Improving Human Activity Recognition Models by Learnable Sparse Wavelet Layer

Haibin Zhao
haibin.zhao@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Yexu Zhou*
yexu.zhou@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Till Riedel
till.riedel@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Michael Hefenbrock
michael.hefenbrock@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Michael Beigl
michael.beigl@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

ABSTRACT

Modern machine learning algorithms for human activity recognition based on artificial neural networks often require a large amount of labelled training data to generalize between human subjects and training contexts. Large degrees of freedom make them susceptible to overfitting and often computationally intensive to implement on portable hardware. In this work, we introduce wavelet-based learnable filters as a feature extraction layer that greatly improves the generalization capability of the detector model. Our evaluations on six benchmark datasets show significant improvements in macro F_1 score when our wavelet-based learnable filter layer is prepended to three state-of-the-art human activity recognition models. As a side effect, in many cases we could drastically reduce the required model size to achieve competitive performance on the benchmark dataset, which is an important requirement for use in wearable computing.

CCS CONCEPTS

• **Information systems** → *Extraction, transformation and loading*; • **Human-centered computing** → **Human computer interaction (HCI)**; *Ubiquitous and mobile computing*; • **Computing methodologies** → **Supervised learning by classification**.

KEYWORDS

human activity recognition, learnable filters, wavelet analysis

ACM Reference Format:
Haibin Zhao, Yexu Zhou, Till Riedel, Michael Hefenbrock, and Michael Beigl. 2022. Improving Human Activity Recognition Models by Learnable Sparse Wavelet Layer. In *The 2022 International Symposium on Wearable Computers (ISWC '22), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3544794.3558461

*Corresponding author.

ISWC '22, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9424-6/22/09…$15.00
https://doi.org/10.1145/3544794.3558461

1 INTRODUCTION

With the growing attention to health, human activity recognition (HAR) plays an increasingly important role in daily life. Inspired by the wavelet analysis such as the work [26], we leverage the wavelet filters to the HAR models. Compared to the data-oriented *convolutional kernels* (*CKs*), wavelets are designed from dominant knowledge and don't need to be learned from data. Meanwhile, necessary properties for general signal-filtering such as bi-orthogonality are held by the wavelets. We believe that, combined with a learning scheme, wavelets can expose superior performance.

Currently, the majority of feature extraction is done by the *CKs* in convolutional neural network (CNN) [10]. However, the *CKs* are usually trained from random initialization and driven only by the objective function. Therefore, the final values of *CKs* are discouraged from a full exploration of the entire search space and tend to overfit the training data. This is often the core reason for suboptimal inter-subject generalizability. To address this overfitting problem, either the quantity or quality of data should be improved. Unfortunately, both ways are challenging: First, labeling for HAR data collected in natural settings is commonly expensive, as labeling often relies on auxiliary information like video, from which people can recognize the activity of subjects. This difficulty of annotation hinders HAR datasets from large natural setting. Moreover, the large diversity between individuals impedes labeled HAR data being sufficiently representative.

We therefore believe that by combining wavelets with HAR models, we can better deal with the deficiencies of available data and which should be observable in superior performance (macro F_1 score) in Leave-One-Subject-Out (LOSO) Cross-Validation.

Several works combining wavelets and machine learning algorithms have already been done. In [29] and [19], wavelets are combined with CNN for HAR tasks. However, only few pre-selected wavelets are utilized (1 wavelet in [29] and 7 wavelets in [19]), besides, the wavelets in these works are not learnable. A framework for learnable "wavelet" filters is proposed in [23], nevertheless, only the form of wavelet transformation (i.e., correlation & down-sampling) are respected, the necessary properties of wavelets such as bi-orthogonality and energy conservation are not guaranteed.

In our work, we extend state-of-the-art (SOTA) HAR models by a *learnable sparse wavelet layer*, which functions as a feature extraction layer. The *learnable sparse wavelet layer* consists of multiple learnable filters based on wavelet primitives followed by filter pruning. With this approach, the performance of the HAR model can be improved. Particularly, the improvement is noticeable when the model size is small. This advantage enhances the real-time HAR and facilitates the deployment of HAR models to wearable computing hardware with low computational power.

In summary, the contributions of this work are:

- We propose the learnable wavelet layer to extend the HAR models. The layer is composed by multiple wavelets.
- To ensure the sparsity of the learnable wavelet layer and to decrease the computational cost on embedded devices, we introduce the filter pruning.
- We prepend the *learnable sparse wavelet layer* to three SOTA HAR models. Their performances on six benchmark datasets show a great improvement.

2 METHODOLOGY

The *learnable sparse wavelet layer* is a convolutional layer composed by several learnable wavelet filters. To keep the sparsity of this layer, some non-informative filters are pruned during training. In the following, we firstly describe the generation, selection and pre-processing of the wavelets in detail. Then, we explain the implementation of the learnable part of wavelets. Lastly, we introduce the filter pruning for the reduction of computational costs.

2.1 Wavelet Filters

In this work, we chose the expressive and widely used wavelets [1] as the primitives of our learnable filters. There are various mother wavelets representing different underlying information, e.g., the Shannon-wavelet performs as an ideal band-pass filter [11], while the Morlet-wavelet acts more like a low-pass filter which is closely related to human perception [5, 18]. Some filters are beyond frequency-domain filters, such as Daubechies-wavelet [27]. By using these wavelets, generally robust features can be extracted for HAR tasks [26]. Different from the approaches in [29] and [19], we do not pre-specify the utilized wavelets, i.e., we select all the 127 discrete mother wavelets provided by PyWavelets[1] at the beginning. We then sample them to the same length as the sliding window used for activity recognition. To initially reduce the number of mother wavelets without losing their expressiveness, we apply K-means to cluster the mother wavelets w.r.t. both the temporal and the frequency domain. The distance between the i-th and the j-th mother wavelet is defined as $\|\boldsymbol{f}_i-\boldsymbol{f}_j\|_2+\|\mathcal{F}_i-\mathcal{F}_j\|_2$, where $\boldsymbol{f}_i$ and $\mathcal{F}_i$ denote the i-th wavelet as well as its Fourier transformation. The number of clusters can be determined by the silhouette coefficient [25]. In order to respect the law of energy conservation in filtering, we normalize the wavelets by

$$\tilde{\boldsymbol{f}}_i = \begin{cases} \boldsymbol{f}_i, & E_i \leq 1, \\ \boldsymbol{f}_i/E_i, & E_i > 1, \end{cases} \qquad E_i = \left| \sum_t \boldsymbol{f}_i[t] \right|, \tag{1}$$

where $\boldsymbol{f}_i[t]$ denotes the value of the t-th element in the i-th wavelet. We apply this normalization not only to the mother wavelets, but all wavelets after the temporal scaling (see Section 2.2).

[1] https://pywavelets.readthedocs.io

2.2 Learnable Wavelets

To make the selected wavelets learnable without losing their functional properties, we introduce a temporal scaling factor $\boldsymbol{k}$ as a learnable parameter. Since we utilized discrete wavelets in this work, the temporal scaling of filters can only be achieved by down-sampling, therefore, $\boldsymbol{k}$ cannot be optimized by gradient-based methods, as the gradient of a scaled wavelet w.r.t. the temporal scaling factor $\nabla_{\boldsymbol{k}} f[\boldsymbol{k}t]$ does not exist. Therefore, we introduce another learnable parameter $\boldsymbol{w}$ that indicates the informativeness of each filtered signal in the HAR model. The informative factor $\boldsymbol{w}$ will be multiplied with the corresponding filtered signals. Due to the existence of the informative factor $\boldsymbol{w}$, we are able to optimize the scaling factor $\boldsymbol{k}$ by dropping the scaling factors related to non-informative filtered signals (see next subsection for more details).

In our work, we implement the temporal scaling by down-sampling the mother wavelets by power-of-two scaling factors and normalize scaled wavelets after the down-sampling by Equation 1.

2.3 Filter Pruning

As mentioned in the last section, we aim to exclude non-informative filters. Similar problems have been studied in the field of *neural architecture search* [7], namely the neural network pruning. In this work, we use the similar strategy as [16] and [14]. Specifically, we first train the HAR model as well as the informative factors. After the training, the non-informative input signals as well as the corresponding wavelets can be removed. After the removal, we re-train the model again to mitigate the changes caused by the removal of temporal scaling factors.

ℓ_1 regularization. To reduce the number of remaining scaling factors, we add a penalty term to the objective function to encourage more zero-valued informative factors. Ideally, this penalty term should be $\ell_0(\boldsymbol{w})$, indicating the number of non-zero elements in $\boldsymbol{w}$, also known as the *sparsity* of $\boldsymbol{w}$. However, $\ell_0(\cdot)$ is ill-conditioned [30] and can not be solved by gradient-based optimization. Coincidently, it has been proven that, $\ell_0(\boldsymbol{w})$ can be usually substituted by the sum of the absolute values of all the elements in the vector $\boldsymbol{w}$ [6], i.e., $\ell_1(\boldsymbol{w}) = \sum_i |\boldsymbol{w}_i|$.

Pruning & fine-tuning. After the training, we remove all informative factors lower than a threshold and as well as the corresponding wavelets. Figure 1 shows the comparison of the filtering before and after pruning. Since the informativeness of the pruned signals is not exactly zero, the performance of the model usually becomes worse after pruning. Therefore, the model will be fine-tuned to adapt to the pruned input. Moreover, since the ℓ_1 regularization behaves as a trade-off between the sparsity of the informative factor and the performance of the model, we remove this term in fine-tuning.

3 EXPERIMENTS

To verify the improvement of the *learnable sparse wavelets layer* for the HAR models, we grafted the layer on three SOTA HAR models and perform the experiment on six benchmark datasets (see Section 3.1). The experiment[2] is executed with NVIDIA® A100 GPU with 40 GB memory. The result is reported in Section 3.2.

[2] The code is available at https://github.com/teco-kit/ISWC22-HAR

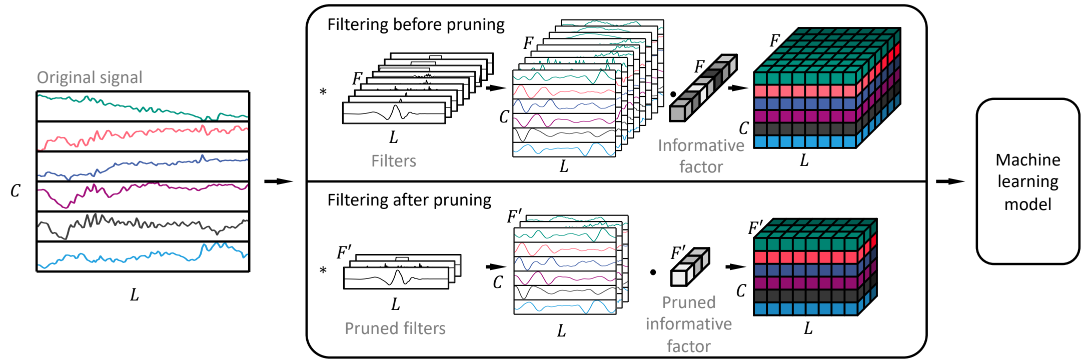

Figure 1: Pipeline of the proposed algorithm. Left: exemplary original data. Middle: the proposed *learnable sparse wavelet layer*. Right: a machine learning model for HAR. C = number of channels, L = length of sliding window, F = number of initially selected filters, F' = number of filters after pruning.

3.1 Experiment Setup

In this section, benchmark datasets, baseline models, the implementation details as well as the evaluation metrics are described.

Datasets. We select six benchmark datasets for the experiment and use the same setup as described in [8, 15, 24]. More specific information about the datasets are:

- **Opportunity** [4] with 79 input channels (including A, G, and M)[3] and 18 classes on 4 subjects. The sampling frequency is 30 Hz, while the length of the sliding window is 1 s.
- **Skoda** [28] with 30 input channels (including only A) and 10 classes on 1 subject. The frequency is down-sampled to 33 Hz, while the length of the sliding window is 2.56 s.
- **PAMAP2** [22] with 18 input channels (including A and G) and 12 classes on 9 subjects. The frequency is down-sampled to 33 Hz, while the length of the sliding window is 5.12 s.
- **DSADS** [3] with 6 input channels (including A and G) and 12 classes on 30 subjects. The sampling frequency is 25 Hz, while the length of the sliding window is 5 s.
- **Daphnet** [2] with 9 input channels (including only A) and 2 classes on 10 subjects. The sampling frequency is 64 Hz, while the length of the sliding window is 1 s.
- **WISDM** [13] with 3 input channels (including only A) and 6 classes on 36 subjects. The sampling frequency is 20 Hz, while the length of the sliding window is 5 s.

Baselines. To verify the effectiveness of the *learnable sparse wavelet layer*, we apply the layer to three SOTA benchmark HAR models, DeepConvLSTM [21], Multibranch CNN (MCNN) [20] and Self-Attention HAR (SA-HAR) [17]. These three models are chosen as they represent three typical structures of HAR models. MCNN is a purely convolution-based model which leverages late-fusion techniques to optimally fuse multimodal sensor data. DeepConvLSTM is a hybrid model of CNN and LSTM that combines the advantages of the CNN and LSTM. SA-HAR is a purely self-attention based HAR model without any recurrent structures. It utilized sensor modality attention, self attention blocks and global temporal attention to extract the inter- and intra-modality features. We evaluate the baseline models with different model sizes by employing the width scaling method [9]. With the model scaling factor α, both the number of input and output channels are modified. When, e.g., $\alpha = 0.5$, the number of the *CKs* and the number of parameters in each *CK* are halved. Therefore, the model size will be quadratically reduced to around $\alpha^2 = 0.25$ of the original model size.

Training & Validation. As the optimizer, we select Adam [12] with default settings and the learning rate being 10^{-4}. The learning rate is multiplied by 0.1 with 5-epoch patience on the valid loss. The maximal training epochs are set to 150, while the early-stop strategy is adopted with 15-epoch patience on valid loss. The size of each mini-batch is 256. The above settings are also identically used in fine-tuning process. Regarding the sparsity of the *learnable sparse wavelets layer*, we pruned the number of wavelets down to 50 for all datasets. Note that, this number can also be tuned as a hyperparameter. Moreover, for all datasets except **Skoda**, we do LOSO Cross-Validation. Since there is only one subject in **Skoda**, 5-fold Cross-Validation is performed. For each dataset, we run three approaches (baseline & baseline + *learnable sparse wavelet layer* & baseline + learnable wavelet layer without pruning) with four model scaling factors ($\alpha \in \{0.25, 0.5, 0.75, 1\}$) for 5 different random seeds (ranging from 1 to 5).

Evaluation. After training, we use test data to evaluate the trained models. Same as work [17], we use window-wise and sample-wise data split for training and testing respectively. I.e., the stride of sliding windows for training and validation data equal 50% window size, whereas the stride for the test data is only 1. As evaluation metric, we use macro F_1 score to suppress the influence of unbalanced classes in the datasets. For each random seed, the macro F_1 scores of all Cross-Validations are averaged for all subjects. After the five runs, the expectation of averaged macro F_1 score w.r.t. random seed is calculated (see Section 3.2). Moreover, we calculate the *floating point operations* for each model on each dataset.

3.2 Result

From Figure 2 we conclude preliminary that, the *learnable sparse wavelet layer* can improve the overall performance of the baseline HAR models. Moreover, the introduction of pruning significantly reduces the computational cost without reducing (or occasionally even improving) performances. Additionally, we notice that, for the

[3] A = accelerometer, G = gyroscope, M = magnetrometer

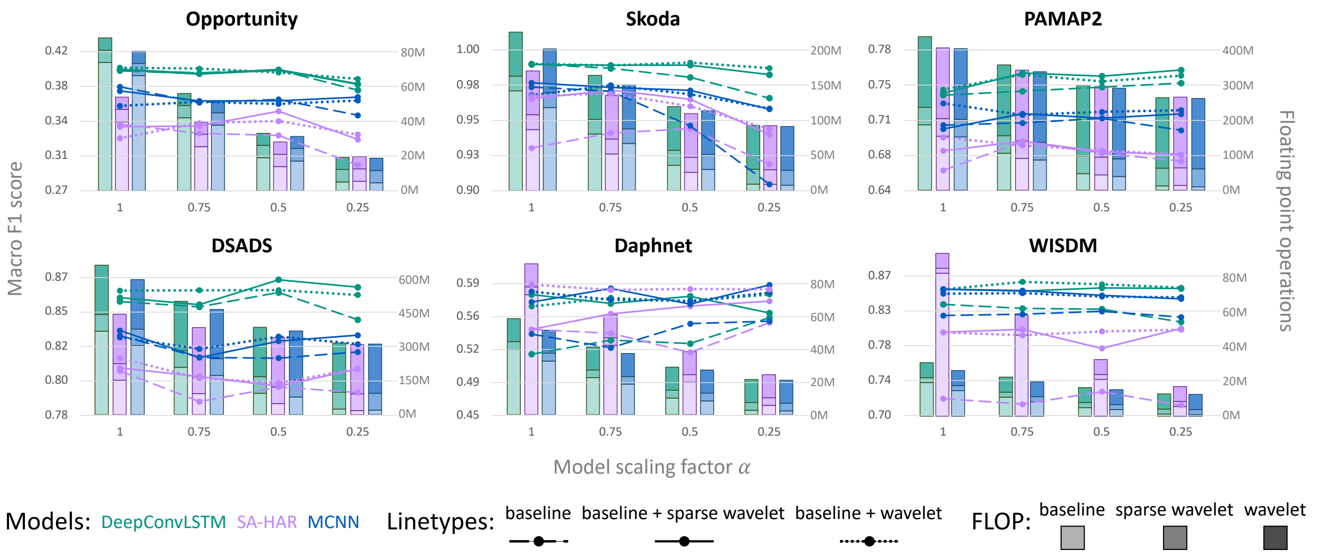

Figure 2: Result of the experiment. Different colors indicate different HAR models (green for DeepConvLSTM, purple for SA-HAR, and blue for MCNN). Different linetypes denote the macro F_1-scores from different setups (dash lines for baselines, solid lines for *learnable sparse wavelet layers*, and dot lines for learnable wavelet layers without pruning). The bars with different intensities refer to the number of floating point operations required by different setups, namely light colors for baselines, normal colors for *learnable sparse wavelet layers*, and dark colors show learnable wavelets without pruning.

datasets **Opportunity**, **Skoda**, **DSADS**, and **WISDM**, the performance of baselines deteriorate as the model sizes decrease, i.e., less information can be learned when the models are smaller. In this case, the information extracted by learnable wavelets compensates significantly for the lack of model size in those cases. Particularly, the contribution of learnable wavelets increases as the model size scales down. Conversely, as the model is close to the saturation size, learnable wavelets may not provide additional information to improve the model (e.g., **Opportunity** with $\alpha = \{1,\ 0.75,\ 0.5\}$). Regarding **Daphnet**, we speculate that the baselines overfit, i.e., with growing model sizes, the generalizability (which is reflected by the performance on test data) decreases. As hypothesized in Section 2.1, learnable wavelets, due to their non-data-oriented nature, should not overfit training data to an equal extend, and thus, may be able to remedy the generalizability of the models to some extent.

3.3 Discussion

From the experiments, we can see that the learnable wavelets generally enhance the performance of the baseline HAR models. We believe that the improvement can be summarized in two aspects that support our initial hypothesis: When the model tends to overfit the training data due to large size, learnable wavelets do not overfit. In this way, the generalization ability of the model can be ameliorated. In contrast, when the model size is too small and therefore lacks learning capability, the rich representational ability of the wavelets can help extract more robust and useful information and consequently improve the performance. Particularly, the improvement gets more noticeable when the model size is smaller. This advantage renders our approach a strong candidate for the deployment of HAR models on hardware with limited computing capacity, such as wearable devices.

4 CONCLUSION

In this work, we proposed the *learnable sparse wavelets layer* by leveraging the superior properties of wavelets. To make the wavelets capable of learning without losing the ability to extract generally useful features and necessary properties for general signal-filtering, we designed the temporal scaling factors $\boldsymbol{k}$ and informative factor $\boldsymbol{w}$ as learnable parameters. Our hypothesis is supported by our experiment that, the *learnable sparse wavelets layer* extracts rich and general information for the subsequent HAR model, and thus, the performance can be improved. Furthermore, the proposed layer is more pronounced when the model is the smaller, this facilitates the deployment of the HAR models on wearable devices.

Future work. In this work, the temporal scaling factor $\boldsymbol{k}$ can not be optimized by gradient-based approaches due to the discreteness of the mother wavelets $f[t]$. In the future, analytical mother wavelets $f(t)$ should be used, so that the gradient w.r.t. temporal scaling factors $\nabla_{\boldsymbol{k}} f(\boldsymbol{k}t)$ exists. Moreover, the *learnable sparse wavelet layer* functions as a data pre-processing that extract more useful and general information. To further decrease the computational cost of HAR, more efficient HAR models can be designed. Generally, more ablation studies are expected to fully understand the positive effects of learnable wavelets in HAR models.

ACKNOWLEDGMENTS

This work has been partially supported by the Carl-Zeiss-Foundation as part of "stay young with robots" (Jubot) project, the German Ministry of Research and Education as part of the SDIL (01IS19030A), and the Ministry of Economic Affairs, Labour and Tourism Baden-Württemberg as part of CC-KING.

REFERENCES

[1] Ali N Akansu, Wouter A Serdijn, and Ivan W Selesnick. 2010. Emerging Applications of Wavelets: A Review. *Physical communication* 3, 1 (2010), 1–18.

[2] Marc Bachlin, Daniel Roggen, Gerhard Troster, Meir Plotnik, Noit Inbar, Inbal Meidan, Talia Herman, Marina Brozgol, Eliya Shaviv, Nir Giladi, et al. 2009. Potentials of Enhanced Context Awareness in Wearable Assistants for Parkinson's Disease Patients With the Freezing of Gait Syndrome. In *2009 International Symposium on Wearable Computers*. IEEE, 123–130.

[3] Billur Barshan and Murat Cihan Yüksek. 2014. Recognizing Daily and Sports Activities in Two Open Source Machine Learning Environments Using Body-worn Sensor Units. *Comput. J.* 57, 11 (2014), 1649–1667.

[4] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. 2013. The Opportunity Challenge: A Benchmark Database for On-body Sensor-based Activity Recognition. *Pattern Recognition Letters* 34, 15 (2013), 2033–2042.

[5] John G Daugman. 1985. Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-dimensional Visual Cortical Filters. *JOSA A* 2, 7 (1985), 1160–1169.

[6] David L Donoho and Michael Elad. 2003. Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via ℓ_1 Minimization. *Proceedings of the National Academy of Sciences* 100, 5 (2003), 2197–2202.

[7] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. 2019. Neural Architecture Search: A Survey. *The Journal of Machine Learning Research* 20, 1 (2019), 1997–2017.

[8] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. *arXiv preprint arXiv:1604.08880* (2016).

[9] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861* (2017).

[10] Manjunath Jogin, MS Madhulika, GD Divya, RK Meghana, S Apoorva, et al. 2018. Feature Extraction Using Convolution Neural Networks (CNN) and Deep Learning. In *2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*. IEEE, 2319–2323.

[11] Gerald Kaiser and Lonnie H Hudgins. 1994. *A Friendly Guide to Wavelets*. Vol. 300. Springer.

[12] Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* (2014).

[13] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity Recognition Using Cell Phone Accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.

[14] Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2018. DARTS: Differentiable Architecture Search. *arXiv preprint arXiv:1806.09055* (2018).

[15] Shengzhong Liu, Shuochao Yao, Jinyang Li, Dongxin Liu, Tianshi Wang, Huajie Shao, and Tarek Abdelzaher. 2020. GlobalFusion: A Global Attentional Deep Learning Framework for Multisensor Information Fusion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–27.

[16] Zhuang Liu, Jianguo Li, Zhiqiang Shen, Gao Huang, Shoumeng Yan, and Changshui Zhang. 2017. Learning Efficient Convolutional Networks Through Network Slimming. In *Proceedings of the IEEE international conference on computer vision*. 2736–2744.

[17] Saif Mahmud, M Tonmoy, Kishor Kumar Bhaumik, AK Mahbubur Rahman, M Ashraful Amin, Mohammad Shoyaib, Muhammad Asif Hossain Khan, and Amin Ahsan Ali. 2020. Human Activity Recognition from Wearable Sensor Data Using Self-attention. *arXiv preprint arXiv:2003.09018* (2020).

[18] Stephane Mallat et al. 2009. A Wavelet Tour of Signal Processing: the Sparse Way. *AP Professional, Third Edition, London* (2009).

[19] Yaowen Mei, Ting Jiang, Xue Ding, Yi Zhong, Sai Zhang, and Yang Liu. 2021. Wiwave: Wifi-based Human Activity Recognition Using the Wavelet Integrated CNN. In *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. 100–105. https://doi.org/10.1109/ICCCWorkshops52231.2021.9538931

[20] Sebastian Münzner, Philip Schmidt, Attila Reiss, Michael Hanselmann, Rainer Stiefelhagen, and Robert Dürichen. 2017. CNN-based sensor fusion techniques for multimodal human activity recognition. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 158–165.

[21] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1 (2016), 115.

[22] Attila Reiss and Didier Stricker. 2012. Introducing A New Benchmarked Dataset for Activity Monitoring. In *2012 16th international symposium on wearable computers*. IEEE, 108–109.

[23] Maria Ximena Bastidas Rodriguez, Adrien Gruson, Luisa Polania, Shin Fujieda, Flavio Prieto, Kohei Takayama, and Toshiya Hachisuka. 2020. Deep Adaptive Wavelet Network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*.

[24] Charissa Ann Ronao and Sung-Bae Cho. 2016. Human Activity Recognition with Smartphone Sensors Using Deep Learning Neural Networks. *Expert systems with applications* 59 (2016), 235–244.

[25] Peter J Rousseeuw. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[26] Yiming Tian, Jie Zhang, Jie Wang, Yanli Geng, and Xitai Wang. 2020. Robust Human Activity Recognition Using Single Accelerometer via Wavelet Energy Spectrum Features and Ensemble Feature Selection. *Systems Science & Control Engineering* 8, 1 (2020), 83–96.

[27] Cédric Vonesch, Thierry Blu, and Michael Unser. 2007. Generalized Daubechies Wavelet Families. *IEEE Transactions on Signal Processing* 55, 9 (2007), 4415–4429.

[28] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini, and Gerhard Tröster. 2008. Activity Recognition from On-body Sensors: Accuracy-power Trade-off by Dynamic Sensor Selection. In *European Conference on Wireless Sensor Networks*. Springer, 17–33.

[29] Dongwen Zhang, Lin Zhang, Qingwu Yi, Lu Huang, and Guanghua Zhang. 2021. Human Activity Recognition Based on Wavelet-CNN Architecture. In *2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT)*. 77–85. https://doi.org/10.1109/ACAIT53529.2021.9731338

[30] Haibin Zhao, Christopher Funk, Benjamin Noack, Uwe Hanebeck, and Michael Beigl. 2021. Kalman Filtered Compressive Sensing Using Pseudo-Measurements. In *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 1–8.