DEBARE: Deep-learning based activity recognition on the edge

Clayton F. S. Leite,[1] Antti Parviainen,[1] Pertti Kasanen,[2] Yu Xiao,[1*] Sami Jokela[2]

[1]Aalto University, Department of Communications and Networking, Konemiehentie 2, 02150 Espoo, Finland
[2] HitSeed Oy, Tekniikantie 2, 02150 Espoo, Finland
*Corresponding author: yu.xiao@aalto.fi

ABSTRACT

Human activity recognition enables pervasive technologies such as health monitoring devices, smart home appliances, fitness trackers, stroke rehabilitation apparatus etc. Deep learning methods are currently gathering momentum in human activity recognition as they often provide leading performance scores. However, these methods are resource-demanding, which makes them difficult to port into low-cost and compact computing devices that are directly connected to the sensors. For this reason, these methods have to be deployed to higher performance computers, therefore requiring a wired or wireless connection through which the data is to be transmitted for processing. Data transmission presents serious drawbacks: 1) it significantly increases energy consumption of the device, 2) it raises privacy-related concerns, and 3) limits the operation space of the device in the case of a wired connection or has to rely on data networks. In this project, we proposed methods that considerably reduce the computational expense and memory footprint of deep learning methods for human activity recognition to enable their use in resource-constrained devices. At the same time, these methods deliver state-of-the-art performance scores. We then developed a pipeline that includes data acquisition and processing, training of our proposed light-weight deep learning algorithm, its deployment to a microcontroller, real-time inference, remote monitoring and over-the-air updates.

Keywords: human activity recognition, deep learning, embedded devices, edge computing

1. INTRODUCTION

Human activity recognition (HAR) is a key component in the development of ubiquitous technologies such as fitness tracking, stroke rehabilitation, employee training, gesture-based augmented and virtual reality and smart home appliances. The aim of HAR is detecting the types of activities a person is executing from sensory data. Due to a giant leap in accuracy, deep learning (DL) methods are gradually becoming the norm in processing multimodal data, such as HAR sensory data. Moreover, specific domain knowledge is not required when designing DL methods as opposed to counterpart methods. However, the computational expense and memory footprint of DL methods are drawbacks that render their implementation in mobile and low-budget devices difficult. It is essential to address this issue in order to expand the benefits of DL algorithms to ubiquitous HAR.

In this project, we propose an end-to-end solution that enables ubiquitous resource-efficient DL based HAR for smart sensors. Our solution allows DL algorithms to be executed on energy efficient smart sensors significantly reducing the amount of data to be transmitted for remote processing. This has three key benefits:

- Because processing is more energy-efficient than transmitting our system doesn't need large batteries and is therefore suited for mobile use.
- Because no personal data has to leave the potentially always-on device our system is more secure and private.
- Because our system can operate independently of data networks it is suited for ubiquitous use.

Moreover, we invented novel methods to significantly reduce the complexity of DL algorithms for HAR at the same time as delivering a state-of-the-art accuracy. Additionally, we deployed our solution to HitSeeds state-of-the-art energy-efficient microcontroller (MCU) and performed diverse experiments that showed that our proposed solution is able to reach state-of-the-art accuracy in HAR while utilizing considerable less computational expense and memory footprint.

Our pipeline includes a data acquisition and processing system, the training of our proposed deep learning solution, its deployment to the MCU, real-time inference system, remote monitoring system and over-the-air (OTA) update system. This pipeline has been tested on an experimental smart glove meant for HAR. The system was able to collect training data for activity recognition from the smart-glove and train and deploy the deep

learning model to the MCU. The activity recognition performs energy efficiently and real-time on the smart-glove.

2. STATE OF THE ART

2.1 **Deep learning solutions for HAR**

The literature in DL for HAR has mainly focused on delivering ever-increasing recognition scores [4, 7, 8, 10] by utilizing an enormous variety of DL architectures such as long short-term memory layers, convolutional layers, inception blocks and residual blocks. These works have succeeded in their goal and, for this reason, DL techniques are gaining popularity in HAR. However, the computational requirements of the methods proposed in these works are far beyond the specifications of common off-the-shelf microcontrollers that are appropriate for ubiquitous HAR due to their low-cost and compact size.

Aware of the high resource utilization of DL in HAR, researchers have proposed compression methods [1, 5, 11] in order to reduce the computational complexity and memory footprint. Even though the compression methods have been able to significantly improve the resource efficiency of DL for HAR, they have not addressed three long-overlooked redundancies: 1) long sliding windows, 2) overlapping sliding windows, and 3) repetitive predictions of the same activity. These redundancies are discussed in technical details in our published article [6]. Combining compression methods and the removal of these overlooked redundancies has the potential to achieve even higher computational efficiency.

2.2 **HAR on mobile devices**

Reliance on a companion computer to perform activity recognition is one of the main drawbacks of mobile HAR as it limits the everyday usability of the system by requiring a always-on connection. A wired connection to the companion computer makes them suitable for stationary applications only. On the other hand, systems with a wireless connection are more mobile, but because they need to transfer large amounts of data they also need high bandwidth radios with high energy consumption. Hence, these systems have either a short operating time or must carry bulky and heavy batteries.

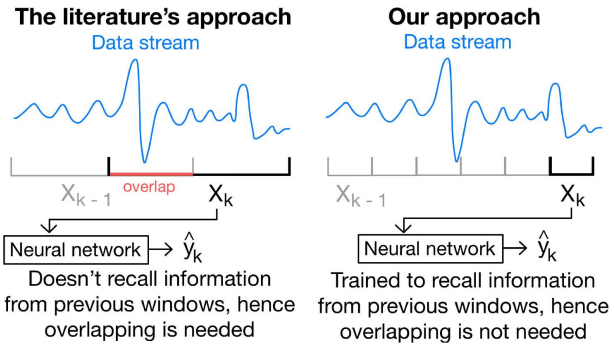

Fig. 1. The approach of short non-overlapping sliding windows (on the right) compared to the literature's approach (on the left). x_k is the k-th sliding window since the beginning of the data stream.

3. BREAKTHROUGH CHARACTER OF THE PROJECT

3.1 **From the perspective of deep learning for HAR**

It is a widespread practice to use long sliding windows for HAR because having an appreciable amount of temporal information is beneficial towards a more accurate classification. However, the size of the input of a neural network is proportional to its computational complexity and memory consumption. Hence, the first challenge is to guarantee high classification accuracy despite utilizing a short window of data.

The second challenge is to drop the need for overlapping sliding windows. In our work [6], we investigated that overlapping sliding windows are used to both provide the neural network with wider temporal context and to generate more frequent predictions. However, utilizing them also means that the same piece of information is processed more than once, which is a computational redundancy.

The third solved challenge is related to the activities that occur for prolonged periods of time (e.g. walking). As detailed in [6], processing long-lasting activities leads to redundancy in computation. To address this, we proposed to simplify the recognition algorithm during the execution of such activities.

The three aforementioned challenges are addressed by 1) altering the structure of the DL neural networks compared to the state-of-the-art and 2) introducing a different manner for training the DL neural network. Figure 1 illustrates a comparison between our work and the literature in DL for HAR in relation to the first two challenges.

3.2 From the perspective of mobile HAR

Tab. 1. Comparison between our system and the state-of-the-art.

	Our system	State-of-the-art
Not reliant on companion devices	x	
Privacy preserving	x	
All-day battery life	x	
Suitable for real-time use	x	x
Suitable for mobile use	x	x
OTA system updates	x	x
Not reliant on data networks	x	x
Suitable for high-end industrial use		x

We have addressed the two main limiting factors by designing an energy-efficient system that doesn't rely on a companion computer. By porting Google's TensorFlow Lite library to our Smart Sensor we are able perform the activity recognition deep learning algorithm directly on the smart sensor and not reliant upon a companion computer. This makes the system highly mobile. Additional benefits of performing HAR directly on the smart sensor is that we don't need to transfer large amounts of data using radios. This not only improves the energy consumption of the system but also the privacy since no raw data leaves the device. Table 1 compares our system with the state-of-the-art.

4. PROJECT RESULTS

4.1 Validating our light-weight deep learning solution

We start by validating our deep learning solution. Before presenting the results, we detail the utilized datasets, the chosen metrics and additional methods used to compare our solution with.

Datasets. We have selected two public datasets related to daily-life activities: Opportunity [2] and PAMAP2 [9].
Metrics. We report the results with 4 distinct metrics: F1-score, number of parameters of the neural network (NPA), total computational expense (TCE), and the prediction delay to detect newly incoming activities (PDNA). The F1-score indicates the accuracy in the detection of the activities. NPA gives a measure of the memory footprint. TCE is a metric that quantifies how computationally expensive an algorithm is. Finally, PDNA tells how much data (in seconds) is needed to be processed in order to correctly predict an activity. Further details on these metrics can be found in [6].

Tab. 2. Experimental results for both datasets considered. Results presented in **bold** are better. The PDNA metric was not computed for the InnoHAR method. However, it is similar to the Baseline CNN-LSTM method, since both methods utilize long and overlapping sliding windows.

Methods	F1-score	NPA	TCE (GFLOPs)	PDNA (s)
PAMAP2				
Baseline CNN-LSTM	0.876	83.65K	33.77	6.16
Ours SN	0.912	**51.75K**	4.60	**1.68**
Ours SNR	0.882	53.08K	**3.99**	2.13
InnoHAR [10]	**0.935**	34.91M	1887.58	-
Opportunity				
Baseline CNN-LSTM	0.895	118.46K	35.58	1.03
Ours SN	**0.959**	**84.91K**	10.07	**0.58**
Ours SNR	0.893	87.06K	**9.06**	0.61
InnoHAR [10]	0.946	6.20M	873.54	-

Additional methods for comparison. We put our work in perspective with two others in Table 2. First, a simple neural network made of convolutional and long short-term memory layers - named *Baseline CNN-LSTM*. Second, InnoHAR [10] which is a state-of-the-art neural network for HAR. Our work is represented by two methods named *Ours SN* and *Ours SNR*. The former addresses the first two challenges described in Section 3, whereas the latter addresses all challenges.

Discussion. Table 2 delineates the results of our experiments. In short, our work - both *SN* and *SNR* methods - is able to deliver similar or even superior F1-score compared to the state-of-the-art, while utilizing 2-3 orders of magnitude fewer network parameters and lighter computational expense. The PDNA metric shows that fewer data is needed to be processed before correctly predicting a class. More detailed results and discussion can be found in [6].

4.2 Applying our solution to physical devices

Our Smart Sensor operates in two main processes (Figure 2): Data Acquisition (1.1) and Realtime HAR (1.2).
In Data Acquisition mode our smart sensor is able to collect training data (1.1) from the smart glove and send it to the Cloud using a cellular LTE connection (2.1). We have also built a data acquisition helper (4.) that gives the user instructions on how to use the data-acquisition mode. The neural network can be trained by downloading the data to a PC (3.1) or directly in the

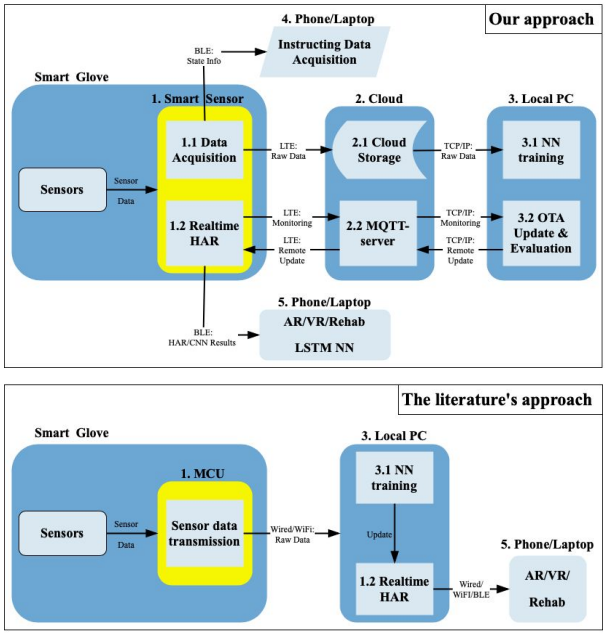

Fig. 2. The authors Deep-learning based activity recognition on the edge (DEBARE) pipeline compared to the literature's approach.

cloud server. The neural network weights in the smart sensor can also be updated remotely, if necessary, through an MQTT server (3.2). This system is significantly more mobile, privacy preserving and user friendly than the literature's approach (Figure 2) which relies on a companion PC where all of the raw sensor data is transferred for further processing.

In real-time HAR mode (1.2) the main advantage of our system is the ability to run a neural network already on the smart sensor (1.2). Running the neural network on the smart sensor is significantly more energy efficient compared to transferring the raw data with radios to a laptop and then running the neural network on the laptop. Due to low energy consumption, our system doesn't require large batteries for all-day operation and is more mobile because it's not reliant on a companion PC. These are major advantages for every-day users. Currently, we are able to run CNN-networks directly on the smart-sensor but not more sophisticated LSTM-networks. Google's TensorFlow Lite team hasn't yet added LSTM support for microcontrollers but this is expected in the near future. Currently, our system is, however, able to support a split CNN-LSTM system by running the CNN part on the smart-sensor and transferring the results to a smartphone or laptop using BLE where the LSTM part is run. This is also more energy-efficient than transferring all of the raw sensor data because the CNN network compresses the data significantly and therefore less data has to be transferred using radios. Additionally, our system allows remote usage monitoring and usage evaluation through the MQTT-server (2.2). This is important for example in rehabilitation applications.

5. FUTURE PROJECT VISION

5.1. Technology Scaling

We plan to increase the Technology Readiness Level (TRL) of DEBARE from 4 to 5-7 in ATTRACT Phase 2 through the following R&D activities: 1) Extension of the platform to support varying deep neural network architectures and a wider range of gestures; 2) Improving the electronic design and building support for different domain specific needs; 3) Platformization of the solution via open API.

5.2. Project Synergies and Outreach

We have identified a few related ATTRACT-projects where we could supplement their functionality with our gesture recognition, such as RPM3D, PRIOS and PRIMELOC. In the next stage we need to discuss further potential collaboration. We are also seeking to collaborate directly with research groups and industrial partners related to our solution.

We plan to disseminate the results and activities through the networks of each single partner in the consortium. For instance, academic partners will disseminate project results through open access publications and presentations in international academic conferences and local events such as Slush. Industrial partners will present the results at interactive exhibitions. The consortium will also create strategy and tools for social media and general media channels.

5.3. Technology application and demonstration cases

Both AR and VR have a huge potential to change our work and lives, but operating with real objects present a number of unsolved challenges related to data collection, analysis and real-time transfer. Our solution allows DL based activity recognition to be embedded directly in the haptics interfaces (e.g. smart gloves) that we can flexibly update when needs change. This enables real-time interaction with physical/virtual objects in various cases and domains.

For Phase 2, we plan to build solutions for two main use cases via pilots.

1) **Preventive/ predictive maintenance** is the most promising B2B use case in which we have

collaborated e.g. with KONE corporation that is one of the largest elevator manufacturers in the world. Predictive maintenance market size is estimated to grow to over 20B€ by 2024 according to Statista with the annual growth rate of around 30% and this is just one of the potential domains for our solution. Our main business case in this domain is to merge together maintenance operations done to the physical equipment with the digital information of the target. For example in training we can measure the tool use and teach the right gestures to serve the equipment. We can also improve the safety and lengthen the equipment lifecycle by monitoring tools, gestures, applied forces and so on.

2) **Digital rehabilitation** such as virtual reality for stroke rehabilitation is another promising use case. Our solution can be utilized for building smart garments for implementing gesture-based interaction with virtual reality, and tracking physical activities in daily life. As healthcare cases often require complex and lengthy testing and certification processes, we see healthcare and related cases more of a longer term partnering opportunity with domain experts.

Another potential application area is entertainment, e.g. gaming. There we can provide the bridge between the digital and physical world so that the capabilities of the interfaces can be tailored and upgraded based on the skill level and task at hand.

5.4. Technology commercialization

We will look for partners to scale up manufacturing and adapting the results to real-life cases.

We will continue with our existing industrial partners in the domain of preventive maintenance and to extend this basis further. With these domain partners we can tailor the solution towards recognized and existing high value use cases and to seek further utilization of the results. These companies will also help to incorporate the solution in their commercial grade tool sets they are currently offering or operating within their daily businesses.

In healthcare and sports we will collaborate with experts that are able to fulfill the domain specific demands e.g. in relation to testing and certifications that can be lengthy and complex.

We will seek public funding e.g. from EU and national instruments, such as Finnish research commercialization grants as well as continue our collaboration with instances such as EIT Digital or EIC's SME Instrument where HitSeed is also a member of, but we will also seek funding from our partners and from our extensive network of angels and VCs that we have worked with in the past.

5.5. Envisioned risks

We envision risks such as low user acceptance of new technology and high costs of manufacturing. We plan to mitigate these risks by following human-centered design processes such as involving end users from the beginning of the co-design process, and collaborating with manufacturing partners to develop low cost solutions.

5.6. Liaison with Student Teams and Socio-Economic Study

During Phase 1, an assignment related to DEBARE was given to a MSc. level student team in the Product Development Project (PDP) course at Aalto Design Factory. During Phase 2, we plan to continue the collaboration with the PDP course, and will offer more MSc. thesis topics. Prof. Yu Xiao from Aalto University will serve as the coordinator for this activity.

Regarding the expert-driven socio-economic study in Phase 2, the DEBARE consortium can contribute with information collected from co-design workshops with domain specific partners, such as manufacturing companies and healthcare providers, as well as results of pilot studies.

6. ACKNOWLEDGEMENT

This project has received funding from the ATTRACT project funded by the EC under Grant Agreement 777222.

7. REFERENCES

[1] C. Alippi, S. Disabato, and M. Roveri. 2018. Moving Convolutional Neural Networks to Embedded Systems: The Alexnet and VGG-16 Case. In IPSN ’18 (Porto, Portugal). IEEE Press, Piscataway, NJ, USA, 212–223.

[2] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. del R. Millán, and D. Roggen. 2013. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. Pattern Recognition Letters 34, 15 (2013), 2033 – 2042.

[3] B. Fang, X. Zeng, and M. Zhang. 2018. NestDNN: Resource-Aware Multi-Tenant On-Device Deep Learning for Continuous Mobile Vision. In MobiCom ’18 (New Delhi, India). ACM, 115–127.

[4] Y. Guan and T. Plötz. 2017. Ensembles of Deep LSTM Learners for Activity Recognition Using Wearables. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1, 2, Article 11 (June 2017), 28 pages.

[5] N. D. Lane, S. Bhattacharya, A. Mathur, C. Forlivesi, and F. Kawsar. 2016. DXTK: Enabling Resource-efficient Deep Learning on Mobile and Embedded Devices with the DeepX Toolkit. In MobiCASE'16. 98–107.

[6] C. F. S. Leite and Y. Xiao. 2020. Improving Resource Efficiency of Deep Activity Recognition via Redundancy Reduction. In Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications (HotMobile '20). Association for Computing Machinery, New York, NY, USA, 33–38. DOI:https://doi.org/10.1145/3376897.3377859

[7] J. Long, W. Sun, Z. Yang, O. I. Raymond, and B. Li. 2019. Dual Residual Network for Accurate Human Activity Recognition. CoRR abs/1903.05359 (2019). arXiv:1903.05359

[8] F. J. Ordóñez and D. Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors 16, 1 (2016).

[9] A. Reiss and D. Stricker. 2012. Introducing a New Benchmarked Dataset for Activity Monitoring. In ISWC'16. 108–109.

[10] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan. 2019. InnoHAR: A Deep Neural Network for Complex Human Activity Recognition. IEEE Access 7 (2019), 9893–9902.

[11] S. Yao, Y. Zhao, H. Shao, S. Liu, D. Liu, L. Su, and T. Abdelzaher. 2018. FastDeepIoT: Towards Understanding and Optimizing Neural Network Execution Time on Mobile and Embedded Devices. In Sensys '18 (Shenzhen, China). ACM, New York, NY, USA, 278–291.