WiFi-Based Human Activity Recognition using Convolutional Neural Networks

Parisa Fard Moshiri[1], Mohammad Nabati[1], Reza Shahbazian[1], Seyed Ali Ghorashi[1, 2]
1: Cognitive Telecommunication Research Group, Department of Electrical Engineering,
Shahid Beheshti University G. C., Tehran, Iran.
2: Department of Computer Science & Digital Technologies, School of Architecture, Computing and Engineering,
University of East London, London, UK.
Email: p.fardmoshiri@mail.sbu.ac.ir, mo.nabati@Mail.sbu.ac.ir, r_shahbazian@sbu.ac.ir, a_ghorashi@sbu.ac.ir.

***Abstract*—Human activity recognition (HAR) as an emerging technology can have undeniable impacts in several applications, including health monitoring, context-aware systems, transportation, robotics, and smart cities. Among the main research methods in HAR, which are sensor, image, and WiFi-based, the last one has attracted more attention due to the ubiquity of WiFi devices. WiFi devices can be utilized to recognize daily human activities such as running, walking, and sleeping. These activities affect WiFi signal propagation and can be further used to identify human activities. This paper proposes a Deep Learning (DL) method for activity recognition tasks using WiFi channel state information (CSI). A new model is developed in which CSI data are converted to grayscale images. These images are then fed into a Convolutional Neural Network (CNN) with 2-dimensional convolutional layers for activity recognition. We take advantage of CNN's high accuracy on image classification along with WiFi-based preponderance. The experimental results demonstrate that our proposed approach achieves good performance in HAR tasks.**

Keywords—Activity Recognition, Channel State Information, Convolutional Neural Network, Deep Learning, WiFi.

I. INTRODUCTION

Recent years have noticed rapid progress of human activity recognition (HAR) techniques due to their copious potential in academia and industrial applications such as smart homes, health diagnosis, transportation, security, robotics, and indoor pedestrian tracking [1]. HAR researches divide into three approaches: Sensor-based, Image-based, and WiFi-based. Sensor-based methods require users to put sensors on, making users inconvenient, bounding sensor-based application environments, and causing long-time monitoring unavailable [2]. The light source's consistency and stability put obstacles in the way of Image-based systems [3]. WiFi-based approaches draw more attention since they can achieve more accurate monitoring without disturbing participants in sensor-based methods or encountering dark room problems in image-based processes.

Moreover, WiFi-based systems impose no additional cost since indoor spaces such as home, workplace, and shopping mall are usually filled up with background wireless signals. Meanwhile, many new HAR technologies are continuously emerging due to their device-free character, i.e., based on light and radio-frequency [4].

The basic idea for WiFi-based HAR is straightforward. Human body movement will affect the surrounding WiFi signals, and the reflected signals by a specific activity obtain different characteristics. Different human activities can be distinguished by analyzing the signal patterns [5]. To discern a user's activity, the received signal changes must be determined by measuring the physical layer characteristics over the wireless channel, including the received signal strength (RSS) and channel state information (CSI). RSS can be measured easily with the majority of devices, whereas CSI can only be measured with specialized hardware such as Intel 5300 network interface card (NIC), raspberry pi, Atheros 9580 [6], and Atheros 9390 [7]. While location estimations of target users using carry-on WiFi devices have historically one of the most representative RSS-based applications, it is unreliable in HAR applications due to the inability to capture small signal fluctuations caused by movement. To achieve more accurate and reliable HAR systems, CSI is usually preferred, which represents more information about scattering, fading, and power decay with distance [8].

As depicted in Fig. 1, different motions have different effects on CSI values; hence they will have different patterns. To extract pattern features, Deep Learning (DL) algorithms can be exploited. As there are different types of DL algorithms, choosing a reasonable neural network depends on the implemented tasks and types of features.

Zhenghua Chen et al. used a DL-based approach, i.e., bi-directional Long Short-Term Memory (LSTM) in WiFi-based research [9], Abdu Gumaei et al. used deep simple recurrent units (SRUs) and gated recurrent units (GRUs) neural network model, in a multi-model sensor-based task [10], and Jichao Liu et al., exerted deep CNN-LSTM for video-based group activity recognition [11].

In this article, we propose a new WiFi-based HAR system that makes capital out of CNN's outstanding classification performance, especially in image processing. The CSI samples have been converted into images before feeding to the neural network. In this case, not only do we take advantage of the complete specifications of Wifi-based systems, but also we use one of the most accurate neural networks in classification.

The contribution of this study is summarized as follows:

- Proposing a new HAR method in which we convert raw CSI data to images
- Using the CNN model to classify the generated images from raw CSI data

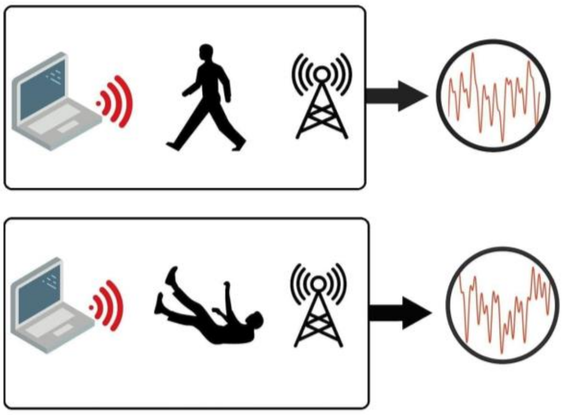


Fig. 1. Two activities (Walk and Fall) have different effects on the CSI signal.

The remaining contents are organized as below; Section II presents the related HAR research. It describes the different approaches in the HAR task. Then in Section III, we illustrate CSI by discussing the system model. The main contributions of this research are summarized in Section IV. In Section V, experimental results are reported based on one of the CSI benchmarks. Finally, conclusions are discussed in Section VI.

II. RELATED WORK

The related works on HAR can be organized into three categories as follows:

A. Sensor-based HAR

The HAR sensor-based method has the advantage of persistent monitoring and has less dependency on light than image-based approaches [12]. To this end, in [12], the authors used an end-to-end CNN to estimate three arm movements by wearing a wrist accelerometer sensor. Many open source sensor-based databases for HAR used a single built-in accelerometer of the smartphone at a low sampling rate. Training an accurate classifier with these data is complicated since they have insufficient quantity and asymmetric distribution. To improve this situation, it is necessary to increase participants' contributions. In [13], the experimental data are collected with the help of 100 participants from accelerometer, gyroscope, and magnetometer sensors. Each participant did seven daily routines activities, including Going Upstairs and Downstairs, Standing, Running, Walking, Bicycling, and Swinging. To further ameliorate the recognition accuracy, the deep CNN architecture was proposed, in which the raw signal was directly used. The proposed CNN scales invariable specifications of the sensor and reached an accuracy of 96%.

B. WiFi-based HAR

Due to WiFi devices' ubiquity, addressing light consistency in image-based and users' disturbance in sensor-based approaches, and technically more accurate monitoring, WiFi-based HAR has attracted significant attention. CSI fingerprints have been widely employed for different types of HAR and indoor localization. As for activity recognition, in [14] and [15], CSI fingerprints are used in elderly-care systems for user's falling detection. Further, in [16], researchers find that CSI fingerprints can be used to diagnose users' typing state by their smartphones. In [17] and [18], CSI fingerprints are used for hand sign recognition. Recently, applying DL algorithms to WiFi-based HAR assisted by signal processing techniques has been researchers' interest, as DL-based HAR methods can automatically improve feature extraction, thereupon improving recognition performance [19]. In [20], researchers utilized CSI fingerprints for both HAR and indoor localization tasks by applying 1-D CNN to sweep along the time axis of their collected CSI data to capture the temporal information of CSI fingerprints. In [21], the authors exploited recurrent neural network (RNN) for the feature extraction task. Moreover, the authors in [22] used RNN for feature extraction and proposed a feature enhancement scheme to enhance CSI data quality.

An autoencoder long-term recurrent convolutional network was developed in [23] to extract high-level features. However, these systems are delicate to the phase shift caused by the timing offset. In other words, as authors in [24] declared, "A slight mismatch in the phase domain of CSI can lead to notable performance degradation." To calculate and recoup the timing offset more accurately, the authors in [24] introduced a CSI-based HAR system, in which an activity filter-based deep learning network with improved correlation features was applied on the CSI. Their scheme can also improve the identification accuracy for similar activities by utilizing an Activity Filter (A.F.). In the A.F. method, the enhanced CSI correlation features obtained from CSI compensation and enhancement are used to recognize similar activities. As their experimental results demonstrated, their method offers faster training and higher accuracy than comparable state-of-art HAR methods.

C. Image-based HAR

In image and video processing, HAR is also an important research direction. In image-based HAR systems, regions corresponding to a person must be segmented from the rest of an image sequence. In many researches, background estimation has been used for motion segmentation since image segmentation could reduce the searching space and provide contextual suggestions in complex video sequences. As the background scene is not totally static, in [25], in the training process, the background scene framework is statistically learned using the redundancy of pixel intensities. Without processing the entire video, their model obtained an initial video frame in which the moving humans are segregated from the background. After training, they made boxes of each part corresponding to behavior, using the vision BlobAnalysis object, in which the detected foreground was further filtered by the object. In the final step, Finally, they used the Lucas-Kanade approach to process the remaining video frames. The proposed approach in [25] only recorded and analyzed video frames when motion was detected. As a result, the number of video frames recorded was reduced.

As wearing sensors is inconvenient for users, and the image-based approach depends on light constancy, we consider an improved CSI-based HAR, which takes advantage of CNN's high potential in the classification task. The WiFi-based HAR, combined with the CNN model, has yielded promising results for activity recognition. In other words, our model follows a signal conversion idea and proposes an image processing model based on a DL network. More precisely, grayscale images are developed from the amplitude of CSI data and feed into our proposed CNN, which will be clarified in Section IV.

III. SYSTEM MODEL

The channel information of a wireless communication link is defined by the CSI metric. CSI can be approximated and parsed from the Physical layer using Orthogonal Frequency-Division Multiplexing (OFDM) technology in the IEEE 802.11n/ac specifications. It describes signal dispersion, fading, environmental, and distance attenuation in each transmission path during signal propagation from the transmitter to the receiver. As a result, CSI offers numerous benefits such as increased transmission information, increased stability, and less environmental impact. The wireless channel can be described in the frequency domain as follows:

$$Y = H \times X + N \tag{1}$$

Where H is the CSI matrix, Y is the received signal vector, and X is the transmitted signal vector, respectively, and N is an additive white Gaussian noise vector. According to (1), H can be described as:

$$H(j) = |H(j)| e^{iSin \angle H(j)} \tag{2}$$

Where H(j) represents the quantity of CSI for the j^{th} subcarrier, including the amplitude (|H(j)|) and phase of CSI (∠H(j)).

Thirty subcarriers can be measured from each Intel 5300 NIC antenna. We can get nine data streams from each CSI packet with thirty subcarriers by using three antennas. According to Multiple-Input Multiple-Output (MIMO) technology, the CSI can be expressed with matrix form as:

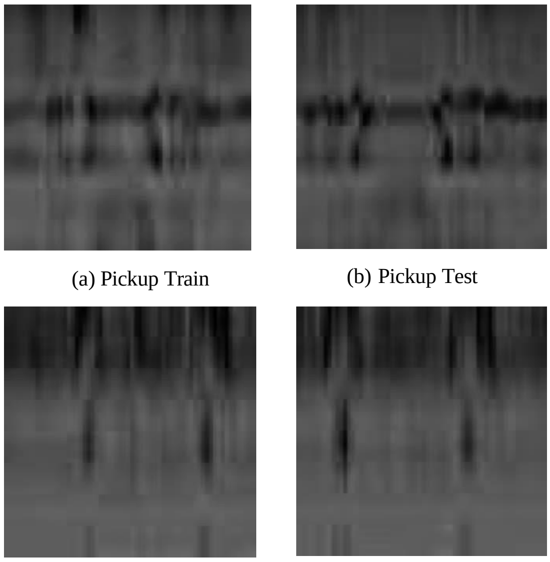
$$CSI = \begin{bmatrix} H_{1,1} & \cdots & H_{1,30} \\ \vdots & \ddots & \vdots \\ H_{9,1} & \cdots & H_{9,30} \end{bmatrix} \tag{3}$$

MIMO technology enhances diversity by identifying transmissions from different paths. Also, it results in intensifying multiplexing gain, and at the same time, reducing the co-channel interference [21].

IV. PROPOSED METHOD

A. *Pre-processing*

The CSI dataset provided by Siamak Yousefi et al. [21] has been used, which includes the amplitude and phase of CSI. In their experiment, they collected CSI for seven different activities at a 1 kHz sample rate. These activities are typical and representative activities in daily life, including lying down (bed), falling, walking, running, sitting down, picking up, and standing up. We exert the CSI amplitude, which is the first 90 columns of the CSV files (2^{nd} column up to 91^{st}, since the first column shows timestamp). A person did each activity within 20 seconds, while in the beginning and at the end of the period, the person didn't do any activity. Each CSV file contains no activity and activity raw data. According to the labeled files provided by these researchers, we can export activity rows and save them in matrices. As CSI data are noisy and time-series, smoothing methods are used to remove random variations that appear as abrasiveness in a raw data plot. There are two smoothing methods: the average method and the exponential method. In this research, based on the fact that close-in-time CSI data usually have similar values, the moving average method is applied to raw data to remove random variation or noise. As the numeric values in a grayscale image pixels matrix are uniformly changed from zero (black pixels) to 255 (white pixels), smoothed data values must be normalized to make 64x64 grayscale images for all activities. Some of these images are depicted in Fig. 2, and the proposed pre-processing technique is depicted in Fig. 3.



(a) Pickup Train (b) Pickup Test

(c)Walk Train (d)Walk Test

Fig. 2. Test and train images for pickup and walk activity.

B. *CNN architecture*

CNN is a DL algorithm that can differentiate images from each other. Comparing to other classification networks, the pre-processing required in this network is much lower. Moreover, CNN itself can learn required filters or characteristics without user help. Convolution function extracts the features from the input image. The first ConvLayer captures the low-level features such as edges, color. As we can use more than one ConvLayer in the network structure, the network can also capture the high-level features, resulting in high recognition accuracy. Generally, a CNN structure is depicted in Fig. 4. After Convolutions and activation functions, a pooling layer is applied and this process can be repeated as many times as needed, depending on the dataset. The main goal of this operation is feature extraction from images and fed these features into a neural network.

As depicted in Fig. 5, the proposed CNN model consists of five types of layers: 1) an input layer, the grayscale images made from CSI data; 2) 2D-Convolutional layers for feature extraction; 3) max-pooling layers as an applied filter to feature maps, resulting in features size reduction and improvement in the robustness of analyzed features; 4) fully connected layer (dense) to integrate all non-linear combinations of extracted features; 5) an output layer with a softmax as an activation function representing a categorical distribution over seven different activities.

After each ConvLayer, we apply ReLU (Rectified Linear Unit) ,which is determined as ReLU(x) = max(0, x) for preventing from vanishing gradient issue. According to this

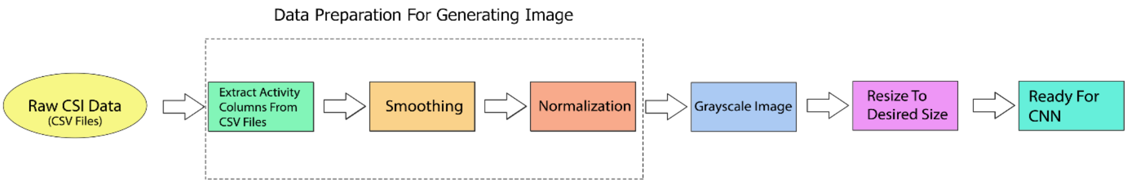


Fig. 3. Pre-processing technique.

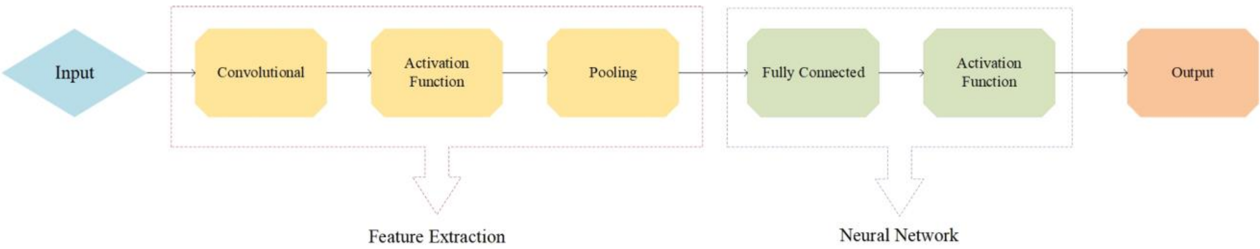


Fig. 4. Main CNN architecture

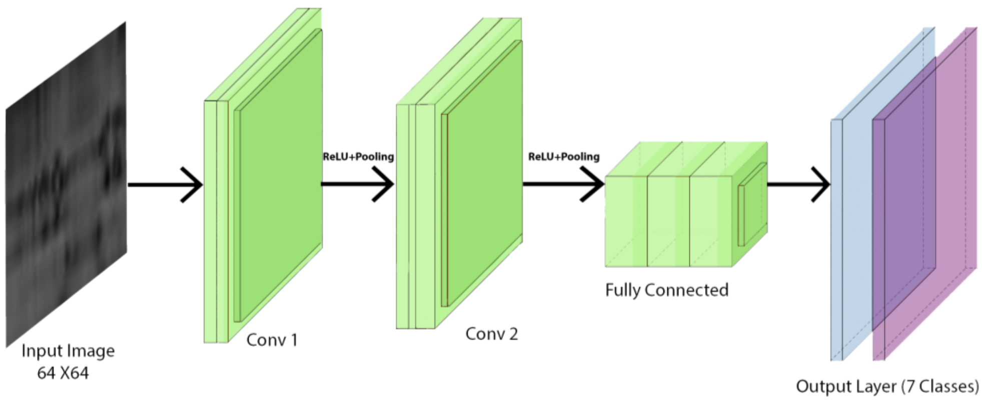


Fig. 5. The proposed CNN architecture.

function, the result of negative input quantities is zero (meaning that these neurons are deactivated). After the first ConvLayer with ReLU activation function, Batch Normalization (B.N.) has been used to make the network more stable during training. B.N. makes the variable mean and standard deviation estimations more stable across mini-batches and respectively closer to 0 and 1. To reduce overfitting, dropout layers have been applied between hidden layers. In this research, we use the max-pooling layer after each ConvLayer. When features have been detected in the ConvLayer, max-pooling layer, as a convolution process, downsample feature maps and helps in extracting low-level features. Before our Fully-Connected layer, the last layer's output should be flattened, which results in a long feature vector (1-dimensional array).

V. Experimental Results

The dataset used in this research, [21], includes seven common activities in our daily life. For classification, we use the CNN model, implemented on Keras and accelerated by Geforce RTX 2060. The raw CSI amplitude data is a 90-dimensional vector (3antennas and 30 sub-carriers), converted into an image for each activity. For each activity, depending on the period of no activity mentioned in the labeled file, we select activity rows and save them in a matrix. The matrix is converted to a grayscale image. This process has been applied through all the CSV files. These images are reshaped to the desired size to agree with CNN's input layer (64x64). Our evaluated method has achieved an accuracy of 89.22% during 200 epochs. Compared to our previous LSTM model [2], in which the accuracy is 84.3%, we enhance activity recognition accuracy up to almost 5%. In [21], they used Random Forest (RF) with 100 trees for classification with an accuracy of

64.66%. They also applied the Hidden Markov Model and improved accuracy up to 73.33% compared to RF. These two networks were implemented on six activities (all of the activities except the Pickup activity). The comparisons are depicted in Fig. 6.

As mentioned in Section I, different activities have different CSI values, resulting in different recognition accuracy. To describe our proposed classifier's performance, we employ a confusion matrix (or error matrix), in which rows represent predicted classes and columns represent actual classes. The activities involving larger body movement, i.e., fall, walk, and run, have higher recognition accuracy (see Fig. 7), since they have more influence on CSI signals' characteristics. The fall activity is highly considerable, especially for elderly healthcare systems [26]. Our proposed CNN model achieves a recognition accuracy of 91.29% for this activity, thereby profiting many healthcare applications.

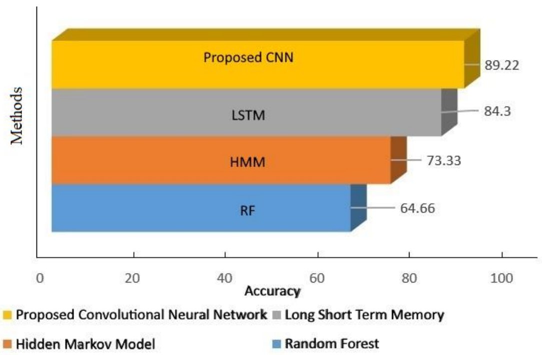


Fig. 6. Activity Recognition results' in various methods.

Fig. 7. Confusion matrix of activity recognition.

VI. Conclusion

The ubiquity of WiFi devices in recent years and their good performance in activity recognition alongside the outstanding performance of CNN in classification tasks encourage us to design a new Wifi-based HAR method. The work's novelty is the conversion of WiFi signal to image and taking advantage of the high performance of CNN methods in image classification. To this end, raw CSI data has been denoised and converted into images for each activity. These images have been fed into a 2-D Convolutional layer. Our proposed model can benefit elderly health monitoring systems since it achieves high recognition accuracy for one of the most important activities in this field, Fall activity. In the future, the raw CSI data will be converted to different types of images, different mappings will be examined, and accordingly, improved optimized CNN will be designed.

References

[1] S. Liu, Y. Zhao, F. Xue, B. Chen, and X. Chen, "DeepCount: Crowd Counting with WiFi via Deep Learning," *arXiv*, Mar. 2019.

[2] P. F. Moshiri, H. Navidan, R. Shahbazian, S. A. Ghorashi, and D. Windridge, "Using GAN to Enhance the Accuracy of Indoor Human Activity Recognition," *arXiv*, Apr. 2020.

[3] A. Aslam, S. Hasan, and E. Curry, "Poster: Challenges with image event processing," in *DEBS 2017 - Proceedings of the 11th ACM International Conference on Distributed Event-Based Systems*, pp. 347–348, 2017.

[4] M. Nabati, S. A. Ghorashi, and R. Shahbazian, "Joint Coordinate Optimization in Fingerprint-Based Indoor Positioning," *IEEE Communications Letters*, 2020.

[5] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "WiAct: A Passive WiFi-Based Human Activity Recognition System," *IEEE Sens. J.*, vol. 20, no. 1, pp. 296–305, Jan. 2020.

[6] Y. Xie, Z. Li, and M. Li, "Precise Power Delay Profiling with Commodity WiFi," *IEEE Trans. Mob. Comput.*, vol. 18, no. 6, pp. 1342–1355, Jun. 2019.

[7] S. Sen, J. Lee, K. H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *MobiSys 2013 - Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 249–262, 2013.

[8] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless Sensing for Human Activity: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 3, pp. 1629–1645, Jul. 2020.

[9] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM," *IEEE Trans. Mob. Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.

[10] A. Gumaei, M. M. Hassan, A. Alelaiwi, and H. Alsalman, "A Hybrid Deep Learning Model for Human Activity Recognition Using Multimodal Body Sensing Data," *IEEE Access*, vol. 7, pp. 99152–99160, 2019.

[11] J. Liu, C. Wang, Y. Gong, and H. Xue, "Deep fully connected model for collective activity recognition," *IEEE Access*, vol. 7, pp. 104308–104314, 2019.

[12] M. Panwar et al., "CNN based approach for activity recognition using a wrist-worn accelerometer," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 2438–2441, 2017.

[13] R. Zhu *et al.*, "Efficient Human Activity Recognition Solving the Confusing Activities Via Deep Ensemble Learning," *IEEE Access*, vol. 7, pp. 75490–75499, 2019.

[14] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–25, Jan. 2018.

[15] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices," *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[16] M. Li *et al.*, "When CSI meets public WiFi: Inferring your mobile phone password via WiFi signals," in *Proceedings of the ACM Conference on Computer and Communications Security*, vol. 24, pp. 1068–1079, 2016.

[17] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger: Talk to your smart devices with finger-grained gesture," in *UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 250–261, 2016.

[18] F. Wang, J. Han, S. Zhang, X. He, and D. Huang, “CSI-Net: Unified Human Body Characterization and Pose Recognition,” *arXiv*. arXiv, 06-Oct-2018.

[19] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, “TW-See: Human activity recognition through the wall with commodity WiFi Devices,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 306–319, Jan. 2019.

[20] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, “Joint activity recognition and indoor localization with WiFi fingerprints,” *IEEE Access*, vol. 7, pp. 80058–80068, 2019.

[21] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, “A Survey on Behavior Recognition Using WiFi Channel State Information,” *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[22] Z. Shi, J. A. Zhang, R. Xu, and G. Fang, “Human Activity Recognition Using Deep Learning Networks with Enhanced Channel State Information,” in *2018 IEEE Globecom Workshops, GC Wkshps 2018 - Proceedings*, 2019.

[23] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, and C. J. Spanos, “DeepSense: Device-Free Human Activity Recognition via Autoencoder Long-Term Recurrent Convolutional Network,” in *IEEE International Conference on Communications*, 2018.

[24] Z. Shi, J. Andrew Zhang, R. Y. Xu, and Q. Cheng, “WiFi-based activity recognition using activity filter and enhanced correlation with deep learning,” in *2020 IEEE International Conference on Communications Workshops, ICC Workshops 2020 - Proceedings*, 2020.

[25] S. Sen, M. Dhar, and S. Banerjee, “Implementation of human action recognition using image parsing techniques,” in *2018 Emerging Trends in Electronic Devices and Computational Techniques, EDCT 2018*, pp. 1–6, 2018.

[26] Y. Wang, K. Wu, and L. M. Ni, “WiFall: Device-Free Fall Detection by Wireless Networks,” *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.