Are Accelerometers for Activity Recognition a Dead-end?

Catherine Tong[†], Shyam A. Tailor[†], Nicholas D. Lane[†◇]
[†]University of Oxford [◇]Samsung AI

ABSTRACT

Accelerometer-based (and by extension other inertial sensors) research for Human Activity Recognition (HAR) is a dead-end. This sensor does not offer enough information for us to progress in the core domain of HAR—to recognize everyday activities from sensor data. Despite continued and prolonged efforts in improving feature engineering and machine learning models, the activities that we can recognize reliably have only expanded slightly and many of the same flaws of early models are still present today. Instead of relying on acceleration data, we should instead consider modalities with much richer information—a logical choice are images. With the rapid advance in image sensing hardware and modelling techniques, we believe that a widespread adoption of image sensors will open many opportunities for accurate and robust inference across a wide spectrum of human activities.

In this paper, we make the case for imagers in place of accelerometers as the default sensor for human activity recognition. Our review of past works has led to the observation that progress in HAR had stalled, caused by our reliance on accelerometers. We further argue for the suitability of images for activity recognition by illustrating their richness of information and the marked progress in computer vision. Through a feasibility analysis, we find that deploying imagers and CNNs on device poses no substantial burden on modern mobile hardware. Overall, our work highlights the need to move away from accelerometers and calls for further exploration of using imagers for activity recognition.

CCS CONCEPTS

• **Computer systems organization → Embedded hardware**; • **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**.

KEYWORDS

Activity Recognition, Imager Sensors, Accelerometers.

ACM Reference Format:
Catherine Tong, Shyam A. Tailor, Nicholas D. Lane. 2020. Are Accelerometers for Activity Recognition a Dead-end?. In *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications (HotMobile '20), March 3–4, 2020, Austin, TX, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3376897.3377867

HotMobile '20, March 3–4, 2020, Austin, TX, USA

ACM ISBN 978-1-4503-7116-2/20/03…$15.00
https://doi.org/10.1145/3376897.3377867

1 INTRODUCTION

For a long time, the default sensors for Human Activity Recognition (HAR) have been accelerometers—low-cost, low-power and compact sensors which provide motion-related information. In the past decades, waves of accelerometer-based (and by extension other inertial sensors) research have enabled HAR studies ranging from the classification of common activities to the objective assessment of their quality, as seen in exciting applications such as skill [21] and disease rehabilitation [27] assessment at scale. However, despite these success stories, we believe that accelerometers can lead us no further in achieving the primary task at the very core of HAR—to recognize activities reliably and robustly.

The reliable recognition of activities, first and foremost, requires data to be collected from information-rich and energy-efficient sensors. In 2004, Bao and Intille [4] published a landmark study to recognize daily activities from acceleration data. 15 years on, we have not moved far from identifying activities much more complex than 'walking' or 'running' with mobile sensors [33]. This slow process is not due to a lack of data (accelerometers are present in almost all smart devices), nor a lack of feature engineering and algorithmic innovation (a great variety of accelerometer-based HAR works exists, including deep learning solutions), but the quality of the sensor data itself.

It is time to rethink this default HAR sensor and move towards a modality with richer information, in order to identify more activities more robustly. While accelerometers capture motion-related information useful for distinguishing elementary activities, a significant portion of our activities are not characterised by their motions but by their precise context, e.g. social setting or objects of interaction. Recognising these activities unobtrusively is of great interest to many research communities such as psychology and health sciences but currently we are unable to do this with acceleration data. Moving forward, can we rely on such a sensor which inherently lacks the dimension to capture non-motion-based information? Through inspecting years of progress in accelerometer-based HAR research, our answer is no. Our over-reliance on accelerometer-based data is contributing to a bottleneck in inference performance, a constrained list of recognisable activity classes and alarming confusion errors. To overcome this bottleneck, we should instead adopt a sensing modality with much richer information—the most natural choice are *images*.

Against the backdrop of a rapid sophistication of both learning algorithms and sensor hardware, imagers—low-form factor visual sensors—stand as a strong candidate to replace accelerometers as the default sensor in a new wave of HAR research. Notably, the idea that images are information-rich is not new: both stationary and wearable cameras have been used in numerous studies for visual-based action recognition [35], and for providing labelable ground

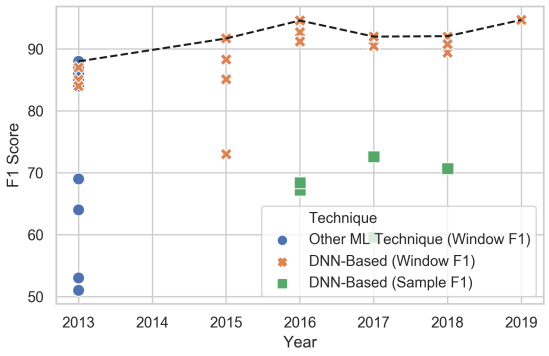

Figure 1: Opportunity gesture recognition, up to two scores plotted per publication per metric. We plot both the window-based F1 metric, and the sample-based F1 metric, which more recent publications prefer. We observe that deep learning has not yielded significant improvements to F1 score, despite its obvious success in other fields. Sample-based F1, believed to be a more accurate proxy to performance, is still not good enough to accurately disambiguate different activities that a user might perform.

truth and additional contexts [10]; The superiority of cameras compared to other sensors, including accelerometers, for high-level activity recognition has also been argued empirically [24]. Critically, previous barriers to the adoption of imagers have now been overcome by substantial advancement seen in recent years: on one hand, powerful and efficient deep learning algorithms have been developed for image processing [14, 23, 32]; on the other hand, there are now miniature image sensors with low energy requirement [1]. Our feasibility analysis suggests that the cost, size and energy requirements of imagers are no longer barriers to their adoption.

What we are proposing, adopting imagers as default sensors for activity recognition, is not to reject the use of accelerometers in tangential problems such as those assessing the level or quality of activities. It is also not to dismiss a multi-modality scenario, although we view the multitude of common low-dimensional sensors (e.g. magnetometer, gyroscope) as having only a supplementary role supporting the rich information collected by imagers. We acknowledge privacy concerns in processing images—but the imagers we envisioned will not be functioning as a typical camera where images are taken for record or pleasure, but as a visual sensor whose sole function is to sense activities; as a result, any images captured will only be processed locally on device.

It is our belief that imagers are the best place to focus our research energy in order to achieve significant progress in HAR—towards recognizing the full spectrum of human activities. Our findings demonstrate that imagers offer a far superior source of information than accelerometer sensors for activity recognition, and currently the algorithmic and hardware advancements are in place to make their adoption feasible. If we place our research focus on imagers, we anticipate further enhancement that could lead to their widespread deployment and a vast array of sensing opportunities in HAR. We hope our work will lay the foundation for subsequent research exploring image-based HAR.

2 HOW FAR HAVE WE COME?

We begin our case for imager-based activity recognition in place of acceleration-based sensing by considering where our attention to accelerometers has got us in activity recognition. In the remaining discussion, we refer to the problem of HAR exclusively as the recognition of activities from sensor data through the use of machine learning models. We review accelerometer-based HAR, summarise current challenges and finally examine how progress in HAR has stalled as a result.

A Brief History. Accelerometers respond to the intensity and frequency of motions, allowing them to provide motion-based features especially useful for characterising repetitive motions with a short time window [5]. Activities such as sitting, standing, walking can be recognized efficiently using accelerometer data. The use of accelerometer data to recognize human activities have spanned almost two decades, with many initial works focused on recognising ambulation and posture [25]. In 2004, Bao and Intille used multiple accelerometers to recognise 20 activities in a naturalistic setting, extending to daily activities such as watching TV and folding laundry [4]. Since then, numerous accelerometer-based activity recognition works have followed [8], and accelerometers are perhaps the most frequently used body-worn sensor in HAR. The ubiquity of accelerometer-based activity recognition works is further cemented by the inclusion of accelerometers in almost all publicly-available datasets [29, 30, 38], which are used as benchmarks and thus define the scope of baseline activity recognition tasks. Beyond activity recognition, accelerometers have been used in applications including monitoring of physical activities, breathing, disease rehabilitation [27], driving pattern analysis [18] and skill assessment [21].

Common Struggles. With their motion-based features, accelerometers are most useful for identifying activities with characteristic motions, such as walking patterns. However, even within the regime of elementary actions, accelerometer-based sensing often struggles to overcome key challenges such as individual and environmental variations, sensor diversity and sensor placement issues, to which many common activities are prone [22]. In turn, additional resources, often in the form of collecting more data or developing more complex models, are required to overcome the confusion between activities or to provide person-specific training.

A fundamental challenge arises when accelerometers are used to recognise more complex activities without distinctive motions: Accelerometers inherently lack the dimension to caption any important contextual and social information beyond motion. Thus, the recognition of the full spectrum of human activities is unattainable with accelerometers.

Another widely recognised challenge with accelerometers is the difficulty of labelling the data collected, which prohibits the development of large datasets useful for deep learning techniques. Unless there is a clear ground truth was obtained closed to data collection (by recollection or additional visual data source), it is nearly impossible to label the data.

Stalled Progress. We perform the following analysis to answer our core question: Have we really made progress with accelerometer-based activity recognition or has it stalled? We review current

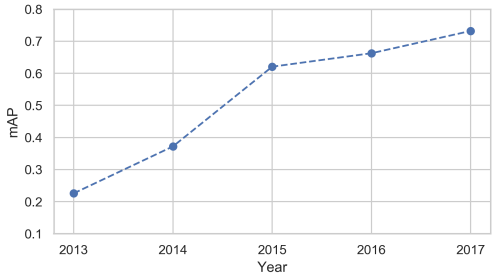

Figure 2: Best mAP achieved on ImageNet yearly between 2013–2017. Unlike Figure 1, we observe that performance for computer vision tasks—widely recognised as being challenging—have seen dramatic improvements in recent years thanks to deep learning.

models and datasets, in particular, the 18-Class Gesture Recognition task from the OPPORTUNITY Activity Recognition Dataset [7]. We make three key observations which indicate stalled progress:

Accuracy is not improving. Though our learning models have matured, activity recognition has seen no dramatic jump in performance. Figure 1 shows the weighted F1-score reported for the Opportunity gesture recognition task following its publication. Despite an initial increase in modelling power brought by Deep Neural Networks (DNN), introducing increasingly complex models has only led to scattered performance improvements in the past 5 years.

Limited Activity Labels. The variety of activity labels present in publicly-available HAR datasets is limited and confined to low to medium-level activities lacking any fine-grained or contextual descriptions. The labelling of the activities has perhaps been restricted to activities that motion-based models have some hope of distinguishing, which means we can never expect activity classes at the level of differentiating eating candy versus taking pills. In most datasets, a large percentage of activities belong to a 'null' or 'other' class which, prospectively, we have no chance of understanding what was actually performed from the acceleration traces.

Alarming confusions. Current inertial-based HAR models are far from robust and reliable. Lack-of-common-sense errors are commonplace, e.g. confusing 'drink from cup' with 'opening dishwasher' in Opportunity gesture recognition [37]. The performance of activity recognition measured by different metrics can also be drastically different—more recent publications consider sample-based measurements to be more relevant (shown in green in Figure 1) which only give the state-of-the-art F1-scores at low 70s [13], not to mention that these models already use many more sensors than are realistically deployable, as Opportunity collects data from accelerometers, gyroscopes, magnetometers at multiple body locations.

3 ARE IMAGERS THE ANSWER?

It appears we have reached the end of the road for HAR with accelerometers. We believe solutions exist in sensors that capture far richer observations about activities and context. We believe that *imagers*, in an embedded form which model images locally, are well-positioned to further HAR by enabling granular and contextual activity recognition. In what follows, we support our argument by pointing out that images are a rich source of information, and that the progress in computer vision provides a foundation from which we could build efficient, cheap and accurate imager-based HAR.

Not the first time. The idea that images are informative is not new, as supported by the vast number of activity recognition works which have used the visual modality in the form of wearable cameras, such as glasses [36], head [12], wrist [24], with notable wearable devices such as the SenseCam [16]. Activities recognizable using images include fine-grained activities (making tea vs coffee), social interactions [12] and context. [8] includes an extensive review of wearable camera-based activity recognition works. Other than wearable cameras, visual HAR based on stationary cameras has been extensively studied by the computer vision community and we refer readers to [39] for a survey in this area.

Information-Rich Images. Image data contains substantially more information than acceleration traces. Visual details captured in images may be high-level scene features quantifying the motion or contextualizing the scene, or they may be fine-grained details which specify an object that a person is interacting with [8]. The rich information provided by images not only enables more granular definitions of activity classes, it also provides abundant context for open-ended study of the subjects' behaviour.

We illustrate this richness in information by conducting a simple comparison between the data captured in the two forms: images and acceleration traces. We mounted two smartphones on a subject's dominant hand, each capturing one modality; we use the app *Accelerometer* [11] for recording acceleration, and the phone's video function for taking images. Figure 3 juxtaposes the images with the acceleration traces when two different activities were performed: typing on a phone and typing on a computer. The differences in the acceleration traces are hardly discernible nor intuitive as compared to that presented by the images. In addition, we can visually infer information such as objects, brands, environment, social setting simply from looking at these images; such additional information can be easily interrogated from the images through further processing and learning.

Deep Learning. More than ever, we have the image processing technology to be able to make use of the rich information in images, especially with the advances in Convolutional Neural Networks (CNNs). Figure 2 shows the yearly mean average precision (mAP) reported in the ImageNet visual recognition challenge [9], a benchmark in the computer visual community. There is an impressive and steady improvement in performance since the introduction of CNNs, and today the model performance has already surpassed that of human labelling; this is in stark contrast to the stalled progress seen in HAR. More than any other modality, the pairing of images and CNNs makes it practical to extract powerful information reliably enough to make the inference of complex human activities viable. In the realm of visual HAR with stationary cameras, state-of-the-art performance on the challenging (400-class) Kinetics dataset [19] has achieved top-1 accuracy exceeding 60%, although there is work to be done in leveraging these models for imager-based HAR, as will be discussed in Section 5.

Now is the time. Finally, it is possible to accelerate CNNs substantially—to the point that on-device inference is viable, which is precisely

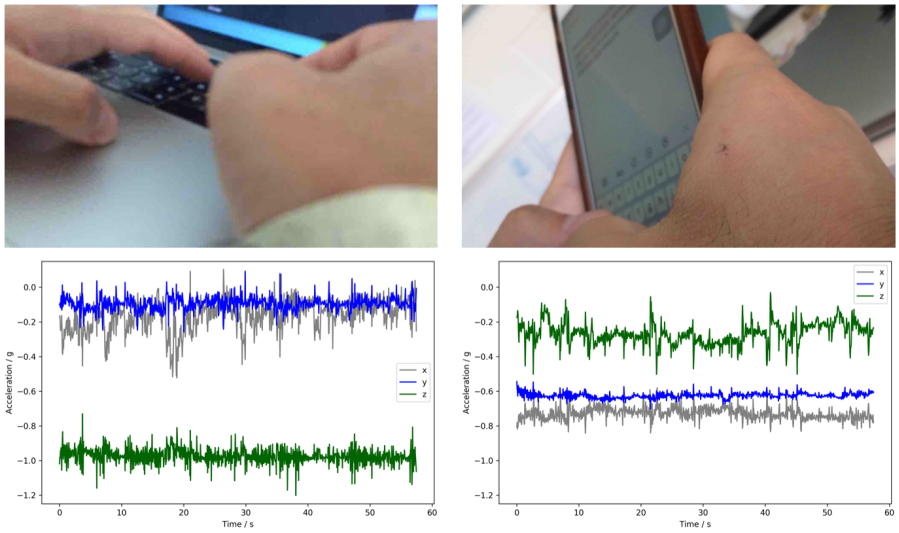

Figure 3: Image and accelerometer data collected from devices located at the wrist. Left: using a laptop, right: using a phone. The dominant difference in the acceleration data alignment is caused by differences in hand orientation, which can hardly be a robust feature for accurate predictions when variations within and between persons are considered. On the other hand, the differences between the two images are clear and comprehensible. A modern CNN would likely be able to disambiguate between these two images classes.

what is needed for the required combination of images, CNNs and tiny devices. We present a detailed feasibility analysis in Section 4.

Multi-sensor? We do not reject a multi-sensor scenario, especially if ample resources are available in terms of energy and computational costs. However, we view the multitude of common low-dimensional sensors (e.g. magnetometer, gyroscope) as having only a supplementary role supporting the rich information collected by imagers. Imagers remain the best choice amongst common inertial and non-inertial sensors (including microphones) due to the rich information they provide that can be used for disambiguating activities and avoiding confusions.

Privacy. We envision imagers to have only one function: to visually sense activities, and so any image captured would only be processed *locally* on the device. Running on-device recognition models prevents private visual information from ever leaving the local device end to the cloud. Concurrently, techniques developed in extreme-low-resolution activity recognition [31] may be incorporated to achieve privacy by using low-resolution hardware (which further reduces computational costs).

While prior works using image-based approaches for activity recognition exist, they largely came before CNNs, as well as other efficient on-device technologies—two key pieces to the viability of image-based HAR which were previously missing, and without which it was not worth pushing for mainstream adoption of imagers for HAR. *Now* is the ideal time to move away from accelerometers and adopt imagers for activity recognition, due to the maturity of these complementary components. As a community, we now have the techniques to build models that allow us to understand images at a higher level than ever before, and the ability to run these models on smaller and more efficient devices than those envisaged even 5 years ago.

4 IS IT PRACTICAL TO USE IMAGERS?

Conventional wisdom suggests an image-based wearable is impractical. In this section, we show that this belief has been outdated by advances in image sensing, microprocessors and machine learning.

Image Sensing Technology. The current power consumption associated with collecting images is sufficiently low that it does not represent a bottleneck. Over the past decade, both academia and industry have contributed to a substantial reduction in the power consumption of image sensors, and it is on the trajectory towards increasing efficiency by another order of magnitude [15].

Low form-factor imagers can now be purchased easily off-the-shelf. One such imager, capable of capturing grayscale QQVGA images at 30 frames per second (FPS), can do so using less than 1mW [1]. We can expect a lower frame rate for activity recognition as there is no need to capture images at a frequency as high as 30FPS, hence power consumption can be reduced even further.

On-Device Image Recognition. On-device modelling is vital to imager-based HAR: not only is it more efficient, it is also a solution to privacy concerns. A number of techniques have been proposed recently to allow neural networks to run on resource-constrained platforms such as wearable devices [14, 23, 32]. With techniques such as quantization and depthwise-separable convolutions it is now possible to run CNNs such as MobileNet [17] which obtain acceptable accuracy on images from ImageNet on ARM Cortex-M hardware, with sufficiently low latency for real-time use. We also expect the inference performance on low-power microcontrollers to improve rapidly: specialised RISC-V devices targeting this market are already available [2], and ARM have announced several features that will vastly improve their next generation of microcontrollers in this area [3].

Quantity	Value
Latency of MobileNet (α = 0.25, 160×160 input)	165ms
Microcontroller and camera run power	170mW
Microcontroller sleep power	2mW
Battery capacity	0.56Ah
Safety factor	0.7
Runtime	**13 hours**

Table 1: Summary of battery life calculation using an STM32H7 microcontroller and a 3.7V 150mAh battery, which represent realistic hardware choices for a wearable device.

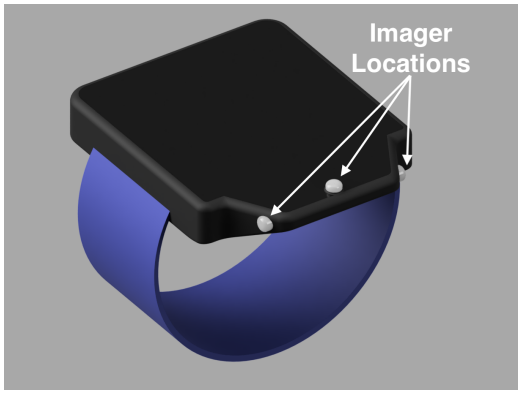

Figure 4: Conceptual image of a wrist-worn device with multiple embedded image sensors facing different directions. This device incorporates three sensors in a small protrusion along one side of the device in order to alleviate the problem of occlusion by clothing. This device is one of many potential form factors for imager-based HAR, which in this case takes a similar style to the common smartwatch.

Estimating Power Consumption. We provide an estimate of the expected battery life using existing hardware in Table 1. In our calculations, we consider the cost of deploying a MobileNet model onto a high-performance microcontroller (STM32H7), which proceed as follows: According to [6], it takes 165ms to run a MobileNet with a width multiplier 0.25 and input shape 160×160×3. The microcontroller consumes ~170mW when running at maximum frequency, and under 2mW during sleep. This comes to an average power consumption of ~30mW if processing at 1 fps. We also assume a realistic battery for a wearable device would be a 150mAh 3.7V LiPo. With a safety factor of 0.7 to account for further losses, we obtain a battery life of *13 hours of continuous monitoring*, including the cost of capturing images.

Since the considered microcontroller is not designated as a low-power model, the power consumption may be assumed to reduce further using lower power manufacturing nodes. Also, the power consumption is dominated by the inference latency, which is likely to decrease over time with the integration of special-purpose accelerators or new instruction set extensions. However, even without these factors, we still obtain a useful battery lifetime, which could be extended by duty cycling readings where appropriate.

Image Acquisition Many statistical and learning-based methods exist to overcome low-light level and blurring issues. Using other imaging modalities (e.g. IR, depth) could also alleviate these. One may also install multiple imagers at different viewpoints, though additional strains on resources are expected due to increases in model size, memory pressure and inference time. A simple concatenation pipeline for fusing images with CNNs would mean inference time scaling linearly with the number of cameras used.

Form Factor. We believe imagers would be best placed in a form factor that is wearable and socially-acceptable while allowing for unobstructed capture of images. The choice of form factor is closely related to the orientation of imagers, the availability of continuous image streams as well as lighting conditions. Glasses are a good candidate as they can provide an unobstructed view of the wearer's current focus, which might be difficult in other cases for which specific design considerations are needed. The smartwatch is a popular form of body-worn device that imagers could be integrated into; here, the major challenge would be occlusion by clothing; this issue can be mitigated by carefully designing the device so that the sensors are placed as far down the arm as possible, along with using fish-eye lenses to increase field of view; Figure 4 shows an artist's conception of one such device, which is one of many potential form factors for imager-based HAR.

Not All Scenarios. While our discussion of imagers for activity recognition has mainly existed in the context of the core HAR task of identifying activities, we acknowledge that there are certain domains adjacent to activity recognition for which imagers might be harder to use or perform worse than accelerometers. In particular, imagers may fall short for sensing applications looking to measure the physical level of activities (e.g. detecting tremble levels in freezing of gait episodes of Parkinson's patients [26]), or to assess the quality of activities (e.g. skill level [21]).

5 THE ROAD AHEAD

In this section, we present a research agenda which will make imager-based activity recognition viable.

Modelling. We discuss key issues in modelling imagers.

From images to activities. Imager-based HAR methodologies are not extensively studied, though there are many shared similar challenges with existing fields, such as egocentric image processing, odometry and object recognition. A key problem is to effectively model sparse snapshots of activities which also have a temporal relationship. An approach, given the sophistication of visual object recognition, is a multi-step approach: recognize objects then activities [28]. Much can also be learnt from visual action recognition methods (e.g. [34]), although activities of interest in HAR might differ, and these models typically consume high-frame-rate videos not likely to be available from imager-embedded devices.

Training. A core challenge is the lack of imager-generated datasets annotated for activities, especially from various ego-centric viewpoints from the body (e.g. wrist, belt, foot). Large image datasets used in object recognition (e.g. ImageNet [9]) or in visual action recognition (e.g. Kinetics [19]) present potential opportunities for

transfer learning from these respective domains to that of naturalistic egocentric images; such transfer learning schemes will be highly fruitful by leveraging non-sensitive, standardized datasets.

Multi-views. To effectively combine data generated from multiple imagers embedded on a device, one might leverage multi-camera fusion from robotics [20], though accomplishing this from an egocentric perspective with sporadic actions is a challenge.

Community efforts. The availability of accessible imager datasets is vital to facilitating modelling efforts in imager-based HAR. The ideal scenario is the development of such a dataset, with the community also defining a wide scope of activity labels, so that models can be built to recognize activities of different complexity at different stages of development. Doing so while addressing privacy concerns of collecting such datasets would be another open challenge.

Image Sensors. Our feasibility analysis had been run with off-the-shelf image sensors sub-optimal for imager-based HAR. The camera selected also provides better specifications than necessary: both the frame rate and image size were larger than what current hardware can process on-device. It will be possible to build cameras that more closely match the target application, with lower power consumption and manufacturing costs, e.g. adopting voltage reduction [15].

Hardware Design. Finally, machine learning on low-power devices is an area that is receiving significant attention from both the academic and industrial community. Microcontrollers incorporating extensions or accelerators suited for machine learning workloads are already becoming a reality [2, 3], which would cause a significant reduction in inference latency. There is concurrent work on how to train networks that utilise this hardware optimally.

6 CONCLUSION

Accelerometers are a dead-end for activity recognition: they do not offer enough information and our reliance on them has led to a stalled progress in HAR. In order to recognize the full spectrum of human activities we should adopt imagers as the default HAR sensor, since it is now possible to exploit the information-richness of images with the rise of energy-efficient CNNs. Collectively, our study argues that now is the time to pursue HAR using imagers, and calls for further exploration into overcoming the existing challenges for imager-based HAR.

ACKNOWLEDGEMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant No.: DTP (EP/R513295/1) and MOA (EP/S001530/), and Samsung AI. We would like to thank Dr. Petteri Nurmi and other anonymous reviewers for their feedback, and William Ip for facilitating the experiments.

REFERENCES

[1] HM01B0 « Himax Technologies, Inc. https://www.himax.com.tw.
[2] Kendryte - Kendryte. https://kendryte.com/.
[3] Arm Enables Custom Instructions for Embedded CPUs. https://www.arm.com/company/news/2019/10/arm-enables-custom-instructions-for-embedded-cpus.
[4] L. Bao and S. S. Intille. 2004. Activity Recognition from User-Annotated Acceleration Data. In *Pervasive*.
[5] A. Bulling et al. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 1–33.
[6] A. Capotondi. 2019. EEESlab/Mobilenet_v1_stm32_cmsis_nn. Energy-Efficient Embedded Systems Laboratory.
[7] R. Chavarriaga et al. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* (2013).
[8] M. Cornacchia et al. A survey on activity detection and classification using wearable sensors. *IEEE Sensors Journal* (2016)
[9] J. Deng et al. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*
[10] A. R. Doherty et al. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *International Journal of Behavioral Nutrition and Physical Activity* (2013)
[11] DreamArc. 2019. Accelerometer App.
[12] A. Fathi et al. Social interactions: A first-person perspective. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*
[13] Yu Guan et al. Ensembles of deep LSTM learners for activity recognition using wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2017)
[14] S. Han et al. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149* (2015).
[15] S. Hanson et al. A 0.5 V Sub-Microwatt CMOS Image Sensor With Pulse-Width Modulation Read-Out. *2010 IEEE JSSC*
[16] S. Hodges et al. 2006. SenseCam: A retrospective memory aid. In *International Conference on Ubiquitous Computing*. Springer.
[17] A. G. Howard et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* (April 2017).
[18] D. A. Johnson and Mohan M Trivedi. 2011. Driving style recognition using a smartphone as a sensor platform. In *2011 IEEE ITSC*, 1609–1615.
[19] W. Kay et al. 2017. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950* (2017).
[20] B. Keyes et al. Camera placement and multi-camera fusion for remote robot operation. In *Proceedings of the IEEE SSSR (2006).*
[21] A. Khan et al. 2015. Beyond activity recognition: skill assessment from accelerometer data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1155–1166.
[22] N. D. Lane. 2011. *Community-Guided Mobile Phone Sensing Systems.* Ph.D. Dissertation. Dartmouth College.
[23] N. D. Lane et al. Squeezing deep learning into mobile and embedded devices. *IEEE Pervasive Computing* 16, 3 (2017), 82–88.
[24] T. Maekawa et al. Object-Based Activity Recognition with Heterogeneous Sensors on Wrist. In *Pervasive Computing* (2010)
[25] J. Mantyjarvi et al. Recognizing human motion with multiple acceleration sensors. In *IEEE International Conference on Systems, Man and Cybernetics* (2001)
[26] K. Niazmand et al. Freezing of Gait detection in Parkinson's disease using accelerometer based smart clothes. In *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS).*
[27] S. Patel et al. A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation* 9, 1 (2012), 21.
[28] M. Philipose et al. Inferring activities from interactions with objects. *IEEE Pervasive Computing* 3, 4 (2004), 50–57.
[29] A. Reiss and D. Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th International Symposium on Wearable Computers*
[30] D. Roggen et al. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh international conference on networked sensing systems (INSS)*
[31] M. S. Ryoo et al. 2018. Extreme low resolution activity recognition with multi-siamese embedding learning. In *Thirty-Second AAAI Conference on Artificial Intelligence.*
[32] V. Sze et al. Efficient processing of deep neural networks: A tutorial and survey. *Proc. IEEE* 105, 12 (2017), 2295–2329.
[33] J. Wang et al. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters* (2019)
[34] L. Wang et al. 2015. Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition.*
[35] D. Weinland et al. A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding* 115, 2 (2011), 224 – 241.
[36] J. Windau and L. Itti. Situation awareness via sensor-equipped eyeglasses. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*
[37] J. Yang et al. 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Twenty-Fourth International Joint Conference on Artificial Intelligence.*
[38] P. Zappi et al. 2008. Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In *European Conference on Wireless Sensor Networks*
[39] H. Zhang et al. 2019. A Comprehensive Survey of Vision-Based Human Action Recognition Methods. *Sensors* 19, 5 (2019), 1005.