A Hybrid Residual Neural Network with Attention Mechanism for Activity-based User Identification Using Smartwatch Sensors

Sakorn Mekruksavanich[1*] and Anuchit Jitpattanakul[2]

[1] Department of Computer Engineering, School of Information and Communication Technology
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th
[2] Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics
Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th

Abstract

The rise in prominence of smartwatches has necessitated the inclusion of user identification as a crucial element in safeguarding confidentiality and safety. The current research examined the potential utilization of smartwatch sensors' data for user identification and provided recommendations for the most effective application aspects. Based on prior findings, it is evident that deep learning methodologies have shown superior identifying effectiveness, as indicated by a substantial array of machine learning indicators of effectiveness. In order to improve the accuracy of identification, this study proposes utilizing a hybrid residual neural network known as Att-ResBiLSTM. This network combines the residual structure and layer normalization with a bidirectional long short-term memory network. The integration described in this study enhances the network's capabilities for extracting features. It also includes an attention method to optimize the final feature information. As a result, the accuracy and stability of human identification based on smartwatch sensor data are improved. In order to assess the efficacy of the proposed Att-ResBiLSTM network, we conducted thorough testing using a publicly available dataset known as the SP-SW HAR dataset. This dataset encompasses a wide range of activity data obtained from sensors embedded in smartwatches. The outcomes were exceedingly promising, with exceptional overall identification accuracy of 98.13% and an F1-score of 98.09%. The empirical findings indicate that the utilization of this approach significantly improves the efficacy of user identification through the utilization of smartwatch sensors.

Keywords: Smartwatch Sensors, Hybrid Residual Neural Network(Att-ResBiLSTM), Attention Mechanism

1 Introduction

In contemporary society, intelligent wearable devices have emerged as an essential element in individuals' daily routines, serving many purposes beyond communication. One noteworthy application may be observed in the Internet of Things (IoT) context, wherein smartphones and smartwatches leverage their data-sensing, representation, and edge-computing characteristics to facilitate the implementation of IoT-based solutions [1, 2].

The 7th International Conference on Mobile Internet Security (MobiSec'23), Dec. 19-21, 2023, Okinawa, Japan, Article No.61
*Corresponding author: Department of Computer Engineering, Phayao University, Phayao, Phayao 56000, Thailand, Tel: +66-054-466666-2326

Smartphones have become an indispensable component of our daily routines in several residential and professional settings, frequently employed to access security systems that rely on cloud-based technology. Smartwatches present a compelling platform for implementing reliable identity verification systems, mainly when smartphones are susceptible to theft or damage. This can be achieved by utilizing cloud-based services, including online banking. In connecting mission-critical Internet services through cloud-based or other data sources, it is imperative to accurately ascertain the authentic individual who consistently engages in this activity. Implementing automated identifying systems that cannot be bypassed is necessary [3].

In the past decade, learning approaches, specifically machine learning (ML), have been utilized to attain favorable results in user identification based on biometrics. Under controlled conditions, various ML algorithms have demonstrated the ability to produce satisfactory outcomes [4]. The efficacy of conventional ML models is heavily contingent upon the approach employed in manually extracting and selecting characteristics by humans to operate them.

Currently, deep learning (DL) techniques have achieved success during research related to user identification. One of the critical aspects of DL is its capacity to autonomously detect and categorize information with enhanced accuracy, hence impacting investigations related to user identification [5]. Deep neural networks can autonomously acquire selective features from unprocessed input, demonstrating considerable potential in the analysis of various datasets and exhibiting a notable capacity for generalization. Many fundamental and advanced DL models have been presented to harness the potential of DL methodologies. These models seek to address the limitations of classic ML techniques while taking advantage of the diverse categories of features present in different hierarchies.

In order to improve the precision of identification, this study proposes utilizing a hybrid residual neural network known as Att-ResBiLSTM. This network combines the residual architecture and layer normalization with a bidirectional long short-term memory network (BiLSTM). The integration described in this study boosts the network's potential in extracting features. It also includes an attention mechanism for improving the final feature information. As a result, the accuracy and stability of user identification based on smartwatch sensor data are improved.

2 Related Works

2.1 User Identification Based on Smartwatch Sensor Data

In the past decade, there has been a growing consideration of incorporating wearable sensors into tracking mechanisms that are based on smartwatches. As an illustration, an idea for a system has been put out by [6] to openly and consistently ascertain the identification of a person. Nevertheless, there is a lack of substantial proof to support the notion that the person's conduct has undergone sufficient changes to justify its obvious categorization in most situations. The approach proposed by Luca et al. [7] involves intentionally determining the gap between pattern traces by utilizing the dynamic temporal warping mechanism. Most of the 22 atypical touch patterns demonstrated by Sae-Bae et al. [8] include the simultaneous utilization of all five fingers. In the research investigation [9], the authors employed k-nearest neighbors and support vector machines as classification algorithms to identify the 22 analytical features derived from touch recordings.

Consequently, each reaction is associated with two fundamental attributes, namely time and space, by the premise of the behavior-based model. For instance, one could examine actions akin to those delineated by [10] in pursuing user identification. The concept of multimodel continuous user identification was proposed in previous research conducted by [11]. An alternative architectural approach for continuous user identification was suggested in a scholarly article [12], which utilizes prior smartphone data and location information.

To a certain extent, all of the responsibilities above necessitate supplementary information and a means of user authentication. Casale et al. [13] proposed a method for user identification based on gait, which utilizes an unobtrusive biometric pattern to mitigate the problems above. The construction of a four-layered framework was based on applying the geometric principle known as a convex hull. One notable drawback is its limited functionality, as it only operates in specified geographical areas. Rong et al. [14] utilized electronic gadgets that relied on gait data captured by a three-dimensional accelerometer. The accelerometer was conveniently affixed to the individual's posterior waist in these studies. Mantyjarvi et al. [15] created a user identification technique incorporating three key factors: data distribution statistics, correlation, and time-frequency characteristics. Simultaneously, individuals were intentionally instructed to ambulate at varying speeds, including slow, moderate, or fast. One notable limitation of their research is the constraint that only one individual can engage in walking activities at any given moment, with a relatively limited range of variations.

2.2 DL Approaches

The domain of DL has exhibited its efficacy in domains of gait identification. This encompasses the tasks involving motion identification [16], video categorization [17], and face recognition [18]. Nevertheless, training DL models necessitates a substantial volume of data, which poses a challenge in gait data collection due to the constraints imposed by portable electronics regarding battery life and processing capabilities. In contrast to conventional ML approaches, DL techniques are less commonly utilized in addressing gait-based implicit authentication issues. In their study, Gadaleta et al. [19] introduced a framework for user authentication called IDNet, which utilizes movement information collected from smartphones. The utilization of IDNet involves exploiting CNN to extract general features while still employing one-class SVM as the classifier. The convolutional neural network (CNN) takes the original gait signal as its input directly. The suitability of this approach for CNN implementation is limited due to CNN's inherent limitations in efficiently interpreting one-dimensional information. The user authentication scheme developed by Giorgi et al. [20] involves the utilization of inertial sensors and employs a recurrent neural network (RNN) for DL-based classification. Nevertheless, there is a substantial disparity in the outcomes between known identities and unknown identities.

The concept of cross-channel interactivity offers clear benefits in enhancing the efficacy of DL methodologies, in contrast to certain contemporary scholarly studies that have been put forth. The squeeze-and-excitation (SE) network, proposed by Hu et al. [21], aims to calibrate channel feature reactions. Chen et al. [22] and Dai et al. [23] employed channel-wise attention mechanisms in their respective semantic segmentation and image captioning studies. Wu and He [24] employed the technique of group normalization. The above model might be interpreted as a specialized model incorporating channel-wise connectivity. Nevertheless, how they engage with one another across several channels could be more complex, as it solely involves the calculation of the mean and standard deviation of feature maps. In computer vision, Yang et al. [25] introduced the concept of cross-channel interaction inside a single layer. This approach facilitates communication between channels within the same layer, enhancing efficiency.

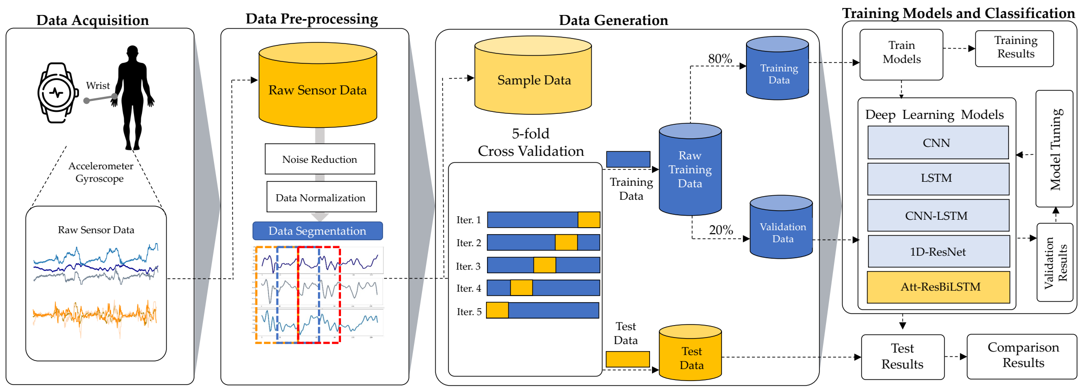

Figure 1: The proposed framework of activity-based user identification using smartwatch sensors

3 The Activity-base User Identification Framework

The methodology employed in this study for user identification is the activity-based approach. It encompasses four primary stages, including data acquisition, pre-processing, generation, and model training and assessment, as depicted in Figure 1.

3.1 SP-SW HAR Dataset

The SP-SW HAR dataset [26] is an openly available Human Activity Recognition (HAR) reference dataset. This dataset comprises sensor data from smartphones and smartwatches, explicitly utilizing the inertial measurement unit (IMU) technology. Data about linear acceleration and angular velocity were gathered from a sample of 23 individuals ranging in age from 23 to 66. The sample comprised ten female participants and fifteen male participants. The participants were provided with a smartphone and a smartwatch, both of which were equipped with the program. They were not subjected to any limitations and were free to engage in their regular daily activities. Per the data-collecting methodology, participants were instructed to categorize behaviors frequently seen in individuals' everyday routines. These actions included seated, sitting, standing, bending, and strolling. Figure 2 displays examples of accelerometer and gyroscope data.

This study primarily centered on using smartwatch sensor data due to several factors. The smartwatch is a prevalent technology utilized in everyday life. In addition, the smartwatch has enhanced computational capabilities for executing exceptionally effective DL models.

3.2 Data Pre-processing

The smartwatch's raw sensor data underwent two pre-processing techniques: noise removal and data standardization. The pre-processed sensor data underwent segmentation employing fixed-width sliding windows of 2 seconds, with a non-overlapping measure, as depicted in Figure 3.

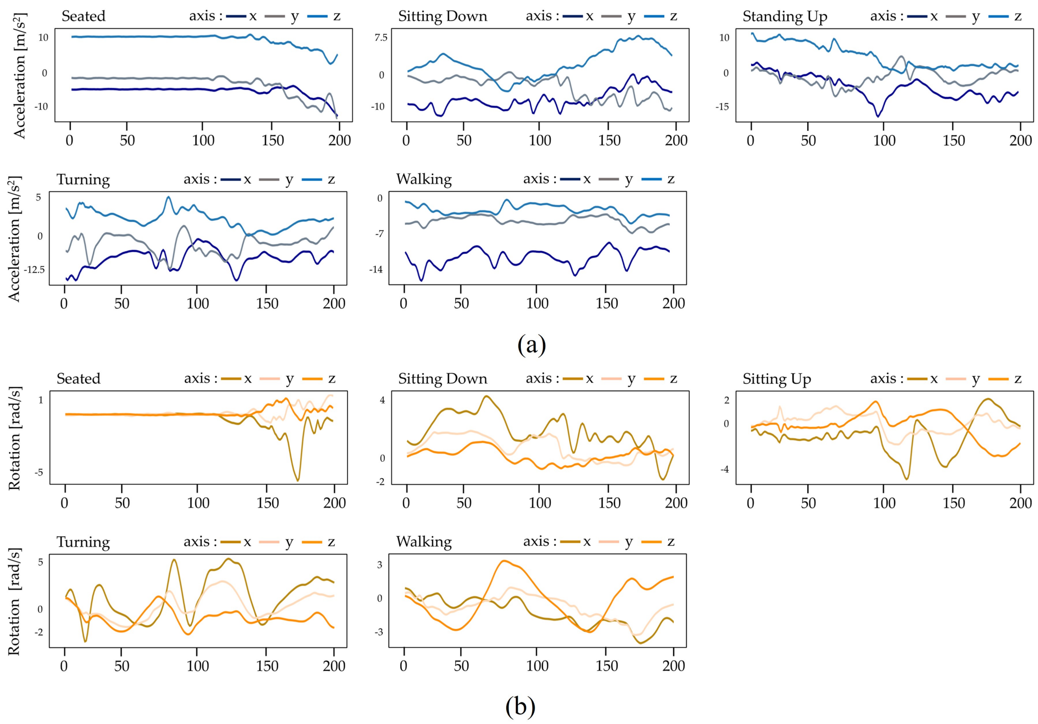

Figure 2: Some sample data of five activities from the SP-SW HAR dataset: (a) accelerometer data (b) gyroscope data

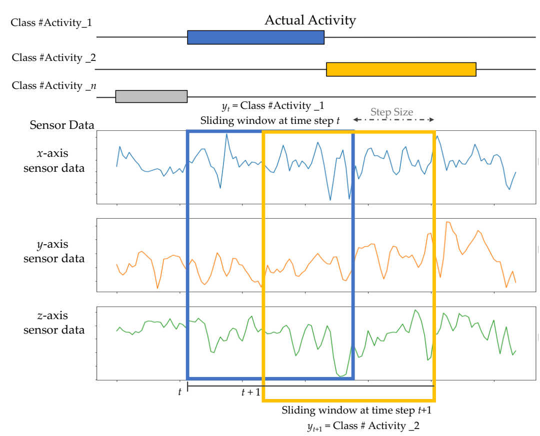

Figure 3: Data segmentation using a fixed-width sliding window used in this work

3.3 Training DL Models

In this study, we have introduced a novel hybrid residual neural network known as Att-ResBiLSTM, which aims to address the issue of user identification. The network design of

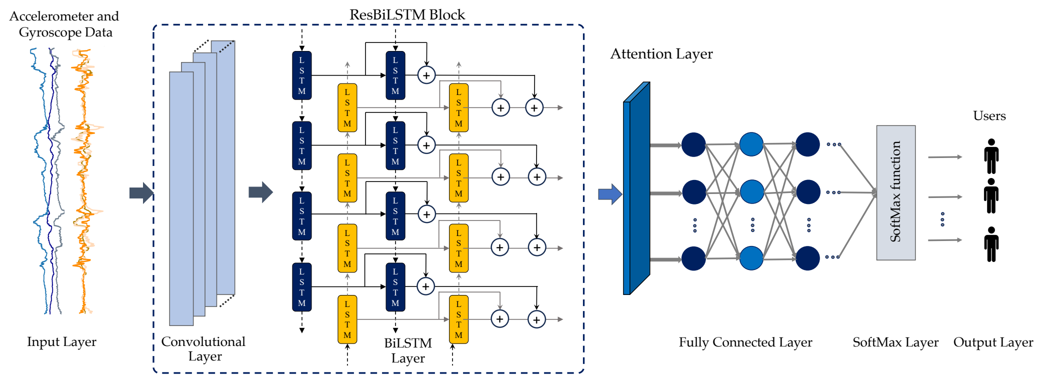

Figure 4: The proposed Att-ResBiLSTM model for user identification

the model is depicted in Figure 4 and comprises three essential elements: convolutional layers, ResBiLSTM, and an attention layer. The convolutional layer's initial element determines spatial characteristics from the pre-processed input. By manipulating the magnitude of the convolution kernel, it significantly diminishes the duration of the temporal sequence. This feature allows the model to decrease the duration required for identification. Subsequently, the enhanced ResBiLSTM network is employed to extract temporal features from the data that has been subjected to convolutional layer processing. The ResBiLSTM module is designed to improve the model's capacity for recognizing long-term dependencies in time series data by using the advantages of BiLSTM and integrating residual connections. Incorporating this component enhances the model's capacity to comprehend intricate temporal patterns and accuracy of identification. In order to enhance the efficiency of the ultimate identification characteristics, we propose incorporating the attention mechanism. The present mode facilitates the computation of weights for the feature information produced by the ResBiLSTM network, thereby enabling the model to concentrate discerningly on the most informative segments of the input data. The attention mechanism plays a crucial role in enhancing the ability to discriminate the model and improving user identification accuracy by stressing the most pertinent variables.

4 Experiments and Results

All experiments conducted in this study were executed on the Google Colab Pro+ platform utilizing a Tesla V100. The Python programming language utilizes libraries like Python, TensorFlow, Keras, Scikit-Learn, Numpy, and Pandas. In this study, we perform experiments to assess the identification capabilities of the presented Att-ResBiLSTM model and compare its performance with standard DL techniques.

The data presented in this study were obtained using a 5-fold cross-validation process. This study conducted a series of investigations to assess the effectiveness of recognizing the suggested Att-ResBiLSTM model. Various measures were used to assess the model's effectiveness, including accuracy, precision, recall, and F1-score. The results of these evaluations are presented in Table 1.

The experimental findings demonstrate that the Att-ResBiLSTM model attained the highest level of accuracy at 98.12% and the highest F1-score at 98.09% when utilizing accelerometer and gyroscope data for analyzing walking motion. The comparison results demonstrate that

Sensor	Identification Performance (%mean±%Std.)			
	Accuracy	**Precision**	**Recall**	**F1-score**
Acc.	96.50%(±0.66%)	96.66%(±0.62%)	96.53%(±0.60%)	94.48%(±0.63%)
Gryo.	96.46%(±1.30%)	96.81%(±1.07%)	96.55%(±1.13%)	96.57%(±1.15%)
Acc. + Gyro.	98.12%(±0.67%)	98.36%(±0.55%)	98.00%(±0.79%)	98.09%(±0.70%)

Table 1: Performance metrics of the proposed Att-ResBiLSTM model using different smartwatch sensor data

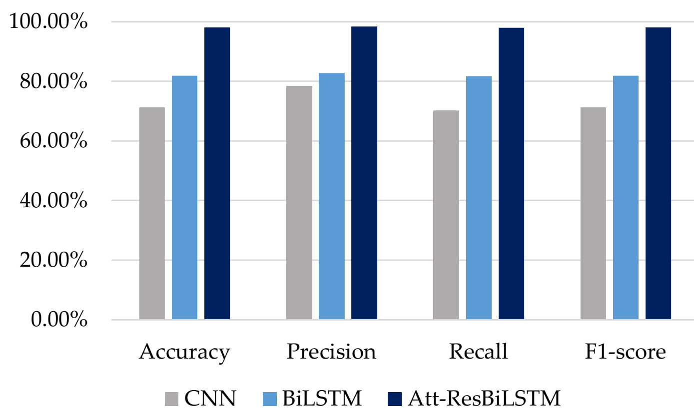

Figure 5: Comparative results of baseline DL models and the proposed Att-ResBiLSTM used in this work

the Att-ResBiLSTM model, as given in Figure 5, exhibits superior performance compared to the benchmark CNN and BiLSTM models.

5 Conclusion and Future Works

This study introduces an innovative model called Att-ResBiLSTM, a hybrid residual neural network incorporating a mechanism for attention. The suggested model incorporates the benefits of the squeeze-and-excitation module in order to take into account the channel-wise information of each convolutional layer. The suggested DL approach was assessed using a publicly available standard dataset of wearable sensor data, specifically the SP-SW HAR dataset. The investigation results indicate that the Att-ResBiLSTM model performs better than other computational models, achieving an accuracy of 98.12% and the highest F1-score of 98.09%. Furthermore, this study conducted a comparative analysis to assess the efficacy of accelerometer data compared to gyroscope data. Based on the results, integrating accelerometer data and gyroscope data exhibited superior performance in the recognition procedure.

Further study endeavors may validate the suggested DL models in additional comparison data sets exhibiting diverse physical activity patterns. Additional enhancements to efficiency could be achieved through the advancement of intricate and lightweight DL networks, as well as the creation of innovative data representations rooted in time-frequency analytics.

6 Acknowledgments

This research project was supported by Thailand Science Research and Innovation Fund; University of Phayao under Grant No. FF67-UoE-Sakorn; National Science, Research and Innovation Fund (NSRF); and King Mongkut's University of Technology North Bangkok with Contract no. KMUTNB-FF-67-B-09.

References

[1] Eyhab Al-Mahadeen, Mansoor Alghamdi, Ahmad S. Tarawneh, Majed Abdullah Alrowaily, Malek Alrashidi, Ibrahim S. Alkhazi, Almoutaz Mbaidin, Anas Ali Alkasasbeh, Mohammad Ali Abbadi, and Ahmad B. Hassanat. Smartphone user identification/authentication using accelerometer and gyroscope data. *Sustainability*, 15(13), 2023.

[2] Qiuhan Meng, Pengran Lu, and Songye Zhu. A smartphone-enabled iot system for vibration and noise monitoring of rail transit. *IEEE Internet of Things Journal*, 10(10):8907–8917, 2023.

[3] Sakorn Mekruksavanich and Anuchit Jitpattanakul. Deep residual network for smartwatch-based user identification through complex hand movements. *Sensors*, 22(8), 2022.

[4] Muhammad Ahmad, Mohammed Alqarni, Asad Khan, Adil Khan, Sajjad Chauhdary, Manuel Mazzara, Tariq Umer, and Salvatore Distefano. Smartwatch-based legitimate user identification for cloud-based secure services. *Mobile Information Systems*, 2018:1–14, 08 2018.

[5] Sakorn Mekruksavanich and Anuchit Jitpattanakul. Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models. *Electronics*, 10(3), 2021.

[6] Baljit singh Saini, Navdeep Kaur, and Kamaljit Bhatia. Authenticating mobile phone user using keystroke dynamics. *International Journal of Computer Sciences and Engineering*, 6:372–377, 12 2018.

[7] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. Touch me once and i know it's you! implicit authentication based on touch screen patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 987–996, New York, NY, USA, 2012. Association for Computing Machinery.

[8] Napa Sae-Bae, Nasir Memon, and Katherine Isbister. Investigating multi-touch gestures as a novel biometric modality. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 156–161, 2012.

[9] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, 2013.

[10] Cristiano C. Rocha, João Carlos D. Lima, M. A. R. Dantas, and Iara Augustin. A2best: An adaptive authentication service based on mobile user's behavior and spatio-temporal context. In *2011 IEEE Symposium on Computers and Communications (ISCC)*, pages 771–774, 2011.

[11] Munish Sabharwal. Multi-modal biometric authentication and secure transaction operation framework for e-banking. *Int. J. Bus. Data Commun. Netw.*, 13(1):102–116, jan 2017.

[12] Markus Jakobsson, Elaine Shi, Philippe Golle, and Richard Chow. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX Conference on Hot Topics in Security*, HotSec'09, page 9, USA, 2009. USENIX Association.

[13] Pierluigi Casale, Oriol Pujol, and Petia Radeva. Personalization and user verification in wearable systems using biometric walking patterns. *Personal and Ubiquitous Computing*, 16:563–580, 2011.

[14] Liu Rong, Zhou Jianzhong, Liu Ming, and Hou Xiangfeng. A wearable acceleration sensor system for gait recognition. In *2007 2nd IEEE Conference on Industrial Electronics and Applications*, pages 2654–2659, 2007.

[15] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Makela, and H.A. Ailisto. Identifying users of portable devices from gait pattern with accelerometers. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 2, pages ii/973–ii/976 Vol. 2, 2005.

[16] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, jul 2006.

[17] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[18] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.

[19] Matteo Gadaleta and Michele Rossi. Idnet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognition*, 74:25–37, 2018.

[20] Giacomo Giorgi, Andrea Saracino, and Fabio Martinelli. Using recurrent neural networks for continuous authentication through gait analysis. *Pattern Recognition Letters*, 147:157–163, 2021.

[21] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[22] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T. Chua. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6298–6306, Los Alamitos, CA, USA, jul 2017. IEEE Computer Society.

[23] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 764–773, 2017.

[24] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[25] Jianwei Yang, Zhile Ren, Chuang Gan, Hongyuan Zhu, and Devi Parikh. Cross-channel communication networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[26] Miguel Matey Sanz. Smartphone and smartwatch inertial measurements from heterogeneous subjects for human activity recognition., October 2023. The dataset in this repository is licensed under a Creative Commons Attribution 4.0 International License. All contained code in the 'utils' package is licensed under the Apache License 2.0.