinctive modification was introduced by incorporating a Long Short-Term Memory (LSTM) block preceding the output layer. Subsequently, Spiking Neural Network (SNN) models were derived from these architectures, wherein the conventional ReLU activation function was replaced with the Leaky-Integrate-and-Fire (LIF) mechanism.

3.2 Datasets

- UCI-HAR contains 10.3k instances collected from 30 subjects. It involves 6 different activities including walking, walking upstairs, walking downstairs, sitting, standing, and lying. The sensors are the 3-axis accelerometer and 3-axis gyroscope (both are 50Hz) from Samsung Galaxy SII.
- UniMB SHAR contains 11.7k instances collected from 30 subjects. It involves 17 different activities including 9 kinds of daily living activities and 6 kinds of fall activities. The sensor is the 3-axis accelerometer (maximum 50Hz) from Samsung Galaxy Nexus I9250.
- HHAR contains 57k instances collected from 9 subjects. It involves 6 daily activities including biking, sitting, standing, walking, stair up, and stair down. The sensors are accelerometers from 8 smartphones and 4 smart watches

3.3 Hyperparameters

Authors implement the SNNs and existing ANNs with the PyTorch framework, Adam optimizer. All models are trained for 60 epochs, with batch size 128. For all three HAR datasets, authors split them to 64% as the training set, 16% as the validation set, and 20% as the test set. Test accuracy is reported when the model reaches the best validation accuracy. These datasets only have one label for each input sample, top-1 accuracy is the same as the F-1 score.

3.4 Experimental setup and code

The experimental configuration entailed the utilization of the PyTorch and TorchVision libraries, implemented in the Python programming language. Access to the comprehensive repository is facilitated through the following hyperlink: GitHub repo.

3.5 Computational requirements

Although the authors did not explicitly specify the hardware configuration used in training the models, careful inspection of the code revealed the utilization of NVIDIA GPUs with CUDA. In order to replicate the experiment on our hardware infrastructure, we judiciously installed library versions harmonious with the Metal framework for optimal hardware acceleration.
The training was conducted on a machine with the specificatios below. Table 1 also presents the total training duration (in hours) for each model.

- Computer: Macbook Pro
- Processor: M1 3.2GHz with 8 core GPU
- Memory (RAM): 16GB unified storage

	FCN	DCL	SFCN	SDCL
Uci-HAR	0:09:20	1:28:07	0:23:34	7:29:51
SHAR	0:11:11	0:56:59	0:20:48	2:36:17
HHAR	1:06:39	5:10:45	1:02:45	11:34:45

Table 1. Training Times for Machine Models on Different Datasets

4 Results

In our experimental replication, adhering to the authors' prescribed architectures and hyperparameters, we achieved results that closely mirrored the original findings. Notably, the Spiking Neural Networks (SNNs) exhibited superior or at least, similar performance compared to their Artificial Neural Network (ANN) counterparts. Moreover, the SNNs demonstrated higher average sparsity scores and reduced energy consumption, suggesting their potential for enhanced efficiency in neural network operations.

4.1 Results details

After training the selected models viz: FCN, DCL, SFCN and SDCL, we got the results which are summarized in Table 3, from which we find the SNNs having higher accuracy than the ANNs.
For example, on the UniMB SHAR dataset, SpikeCNN has a 3.2% average accuracy improvement over its artificial CNN counterpart. Even more remarkably, the SpikeDeepConvLSTM (SpikeDCL) on the UCI-HAR dataset reaches 98.4% accuracy, which is 2.2% higher than DCL. Considering the accuracy is approaching 100%, the 2.2% improvement would be very significant. For UCI-HAR and HHAR datasets, we find SpikeCNN has similar accuracy to CNN, instead, the SpikeDeepConvLSTM consistently outperforms DeepConvLSTM, indicating that SNNs can be more coherent with the LSTM layer. Regarding the standard deviation of accuracy, we find that SNNs are usually more stable than ANNs, except for only one case, SpikeCNN on UCI-HAR.

	FCN	DCL	SFCN	SDCL
Uci-HAR	96.7%	96.2%	96.7%	98.4%
SHAR	90.5%	75.0%	93.7%	90.9%
HHAR	94.8%	96.5%	96.0%	96.0%

Table 2. Accuracy comparison for the Machine Models on Different Datasets

Additionally, we provide the original results obtained in the authors' experiment in Table 3 for comparative analysis. The comparison reveals that our models performed on par with the authors' reported values, exhibiting a marginal variance of approximately ±2% except in the DCL on the SHAR dataset where our model performed considerably worse.

	FCN	DCL	SFCN	SDCL
Uci-HAR	96.29%	97.87%	96.40%	98.86%
SHAR	92.38%	90.78%	94.04%	92.08%
HHAR	96.19%	97.15%	96.20%	97.52%

Table 3. Author's accuracy comparison for the Machine Models on Different Datasets

5 Discussion

The paper delineated its objectives with clarity, employing a well-defined methodology to address the problem at hand. The exploration of Spiking Neural Networks (SNN) introduces a prospect for more efficient and temporally sensitive models, particularly suited for tasks aligning with the specified criteria. Our study substantiates the assertions made by the authors, demonstrating models that closely approximate the performance of the author's model. However, an intriguing question emerges regarding the potential influence of the models' training on their subsequent inference capabilities on low power and wearable devices — a facet not explicitly addressed within the scope defined by the authors.

5.1 What was easy

The authors distinctly outlined the experiment's scope and methodology, facilitating a clear and well-defined replication process. Additionally, the accompanying code repository featured comprehensive documentation detailing the necessary software and datasets essential for conducting the experiment. The authors' implementation of a standardized data processing pipeline further enhanced the efficiency of utilizing all three designated datasets during the training phase, ensuring a seamless execution of the study's methodologies.

5.2 What was difficult

Significant challenges surfaced while adapting the codebase to function on a different hardware configuration than that employed by the original authors. Overcoming this obstacle involved updating libraries to their latest versions and resolving consequential dependency issues. Particularly noteworthy is the decision, in the case of the SDCL models, to strategically reduce the number of training epochs due to prolonged training durations, aimed at expediting the results without significantly impacting the replication study's outcomes.

5.3 Communication with original authors

An email inquiry was initiated with the authors to solicit insights into the comparative analysis of energy and computation efficiency between the SNN architecture models during both training and inference. Regrettably, as of the present report, no response has been received from the authors.