Sensor Data Processing

Light Residual Network for Human Activity Recognition using Wearable Sensor Data

Francisco M. Calatrava-Nicolás[1], Oscar Martinez Mozos[1*]

[1] *AI for Life, Centre for Applied Autonomous Sensor Systems, Örebro University, 70182, Sweden*
[*] *Member, IEEE*

Manuscript received June 7, 2017; revised June 21, 2017; accepted July 6, 2017. Date of publication July 12, 2017; date of current version July 12, 2017.

Abstract—This paper addresses the problem of Human Activity Recognition (HAR) of people wearing inertial sensors using data from the UCI-HAR dataset. We propose a light residual network which obtains an F1-Score of 97.6% that outperforms previous works, while drastically reducing the number of parameters by a factor of 15, and thus the training complexity. In addition, we propose a new benchmark based on leave-one(person)-out cross-validation to standardize and unify future classifications on the same dataset, and to increase reliability and fairness in the comparisons.

Index Terms—Sensor data processing, deep learning, human activity recognition, residual network, inertial sensors

I. INTRODUCTION

Human Activity Recognition (HAR) is the problem of identifying the activities carried out by a person by collecting and analyzing data from different cues [1], [2] such as wearable [3], [4] or environmental sensors [5]–[7]. Applications of HAR include the recognition of Activities of Daily Living (ADL) [4], [8], [9], surveillance [6], [10], Human-Robot Interaction (HRI) [11]–[13], autonomous vehicles [14], [15], and remote healthcare [16]–[18], among others.

This paper focuses on the HAR problem using data from wearable inertial sensors applied to the classification of ADL. Recent approaches to this problem have applied deep learning techniques obtaining high classification rates [19]–[27]. However, deep learning approaches demand high computational power, long training times, and high energy consumption [28], [29]. A reduction in those demands is desirable, in particular for applications based on wearables, smartphones, or the Internet of Things (IoT) technologies [30].

In this paper, we introduce a light residual network for the HAR problem when using temporal data from wearable inertial sensors. Our architecture is a modification of the ResNet18 [31], in which we have reduced the number of residual blocks and have adapted the kernels to the 1-dimensional nature of the temporal signals provided by inertial sensors. As a result, our model drastically reduces the number of trainable parameters from several million to 234, 950, thus improving efficiency in performance and reducing complexity. The full architecture is shown in Fig. 1.

We tested our approach in the popular ADL dataset UCI-HAR [4], which is one of the most cited in the UC Irvine Machine Learning Repository [32]. Our classification results outperform previous approaches while reducing the complexity of the model, the training time per epoch, and thus, the energy consumption.

Corresponding author: Francisco M. Calatrava-Nicolás (e-mail: francisco.calatrava-nicolas@oru.se).
Associate Editor: Alan Smithee.
Digital Object Identifier 10.1109/LSENS.2017.0000000
This work was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

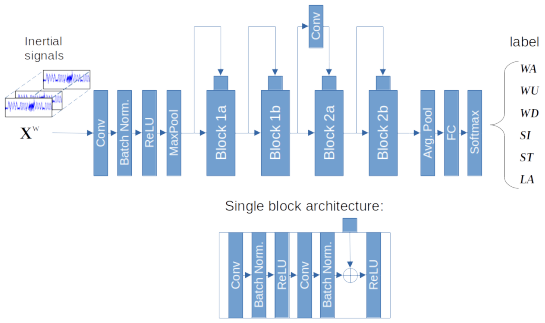

Fig. 1. Light residual architecture. The input signals window is represented by tensor $\mathbf{X}^w$ (c.f Sect III). Four residual blocks share a common architecture (bottom diagram). The output is one of the six activities: walking (WA), walking upstairs (WU), walking downstairs (WD), sitting (SI), standing (ST), and laying (LA).

Previous works using the UCI-HAR dataset use the original fixed division of participants for training and testing, which limits the details on performance. Therefore, we propose a benchmark based on leave-one-out cross-validation (LOOCV) by iteratively leaving one participant out for testing while training with the rest, which is a more standardized way of comparing results in these kinds of problems. We think this benchmark will unify comparisons better, and will increase their reliability and fairness.

In summary, the contributions of this paper are three-fold. First, we present a light architecture that outperforms previous deep learning works on the UCI-HAR dataset. Second, our simplified architecture drastically reduces the complexity of the model. And third, we propose a standard benchmark in order to unify future comparisons.

II. RELATED WORK

Last approaches addressing the HAR problem are mostly based on deep learning techniques [33], [34] including Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) or

TABLE 1. Parameters for the light residual model

Layer	Output Size	Kernel Size	Stride	Padding
Input	$9 \times 1 \times 128$	-	-	-
Conv1	$64 \times 1 \times 64$	1×5	2	2
Maxpool	$64 \times 1 \times 32$	1×3	2	1
Block 1a	$64 \times 1 \times 32$	1×3	1	1
Block 1b	$64 \times 1 \times 32$	1×3	1	1
Block 2a	$128 \times 1 \times 16$	1×3	2	1
Block 2b	$128 \times 1 \times 16$	1×3	1	1
Avgpool	$128 \times 1 \times 1$	-	-	-
FC	6	-	-	-

Gated Recurrent Units (GRUs) [20], [23], [25]; Convolutional Neural Networks (CNNs) [24]; a combination of CNN and RNN [22]; or a combination of the previous models with additional techniques like attention mechanisms, or residual connections [19], [21], [26], [27].

In particular, the work in [24] presents a comparison between automatic features from a CNN with Human Crafted Features (HCFs), and confirms that CNN-based features provide performances comparable to the best set of HCFs. Moreover, authors in [26] present a CNN residual network with a modified inception module to improve the predictions. Also, a bidirectional LSTM is proposed in [20] to explore the impact of temporal features in the classification performance. Similarly, the work in [23] defines a model based on stacked LSTM which improves performance. In [21], a residual bidirectional LSTM is proposed for a better temporal feature extraction, thus improving the classification while avoiding the vanishing gradient problem.

The work in [22] introduces a hybrid CNN+LSTM model for temporal feature extraction, obtaining better results than models based on LSTM, LSTM+Dense layers, and CNN+LSTM+Dense layers. In [25], a comparison among different hybrid models showed the best results using a combination of CNN+LSTM with a self-attention mechanism which keeps a good balance between performance and the number of parameters. Still, our light model has fewer parameters than [25] while outperforming the classification.

A parallel two-branch model is presented in [19] where the first branch uses residual attention blocks for the spatial feature extraction, and the second branch applies bidirectional GRU with self-attention for the temporal features. While the classification rates are high, the number of parameters remains very high (1.6 million). Finally, the work in [27] proposes a multi-frequency channel attention framework combined with residual networks and obtains the best classification results so far. In comparison, our light model slightly outperforms [27] while reducing its number of parameters by a factor of 15.

Our light architecture outperforms all previous methods while drastically reducing the number of trainable parameters to 234, 950, thus improving efficiency in performance and reducing complexity.

III. HAR USING INERTIAL SENSORS

In this work, we address the HAR problem using temporal data from inertial sensors worn by a person while carrying on different ADLs. For this, we use the UCI-HAR dataset [4] which contains inertial data from different participants executing different activities. We focus on the six activities in the dataset: walking (WA), walking upstairs (WU), walking downstairs (WD), sitting (SI), standing (ST), and laying (LA). The dataset contains data from 30 participants that

TABLE 2. Comparison with previous works.

Approaches	F1-Score (%)	Accuracy (%)	Params.
GRU [19] (2023)	89.2	89.2	-
CNN-LSTM Self-Att. [25] (2022)	90.9	93.1	634,188
Res-BiLSTM [21] (2018)	91.5	91.6	-
CNN-LSTM [22] (2020)	-	92.1	-
Bi-LSTM [20] (2019)	92.7	92.7	-
Stacked LSTM [23] (2019)	93.1	93.1	-
CNN [24] (2019)	93.5	-	-
iSPLInception [26](2021)	95.0	95.1	1,327,754
GRU+Attention [19] (2023)	95.8	96.0	1,600,000
CNN-DCT [27] (2023)	97.1	-	930,000
ResNet-DCT [27] (2023)	97.5	-	3,540,000
Our model	**97.6**	**97.6**	**234,950**

TABLE 3. Confusion matrix for the first experiment.

		Predicted label						
		WA	WU	WD	SI	ST	LA	Sup.
Actual label	WA	**96.2%**	0.6%	3.2%	0%	0%	0%	
		477	3	16	0	0	0	496
	WU	0%	**99.8%**	0.2%	0%	0%	0%	
		0	**470**	1	0	0	0	471
	WD	0%	0.5%	**99.5%**	0%	0%	0%	
		0	2	**418**	0	0	0	420
	SI	0%	0.4%	0%	**91.6%**	7.0%	1.0%	
		0	2	0	**450**	34	5	491
	ST	0%	0%	0%	1.5%	**98.5%**	0%	
		0	0	0	8	**524**	0	532
	LA	0%	0%	0%	0%	0%	**100%**	
		0	0	0	0	0	**537**	537

wore a smartphone on their waist while performing the activities. The temporal signals were obtained from the inertial sensors inside the smartphone and were composed of nine 1-dimensional signals: tri-axial acceleration from the accelerometer, tri-axial estimated body acceleration, and tri-axial angular velocity, all sampled at a frequency of 50Hz. The signals were pre-processed using a Butterworth low-pass filter to separate body acceleration and gravity. Each 1-dimensional temporal signal is divided into windows of 2.56 seconds containing 128 samples each, with an overlapping of 50%. In total, the dataset contains 10299 signal windows of 128 samples each [4].

The signal from each inertial sensor is represented as follows:

$$\mathbf{x}_c^w = \{x_1, ..., x_{|S|}\}, \tag{1}$$

where $\mathbf{x}_c^w$ indicates the sample vector from window w and cue c, with $S = \{1, ...128\}$, and $c \in C = \{1...9\}$. Thus, each input to our architecture is composed of nine parallel 1-dimensional window vectors in the form:

$$\mathbf{X}^w = \{\mathbf{x}_1^w, ..., \mathbf{x}_{|C|}^w\}, \tag{2}$$

which translates into a tensor $\mathbf{X}^w$ with dimensions $(|C| \times 1 \times |S|)$, i.e. $(9 \times 1 \times 128)$, corresponding to (depth, height, width). Thus, our classification problem consists of labeling each tensor $\mathbf{X}^w$ into one of the six activities $\{WA, WU, WD, SI, ST, LA\}$, as shown in the input and output of Fig.1.

IV. LIGHT RESIDUAL NETWORK

We propose a light CNN with residual connections based on the ResNet18 [31]. Our model reduces the number of residual blocks to four in order to keep a balance between classification performance and model complexity. Moreover, instead of 2-dimensional kernels, we defined 1-dimensional kernels that adapt better to the 1-dimensional

TABLE 4. Classification results for the LOOCV benchmark.

Person out	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Precision	100.0 ± 0.0	99.4 ± 1.5	98.6 ± 2.5	98.4 ± 2.4	93.2 ± 14.3	98.7 ± 3.3	98.3 ± 4.2	97.3 ± 6.7	94.4 ± 5.5	83.8 ± 17.2	100.0 ± 0.0	97.9 ± 5.2	99.7 ± 0.7	71.8 ± 40.6	99.7 ± 0.8
Recall	100.0 ± 0.0	99.3 ± 1.8	98.4 ± 3.1	98.4 ± 2.6	92.2 ± 14.8	98.5 ± 3.7	97.9 ± 5.1	97.2 ± 6.8	94.5 ± 7.5	81.0 ± 20.8	100.0 ± 0.0	97.1 ± 7.2	99.7 ± 0.8	79.6 ± 40.0	99.7 ± 0.7
F1-Score	100.0 ± 0.0	99.3 ± 1.1	98.5 ± 2.3	98.4 ± 2.5	91.5 ± 11.5	98.5 ± 2.3	98.0 ± 3.1	97.0 ± 4.6	94.4 ± 6.3	79.3 ± 11.5	100.0 ± 0.0	97.2 ± 4.4	99.7 ± 0.5	73.8 ± 38.4	99.7 ± 0.5
Support	347	302	341	317	302	325	308	281	288	294	316	320	327	323	328
Person out	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Precision	94.4 ± 9.8	97.7 ± 4.3	99.4 ± 1.4	99.8 ± 0.5	100.0 ± 0.0	98.0 ± 4.9	99.7 ± 0.6	97.3 ± 6.6	100.0 ± 0.0	93.7 ± 10.9	100.0 ± 0.0	99.8 ± 0.5	95.8 ± 10.4	99.5 ± 1.3	100.0 ± 0.0
Recall	94.3 ± 10.4	97.2 ± 5.6	99.5 ± 1.1	99.8 ± 0.6	100.0 ± 0.0	97.6 ± 5.8	99.7 ± 0.7	96.8 ± 7.8	100.0 ± 0.0	93.4 ± 10.2	100.0 ± 0.0	99.8 ± 0.6	93.8 ± 15.3	99.5 ± 1.3	100.0 ± 0.0
F1-Score	94.1 ± 9.2	97.4 ± 4.2	99.5 ± 0.8	99.8 ± 0.3	100.0 ± 0.0	97.7 ± 3.6	99.7 ± 0.4	96.8 ± 5.0	100.0 ± 0.0	93.3 ± 9.8	100.0 ± 0.0	99.8 ± 0.3	93.7 ± 10.1	99.5 ± 0.8	100.0 ± 0.0
Support	366	368	364	360	354	408	321	372	381	409	392	376	382	344	383

TABLE 5. Global confusion matrix for the LOOCV benchmark.

		Predicted label						
		WA	WU	WD	SI	ST	LA	Sup.
Actual label	WA	**95.7%**	3.5%	0.6%	0.2%	0.0%	0.0%	
		1648	60	11	3	0	0	1722
	WD	0.6%	**98.3%**	1.0%	0.0%	0.0%	0.0%	
		10	**1518**	16	0	0	0	1544
	WU	0.0%	0.0%	**99.9%**	0.1%	0.0%	0.0%	
		0	0	**1404**	2	0	0	1406
	SI	0.0%	0.2%	0.0%	**92.3%**	7.4%	0.1%	
		0	3	0	**1641**	132	1	1777
	ST	0.0%	0.0%	0.0%	5.2%	**94.8%**	0.0%	
		0	0	0	99	**1807**	0	1906
	LA	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**	
		0	0	0	0	0	**1944**	1944

input signals (see Sect. III). As a consequence, the total number of trainable parameters in our model was reduced to 234, 950.

Our architecture is shown in Fig. 1. The input contains the raw signal data in tensor $\mathbf{X}^w$ (c.f. Sect.III). The initial structure includes a convolutional, batch normalization, and a ReLU layer, with a final max pooling layer. Afterward, we have four residual blocks (1a, 1b, 2a, and 2b), each composed of two convolutional layers (see bottom of Fig 1). Blocks 1a and 1b contain 64 filters, while blocks 2a and 2b contain 128 filters. Finally, the features are transformed using an average pooling and a fully connected layer into the six output probability labels, that are discretized using a softmax function. Table 1 shows the most important parameters.

Like in some previous works, e.g. [19], [24], our input tensor has dimensions $(9, 1, 128)$ which adapts better to the 1-dimensional nature of the input data. This also allows us to simplify the architecture, to reduce the parameters, and to prioritize the correlations among all signal cues. Some other works, like [20], used a tensor with dimensions $(1, 128, 9)$ to adapt to their specific architectures, but those architectures grow in complexity and number of parameters.

V. EXPERIMENTS

We tested our light architecture on the UCI-HAR dataset [4]. We trained with a learning rate of 0.0008 that increased by a factor of 0.4 every 50 epochs, and a batch size of 16. We applied 300 epochs and used the Adam optimizer with a weight decay value of 0.0001. We implemented our model on a GPU Quadro RTX 6000 using Pytorch 1.10.1+cu102 on Ubuntu 18.04.6 LTS.

In the first experiment we compared our classification results with previous works. For a fair comparison, we kept the training and test conditions presented in the original UCI-HAR dataset [4], i.e., 70% of participants for training and 30% for testing, and the same distribution of participants. We included works that confirmed the same experimental conditions [19]–[27].

Table 2 compares our average F1-Score, accuracy, and number of parameters with previous works, with our model outperforming them (entries marked with "-" indicate that the value was not made available on the corresponding paper). These results show that our light model obtains better results using much fewer parameters and thus reducing the complexity of the model.

In addition, Table 3 details the confusion matrix in this first experiment. The main confusions are between SI and the standing activity ST. This result is in accordance with previous works [19]–[21], [23], [25]. Additionally, our model reduces the confusion among the activities WA, WU, and WD, improving over previous results [21]–[23], [25]. Our light model does not apply temporal relations between consecutive tensors. Instead, we used a 1-dimensional input and focused on finding the optimal number of residual blocks to reduce complexity and increase efficiency. An extra ablation study shows that 4 residual blocks provide the best F1-Score while keeping the number of parameters low (see Appendices in [35]).

In the previous experiments we followed the original experimental conditions presented in [4] to keep a fair comparison with previous works. However, they restricted to only one pair of training and test sets with a fix number of participants. To unify future comparisons we present a benchmark based on Leaving-One-Out Cross-Validation (LOOCV) where iteratively one participant is left out for testing and the rest (29) are used for training. We applied our light model to this benchmark and the global confusion matrix is shown in Table 5. The biggest confusion occurs between SI and ST but in small percentages. Table 2 and Table 5 present similar behaviors, which confirms the robustness of our model in different training and test sets.

Finally, Table 4 shows the average values for the precision, recall, and F1-score metrics among the six activities for each participant in the LOOCV benchmark. We think the low results for Participant 14 can be due to a problem in the data collection. This could be contrasted with other works when using our proposed one-person-out scheme, which is also a reason to support the use of our benchmark. Further results on the benchmark, and the code for our model are available in the Appendices in [35]. We carried out two additional 10-fold cross-validation studies, with and without stratification, with F1-Scores of 97.1% and 96.7% respectively (see Appendices in [35]).

VI. CONCLUSION

We presented a new light architecture for the HAR problem that outperformed previous works. We simplified the deep-learning models to increase their suitability to real life wearable-based applications. Future work will extend our study to new datasets, and will investigate further reductions in the models' complexity.

REFERENCES

[1] L. M. Dang, K. Min, H. Wang, *et al.*, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognit.*, vol. 108, p. 107 561, 2020.

[2] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," *Comput. Vis. Image Underst.*, vol. 117, no. 6, pp. 633–659, 2013.

[3] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 3, pp. 1192–1209, 2012.

[4] D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz, *et al.*, "A public domain dataset for human activity recognition using smartphones," in *Proc. 21st ESANN*, 2013, pp. 437–442.

[5] J. K. Aggarwal and L. Xia, "Human activity recognition from 3d data: A review," *Pattern Recognit. Lett.*, vol. 48, pp. 70–80, 2014.

[6] L. Yeffet and L. Wolf, "Local trinary patterns for human action recognition," in *IEEE ICCV*, 2009, pp. 492–497.

[7] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from rgbd images," in *IEEE ICRA*, 2012, pp. 842–849.

[8] H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *IEEE CVPR*, 2012, pp. 2847–2854.

[9] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 81–94, 2016.

[10] M. Barnachon, S. Bouakaz, B. Boufama, and E. Guillou, "Ongoing human action recognition with motion capture," *Pattern Recognit.*, vol. 47, no. 1, pp. 238–247, 2014.

[11] L. Piyathilaka and S. Kodagoda, "Human activity recognition for domestic robots," in *Field and Service Robotics: Results of the 9th International Conference*, Springer, 2015, pp. 395–408.

[12] A. Roitberg, A. Perzylo, N. Somani, *et al.*, "Human activity recognition in the context of industrial human-robot interaction," in *IEEE APSIPA*, 2014, pp. 1–10.

[13] S. Coşar, M. Fernandez-Carmona, R. Agrigoroaie, *et al.*, "Enrichme: Perception and interaction of an assistive robot for the elderly at home," *Int. J. Soc. Robot.*, vol. 12, pp. 779–805, 2020.

[14] E. Ohn-Bar and M. M. Trivedi, "Looking at humans in the age of self-driving and highly automated vehicles," *IEEE T. Intell. Veh.*, vol. 1, no. 1, pp. 90–104, 2016.

[15] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 900–918, 2019.

[16] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Syst. Appl.*, vol. 137, pp. 167–190, 2019.

[17] X. Zhou, W. Liang, I. Kevin, *et al.*, "Deep-learning-enhanced human activity recognition for internet of healthcare things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6429–6438, 2020.

[18] T. Hossain and S. Inoue, "Sensor-based daily activity understanding in caregiving center," in *IEEE PerCom Workshops*, 2019, pp. 439–440.

[19] Y. Wang, H. Xu, Y. Liu, *et al.*, "A novel deep multi-feature extraction framework based on attention mechanism using wearable sensor data for human activity recognition," *IEEE Sens. J.*, vol. 23, no. 7, pp. 7188–7198, 2023.

[20] F. Hernández, L. F. Suárez, J. Villamizar, and M. Altuve, "Human activity recognition on smartphones using a bidirectional lstm network," in *STSIVA)*, 2019, pp. 1–5.

[21] Y. Zhao, R. Yang, G. Chevalier, *et al.*, "Deep residual bidir-lstm for human activity recognition using wearable sensors," *Math. Probl. Eng.*, vol. 2018, pp. 1–13, 2018.

[22] R. Mutegeki and D. S. Han, "A cnn-lstm approach to human activity recognition," in *ICAIIC*, 2020, pp. 362–366.

[23] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, "Stacked lstm network for human activity recognition using smartphone data," in *EUVIP*, 2019, pp. 175–180.

[24] F. Cruciani, A. Vafeiadis, C. Nugent, *et al.*, "Comparing cnn and human crafted features for human activity recognition," in *IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*, 2019, pp. 960–967.

[25] M. A. Khatun, M. A. Yousuf, S. Ahmed, *et al.*, "Deep cnn-lstm with self-attention model for human activity recognition using wearable sensor," *IEEE J. Transl. Eng. Health Med.-JTEHM*, vol. 10, pp. 1–16, 2022.

[26] M. Ronald, A. Poulose, and D. S. Han, "Isplinception: An inception-resnet deep learning architecture for human activity recognition," *IEEE Access*, vol. 9, pp. 68 985–69 001, 2021.

[27] S. Xu, L. Zhang, Y. Tang, C. Han, H. Wu, and A. Song, "Channel attention for sensor-based activity recognition: Embedding features into all frequencies in dct domain," *IEEE Trans. Knowl. Data Eng.*, pp. 1–15, 2023.

[28] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," *arXiv:2007.05558*, 2020.

[29] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahn, "Estimation of energy consumption in machine learning," *J. Parallel Distrib. Comput.*, vol. 134, pp. 75–88, 2019.

[30] N. D. Lane, S. Bhattacharya, P. Georgiev, *et al.*, "An early resource characterization of deep learning on wearables, smartphones and internet-of-things devices," in *Int. Work. on Internet of Things Towards Applications*, 2015, pp. 7–12.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. arXiv: 1512.03385.

[32] *University of california, irvine, machine learning repository*, https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones, Accessed: 2023-05-22.

[33] E. Ramanujam, T. Perumal, and S. Padmavathi, "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review," *IEEE Sens. J.*, vol. 21, no. 12, pp. 13 029–13 040, 2021.

[34] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, *et al.*, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, 2018.

[35] F. Calatrava-Nicolas and O. M. Mozos, *Light residual network*, github.com/FranciscoCalatrava/Light_Residual_Network.