Person Re-Identification Through Wi-Fi Extracted Radio Biometric Signatures

Danilo Avola, *Member, IEEE*, Marco Cascio, *Graduate Student Member, IEEE*, Luigi Cinque, *Senior Member, IEEE*, Alessio Fagioli, *Graduate Student Member, IEEE*, and Chiara Petrioli, *Fellow, IEEE*

***Abstract*— Person re-identification (Re-ID) is a challenging task that tries to recognize a person across different cameras, and that can prove useful in video surveillance as well as in forensics and security applications. However, traditional Re-ID systems analyzing image or video sequences suffer from well-known issues such as illumination changes, occlusions, background clutter, and long-term re-identification. To simultaneously address all these difficult problems, we explore a Re-ID solution based on an alternative medium that is inherently not affected by them, i.e., the Wi-Fi technology. The latter, due to the widespread use of wireless communications, has grown rapidly and is already enabling the development of Wi-Fi sensing applications, such as human localization or counting. These sensing procedures generally exploit Wi-Fi signals variations that are a direct consequence, among other things, of human presence, and which can be observed through the channel state information (CSI) of Wi-Fi access points. Following this rationale, in this paper, for the first time in literature, we show how the pervasive Wi-Fi technology can also be directly exploited for person Re-ID. More accurately, Wi-Fi signals amplitude and phase are extracted from CSI measurements and analyzed through a two-branch deep neural network working in a siamese-like fashion. The designed pipeline can extract meaningful features from signals, i.e., radio biometric signatures, that ultimately allow the person Re-ID. The effectiveness of the proposed system is evaluated on a specifically collected dataset, where remarkable performances are obtained; suggesting that Wi-Fi signal variations differ between different people and can consequently be used for their re-identification.**

***Index Terms*— Person re-identification, channel state information (CSI), Wi-Fi signal, radio biometric signature.**

I. Introduction

PERSON re-identification (Re-ID) addresses a recognition task across non-overlapping camera views, understanding whether a given person appeared in the same (or a different) location at distinct time instants [1]. Direct evolution from identification approaches, where a person identity is classified into one of those known by the system, in a Re-ID method the input image, called probe, is matched against a gallery of identities so that the most probable one can be retrieved. While difficult, this task naturally enables for the re-identification of people that were never seen before. To achieve this goal, however, it is necessary to correctly model a person's appearance and, as a consequence, existing approaches exploit distinctive and reliable visual features extracted from images and video sequences [2], [3]. Although these features have achieved impressive results in Re-ID over the last decade, especially due to deep learning advances, several challenges are still open, including different viewing angles [4], [5]; illumination changes [6], [7]; background clutter [8], [9]; occlusions [10], [11]; and long-term Re-ID [12], [13], where the person's appearance can drastically change after long periods of time (e.g., weeks). What is more, even though ever improving methodologies are being developed to address these challenging problems [14]–[16], person Re-ID is still considered an open task and, even more aggravating, it also presents a considerable gap between research-oriented and practical scenarios [2]. Nevertheless, great efforts are being made to improve this situation by also investigating techniques based on different prerequisites. Indeed, alternative solutions are already exploiting skeleton information [17] or multiple and diverse technologies such as thermal and infrared images [18], [19], since the re-identification task can be a crucial asset in real application areas such as surveillance and forensics. To expand on this matter, we explore an unorthodox technology that, due to its nature, inherently avoids the aforementioned complications, and introduce a novel approach based on a different medium, i.e., Wi-Fi transmissions.

Wi-Fi is a mature technology that leverages radio signals transmitted by several access points (APs) to enable wireless communication between devices. When traveling between two connected devices, Wi-Fi signals are influenced by objects as well as people along their path, resulting in variations of the signal itself [20], [21]. These changes can be captured via either the received signal strength indicator (RSSI) or the channel state information (CSI) measurements [22]. Despite this, the latter is more stable and can carry more signal information with respect to the former due to the underlying technology principles. In fact, for a given wireless data packet, the

Manuscript received May 11, 2021; revised September 20, 2021 and December 5, 2021; accepted February 21, 2022. Date of publication March 8, 2022; date of current version March 22, 2022. This work was supported in part by the MIUR under Grant "Departments of Excellence 2018–2022" of the Department of Computer Science of Sapienza University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. P. C. Yuen. *(Corresponding author: Alessio Fagioli.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Sapienza University of Rome Ethics Committee.

The authors are with the Department of Computer Science, Sapienza University of Rome, 00198 Rome, Italy (e-mail: avola@di.uniroma1.it; cascio@di.uniroma1.it; cinque@di.uniroma1.it; fagioli@di.uniroma1.it; petrioli@di.uniroma1.it).

Digital Object Identifier 10.1109/TIFS.2022.3158058

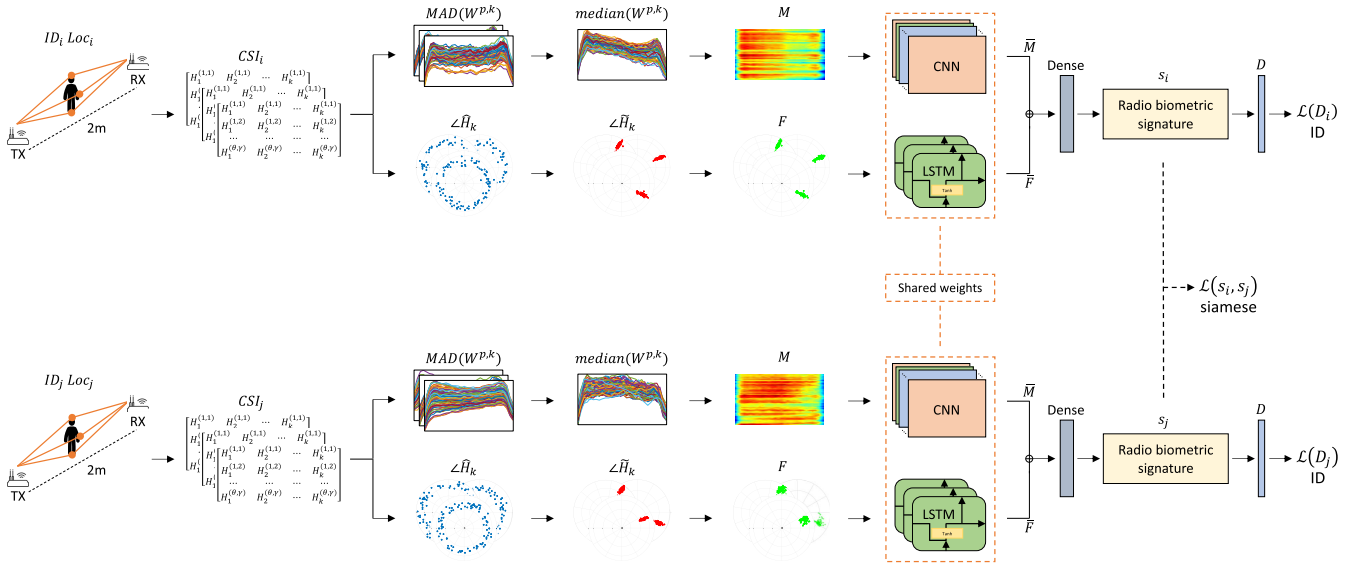

Fig. 1. Proposed model architecture for person Re-ID through Wi-Fi. Starting from a wireless transmission, CSI is estimated and used to extract amplitude heatmaps and phase vectors as radio biometrics. A CNN and LSTM unit are then exploited to build relevant radio biometric signatures, used for the person re-identification.

RSSI is represented by a single value computed at the MAC layer, indicating the relative signal quality; whereas the CSI is measured by employing the orthogonal frequency-division multiplexing (OFDM) transmission technology at the PHY layer, and includes fine-grained signal information defined at the subcarrier level [23]. What is more, this measurement has been proven to be more robust in complex environments, enabling the extraction of relevant signal characteristics, especially in indoor areas [24]. Indeed, among other things, signals amplitude and phase can be retrieved from CSI measurements and, as a matter of fact, several works exploit these radio signal properties to develop useful Wi-Fi sensing applications such as indoor localization [25], [26], tracking [27], or action recognition [28], [29]. A sound strategy to design these Wi-Fi sensing applications generally requires some sort of signal pre-processing to improve the received CSI measurements quality by removing, for example, amplitude outliers [30] and phase offset [31]. Subsequently, either a machine or deep learning approach is usually employed to address the given sensing task [32]. Among the many practical Wi-Fi sensing applications, systems performing person identification, i.e., methods that classify signals into known people identities, indicate that distinct people affect the Wi-Fi signals differently and, therefore, can prove particularly useful for security applications [33]. Moreover, this difference in signal variation between people was also extensively examined by previous studies on both electromagnetic absorption of human bodies and radio waves interactions with biological tissues [34], [35]; where it was shown that wireless propagation around a human body is highly dependent on several characteristics such as skin and other biological tissues conditions, total body water volume, and additional physical attributes including mass and height. As a consequence, due to the high variability of such features, radio biometrics can be extracted from Wi-Fi signals to describe and ultimately recognize a given person [33].

In this paper, inspired by previous signal propagation studies as well as the existing person identification task literature results, we introduce a new Wi-Fi sensing application and design a re-identification system able to recognize both known and unknown persons, i.e., never seen before, via radio biometric signatures computed across the same or different locations. In particular, CSI measurements of Wi-Fi transmissions describe the propagation of a wireless signal from a transmitter (TX) to a receiver (RX) by analysing specific carrier frequencies along multiple paths, and this data can be exploited to implement a Re-ID method, as empirically demonstrated in this paper. Observe that CSI was chosen since it has been proven to carry more information with respect to RSSI [32], therefore resulting in a better choice for a feature sensitive task such as person Re-ID. In more detail, starting from an estimated CSI measurement, signals amplitudes and phases are extracted and processed to improve their quality through established procedures such as the outlier removal [30], for signals amplitudes; and offset removal [31], for signals phases. Subsequently, the sanitized signals are further refined through a median filtering [36] to create a single amplitude heatmap and more stable phases along the various CSI measurements. These features, representing a person's radio biometrics, are then used as input for a novel deep neural network architecture, based on a siamese structure, that was specifically designed to extract meaningful radio biometric signatures through two parallel sub-networks inside each siamese branch. The rationale behind this choice is twofold and a requirement for the person Re-ID task, where the lack of annotated data must be managed [37]–[39]. First, this design is efficient in both supervised and unsupervised feature space learning. Second, the underlying strategy allows to extract invariant feature representations that enable a distance-based Re-ID, where similar identities will be close in the learned feature space, while dissimilar persons will have distant representations [40].

Moreover, a siamese strategy [41] has been widely used to address the vision-based person Re-ID, achieving remarkable results [42], [43]. However, its usage for other wireless sensing applications is fairly new and still being developed [44]. What is more, to the best of our knowledge, currently, there are no other siamese-like pipelines processing Wi-Fi signals addressing the radio domain person Re-ID, therefore encouraging the exploration and design of such a system. To achieve this end, our network comprises two siamese branches, both containing parallel sub-networks based on different models to correctly handle the computed signal amplitude heatmaps and filtered phases, i.e., a convolutional neural network (CNN) and a long-short term memory (LSTM). Afterwards, the output radio biometric signatures are used to re-identify the person across the same or two different locations, while also avoiding classical Re-ID drawbacks since a non-video-based medium is exploited. Notice, however, that even though the latter holds true, different issues can affect Wi-Fi signals. This is especially noticeable in real-world application scenarios characterized by unconstrained environmental properties such as crowded areas, where radio biometrics from different people would be merged, or in critical outdoor weather conditions, where the generation of consistent signatures could be impaired altogether. Nevertheless, by focusing on single individuals in typical constrained surveillance settings [45], it is possible to obtain well-defined radio biometrics that allow to re-identify a given person. For this reason, the proposed Re-ID method could be employed, for instance, in relevant and strategic buildings checkpoints, e.g., airports or banks, to further increase their security by confirming a person's identity. To simulate such scenarios, and due to the lack of existing Wi-Fi collections for the Re-ID task, a dataset was acquired capturing single persons standing in several indoor locations. Moreover, comparably to real-world environments, no shielding mechanism was implemented against interference from other radio signals, since they are inevitable in our highly connected world. Thereafter, similarly to classical Re-ID methods that make use of visual information extracted from images and videos [46]–[48], common Re-ID metrics, such as the mean average precision (mAP) and cumulative matching characteristic (CMC), were employed to accurately evaluate the proposed methodology. As shown by the experimental results, significant performances are obtained on the acquired collection; highlighting the presented pipeline effectiveness and opening up a new classical drawbacks-free person Re-ID task frontier.

Summarizing, the major contributions of this paper are:

- the definition of a completely new Wi-Fi sensing application by implementing a person Re-ID method based exclusively on Wi-Fi signals that can also avoid classical vision-based drawbacks thanks to the different medium nature of the wireless technology;
- the design of a novel architecture, based on a siamese model structure, leveraging parallel sub-networks in each siamese branch to extract meaningful radio biometric signatures from Wi-Fi signals examined at either the same or at different locations;
- the presentation of quantitative and qualitative experiments that first, highlight the effectiveness of signal-based approaches to address the person Re-ID task in constrained ambients, and second, establish a baseline methodology for this new Wi-Fi sensing application.

II. Related Work

Depending on the signal measurement type, Wi-Fi sensing methods can be broadly categorized into two classes, i.e., RSSI and CSI based approaches [49].

Concerning the RSSI measure, it indicates the received power level after any possible transmission loss, thus representing the relative signal quality. Inanimate objects (e.g., furniture) or human presence can influence radio signals and, as a matter of fact, the authors of [50] noticed significant RSSI fluctuations in both line-of-sight (LOS) and non-line-of-sight (NLOS) conditions. Supporting these findings, the RSSI signal quality was successfully employed in heterogeneous tasks such as map reconstruction [51], [52] as well as human localization [53]–[57], tracking [58] and identification [59]. Confirming the inanimate object influence on radio signals, a grid points filling with low rank matrix theory on RSSI fingerprints exchanged between several APs is used in [51], for example, to reconstruct radio maps of indoor environments; while Markov random field modeling for loopy belief propagation of sparse signals is employed, by the authors of [52], to build 3D radio maps of unknown structures using RSSI signals examined by unmanned aerial vehicles. Considering human-focused RSSI applications, instead, a popular and well-explored task is the localization one. In [53], for instance, the best AP, i.e., with the best RSSI signal quality, is selected to achieve indoor localization according to an eight-diagram approach defining the signal propagation direction; while the authors of [54] develop a feature-scaling-based k-nearest neighbors (KNN) algorithm and further refine the RSSI signals via outlier removal to address an analogous task. Further improvements to localization systems are also provided by techniques that can clean up the received signals. For example, in [55], [56], and [57], Gaussian, weighted average, and continuous wavelet transform (CWT) filtering are applied, respectively, to improve the input for the chosen localization algorithms. The received signal quality increment can also be obtained by reducing possible interferences as shown in [58], where a custom communication protocol enables to track humans via RSSI measurements; or by using wearable devices as demonstrated by the authors of [59], that present an RSSI proximity algorithm able to identify several persons. What is more, improved RSSI measurements are also successfully used to address other complex tasks such as human action recognition. Indeed, as described in [60], sanitizing the RSSI through outlier removal and Gaussian filtering, enables a feature fusion approach to obtain significant results on the action recognition task, therefore indicating that the Wi-Fi technology is a good medium for sensing applications and could be also effectively applied to other tasks.

Regarding the CSI measure, it captures richer information about the wireless transmission among communicating APs, contrary to the RSSI that does not provide fine-grained

features except from the relative signal quality of a wireless environment. For instance, CSI can acquire amplitude and phase features for each subcarrier in the OFDM channel [61]. By describing the channel characteristics of a frequency-diverse group of subcarriers, this measurement is more robust to narrowband interference from other signals. Moreover, it is not affected by the automatic power level adjustment algorithm implemented in commodity wireless APs [62]. As a consequence, the channel state information is gaining momentum in the latest years. In [63], for example, the authors show that there is a high correlation between subsequent CSI measurements, and consistently detect falling humans via CSI amplitude. Similarly, [64] and [29] detect falling humans through denoised frequency spectrogram images, in the former, and phase differences, in the latter, that can be both extracted from CSI measurements; thus indicating that CSI is an information-rich measure. Indeed, among others, Wi-Fi signals amplitude [30], [65], phase [31], [66], and frequency [67], obtained from CSI, are also effectively employed in other tasks such as human indoor localization. For example, the authors of [65] apply a fingerprint matching procedure after optimizing a centroid-based algorithm (i.e., KNN) used to examine locations through CSI amplitudes. In [31], instead, indoor locations fingerprints are defined by linearly transformed CSI phases, where offsets deriving from the transmission are removed; while in [67] the indoor localization is addressed through passive radio maps that are analyzed via a probabilistic algorithm detecting anomalies in CSI frequencies. Furthermore, CSI measurements can also be exploited to capture human movements and, consequently, perform action recognition from the received signals. For instance, the authors of [68] use several CSI channels to produce radio images that enable both to localize and to recognize the corresponding human actions. Similarly, in [69], activity recognition is achieved through variance-normalized CSI amplitude waveforms filtered by the principal component analysis (PCA) procedure. In [44], instead, gestures are recognized via a spatiotemporal examination of CSI phases executed by a siamese recurrent neural network; an architecture that extracts meaningful features from the input phases, and which we have extended in the proposed approach since it is already successfully applied in classical vision-based re-id approaches [37], [38], [70]–[72].

Although CSI-based works generally tend to focus on a single Wi-Fi signal characteristic (e.g., amplitude), it is not uncommon to find approaches exploiting more information derived from CSI measurements. Such an example is found in [73], where both CSI amplitude and phase, computed along two distinct RX devices, are used jointly to generate fine-grained human skeleton poses. Indeed, the multiple CSI channels can provide relevant information about people between two, or more, APs and, as a matter of fact, they are employed to define structural biometric features in [33]. In particular, these features represent body pose differences that can be registered and distinguished through the CSI by employing a time-reversal (TR) technique, that ultimately enables to identify a known person. Following a similar reasoning, the authors of [74] perform user identification through CSI frequency shifts associated to gestures, by band-pass filtering the signals and further refining them through PCA. In [75], instead, CSI amplitudes, treated with discrete wavelet transforms (DWT) and statistical profiling (e.g., channel power distribution), are coupled with people's gait allowing for their identification through Wi-Fi signals. In addition, further confirming the CSI measurement effectiveness, filtered CSI amplitudes are exploited in [76] to simultaneously learn several tasks such as action recognition, user tracking and identification, through a deep neural network. Finally, in [77], the authors are able to also extract internal characteristics (i.e., respiration rates) from CSI measurements by improving the SNR of signals associated to breaths, which allow to identify known users; indicating that the CSI measure contains different signal cues for distinct people, and therefore supporting our direct investigation of CSI measurements to describe a person through radio biometric signatures to address the re-identification task.

III. Proposed Method

To achieve person Re-ID from Wi-Fi signals, we design an architecture that expands the two-branch siamese structure to comprise two parallel sub-networks per model branch, as shown in Fig. 1. By following the proposed pipeline, the whole network can exploit both signals amplitude and phase to address the re-identification task. Specifically, the system first performs a CSI estimation step to capture propagation properties of signals influenced by humans standing between the transmitting and receiving APs. The CSI measurement containing the affected signal is then employed to extract amplitudes and phases, which are in turn processed to generate sanitized feature vectors that represent relevant radio biometrics of a given person. In particular, filtered amplitudes are transformed into heatmaps and analyzed through a CNN-based network to capture meaningful signal patterns; calibrated phases are instead processed via an LSTM-based model to describe discriminant temporal changes deriving from life processes such as respiration and heartbeat. Subsequently, the two sub-networks (i.e., CNN and LSTM components) outputs are combined into a single feature vector representing a radio biometric signature that can be used to re-identify a person across the same or at different locations. Notice that the proposed model, due to the identical branches with shared weights, is suitable for finding similarities between comparable inputs and can generate final feature vectors, i.e., radio biometric signatures, that account for possible environment noise derived, for example, from different furniture. Indeed, by following the classical vision-based siamese objective function structure, the proposed method ensures that signals associated to the same person will have similar representations in the feature space; therefore enabling for their Re-ID.

A. CSI Estimation

The first CSI estimation step leverages commodity hardware for the TX and RX APs, fixed in place inside stationary environments to reduce the amount of random ambient noise. In detail, an 802.11n commercial router is used as transmitter (i.e., TX), while an Intel Wi-Fi Link 5300 (IWL5300) network interface card (NIC), connected to a Desktop PC, acts

as receiver (i.e., RX). The latter was chosen since custom firmware and drivers that enable the CSI estimation were implemented in [22], as it is still rather uncommon to use commodity hardware to access CSI estimation. Furthermore, the proposed system exploits a multiple-input and multiple-output (MIMO) technology to take full advantage of the multi-path propagation as a consequence of TX and RX APs integrating $\Gamma = 2$ and $\Theta = 3$ antennas, respectively. Formally, considering a multi-path propagation scenario, the communication channel can be modeled in the time domain as the channel impulse response (CIR) of a linear time-invariant channel filter, defined as:

$$h(\tau) = \sum_i \alpha_i \delta(\tau - \tau_i), \tag{1}$$

where α_i and τ_i are the attenuation factor and propagation delay of the i-th path, respectively; while $\delta(\tau)$ corresponds to the Dirac delta function. Note, however, that the CSI is a frequency-based measurement; thus, for its estimation, the fast Fourier transform (FFT) is applied on the impulse response at the receiver to obtain the corresponding channel frequency response (CFR) complex value [78]. Consequently, the APs communication channel in the frequency domain can be linearly modeled as follows:

$$y = Hx + n, \tag{2}$$

where y is the received signal vector; H represents the CFR; x is the transmitted signal vector; and n indicates the additive white Gaussian noise (AWGN) [79]. From this channel model, the OFDM technology provides a sampled CFR with a subcarrier granularity; therefore, the CSI measurement is computed by including the CFR value from each of them. Specifically, the IWL5300 component uses $K = 30$ OFDM subcarriers sampled from the 20MHz channel which contains 56 subcarriers. For each subcarrier $\kappa \in K$, the frequency response $H_\kappa^{(\theta,\gamma)}$ over the receiving $\theta \in \Theta$ and transmitting $\gamma \in \Gamma$ antennas, can then be represented via the following complex equation:

$$H_\kappa^{(\theta,\gamma)} = |H_\kappa^{(\theta,\gamma)}| e^{j \angle H_\kappa^{(\theta,\gamma)}}, \tag{3}$$

where j indicates the imaginary component; while $|H_\kappa^{(\theta,\gamma)}|$ and $\angle H_\kappa^{(\theta,\gamma)}$ represent the signal amplitude and phase, respectively. The final CSI matrix computed over the frequency response of the K subcarriers, accounting for all transmitting and receiving antennas, is a $\Theta \times \Gamma \times K$ matrix defined as:

$$CSI = \begin{bmatrix} H_1^{(1,1)} & H_2^{(1,1)} & \dots & H_\kappa^{(1,1)} \\ H_1^{(1,2)} & H_2^{(1,2)} & \dots & H_\kappa^{(1,2)} \\ \vdots & \vdots & \vdots & \vdots \\ H_1^{(\theta,\gamma)} & H_2^{(\theta,\gamma)} & \dots & H_\kappa^{(\theta,\gamma)} \end{bmatrix}, \tag{4}$$

where $H_\kappa^{(\theta,\gamma)}$ is a signed 8-bit complex number indicating the κ-th subcarrier CFR value over the $\theta \in \Theta$ and $\gamma \in \Gamma$ antennas. Observe that both amplitude and phase can be retrieved from the CSI matrix, but require further processing to be used by the proposed system as shown in Sections III-B and III-C.

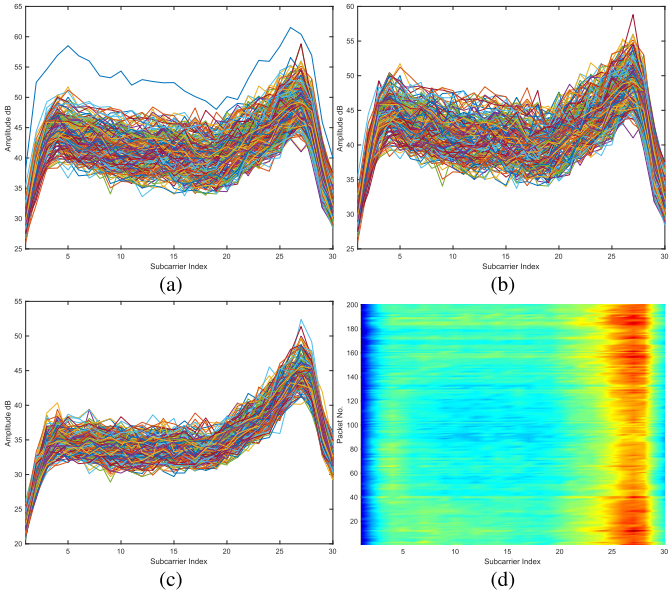

Fig. 2. CSI extracted amplitude processing example. In (a) and (b), the raw and sanitized amplitudes for a single TX-RX antenna pair, respectively. In (c) and (d), the median filtered amplitudes across all antenna pairs, and the corresponding heatmap used as input for the CNN sub-network.

B. Amplitude Sanitization and Heatmap Generation

To prepare clean radio biometrics for the CNN sub-network, we sanitize and transform CSI extracted amplitudes into heatmaps. See that the sanitization step is required since the retrieved amplitudes present noise due to various factors such as furniture material and position, external radio interference, and other environmental conditions, as shown in Fig. 2(a).

Concerning the sanitization procedure, we expand on the method presented in [30]. Specifically, local outliers are first detected through local median values computed over a sliding window of fixed length. Subsequently, these outliers are replaced using the previous non-outlier value to retain consistent amplitude information. In particular, outliers are identified by points resulting more than three local median absolute deviations (MAD) away from the local median within the sliding window applied across packets of each subcarrier. Formally, given a wireless transmission between a TX and RX antenna, amplitudes extracted from CSI measurements of $p \in P$ data packets, and a window size w, the local median is defined as:

$$median(W^{p,\kappa}) = W^{p,\kappa}_{\lceil w/2 \rceil}, \tag{5}$$

$$W^{p,\kappa} = \Big\{ |H_\kappa|^{p-\lfloor w/2 \rfloor}, \dots, |H_\kappa|^{p+\lfloor w/2 \rfloor} : |H_\kappa|^{p-\lfloor w/2 \rfloor} < |H_\kappa|^{p+\lfloor w/2 \rfloor} \Big\}, \tag{6}$$

where $W^{p,\kappa}$ represents an ascending order set containing w neighboring packets amplitude $|H_\kappa|$ of the κ-th subcarrier. Note that the median described in Eq. (5) is computed for all $\Theta \times \Gamma \times K$ antennas and subcarriers combinations, however a single sample is reported for the sake of simplicity. The local MAD used to identify outliers and sanitize the amplitudes is

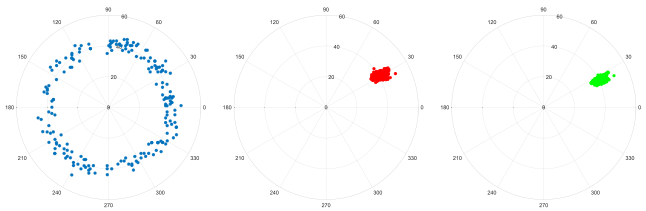

Fig. 3. CSI extracted phase processing example. In (a), (b), and (c), the raw, sanitized and median filtered phases of a single subcarrier, respectively.

then computed as follows:

$$MAD(W^{p,\kappa}) = median(|W_i^{p,\kappa} - median(W^{p,\kappa})|), \quad \forall i, \quad 1 \leq i \leq w. \tag{7}$$

Finally, acceptable local amplitude ranges are defined as:

$$limit^{p,\kappa} = median(W^{p,\kappa}) \pm 3 * MAD(W^{p,\kappa}), \tag{8}$$

and every amplitude resulting outside these limits is replaced with the previous non-outlier value to maintain signal consistency. The produced sanitized signals, for an empirically chosen window size $w = 5$, are shown in Fig. 2(b).

Upon this first processing procedure that enables to reduce artifacts in CSI measurements, a second median filtering is applied over the $\Theta \times \Gamma$ transmissions. The reason behind this decision is twofold. First, it allows to reduce the data dimensionality, and second, it condensates amplitudes characteristics shared among different antennas transmissions, as shown in Fig. 2(c). Notice that this decision was taken since, in general, the sanitized $\Theta \times \Gamma$ transmissions present similar properties. Lastly, the concentrated amplitudes are transformed into a single heatmap M of size $P \times K$, as displayed in Fig. 2(d), representing a person's amplitude radio biometric; which is to be used as input for the CNN sub-network.

C. Phase Sanitization

Similarly to CSI amplitudes, phases also require to be processed due to common issues such as random noise and unsynchronized time clocks between TX and RX APs, that can result, among other things, in random phase offsets, as shown in Fig. 3(a). To address this issue, we calibrate CSI-extracted phases using the linear transformation presented in [31]. Formally, a raw CSI phase $\angle \widehat{H}_\kappa$ measured for the κ-th subcarrier can be expressed as follows:

$$\angle \widehat{H}_\kappa = \angle H_\kappa + 2\pi \frac{m_\kappa}{N} \Delta t + \beta + Z, \tag{9}$$

where $\angle H_\kappa$ indicates the real phase; Δt represents the receiver timing offset, which corresponds to the time interval between signal arrival and detection; β denotes the unknown phase offset; Z is the measurement noise; while m_κ and N correspond to the subcarrier index and fast Fourier transform (FFT) size as specified by the IEEE 802.11n standard. In particular, for the IWL5300 network interface, subcarrier indices range from -15 to 15, while $N = 30$. Moreover, since Δt and β are unknown, the original phase information cannot be directly retrieved. However, since the subcarrier frequency is symmetric, it possible to apply a linear transformation and ignore those parameters by considering the phase across the total frequency band [31]. Specifically, the phase slope a and offset b can be represented via the following equations:

$$a = \frac{\angle \widehat{H}_{30} - \angle \widehat{H}_1}{m_{30} - m_1}, \tag{10}$$

$$b = \frac{1}{30} \sum_{\kappa=1}^{30} \angle \widehat{H}_\kappa. \tag{11}$$

Afterwards, the calibrated phases $\angle \tilde{H}_\kappa$, shown in Fig. 3(b), are computed as follows:

$$\angle \tilde{H}_\kappa = \angle \widehat{H}_\kappa - am_\kappa - b. \tag{12}$$

Once the phase calibration is completed, a median filtering is applied over the $\Theta \times \Gamma$ transmissions, similarly to the amplitude procedure, to reduce data dimensionality and agglomerate typical phase values along the various subcarriers, into a vector F. An example of filtered phases is shown in Fig. 3(c). Finally, the vector F of size $P \times K$, containing the processed phases that capture temporal changes of a signal propagation, is used as input for the LSTM sub-network.

D. Radio Biometric Signatures

To perform person Re-ID we propose an architecture based on a siamese structure, implemented via a two-branch neural network with parallel sub-networks in each branch, that is trained as a feature extractor, and that can learn invariant mappings [80] from the extracted radio-based features, thus resulting in a good choice for the addressed task. In detail, to compute such mappings, the model is composed by two identical branches with shared weights. Moreover, both of the presented architecture branches contain two parallel sub-networks, i.e., a CNN and an LSTM model, to correctly analyze the preprocessed signals. In particular, for the CNN module we followed a VGG-16 structure without its classification component (i.e., up to and including the last max pooling operation), since this model is an effective image pattern extractor [81]. The LSTM sub-network, instead, is implemented via a single recurrent neural network (RNN) layer containing P LSTM units. Moreover, the CNN model takes as input the heatmap M presented in Sec. III-B, representing biometric information derived from amplitudes, and outputs a feature map vector $\bar{M}$; while the LSTM receives as input the F vector introduced in Sec. III-C, containing temporal biometric information, and outputs the feature vector $\bar{F}$. The resulting sub-network outputs, i.e., $\bar{M}$ and $\bar{F}$, are then concatenated and merged together through a dense layer to build what is defined in this work as a radio biometric signature s.

Concerning the model training, the proposed pipeline accepts as input data pairs representing signals associated to the same or different persons. Afterwards, while biometric signatures are being learned by the sub-networks of a branch, the Euclidean distance is applied across the branches resulting outputs via a siamese loss function. This procedure allows to minimize, or maximize, the generated biometric signatures distance for similar, or dissimilar, inputs, respectively. Formally,

given a pair of CSI measurements (CSI_i, CSI_j) as input, biometric signatures s_i and s_j are computed by concatenating the model branches outputs and elaborating them through a dense layer as follows:

$$\begin{aligned} s_i &= w_i(\bar{M}_i \oplus \bar{F}_i) + b_i, \\ s_j &= w_j(\bar{M}_j \oplus \bar{F}_j) + b_j, \end{aligned} \tag{13}$$

where $\oplus$ represents the concatenation operation; while w and b indicate the dense layer weights and bias, respectively. Subsequently, the siamese loss can be defined as:

$$\underset{siamese}{\mathcal{L}}(s_i, s_j) = \begin{cases} \frac{1}{2} \left\| s_i - s_j \right\|^2, & \text{if } i = j; \\ \frac{1}{2} \, max(m - \left\| s_i - s_j \right\|, 0)^2, & \text{if } i \neq j, \end{cases} \tag{14}$$

where $\|\cdot\|^2$ is the Euclidean distance; while m is a margin, empirically set to 2 in this work, that helps the dissimilar signatures separation during the optimization process. What is more, notice that this architecture also enables the Re-ID of unknwon people, i.e., not observed during training, since their radio biometric signatures can still be extracted and compared at test time; where likely matching identities will be associated by lower distances among signatures, in accordance with the reported siamese loss function.

E. Joint Identification and Verification

The siamese loss function is key to the signatures generation, however, on the basis of [82], [83], we further extend the training loss function by following a joint identification and verification strategy that can improve the signatures quality. In particular, at training time, each model branch will predict person identities while the siamese cost described in Sec. III-D is also globally satisfied. Formally, the biometric signature s generated by a given branch is fed to a dense layer with dimension D, i.e., the number of known persons, and the identity loss is then implemented through a categorical cross-entropy function, as follows:

$$\underset{ID}{\mathcal{L}}(D) = -\sum_{d}^{D} y_d \log\left(\frac{\exp(d)}{\sum_{d'}^{D} \exp(d')}\right), \tag{15}$$

where d and y_d correspond to the predicted person identity and ground truth, respectively. Subsequently, to improve Re-ID accuracy, the identification losses computed by the two model branches are also employed in the overall training objective function, as described in the following equation:

$$\mathcal{L} = \underset{siamese}{\mathcal{L}}(s_i, s_j) + \underset{ID}{\mathcal{L}}(D_i) + \underset{ID}{\mathcal{L}}(D_j). \tag{16}$$

Observe that this joint objective function is only used to enable the model to extract good biometric signatures from the input signals. However, at test time, the architecture is only employed as a signature extractor. As a consequence, the identification losses are ignored while the siamese one is replaced by an Euclidean distance to address the re-identification task; where lower distances between two signatures naturally indicate more likely matching identities.

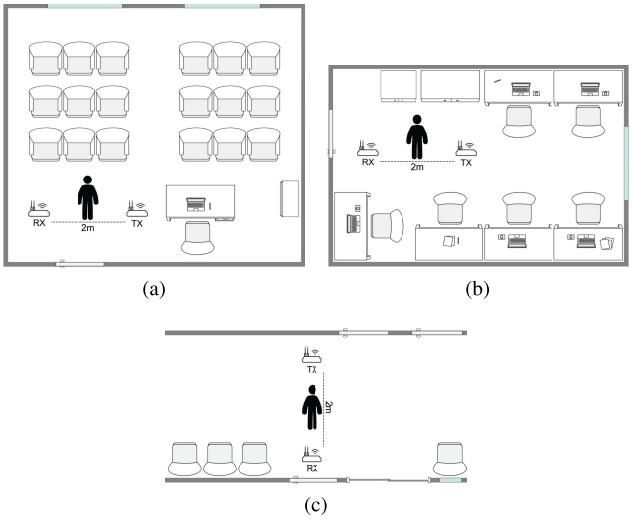

Fig. 4. Rooms configuration for the proposed dataset acquisition protocol. In (a), (b), and (c) the conference hall, office, and indoor hallway, respectively.

IV. Experimental Results

To present a comprehensive evaluation of the proposed methodology, in this section we describe the data collection procedure, necessary due to the lack of public datasets for Wi-Fi Re-ID; relevant implementation details, including the chosen testing protocol and metrics; as well as qualitative and quantitative evaluations for the various system components.

A. Dataset

To compensate for the Wi-Fi Re-ID datasets unavailability, a collection was acquired to assess the presented approach. Specifically, we captured Wi-Fi signals of 35 distinct people, comprising 15 women and 20 men with similar body characteristics, standing between the TX and RX APs, for a total of 525 transmissions. In more detail, the average women height and weight were 165.3 ± 4.6cm and 61 ± 7kg, while the average men measurements corresponded to 176.1 ± 6.7cm and 76 ± 8kg. Furthermore, for each identity, five 3-seconds long transmissions (i.e., spanning over 300 packets) were collected using the 20 MHz channel of a 2.4 GHz Wi-Fi link in 3 different rooms: a conference hall, an office, and indoor hallway. Each room configuration is shown in Fig. 4. In all cases, the TX and RX were fixed and placed 2 metres apart, with no objects in between, and one person at a time was asked to either face toward to or away from the TX while standing between the two devices. Finally, all furniture and environment items were otherwise left untouched, and no shielding mechanism was employed to avoid interferences from other radio signals, which effectively replicates real Wi-Fi networks characteristics where multiple connections propagate across the same area and affect one another.

B. Implementation Details

Regarding the various experimental settings, Wi-Fi signals were preprocessed via the Matlab R2021a software, while several ablation studies were performed on the neural network

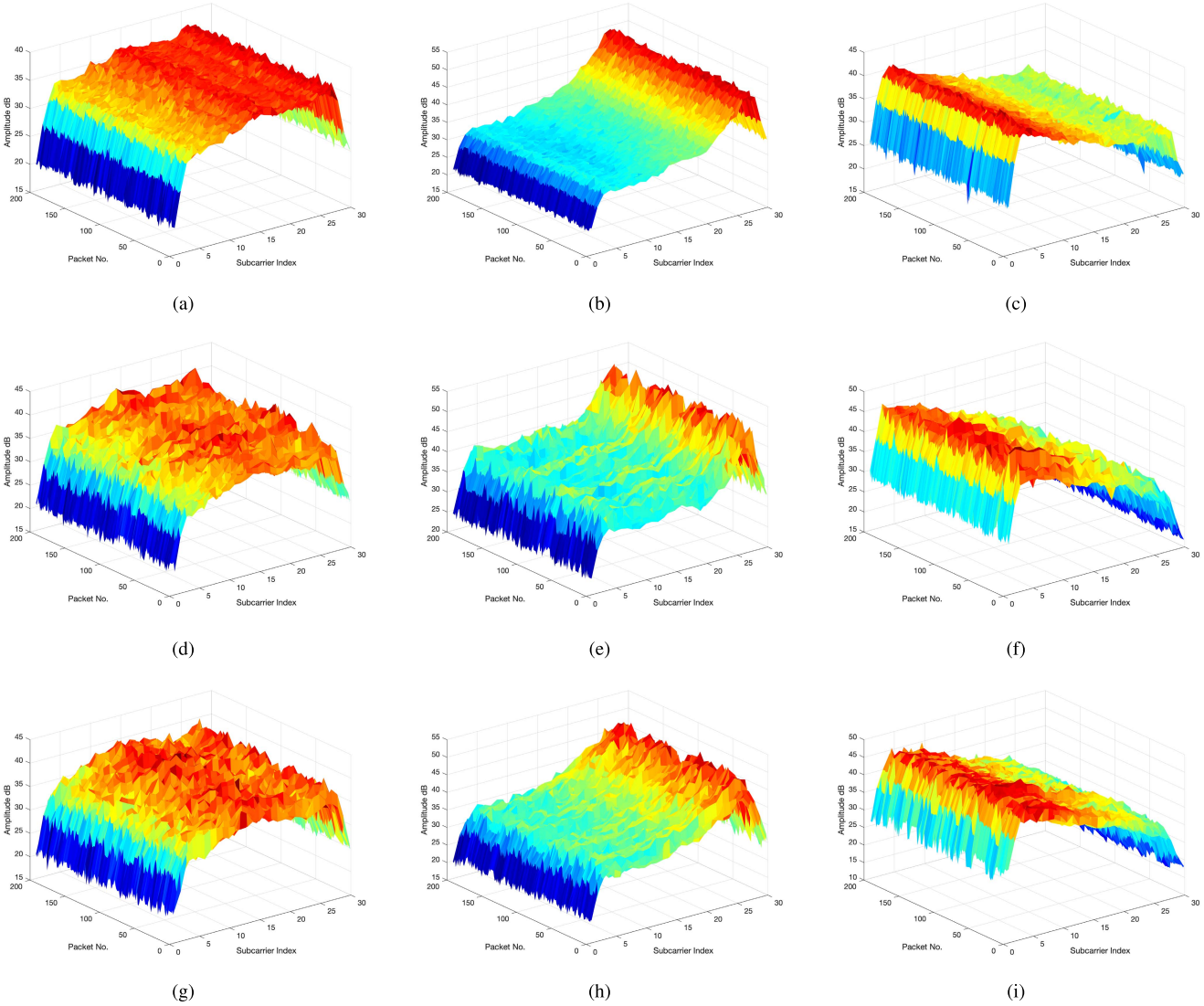

Fig. 5. Heatmaps M-derived 3D surfaces examples for 200-packets acquisitions in the three rooms, i.e., a hallway, office, and conference hall, in the first, second and third row, respectively. Images (a), (d), and (g), correspond to no obstacles between TX and RX APs; while triples (b), (e), (h) and (c), (f), (i) show two different persons of the collected dataset.

component to correctly evaluate the proposed approach. The assessed models followed the same protocol for all tests. Specifically, the dataset was split into two subsets D_1 and D_2. The first one, which enables the model to learn how to extract meaningful signatures via the two siamese branches, contained 20 distinct people, for a total of 300 Wi-Fi transmissions. This collection was used in conjunction with a 10-fold cross-validation procedure using 4/1 random splits per person samples of each room (i.e., 240 and 60 transmissions) for the training and test sets, respectively. The second subset comprises, instead, the remaining 15 identities, counting a total of 225 samples, which were left out to evaluate the system on the re-identification of unknown people. Furthermore, for each fold, every architecture was trained for 200 epochs using the SGD algorithm, with an initial learning rate lr set to 0.1, a weight decay of 5e-4, and a Nesterov momentum of 0.9. Moreover, a scheduler was also implemented to divide the lr by 5 at epochs 60, 120, and 160, so that the gradient update speed would be gradually reduced for more stable signatures updates. Notice that for all experiments we used common person Re-ID metrics, such as the mean average precision (mAP) and cumulative matching characteristic (CMC) curve to represent up to Rank #10 re-identification accuracy. Finally, all networks were implemented through the PyTorch framework and its TorchVision library, while tests were performed using a single GPU, i.e., a GeForce GTX 1070 with 8GB of RAM.

C. Signals Pre-Processing Qualitative Evaluation

Amplitude and phase extracted from CSI measurements of Wi-Fi signals contain several information that can help distinguish different persons. Examples of the resulting pre-processed signals characteristics are shown in Fig. 5, for the amplitudes, and in Fig. 6 and Fig. 7 for the phases.

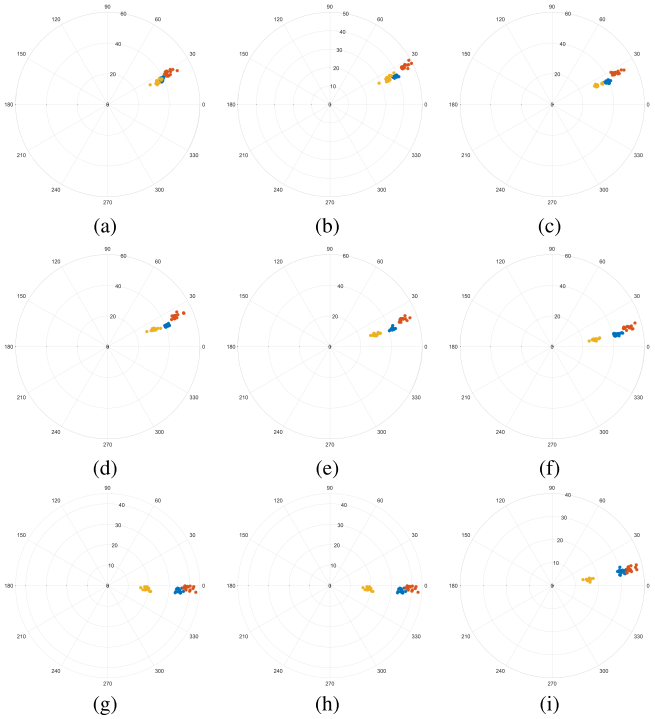

Fig. 6. Processed phases examples for 15-packets acquisitions in the same room, i.e., indoor hallway. In (a) to (i) the resulting F phases for 9 consecutive subcarriers of no obstacles between TX and RX APs and two different persons; shown in blue, red, and yellow, respectively.

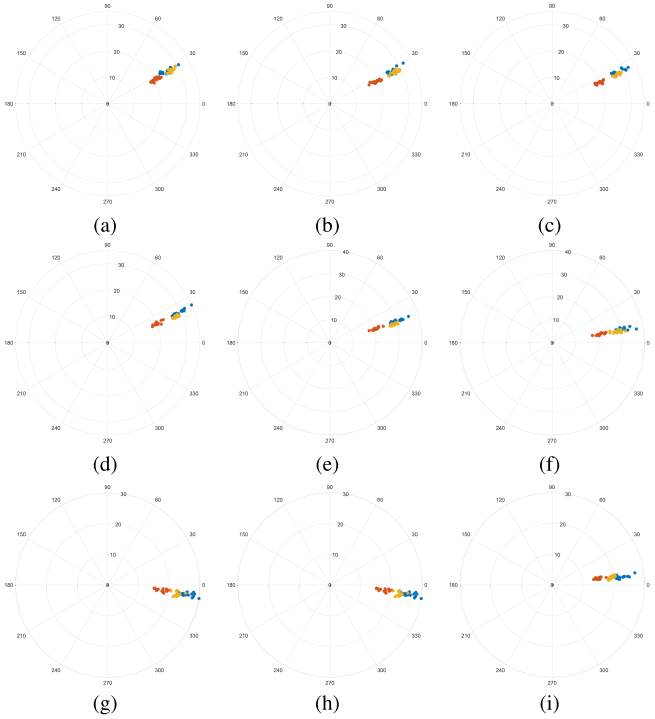

Fig. 7. Processed phases examples for 15-packets acquisitions in the same room, i.e., conference hall. In (a) to (i) the resulting F phases for 9 consecutive subcarriers of no obstacles between TX and RX APs and two different persons; shown in blue, red, and yellow, respectively.

Concerning the amplitudes, as can be seen in the heatmaps-derived 3D surfaces, the general shape is retained across different rooms for the same identity (i.e., first, second and third column in Fig. 5). However, the environment does affect the received signal even after applying the sanitizing procedure, as clearly shown in Fig. 5(a), Fig. 5(b) and Fig. 5(c), where the reported amplitudes are associated to the empty rooms, i.e., only furniture was present and no obstacle was left between the TX and RX APs. As a consequence, the resulting ambient noise will also affect the received amplitude quality when humans stand across the propagated signal. This outcome can be traced back to the random path followed by the signal itself, which is not guaranteed to be the same across multiple transmissions. However, due to the stationary environments, we can successfully mitigate possible random ambient noise by operating directly in the frequency domain, which is not feasible in non-stationary scenarios, where the signal should also be processed in the time domain [79], [84]. While this result might suggest that other techniques, such as the angle of arrival, might further improve the signal processing procedures, the produced heatmaps are still able to describe human presence in a detailed way, especially when many packets are used to build the corresponding image. As a matter of fact, the heatmaps generated in the various rooms show high inter-class and low intra-class shape differences, as can be seen in each Fig. 5 row; therefore indicating that the derived amplitude heatmaps can correctly model characteristics of distinct persons.

Regarding the phases, they can be used to capture temporal cues from Wi-Fi signals, as can be inferred from Fig. 6 and Fig. 7. In more detail, the two images report 15 consecutive filtered sanitized phases of distinct identities for 9 adjacent subcarriers in two different rooms. i.e., indoor hallway and conference hall. As shown, as time evolves, i.e., more packets are analyzed at the RX AP, phases at each subcarrier tend to concentrate on the same spot due to the presented filtering procedure that removes phase offsets. Even more interesting, for different persons, the resulting phases have dissimilar values across the various subcarriers. This outcome suggests that first, different people also have diverse effects on the signal propagation, in accordance with the findings described in [34] and [35] and second, there is a little probability for two dinstict people to have the exact distribution across all 30 subcarriers for several consecutive packets; thus indicating that a sequence-based architecture (e.g., LSTM) could most likely capture temporal shifts associated to different persons. Observe that a small number of packets is reported for each identity to avoid image clutter. However, the same reasoning applies to more subsequent packets, therefore supporting that, similarly to amplitudes, phases can also help to model unique persons and support their discrimination.

D. Wi-Fi Person Identification and Verification Evaluation

To show the effectiveness of the proposed method, we performed several ablation studies concerning the architecture, the generated signature size, as well as the number of consecutive packets to be analyzed from the Wi-Fi transmission to generate meaningful amplitude and phase features.

In relation to the chosen Re-ID model, the first batch of experiments explored the extracted features efficacy in

TABLE I

MODEL CONFIGURATION 10-FOLD CROSS-VALIDATION PERFORMANCE EVALUATION ON DATASET D_1 FOR 300 CONSECUTIVE PACKETS, AND $|s| = 256$. SIAMESE$_A$, SIAMESE$_P$, AND SIAMESE MODELS EXPLOIT AMPLITUDE, PHASE, AND JOINT SIGNAL PROPERTIES, RESPECTIVELY

Model	#Rooms	Rank #1	mAP
Siamese$_A$	1	90.46% ± 4.40%	88.29% ± 6.36%
Siamese$_P$	1	90.12% ± 4.05%	88.17% ± 5.12%
Siamese	1	94.42% ± 0.95%	92.90% ± 2.27%
Siamese$_A$	2	89.78% ± 6.20%	87.96% ± 7.10%
Siamese$_P$	2	89.35% ± 4.87%	87.90% ± 5.56%
Siamese	2	93.99% ± 1.01%	92.79% ± 2.31%
Siamese$_A$	3	88.71% ± 7.24%	86.65% ± 7.51%
Siamese$_P$	3	88.57% ± 5.15%	86.55% ± 5.97%
Siamese	3	93.51% ± 1.04%	92.17% ± 2.47%

TABLE II

SIGNATURE SIZE 10-FOLD CROSS-VALIDATION PERFORMANCE EVALUATION ON DATASET D_1 FOR 300 CONSECUTIVE PACKETS

Model	$\|s\|$	Rank #1	mAP
Siamese	16	56.72% ± 10.24%	50.84% ± 11.20%
Siamese	32	68.80% ± 8.05%	64.76% ± 9.06%
Siamese	64	85.59% ± 4.24%	83.41% ± 5.83%
Siamese	128	93.17% ± 1.12%	91.99% ± 2.63%
Siamese	256	93.51% ± 1.04%	92.17% ± 2.47%
Siamese	512	93.50% ± 1.01%	92.12% ± 2.48%
Siamese	1024	93.45% ± 0.99%	92.10% ± 2.43%

both standalone and joint solutions by designing, respectively, a siamese architecture with single sub-network streams, elaborating either amplitudes or phases, and the presented model. Notice that these experiments were performed by using subsets with increasing complexity generated from dataset D_1. Specifically, the evaluation was performed using signals associated to either the single rooms (e.g., hallway or office), all possible pairs (e.g., hallway and office or office and conference room), and all rooms in dataset D_1 (i.e., as described in Sec. IV-B). The obtained results are summarized in Table I. As shown, all models achieve significant performances for both Rank #1 and mAP metrics, with the full model consistently outperforming the single-subnetwork versions (i.e., Siamese$_A$ and Siamese$_P$) by an ≈5% margin, independently of the number of examined rooms. The reason behind this outcome is twofold. First, the extracted features can capture enough differences to distinguish the 20 identities present in D_1, since each person seems to affect the signal similarly even across distinct rooms, as discussed in Sec. IV-C. Second, amplitudes and phases describe different characteristics due to the chosen representation (i.e., heatmap $\bar{M}$ and temporal sequence $\bar{F}$, respectively), further improving the derived human descriptions when used jointly. Even more interesting, siamese models exploiting phase information attained lower variance across the 10-folds, which is due to temporal information captured from vector $\bar{F}$ by the LSTM unit. Indeed, while heatmaps can still represent different humans in a meaningful way, they can also be subject to higher association errors since they represent a coarse view of signals amplitudes.

Concerning the second round of ablation studies, an evaluation of different signature sizes was performed to assess

TABLE III

PACKETS NUMBER 10-FOLD CROSS-VALIDATION PERFORMANCE EVALUATION ON DATASET D_1 WITH $|s| = 256$

Model	#Packets	Rank #1	mAP
Siamese	10	80.28% ± 9.13%	79.02% ± 9.98%
Siamese	25	85.51% ± 7.99%	82.65% ± 8.65%
Siamese	50	88.50% ± 5.34%	87.23% ± 6.34%
Siamese	100	91.56% ± 3.01%	89.88% ± 4.00%
Siamese	200	93.29% ± 1.10%	92.02% ± 2.54%
Siamese	300	93.51% ± 1.04%	92.17% ± 2.47%

the effectiveness of the fused features $\bar{M}$ and $\bar{F}$. The results obtained on dataset D_1 are reported in Table II. As can be seen, employing higher dimensions for the signature s naturally results in improved performances. This is a direct consequence of the task complexity when multiple identities are present, as their representation cannot be fully described via small signatures. As a matter of fact, for $|s| < 64$, the system performances degrade rapidly and show high variance, confirming that the signature s is not able to capture meaningful characteristics for the unique identities discrimination. Moreover, there are also diminished increase returns in correspondence to bigger s sizes. This behavior is easily explained by the relatively low number of identities at our disposal (i.e., 20 for dataset D_1) which can be characterized by a signature size of 256. Nevertheless, to correctly represent as many unique persons as possible, the chosen signature size is a key component for the proposed system.

Regarding the last group of experiments, tests were performed to evaluate the effectiveness of the extracted features by modifying the number successive packets analyzed for their generation. The results obtained on dataset D_1 are summarized in Table III. As can be seen, performances start converging to a stable percentage from 200 packets, indicating that the corresponding extracted features carry enough information to correctly describe the various identities in D_1. In fact, using the whole sequence of 300 packets results only in slight gains for both Rank #1 and mAP metrics. This outcome confirms the representation capability of the system, that can fully describe the various identities, while also suggesting the extracted features effectiveness. As a matter of fact, significant performances are obtained even by analyzing a lower number of packets (i.e., $P \leq 25$). However, for these configurations there is a higher variance due to the smaller extracted features which might not fully capture distinct traits for more similar radio biometrics. What is more, due to the chosen median procedures, using less than 10 packets results in a performance degradation due to an increased noise in the produced features. Nevertheless, the approach quality for fewer packets is also validated through the CMC curve shown in Fig. 8, where all models attain higher performances (i.e., a score of ≈90%) starting from Rank #3. This result highlights the proposed method effectiveness and its ability to represent unique radio biometric signatures which are suitable for the person re-identification task, as demonstrated in the next section.

E. Wi-Fi Person Re-Identification Evaluation

Real-world person Re-ID scenarios, such as surveillance systems, require models to also re-identify persons with

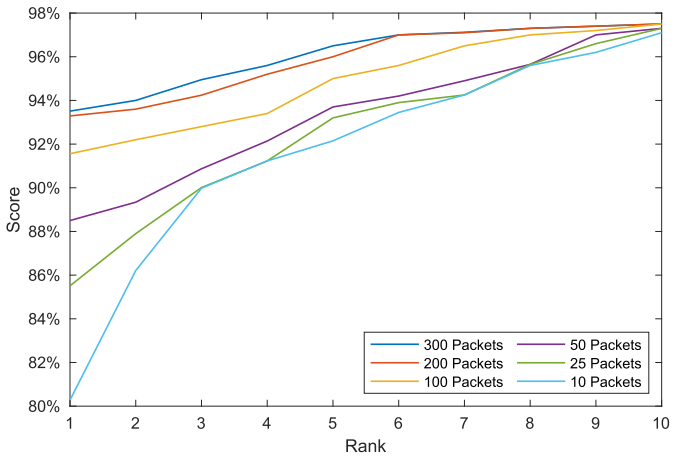

Fig. 8. CMC curve up to Rank #10 computed on dataset D_1 for different packets number.

TABLE IV

MODEL CONFIGURATION 10-FOLD CROSS-VALIDATION PERFORMANCE EVALUATION ON DATASET D_2 FOR 300 CONSECUTIVE PACKETS AND $|s| = 256$. SIAMESE$_A$, SIAMESE$_P$, AND SIAMESE MODELS EXPLOIT AMPLITUDE, PHASE, AND JOINT SIGNAL PROPERTIES, RESPECTIVELY

Model	#Rooms	Rank #1	mAP
Siamese$_A$	1	88.68% ± 3.59%	86.35% ± 4.54%
Siamese$_P$	1	88.18% ± 3.50%	86.05% ± 3.64%
Siamese	1	90.28% ± 1.02%	89.77% ± 2.29%
Siamese$_A$	2	86.72% ± 5.19%	80.12% ± 5.62%
Siamese$_P$	2	86.52% ± 4.24%	80.02% ± 4.78%
Siamese	2	89.42% ± 1.09%	88.58% ± 2.55%
Siamese$_A$	3	84.02% ± 7.25%	78.12% ± 7.70%
Siamese$_P$	3	83.62% ± 5.07%	77.62% ± 5.12%
Siamese	3	88.82% ± 1.29%	87.52% ± 2.67%

different and unknown identities from those seen at training time. Therefore, to correctly evaluate the presented pipeline in such scenarios, we performed a comprehensive assessment for both model configurations and successive packets number on the distinct dataset D_2, by using the D_1-trained models with signature size $s = 256$, introduced in Section IV-D. Specifically, regarding the evaluation on D_2, for each of its 15 unique persons, 1 wireless transmission per room was randomly selected as the gallery, for a total of 45 transmissions; while the remaining 4 samples were used as probes to assess the re-identification capabilities of the system, counting 180 test transmissions. Moreover, since dataset D_2 represents only a small fraction of real world data, tests were performed 10 times using different random selections, and the average performance was reported to ensure statistically stable results.

Concerning experiments on model configurations, the obtained results on dataset D_2 for 300 consecutive packets are reported in Table IV. As shown, the same behavior observed and discussed in Sec. IV-D also applies to the different unique identities of collection D_2. Indeed, by increasing the number of rooms, there is a slight performance decrease for all models, and architectures exploiting phase information have reduced variance for both Rank #1 and mAP metrics most likely due to, as mentioned, the temporal information

TABLE V

PACKETS NUMBER 10-FOLD CROSS-VALIDATION PERFORMANCE EVALUATION ON DATASET D_2 WITH $|s| = 256$

Model	#Packets	Rank #1	mAP
Siamese	10	72.12% ± 12.69%	64.52% ± 13.35%
Siamese	25	74.22% ± 9.34%	70.72% ± 10.23%
Siamese	50	80.95% ± 6.37%	78.75% ± 7.57%
Siamese	100	82.15% ± 4.98%	80.55% ± 5.60%
Siamese	200	85.05% ± 2.70%	83.85% ± 3.69%
Siamese	300	88.82% ± 1.29%	87.52% ± 2.67%

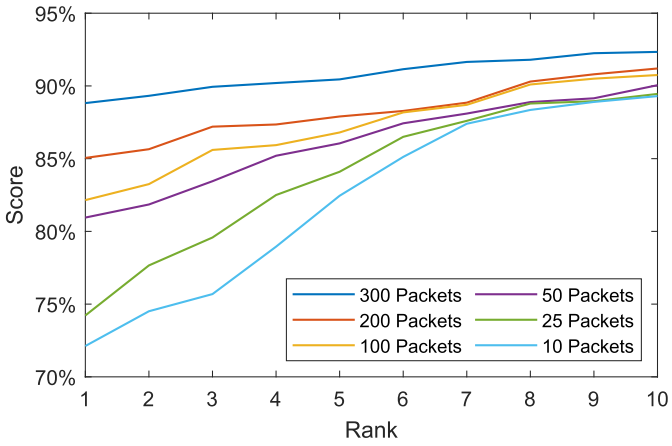

Fig. 9. CMC curve up to Rank #10 computed on dataset D_2 for different packets number.

for the interested feature. Nevertheless, when analyzing the most complex scenario with 3 rooms, the full siamese model achieves significant perfomances; suggesting that even though the architectures have never observed the various identities, they can still extract relevant radio biometric signatures for their re-identification.

In relation to tests on the number of successive packets, results are reported in Table V, while the corresponding CMC curve up to Rank #10 is depicted in Fig. 9. As shown, the best performing model exploits the whole sequence of 300 packets. However, differently from the performances observed in Table III, where models using at least 100 packets had similar scores, for unknown identities there is a greater gap with respect to the maximum amount of recorded transmission packets. The motivation for this outcome is twofold. First and foremost, dataset D_2 has 3 times the number of test samples with respect to D_1, which was purposely built in this way to obtain consistent results over a more complex collection. Second, for real-world scenarios, i.e., where re-identification is performed on unknown people, the proposed model does not execute a training phase and, consequently, does not exploit the joint loss function shown in Eq. (16), which also leverages the specific identities to build more robust signatures. Nevertheless, while interesting performances are already achieved with only 10 successive packets, by increasing this number it is possible to obtain more discriminative radio biometrics and, therefore, improved radio biometric signatures able to mitigate the identity loss absence. Thus, these results confirm the findings presented by [34], [35] on signal variations in correspondence with different biological tissues, and highlight the

Wi-Fi effectiveness in addressing the person re-identification task without classical vision-based drawbacks; consequently opening up a new frontier for surveillance applications where it can be crucial to re-identify unknown persons across different locations.

V. Conclusion

In this paper we introduced a novel person re-identification approach based on radio biometric signatures extracted from Wi-Fi signals. As shown by the results achieved in restricted environments, the proposed siamese architecture with parallel sub-networks, analyzing amplitude heatmaps and phase vectors, can extract meaningful representations, i.e., signatures, that enable the Re-ID of both known and unknown persons thanks to the information carried by the transmitted signals, validating the presented idea and potential application in real-world surveillance scenarios that are typically constrained.

As future work, an extended dataset version with more than 150 distinct identities will be collected and released to offer a benchmark for this unorthodox re-identification approach. In particular, this re-identification dataset will comprise multiple modalities in the form of synchronized tuples. The latter will contain Wi-Fi transmissions, RGB, and depth videos, to enable, on the one hand, a direct comparison between video and wireless modalities on the Re-ID task, and on the other hand, the implementation of multimodal methods that might benefit from the added cross-modality information when re-identifying a person across different locations. Furthermore, the presented pipeline will be used as a baseline approach for the Re-ID from Wi-Fi signals. At the same time, specific video-based and multimodal architectures will be implemented to present a comprehensive benchmark comparing the differences between the various modalities with a focus on their strengths and weaknesses. Moreover, further inquiries will be performed on other signal properties in the time domain (e.g., impulse response or time of arrival) that might be used either in a standalone solution or jointly with those implemented in this work. In addition, different solutions will also be designed to better exploit other characteristics, such as the angle of arrival, to further reduce possible ambient noise and ultimately handle more complex non-stationary environments.

References

[1] Y.-C. Chen, X. Zhu, W.-S. Zheng, and J.-H. Lai, "Person re-identification by camera correlation aware feature augmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 392–408, Feb. 2018.

[2] Q. Leng, M. Ye, and Q. Tian, "A survey of open-world person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 1092–1108, Apr. 2020.

[3] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jan. 26, 2021, doi: 10.1109/TPAMI.2021.3054775.

[4] Z. Feng, J. Lai, and X. Xie, "Learning view-specific deep networks for person re-identification," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3472–3483, Jul. 2018.

[5] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of viewpoint," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 608–617.

[6] A. Wu, W.-S. Zheng, and J.-H. Lai, "Robust depth-based person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2588–2603, Jun. 2017.

[7] Z. Feng, J. Lai, and X. Xie, "Learning modality-specific representations for visible-infrared person re-identification," *IEEE Trans. Image Process.*, vol. 29, pp. 579–590, 2020.

[8] C. Song, Y. Huang, W. Ouyang, and L. Wang, "Mask-guided contrastive attention model for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1179–1188.

[9] S. Zhou, F. Wang, Z. Huang, and J. Wang, "Discriminative feature learning with consistent attention regularization for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8040–8049.

[10] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, "Pose-guided feature alignment for occluded person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 542–551.

[11] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "VRSTC: Occlusion-free video person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7183–7192.

[12] J. Li, S. Zhang, J. Wang, W. Gao, and Q. Tian, "Global-local temporal representations for video person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3958–3967.

[13] Y. Huang, J. Xu, Q. Wu, Y. Zhong, P. Zhang, and Z. Zhang, "Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3459–3471, Oct. 2020.

[14] M. Ye and P. C. Yuen, "PurifyNet: A robust person re-identification model with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2655–2666, 2020.

[15] K. L. Navaneet, R. K. Sarvadevabhatla, S. Shekhar, R. V. Babu, and A. Chakraborty, "Operator-in-the-loop deep sequential multi-camera feature fusion for person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2375–2385, 2020.

[16] H. Li, Y. Chen, D. Tao, Z. Yu, and G. Qi, "Attribute-aligned domain-invariant feature learning for unsupervised domain adaptation person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1480–1494, 2021.

[17] D. Avola, L. Cinque, A. Fagioli, G. L. Foresti, D. Pannone, and C. Piciarelli, "Bodyprint—A meta-feature based LSTM hashing model for person re-identification," *Sensors*, vol. 20, no. 18, p. 5365, Sep. 2020.

[18] M. Ye, X. Lan, Z. Wang, and P. C. Yuen, "Bi-directional center-constrained top-ranking for visible thermal person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 407–419, 2020.

[19] M. Ye, J. Shen, and L. Shao, "Visible-infrared person re-identification via homogeneous augmented tri-modal learning," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 728–739, 2020.

[20] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2018.

[21] O. Bialer, D. Raphaeli, and A. J. Weiss, "A time-of-arrival estimation algorithm for OFDM signals in indoor multipath environments," *Signal Process.*, vol. 169, Apr. 2020, Art. no. 107375.

[22] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.

[23] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–32, 2013.

[24] A. A. Ghany, B. Uguen, and D. Lemur, "A robustness comparison of measured narrowband CSI vs RSSI for IoT localization," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC-Fall)*, Nov. 2020, pp. 1–5.

[25] Q. Li *et al.*, "AF-DCGAN: Amplitude feature deep convolutional GAN for fingerprint construction in indoor localization systems," *IEEE Trans. Emerg. Topics Comput.*, vol. 5, no. 3, pp. 468–480, Jun. 2021.

[26] X. Rao, Z. Li, Y. Yang, and S. Wang, "DFPhaseFL: A robust device-free passive fingerprinting wireless localization system using CSI phase information," *Neural Comput. Appl.*, vol. 32, no. 18, pp. 14909–14927, Sep. 2020.

[27] S. Shi, S. Sigg, L. Chen, and Y. Ji, "Accurate location tracking from CSI-based passive device-free probabilistic fingerprinting," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5217–5230, Jun. 2018.

[28] F. Xiao, J. Chen, X. Xie, L. Gui, L. Sun, and R. Wang, "SEARE: A system for exercise activity recognition and quality evaluation based on green sensing," *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 3, pp. 752–761, Jul. 2020.

[29] 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. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[30] X. Dang, X. Si, Z. Hao, and Y. Huang, "A novel passive indoor localization method by fusion CSI amplitude and phase information," *Sensors*, vol. 19, no. 4, p. 875, Feb. 2019.

[31] X. Wang, L. Gao, and S. Mao, "PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1–6.

[32] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 1–36, Jul. 2019.

[33] Q. Xu, Y. Chen, B. Wang, and K. J. R. Liu, "Radio biometrics: Human recognition through a wall," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 5, pp. 1141–1155, May 2017.

[34] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz," *Phys. Med. Biol.*, vol. 41, no. 11, p. 2251, Nov. 1996.

[35] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues," *Phys. Med. Biol.*, vol. 41, no. 11, p. 2271, Nov. 1996.

[36] T. Huang, G. Yang, and G. Tang, "A fast two-dimensional median filtering algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. APPS-27, no. 1, pp. 13–18, Feb. 1979.

[37] L. Wu, Y. Wang, J. Gao, and X. Li, "Where-and-when to look: Deep Siamese attention networks for video-based person re-identification," *IEEE Trans. Multimedia*, vol. 21, no. 6, pp. 1412–1424, Jun. 2019.

[38] M. Zheng, S. Karanam, Z. Wu, and R. J. Radke, "Re-identification with consistent attentive Siamese networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5735–5744.

[39] X. Rafael-Palou *et al.*, "Re-identification and growth detection of pulmonary nodules without image registration using 3D Siamese neural networks," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101823.

[40] M. Ye, J. Shen, X. Zhang, P. C. Yuen, and S.-F. Chang, "Augmentation invariant and instance spreading feature for softmax embedding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 924–939, Feb. 2022.

[41] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 6, 1993, pp. 737–744.

[42] Y. Guo and N.-M. Cheung, "Efficient and deep person re-identification using multi-level similarity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 2335–2344.

[43] Z. Zhang and H. Peng, "Deeper and wider Siamese networks for real-time visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4586–4595.

[44] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from WiFi: A Siamese recurrent convolutional architecture," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10763–10772, Dec. 2019.

[45] P. Turaga and Y. A. Ivanov, "Diamond sentry: Integrating sensors and cameras for real-time monitoring of indoor spaces," *IEEE Sensors J.*, vol. 11, no. 3, pp. 593–602, Mar. 2011.

[46] S. Zhang, L. Zhang, W. Wang, and X. Wu, "AsNet: Asymmetrical network for learning rich features in person re-identification," *IEEE Signal Process. Lett.*, vol. 27, pp. 850–854, 2020.

[47] D. Avola, M. Cascio, L. Cinque, A. Fagioli, G. L. Foresti, and C. Massaroni, "Master and rookie networks for person re-identification," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2019, pp. 470–479.

[48] S. Zhou *et al.*, "Deep self-paced learning for person re-identification," *Pattern Recognit.*, vol. 76, pp. 739–751, Apr. 2018.

[49] C. Li, Z. Cao, and Y. Liu, "Deep AI enabled ubiquitous wireless sensing: A survey," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–35, Mar. 2022.

[50] E. B. Hamida and G. Chelius, "Investigating the impact of human activity on the performance of wireless networks—An experimental approach," in *Proc. IEEE Int. Symp. WoWMoM*, Jun. 2010, pp. 1–8.

[51] W. Xue, Q. Li, X. Hua, K. Yu, W. Qiu, and B. Zhou, "A new algorithm for indoor RSSI radio map reconstruction," *IEEE Access*, vol. 6, pp. 76118–76125, 2018.

[52] C. R. Karanam and Y. Mostofi, "3D through-wall imaging with unmanned aerial vehicles using WiFi," in *Proc. 16th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2017, pp. 131–142.

[53] W. Xue *et al.*, "Eight-diagram based access point selection algorithm for indoor localization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13196–13205, Nov. 2020.

[54] Y. Fu, P. Chen, S. Yang, and J. Tang, "An indoor localization algorithm based on continuous feature scaling and outlier deleting," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1108–1115, Apr. 2018.

[55] Z. Xu, R. Wang, X. Yue, T. Liu, C. Chen, and S. Fang, "FaceME: Face-to-machine proximity estimation based on RSSI difference for mobile industrial human–machine interaction," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3547–3558, Aug. 2018.

[56] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10639–10651, Sep. 2019.

[57] B. Soro and C. Lee, "Joint time-frequency RSSI features for convolutional neural network-based indoor fingerprinting localization," *IEEE Access*, vol. 7, pp. 104892–104899, 2019.

[58] A. Booranawong, N. Jindapetch, and H. Saito, "A system for detection and tracking of human movements using RSSI signals," *IEEE Sensors J.*, vol. 18, no. 6, pp. 2531–2544, Mar. 2018.

[59] V. Bianchi, P. Ciampolini, and I. De Munari, "RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 2, pp. 566–575, Feb. 2019.

[60] Y. Gu, F. Ren, and J. Li, "PAWS: Passive human activity recognition based on WiFi ambient signals," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 796–805, Oct. 2016.

[61] Z. Zhou, C. Wu, Z. Yang, and Y. Liu, "Sensorless sensing with WiFi," *Tsinghua Sci. Technol.*, vol. 20, no. 1, pp. 1–6, Feb. 2015.

[62] S. Denis, R. Berkvens, and M. Weyn, "A survey on detection, tracking and identification in radio frequency-based device-free localization," *Sensors*, vol. 19, no. 23, pp. 1–59, 2019.

[63] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2016.

[64] T. Nakamura, M. Bouazizi, K. Yamamoto, and T. Ohtsuki, "Wi-Fi-CSI-based fall detection by spectrogram analysis with CNN," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[65] Q. Song, S. Guo, X. Liu, and Y. Yang, "CSI amplitude fingerprinting-based NB-IoT indoor localization," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1494–1504, Jun. 2017.

[66] X. Tong, H. Li, X. Tian, and X. Wang, "Wi-Fi localization enabling self-calibration," *IEEE/ACM Trans. Netw.*, vol. 29, no. 2, pp. 904–917, Apr. 2021.

[67] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "Pilot: Passive device-free indoor localization using channel state information," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2013, pp. 236–245.

[68] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10346–10356, Nov. 2017.

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

[70] D. Chung, K. Tahboub, and E. J. Delp, "A two stream Siamese convolutional neural network for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1983–1991.

[71] E. Nepovinnykh, T. Eerola, and H. Kalviainen, "Siamese network based pelage pattern matching for ringed seal re-identification," in *Proc. IEEE Winter Appl. Comput. Vis. Workshops (WACVW)*, Mar. 2020, pp. 25–34.

[72] N. K. Tagore, A. Singh, S. Manche, and P. Chattopadhyay, "Person re-identification from appearance cues and deep Siamese features," *J. Vis. Commun. Image Represent.*, vol. 75, Feb. 2021, Art. no. 103029.

[73] L. Guo, Z. Lu, X. Wen, S. Zhou, and Z. Han, "From signal to image: Capturing fine-grained human poses with commodity Wi-Fi," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 802–806, Apr. 2020.

[74] C. Li, M. Liu, and Z. Cao, "WiHF: Enable user identified gesture recognition with WiFi," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Jul. 2020, pp. 586–595.

[75] J. Ding, Y. Wang, and X. Fu, "Wihi: WiFi based human identity identification using deep learning," *IEEE Access*, vol. 8, pp. 129246–129262, 2020.

[76] V. Jayasundara, H. Jayasekara, T. Samarasinghe, and K. T. Hemachandra, "Device-free user authentication, activity classification and tracking using passive Wi-Fi sensing: A deep learning-based approach," *IEEE Sensors J.*, vol. 20, no. 16, pp. 9329–9338, 2020.

[77] F. Wang, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "Respiration tracking for people counting and recognition," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5233–5245, Jun. 2020.

[78] X. Shen, L. Guo, Z. Lu, X. Wen, and Z. He, "WiRIM: Resolution improving mechanism for human sensing with commodity Wi-Fi," *IEEE Access*, vol. 7, pp. 168357–168370, 2019.
[79] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based indoor localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1300–1309, Jul. 2013.
[80] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 1735–1742.
[81] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.
[82] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 1988–1996.
[83] Y. Mang, L. Xiangyuan, L. Jiawei, and Y. C. Pong, "Hierarchical discriminative learning for visible thermal person re-identification," in *Proc. Conf. Art. Intel. (AAAI)*, 2018, pp. 7501–7508.
[84] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

Danilo Avola (Member, IEEE) received the M.Sc. degree in computer science from Sapienza University of Rome, Rome, Italy, in 2002, and the Ph.D. degree in molecular and ultrastructural imaging from the University of L'Aquila, L'Aquila, Italy, in 2014. Since 2021, he has been an Assistant Professor with the Department of Computer Science, Sapienza University of Rome, where he leads both the Robotics Vision and Artificial Intelligence Laboratory (TITAN Lab) and the Computer Vison Laboratory (VisionLab). Since 2018, he has been a Research and Development Senior Engineer at WSENSE S.r.l., a spin-off of Sapienza University of Rome, where he leads the Computer Vision Team (CVT), and since 2010, he provides consultation and collaboration to companies engaged in computer science, computer vision, and artificial intelligence research projects. Previously, he was Post-Doctoral Researcher with the Department of Mathematics, Computer Science and Physics (DMIF), University of Udine, Udine, Italy, where he was a Research and Development Senior Engineer at the Artificial Vision and Real-Time Systems Laboratory (AVIRES Lab). His research interests include computer vision, image/video processing, human–computer interaction, Wi-Fi signal processing, EGG signal processing, machine/deep learning, multimodal systems, pattern recognition, event/action/affect recognition, action, scene understanding, body language and face expression interpretation, and robotics (UAVs, AUVs, ROVs, and Humanoids), and he has published over 100 papers on these topics. Since 2011, he is a member of IAPR and CVPL. Among other awards and prizes, he received the Outstanding Paper Award for IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS (TII) in 2020. Currently, he is a Reviewer of top international journals in computer vision, including *Pattern Recognition*, *Pattern Recognition Letters*, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, and *International Journal of Computer Vision*; in addition, he is an associate editor and a guest editor of different ranked international journals.

Marco Cascio (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from the University of Messina, Messina, Italy, in 2015, and the M.Sc. degree *(cum laude)* in computer science from the Sapienza University of Rome, Rome, Italy, in 2018, where he is currently pursuing the Ph.D. degree in computer science. Since 2018, he has been a member of the Computer Vision Laboratory, Department of Computer Science, Sapienza University of Rome. His current research interests include scene analysis, pattern recognition, machine/deep learning, event recognition, object tracking, people re-identification, Wi-Fi signal processing, active vision in surveillance systems, signal processing, and human–computer interaction.

Luigi Cinque (Senior Member, IEEE) received the M.Sc. degree in physics from the University of Napoli, Naples, Italy, in 1983. From 1984 to 1990, he was with the Laboratory of Artificial Intelligence (Alenia S.p.A), working on the development of expert systems and knowledge-based vision systems. He is a Full Professor of computer science with the Sapienza University, Rome, Italy. He is author of more than 200 papers in the national and international journals and conference proceedings. His first scientific interests cover image processing, shape and object recognition, analysis and interpretation of images and their applications, with particular emphasis at content-based retrieval in visual digital archives, and advanced man–machine interaction assisted by computer vision. Currently, his main interests involve distributed systems for analysis and interpretation of video sequences, target tracking, multisensor data, and information fusion. Some of techniques, he has proposed have found applications in field of video-based surveillance systems, autonomous vehicle, road traffic control, human behavior understanding, and visual inspection. He is a member of ACM, IAPR, and CVPL. He currently serves as a reviewer for many international journals (e.g., IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON MEDICAL IMAGING, and *Image and Vision Computing*). He served on the scientific committees of international conferences (e.g., CVPR, ICME, and ICPR) and symposia. He serves as a reviewer for the European Union in different research program.

Alessio Fagioli (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees *(cum laude)* in computer science from the Sapienza University of Rome, Rome, Italy, in 2016 and 2019, respectively. He is currently a Ph.D. Fellow and a member of the VisionLab, Department of Computer Science, Sapienza University of Rome. His current research interests include medical image analysis, machine learning, deep learning, affect recognition, event recognition, object tracking, and human–computer interaction.

Chiara Petrioli (Fellow, IEEE) is currently a Professor of computer science, the Director of the Sensor Networks and Embedded Systems Laboratory (SENSES Lab), Department of Computer Science, University of Rome "La Sapienza." She also leads the Cyber Physical System Laboratory of "La Sapienza" Center for Cyber Intelligence and Information Security, and is a Founding Partner of "La Sapienza" spinoff WSENSE S.r.l. She has been a member of the academic senate and the Chair of the Ph.D. Program in computer science at La Sapienza. Her research interests concern the design and optimization of future wireless, embedded, the IoT, and cyber physical systems. She has published over a 150 papers in prominent international journals and conferences (over 6500 citations; H-index 45). She has been the PI of over 20 national and international research projects, serving as coordinator of three EC projects (FP7 projects GENESI and SUNRISE, EASME ArcheoSub) highlighted as success stories on the Digital Agenda of Europe and featured by international mass media, including RAI SuperQuark and Presa Diretta, Wired USA, *The Guardian*, *Bild* magazine, and *National Geographic*. Her research has resulted in international patents and in award-winning innovative technologies. She is a pioneer of the Internet of Underwater Things, an area on which she has led the development of breakthrough technologies listed in the NT100 Top "Social Global Techs changing our lives 2016." She is a Fulbright Scholar, one of the Inspiring 50 2018, top women in technology and of the N2Women 2019 Stars in Computer Networking and Communications. She is the Chair of the Steering Committee of IEEE SECON and the General Chair of ACM MobiHoc 2019, was the Program Co-Chair of IEEE INFOCOM 2016 and the General Chair of ACM SenSys 2013. She has been a member of the Steering Committee of ACM SenSys, the Executive Committee of ACM SIGMOBILE, and of the ACM Europe Council; an Associate Editor of IEEE TRANSACTIONS ON MOBILE COMPUTING and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Communications* (Elsevier); a Guest Editor of Special Issues for IEEE ACCESS, *Ad Hoc Networks* (Elsevier), and *Physical Communications* (Elsevier); and has been the Program Co-Chair of leading conferences in the field, such as ACM MobiCom and IEEE SECON.