Wearable-based Human Activity Recognition with Spatio-Temporal Spiking Neural Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

We study the Human Activity Recognition (HAR) task, which predicts user daily activity based on time series data from wearable sensors. Recently, researchers use end-to-end Artificial Neural Networks (ANNs) to extract the features and perform the classification in HAR. However, ANN carries huge computation burdens to wearable devices and lacks temporal feature extraction. In this work, we leverage Spiking Neural Networks (SNNs)—an architecture inspired by biological neurons—to HAR tasks. SNNs allow spatio-temporal extraction of features and enjoy low-power computation by binary spikes. We conduct extensive experiments on three HAR datasets with SNNs, demonstrating that SNNs are on par with ANNs in terms of accuracy while reducing up to 94% energy consumption.

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. In this case, Human Activity Recognition (HAR) [1, 2, 3] seeks to predict the user activities using the smart devices' sensors such as accelerometer, gyroscope, electroencephalogram (EEG) sensor, etc. The objective of HAR includes sports injury detection, well-being management, medical diagnosis, smart building solutions [4] and elderly care [5].

Traditionally, researchers use hand-crafted features and simple classifiers for HAR tasks. Yet this type of method requires expert knowledge to get high-quality features. More recently, deep learning has been introduced to use end-to-end feature extraction, as well as classification [6]. They use convolutional layers in the Artificial Neural Networks (ANNs) [7, 8, 9] and optimize the model with gradient descent. However, ANNs use full precision computation and incur low sparsity, bringing huge computation complexity and energy consumption to wearable devices. In addition, ANNs use ReLU neurons that do not consider correlation in time. This choice may be sub-optimal, especially for time series data since it simply adapts the ANN regime from the image domain.

To overcome the above limitations, we utilize Spiking Neural Networks (SNNs) [10, 11, 12, 13, 14] combined with convolutional layers for dealing with time series data in HAR. The HAR can be benefited from SNNs in two aspects: (1) SNNs take advantage of binary spikes (either 0 or 1) and thus enjoy multiplication-free and highly sparse computation to lower energy consumption on time-series data; (2) 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 fire a spike in the current time step. We verify our SNNs on three popular HAR datasets (UCI-HAR [3], UniMB SHAR [15], HHAR [16]) and compare them with ANNs baselines. Our SNNs can deliver the same or even higher accuracy than ANNs while reducing up to 94% energy consumption.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

The rest of this paper is organized as follows: Sec. 2 describes the problem statement and our approach to spatio-temporal SNN. Sec. 3 provides the experiment results and visualization on three HAR datasets. Finally, we give our conclusion remark in Sec. 4.

2 Method

2.1 Notation and Problem Statement

In this paper, vectors/matrices are denoted with bold italic/capital letters (*e.g.* $\boldsymbol{x}$ and $\mathbf{W}$ represents the input vector and weight matrix). Constants are denoted by small upright letters.

Concretely, denote the wearable-based sensor dataset with $\{\mathbf{x}_i\}_{i=1}^{N}$, and each sample $\mathbf{x}_i \in \mathbb{R}^{T\times D}$ is collected when the wearer is doing certain activity $\mathbf{y}_i$, *e.g.* running, sitting, lying, standing, *etc*. Here, data samples are streaming and have T time steps in total. D is the dimension of the sensor's output. As an example, the accelerometer records the acceleration in the (x, y, z)-axis, thus $D = 3$ for the accelerometer data. We are interested in designing an end-to-end model $f(\cdot)$ and optimizing it to predict the accurate activity label $\mathbf{y}$.

2.2 Spiking Neurons

We adopt the well-known Leaky-Integrate-and-Fire (LIF) neurons model for spiking neurons [17], which constantly receives input and outputs spikes through time. Formally, the LIF neuron maintains the membrane potential $\boldsymbol{v}$ through time, and suppose at t-th time step ($1 \leq t \leq T$), the membrane potential receives the pre-synaptic input charge $\boldsymbol{c}^{(t)}$, given by

$$\boldsymbol{v}^{(t+1),\text{pre}} = \tau \boldsymbol{v}^{(t)} + \boldsymbol{c}^{(t)}, \text{where } \boldsymbol{c}^{(t)} = \mathbf{W}\boldsymbol{s}^{(t)}. \tag{1}$$

Here, τ is a constant between $[0, 1]$ representing the decay factor of the membrane potential as time flows, which controls the correlation between time steps. $\tau = 0$ stands for 0 correlation and LIF degenerates to binary activation [18] without temporal dynamics, while $\tau = 1$ stands for maximum correlation and [19, 20] proves that LIF will become ReLU neuron when T is sufficiently large. $\boldsymbol{c}^{(t+1)}$ is the product between weights $\mathbf{W}$ and the spike $\boldsymbol{s}^{(t+1)}$ from last layer. After receiving the input charge, the LIF neuron will fire a spike if the pre-synaptic membrane potential exceeds some threshold, given by

$$\boldsymbol{s}^{(t+1)} = \begin{cases} 1 & \text{if } \boldsymbol{v}^{(t+1),\text{pre}} > V_{th} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where V_{th} is the firing threshold. Note that the spike $\boldsymbol{s}^{(t+1)}$ will propagate to the next layer, here we omit the layer index for simplicity.

If the LIF neurons fire the spike, the membrane potential will be reset. This can be done by either soft-reset or hard-reset, denoted by

$$\begin{cases} \boldsymbol{v}^{(t+1)} = \boldsymbol{v}^{(t+1),\text{pre}} \cdot (1 - \boldsymbol{s}^{(t+1)}) & \texttt{\# Hard-Reset} \\ \boldsymbol{v}^{(t+1)} = \boldsymbol{v}^{(t+1),\text{pre}} - \boldsymbol{s}^{(t+1)} \cdot V_{th} & \texttt{\# Soft-Reset} \end{cases}, \tag{3}$$

where hard-reset sets $\boldsymbol{v}^{(t+1)}$ to 0, while soft-reset subtracts $\boldsymbol{v}^{(t+1)}$ by V_{th}. We choose LIF neurons because the $s^{(t+1)}$ is binary and dependent on input in previous time steps. In our experiments, we will conduct ablation studies on the decay factor, the firing threshold, and the reset mechanism.

2.3 Integrating Spiking Neurons into Network

We 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, we also integrate the pre-synaptic potential charge along this time dimension. For instance, suppose $\boldsymbol{a} \in \mathbb{R}^{n\times c\times T}$ is a pre-activation tensor, where n, c, T represents the batch size, channel number, total time steps, respectively, we set the charge in each time step for LIF as the pre-activation in corresponding time step, *i.e.* $\boldsymbol{c}^{(t)} = \boldsymbol{a}_{:,:,t}$. Then, we stack the output spikes along the time dimension again, *i.e.* $\mathbf{S} = \text{stack}(\{\boldsymbol{s}^{(t)}\}_{t=1}^{T})$, for calculating the pre-activation in next layer.

2.4 Optimization

Although LIF neurons manage to model the temporal features and produce binary spikes, the firing function (Eq. (2)) is discrete and thus produces zero gradients almost everywhere, prohibiting gradient-based optimization. Particularly, the gradient of loss (denoted by L) w.r.t. weights can be computed using the chain rule:

$$\frac{\partial L}{\partial \mathbf{W}} = \sum_{t=1}^{T} \frac{\partial L}{\partial \boldsymbol{s}^{(t)}} \frac{\partial \boldsymbol{s}^{(t)}}{\partial \boldsymbol{v}^{(t),\text{pre}}} \left(\frac{\partial \boldsymbol{v}^{(t),\text{pre}}}{\partial \boldsymbol{c}^{(t)}} \frac{\partial \boldsymbol{c}^{(t)}}{\partial \mathbf{W}} + \sum_{t'=1}^{t-1} \frac{\partial \boldsymbol{v}^{(t),\text{pre}}}{\partial \boldsymbol{v}^{(t)}} \frac{\partial \boldsymbol{v}^{(t)}}{\partial \boldsymbol{v}^{(t'),\text{pre}}} \frac{\partial \boldsymbol{v}^{(t'),\text{pre}}}{\partial \boldsymbol{c}^{(t')}} \frac{\partial \boldsymbol{c}^{(t')}}{\partial \mathbf{W}} \right). \tag{4}$$

Here, all other terms can be differentiated except $\frac{\partial \boldsymbol{s}^{(t)}}{\partial \boldsymbol{v}^{(t),\text{pre}}}$ which brings zero-but-all gradients. To circumvent this problem, we use the surrogate gradient method. In detail, we use the triangle surrogate gradient, given by

$$\frac{\partial \boldsymbol{s}^{(t)}}{\partial \boldsymbol{v}^{(t),\text{pre}}} = \max\left(0, 1 - \left| \frac{\boldsymbol{v}^{(t),\text{pre}}}{V_{th}} - 1 \right| \right). \tag{5}$$

As a result, the SNNs can be optimized with stochastic gradient descent algorithms.

3 Experiments

In this section, we verify the effectiveness and efficiency of our SNNs on three popular HAR benchmarks. We first briefly provide the implementation details of our experiments and then compare our method with ANNs' baselines. Finally, we conduct ablation studies to validate our design choices.

3.1 Implementation Details

We implement our SNNs and existing ANNs with the PyTorch framework [21]. For all our experiments, we use Adam optimizer [22]. All models are trained for 60 epochs, with batch size 128. The only flexible hyper-parameter is the learning rate, which is selected from $\{1e-4, 3e-4, 1e-3\}$ with the best validation accuracy. We use Cosine Annealing Decay for the learning rate schedule. For all three HAR datasets, we split them to 64% as the training set, 16% as the validation set, and 20% as the test set. We report test accuracy when the model reaches the best validation accuracy. Note that these datasets only have one label for each input sample, therefore top-1 accuracy is the same as the F-1 score. The dataset descriptions are shown below:

UCI-HAR [3] 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 [15] 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 [16] 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 (sampling rate from 50Hz to 200Hz).

3.2 Comparison with ANNs

For ANN baselines, we select CNN [23], DeepConvLSTM [24], LSTM [25], and Transformer [26] architectures. For our SNNs, we integrate them into CNN and DeepConvLSTM. We put the network architecture specifications in Appendix. Each result is averaged from 5 runs (random seeds from 1000 to 1004) and includes a standard deviation value. We summarize the results in Table 1, from which we find the SNNs even have higher accuracy than the ANNs. For example, on the UniMB SHAR dataset, the SpikeCNN has a 1.7% average accuracy improvement over CNN. Even more remarkably, the SpikeDeepConvLSTM has much higher accuracy (8.5%). For UCI-HAT and HHAR datasets,

Table 1: Accuracy comparison between different networks on three HAR datasets (DCL=DeepConvLSTM).

Model	CNN	DCL	LSTM	Transformer	SpikeCNN	SpikeDCL
UCI-HAR [3]	96.29±0.12	97.87±0.32	82.41±4.04	96.02±0.27	**96.40±0.15**	**98.63±0.24**
SHAR [15]	92.38±0.51	82.77±1.51	83.87±0.96	83.19±0.74	**94.04±0.34**	**91.28±0.60**
HHAR [16]	96.19±0.14	97.15±0.17	95.59±0.20	95.82±0.16	**96.20±0.09**	**97.52±0.10**

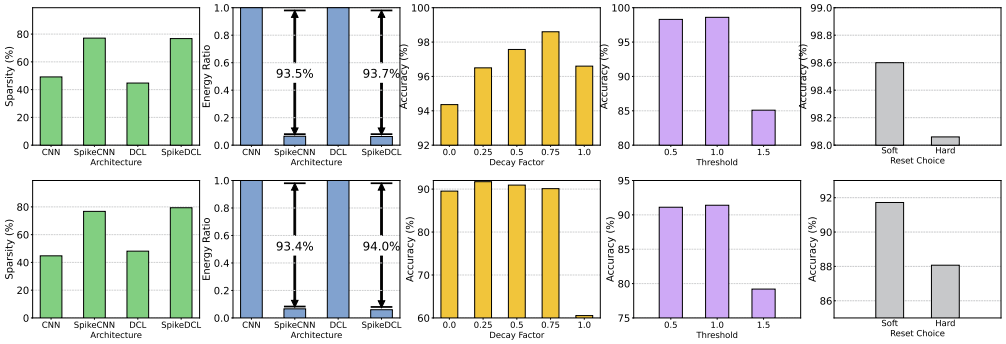

Figure 1: Hardware comparison between ANNs and SNNs (left two columns) and ablation studies of LIF hyper-parameter (right 3 columns). First row: UCI-HAR, second row: SHAR.

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.

We next compare the hardware performance. The first metric is the sparsity (ratio of zeros) of activations. Higher sparsity can avoid more computations with weights in hardware. Here we measure the sparsity either in ReLU (ANNs) or in LIF (SNNs) and visualize them in Fig. 1 first column. We find SNNs usually have 30% higher sparsity than ANNs. We then measure the energy reduction ratio on a hardware accelerator using the 65nm CMOS technology [27] in Fig. 1 second column. Note that we normalize ANNs' energy to 1. It can be seen that SNNs consume up to 94% less energy than ANNs, which could largely promote the battery life in smart devices. In summary, SNNs bring higher task performance due to the LIF neurons, and also energy efficiency due to the binary representation with high sparsity.

3.3 Ablation Studies

Here, we conduct ablation studies for decay factor, threshold, and reset mechanisms on LIF neurons in Fig. 1 right. We test SpikeDCL on UCI-HAR and SHAR and summarize the results below:

1. Neither 0 nor 1 for the decay factor τ is the optimal choice. Note that $\tau = 0$ is a binary activation network, which shows a large decrease than SNNs, which proves that temporal correlation is important. However, the optimal choice of τ also varies with different datasets.
2. For the firing threshold, both 0.5 and 1 is suitable for decent test accuracy. Instead, a larger than 1 threshold will largely decrease the performance.
3. Generally, the soft-reset regime is better than the hard-reset one since it retains some information after the firing of spikes. However, it may incur lower sparsity. Nevertheless, the soft-reset regime still has much lower energy than ANNs.

4 Conclusion

In this paper, we have introduced the Spiking Neural Networks (SNNs) to the HAR tasks, which, to our best knowledge, is the first to do that. Compared to the original Artificial Neural Networks (ANNs), SNNs utilize their LIF neurons to generate spikes through time, bringing energy efficiency as well as temporally correlated non-linearity. Our results show that SNNs achieve competitive accuracy while reducing energy significantly.

References

[1] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3):1192–1209, 2012.

[2] Michalis Vrigkas, Christophoros Nikou, and Ioannis A Kakadiaris. A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2:28, 2015.

[3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra Perez, and Jorge Luis Reyes Ortiz. A public domain dataset for human activity recognition using smartphones. In *Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning*, pages 437–442, 2013.

[4] E Ramanujam, Thinagaran Perumal, and S Padmavathi. Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, 21(12):13029–13040, 2021.

[5] Henry Friday Nweke, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-Garadi. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, 46:147–170, 2019.

[6] Henry Friday Nweke, Ying Wah Teh, Mohammed Ali Al-Garadi, and Uzoma Rita Alo. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, 105:233–261, 2018.

[7] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 2015.

[8] Andrey Ignatov. Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing*, 62:915–922, 2018.

[9] Shaohua Wan, Lianyong Qi, Xiaolong Xu, Chao Tong, and Zonghua Gu. Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, 25(2):743–755, 2020.

[10] Amirhossein Tavanaei, Masoud Ghodrati, Saeed Reza Kheradpisheh, Timothée Masquelier, and Anthony Maida. Deep learning in spiking neural networks. *Neural networks*, 111:47–63, 2019.

[11] Kaushik Roy, Akhilesh Jaiswal, and Priyadarshini Panda. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784):607–617, 2019.

[12] Lei Deng, Yujie Wu, Xing Hu, Ling Liang, Yufei Ding, Guoqi Li, Guangshe Zhao, Peng Li, and Yuan Xie. Rethinking the performance comparison between snns and anns. *Neural Networks*, 121:294 – 307, 2020.

[13] Priyadarshini Panda, Sai Aparna Aketi, and Kaushik Roy. Toward scalable, efficient, and accurate deep spiking neural networks with backward residual connections, stochastic softmax, and hybridization. *Frontiers in Neuroscience*, 14:653, 2020.

[14] Dennis Valbjørn Christensen, Regina Dittmann, Bernabé Linares-Barranco, Abu Sebastian, Manuel Le Gallo, Andrea Redaelli, Stefan Slesazeck, Thomas Mikolajick, Sabina Spiga, Stephan Menzel, et al. 2022 roadmap on neuromorphic computing and engineering. *Neuromorphic Computing and Engineering*, 2022.

[15] Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10):1101, 2017.

[16] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 127–140, 2015.

[17] Ying-Hui Liu and Xiao-Jing Wang. Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. *Journal of computational neuroscience*, 10(1):25–45, 2001.

[18] Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, and Ali Farhadi. Xnor-net: Imagenet classification using binary convolutional neural networks. In *European conference on computer vision*, pages 525–542. Springer, 2016.

[19] Yuhang Li, Shikuang Deng, Xin Dong, Ruihao Gong, and Shi Gu. A free lunch from ann: Towards efficient, accurate spiking neural networks calibration. In *International Conference on Machine Learning*, pages 6316–6325. PMLR, 2021.

[20] Shikuang Deng and Shi Gu. Optimal conversion of conventional artificial neural networks to spiking neural networks. *arXiv preprint arXiv:2103.00476*, 2021.

[21] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[23] Carlos Avilés-Cruz, Andrés Ferreyra-Ramírez, Arturo Zúñiga-López, and Juan Villegas-Cortéz. Coarse-fine convolutional deep-learning strategy for human activity recognition. *Sensors*, 19(7):1556, 2019.

[24] Debadyuti Mukherjee, Riktim Mondal, Pawan Kumar Singh, Ram Sarkar, and Debotosh Bhattacharjee. Ensemconvnet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. *Multimedia Tools and Applications*, 79(41):31663–31690, 2020.

[25] LuKun Wang and RuYue Liu. Human activity recognition based on wearable sensor using hierarchical deep lstm networks. *Circuits, Systems, and Signal Processing*, 39(2):837–856, 2020.

[26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[27] Ruokai Yin, Abhishek Moitra, Abhiroop Bhattacharjee, Youngeun Kim, and Priyadarshini Panda. Sata: Sparsity-aware training accelerator for spiking neural networks. *arXiv preprint arXiv:2204.05422*, 2022.