Integrated Sensing and Communication From Learning Perspective: An SDP3 Approach

Guoliang Li, Graduate Student Member, IEEE, Shuai Wang, Member, IEEE,
Jie Li, Rui Wang, Member, IEEE, Fan Liu, Senior Member, IEEE, Xiaohui Peng, Member, IEEE,
Tony Xiao Han, Senior Member, IEEE, and Chengzhong Xu, Fellow, IEEE

Abstract—Characterizing the sensing and communication performance tradeoff in integrated sensing and communication (ISAC) systems is challenging in the applications of learning-based human motion recognition. This is because of the large experimental data sets and the black-box nature of deep neural networks. This article presents SDP3, a Simulation-Driven Performance Predictor and Optimizer, which consists of SDP3 data simulator, SDP3 performance predictor and SDP3 performance optimizer. Specifically, the SDP3 data simulator generates vivid wireless sensing data sets in a virtual environment, the SDP3 performance predictor predicts the sensing performance based on the curve fitting method, and the SDP3 performance optimizer investigates the sensing and communication performance tradeoff analytically. It is shown that the simulated sensing data set matches the experimental data set very well in the motion recognition accuracy. By leveraging SDP3, it is found that the achievable region of recognition accuracy and communication throughput consists of a communication saturation zone, a sensing saturation zone, and a communication-sensing adversarial zone, of which the desired balanced performance for ISAC systems lies in the third one.

Index Terms—Integrated sensing and communication (ISAC), resource allocation, wireless sensing simulation.

I. INTRODUCTION

INTEGRATED sensing and communication (ISAC) is a promising technology for the next generation cellular system and wireless local area network (WLAN) which enables a huge amount of applications, including security surveillance, health care, and human–computer interaction (HCI) [1], [2]. It is expected to considerably improve the spectral, energy, and hardware efficiencies of wireless systems [3], [4]. Since wireless resource is shared between sensing and communication functionalities in ISAC systems, it is of significant interest to investigate their tradeoff relation. However, the analysis of sensing performance could be challenging.

Generally, most of the sensing tasks can be classified into three categories, including detection, estimation, and recognition. The detection tasks aim to determine the state of a target, such as presence/absence. The detection probability is a typical metric to quantify the detector performance. In [5], an integrated communication and passive radar system was considered, where a generalized likelihood ratio test (GLRT) was proposed for target detection and the corresponding detection probability was approximated in a closed-form formula. Accordingly, the detection probability was maximized subject to a minimum communication rate constraint. Furthermore, the above method was extended to an integrated multistatic radar and communication system with the aim for power allocation optimization in [6].

The estimation tasks acquire the useful parameters, i.e., the distance, velocity, and angle, from the sensed targets. For instance, a dual functional waveform was utilized to minimize the estimation mean square error (MSE) of the target response matrix while ensuring a worst communication performance in [7]. When the closed-form expression of MSE is not attainable, the Cramér–Rao bound (CRB) could be adopted for the sensing performance evaluation, which represents the lower bound of the variance of all the unbiased estimators. For example, in [8], the CRB was minimized subject to the signal-to-interference-plus-noise ratio (SINR) constraint of communication receivers. Furthermore, in [9], and [10], the estimation rate, quantifying the reduction of the uncertainties for the sensing parameters per second, was proposed.

Manuscript received 13 February 2023; revised 20 July 2023; accepted 11 August 2023. Date of publication 29 August 2023; date of current version 6 February 2024. This work was supported in part by the Science and Technology Development Fund of Macao S.A.R (FDCT) under Grant 0081/2022/A2 and Grant 0123/2022/AF3; in part by the Guangdong Basic and Applied Basic Research Project under Grant 2021B1515120067; in part by the National Natural Science Foundation of China under Grant 62171213, Grant 62101234, and Grant U20B2039; in part by the Shenzhen Science and Technology Program under Grant RCBS2020071411456153; and in part by the Young Elite Scientist Sponsorship Program by CAST under Grant YESS20210055. This article was presented in part at the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Lucca, Italy, September 2021 [DOI: 10.1109/SPAWC51858.2021.9593198]. (Corresponding authors: Shuai Wang; Rui Wang.)

Guoliang Li was with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China. He is now with the State Key Laboratory of Internet of Things for Smart City and the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: li.guoliang@connect.um.edu.mo).

Shuai Wang is with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: s.wang@siat.ac.cn).

Jie Li, Rui Wang, and Fan Liu are with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: lij2019@mail.sustech.edu.cn; wang.r@sustech.edu.cn; liuf6@sustech.edu.cn).

Xiaohui Peng and Tony Xiao Han are with the Wireless Technology Lab, 2012 Laboratories, Huawei Technologies Company Ltd., Shenzhen 518129, China (e-mail: pengxiaohui5@huawei.com; tony.hanxiao@huawei.com).

Chengzhong Xu is with the State Key Laboratory of Internet of Things for Smart City and the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: cxxu@um.edu.mo). Digital Object Identifier 10.1109/JIOT.2023.3309837
then the tradeoff analysis between the estimation rate and the communication rate was provided.

The above two sensing tasks are usually processed in the physical layer, while the recognition tasks are usually accomplished in the application layer aided by the machine intelligence. They aim to acquire the semantic information of the sensed targets, e.g., activity recognition [11], [12], hand gesture recognition [13], [14], and gait recognition [15], [16]. For example, in [17], the short time fourier transform (STFT) was adopted to generate spectrograms of the sensing data. Then the support vector machine (SVM) was trained for motion classification, whose accuracy was higher than 90%. By replacing the SVM with a convolutional deep neural network (CNN), it was shown in [18] that a higher classification accuracy could be achievable. However, most of the recognition tasks are based on deep neural network, where the relation between sensing accuracy and sensing resource (e.g., sensing time, power, etc.) can hardly be represented analytically. Moreover, the above works rely on the motion data sets generated from time-consuming and labor-intensive experiments. As a result, there are the following challenges in the sensing and communication tradeoff analysis with the recognition tasks.

1) There is no analytical model of recognition accuracy, which maps the sensing resource (e.g., sensing time, power, etc.) to the recognition accuracy.

2) Even for numerical performance evaluation, there is no low-cost method to generate the motion sensing data sets for the training of recognition models.

In this article, we would like to shed some light on the above open issues. In our preliminary work [1], a data-assisted hybrid channel (DAHC) model was proposed to simulate the received sensing signals. In this article, we shall extend this work and propose a complete simulation-driven analysis framework, namely, Simulation-Driven Performance Predictor and oPtimizer (SDP3), to quantify the tradeoff between recognition accuracy and communication throughput, which consists of SDP3 data simulator, SDP3 performance predictor, and SDP3 performance optimizer as illustrated in Fig. 1.

Compared to the existing studies [5], [6], [7], [8], [9], [10], the novelties and contributions of this article are listed as follows:

1) The SDP3 data simulator, which equipped with a wireless channel model capable of portraying sensing features precisely and mimicks experimental data sets in a virtual environment consisting of human target, communication receivers, and static objects (e.g., walls), is constructed to generate the high-fidelity motion data sets with low cost.

2) The SDP3 performance predictor is developed to capture the sensing performances and establish the analytical models with closed-form expression for the mapping relationship between sensing resources (e.g., sensing time, power, etc.) and recognition accuracy.

3) The SDP3 performance optimizer is proposed to characterize the performance tradeoff in ISAC systems oriented to the recognition task. It is shown that the derived accuracy-throughput (A-T) region consists of a communication saturation zone, a sensing saturation zone, and a sensing-communication adversarial zone, where the last zone achieves the best balance between sensing and communication.

The remainder of the article is organized as follows. The system model considered in this article is introduced in Section II. The SDP3 data simulator is presented in Section III. The mechanisms of the proposed SDP3 performance predictor and optimizer are presented in Section IV. The simulation and experiment results are demonstrated in Section V. Finally, the conclusion is drawn in Section VI.

II. SYSTEM MODEL

An SDP3 framework to characterize the communication and human motion sensing tradeoff is proposed in this article. In order to demonstrate the procedure of the framework,
we consider an time-division-multiple-access-based (TDMA-based) ISAC system located in a conference room as an example, where the conference room follows the same specification as that in IEEE 802.11ay [19].

As illustrated in Fig. 2, the system consists of one ISAC-enabled base station (BS), K communication receivers and one human target to be sensed. Both the radar and communication modules are implemented at the BS, which are multiplexed in time domain. In order to facilitate transmission and sensing scheduling, the time is organized by slots, including the sensing slots and communication slots. The slot duration is $T_s$ and the channel impulse response is assumed to be quasi-static in each slot. The communication modular is enabled in the communication slots for downlink data transmission, and the radar modular is enabled in the sensing slots for the motion recognition of the human target. The sensing and communication tradeoff is studied by adapting the number of sensing and communication slots in every $N$ slots, which are referred to as a scheduling period. The numbers of communication and sensing slots in a scheduling period are denoted as $N_c$ and $N_s$, respectively. Thus, $N_c + N_s = N$. Moreover, in order to better capture the micro-Doppler effect, the sensing slots are not successive. Instead, there are $N_m$ communication slots between two neighboring sensing slots. Large $N_s$ could lead to better resolution of micro-Doppler effect and capture more motion details of human target, at the price of lower communication throughput. The communication and sensing models are elaborated below, respectively.

A. Communication Model

One receiver is selected in one communication slot for downlink transmission. The transmission signal is modulated via orthogonal frequency-division multiplexing (OFDM) with $M$ subcarriers. Let $h_{k,j}(t)$ be the channel impulse response from the BS to the $k$th communication receiver in the $j$th slot.

The channel gain of the $m$th subcarrier from the BS to the $k$th receiver can be written as

$$H_{k,j,m} = \sum_{l=0}^{L-1} h_{k,j}(t - lT_{sp}) e^{-j2\pi ml/M}, \quad 0 \leq m \leq M - 1$$

where $T_{sp}$ is the sampling period of OFDM transceiver [20]. Then, the received signal of the $m$th subcarrier is given by

$$Y_{k,j,m} = H_{k,j,m}X_{k,j,m} + Z_{k,j,m}$$

where $X_{k,j,m}$ and $Z_{k,j,m} \sim CN(0, \sigma_0^2)$ are the transmitted signal and the white Gaussian noise, respectively, and $\sigma_0^2$ is the noise power. As a result, the throughput of the $k$th receiver in the $j$th slot (if this receiver is selected) can be expressed as

$$T_{k,j} = \frac{T_s}{T_o} \sum_{m=1}^{M} \log_2(1 + \gamma_{k,j,m})$$

where $T_o$ is the duration of one complete OFDM symbol (including the cyclic prefix),

$$\gamma_{k,j,m} = \frac{|H_{k,j,m}|^2 P_{k,j,m}}{\sigma_0^2}$$

and $P_{k,j,m}$ is the power allocated to the $m$th subcarrier. Notice that the total power constraint $\sum_{m=1}^{M} P_{k,j,m} \leq P$ should be satisfied in the power allocation.

B. Sensing Model

In each sensing slot, the frequency-modulated continuous wave (FMCW) is broadcasted for human motion recognition, followed by a guarding interval. Let $s(t)$ denote the FMCW and $h_{0,j}(t)$ be the channel impulse response between the radar modular transmitter and receiver of the $j$th slot (if the $j$th slot is sensing slot), the received signal at the radar modular is

$$r_j(t) = h_{0,j}(t) * s(t) + n_j(t)$$

where $n_j(t)$ is the Gaussian noise with average power $\sigma_z^2$, and $s(t)$ follows the power constraint $(1/T)\int_{-T/2}^{T/2} |s(t)|^2 dt = P$.

Based on the received signals of the sensing slots $\{r_j(t)\}_{j \mod (N_m + 1) = 1, \ 1 \leq j \leq N_s(N_m + 1)}$, the

---

1Although there is Doppler effect in the channel due to the motion of sensing target, the phase shift raised by the Doppler effect is negligible in one slot. Moreover, the propagation paths without Doppler shift is dominant in the overall channel impulse response. Hence, it is assumed that the CSI is quasi-static in one slot.
human motion can be recognized via a neural network which should have been trained in advance via a data set collected in the same environment. Note it is inefficient to collect data set of \( \{ r_{j}(t) \mid j \mod (N_{m} + 1) = 1, 1 \leq j \leq N_{s}(N_{m} + 1) \} \) via experiments for various environments, and it is also difficult to investigate the recognition accuracy of neural network with respect to \( N_{s} \) analytically. The study of communication and sensing tradeoff is challenging. In order to address the above issues, in the following Section III, the SDP3 data simulator based on a DAHC model is proposed to generate the data sets of received sensing signals via simulation, which has been adopted as a part of the channel model for WLAN sensing in IEEE 802.11bf [21]. Based on it, a Deep Spectrogram Network (DSN) for motion recognition is proposed, and the approximated expression of recognition accuracy versus the number of sensing slots, denoted as \( A = \Theta(N_{s}) \), is obtained in Section IV, which is referred to as the SDP3 performance predictor. Finally, the tradeoff between the sensing and communication performance is studied via SDP3 performance optimizer in Section IV.

### III. SDP3 Data Simulator

In this section, we first summarize the drawbacks of the existing channel models in sensing data set generation, then propose a novel DAHC model for the purpose of efficient sensing data set generation. Moreover, the human kinematic model for the motion simulation is also introduced.

#### A. Preliminaries

A wireless channel model for sensing performance evaluation should generate consistent channel impulse response spatially and temporally so that the receiver can capture the micro-Doppler effect via the received signals in a time interval. In fact, the following two kinds of consistency in both spatial and temporal domains have been proposed for communication channel model design [27].

1. Large-scale spatial-temporal consistency refers to consistent power fading, delay spreads and angular spreads at two close locations and time instances.
2. Small-scale spatial-temporal consistency refers to consistent delays and angles of rays at two close locations and time instances.

Besides them, the channel model for wireless sensing should be composed of the rays consistent with the environment and the motions of sensing target. Finally, random interference should also be included to model the unpredictable motions except the sensing target.

As summarize in Table I, current channel models can be categorized into statistical models, deterministic models, and quasi-deterministic models. As an example of the statistical model proposed in [22], the non Line-of-Sight (NLoS) rays between the transmitter and the receiver were generated via scattering clusters. The phases, amplitudes and delays of rays were generated via independent distributions. Lack of the spatial-temporal consistency in small scale, this model cannot simulate the micro-Doppler effects due to the nonrigid human motions.

On the other hand, the quasi-deterministic models could maintain the spatial-temporal consistency in both large and small scales, which are adopted in many existing industrial standards for communication systems [19], [25], [26], [27]. In [25], [26], and [27], the large-scale channel parameters, including delay spread, angular spreads, Ricean K-factor, and so on, were modeled via semi-deterministic methods.

#### Table I: Comparison of Existing and Proposed Channel Models

| Type | Literature | Methodology | Application Scenario | Spatial-Temporal Consistency | Motion Consistency | Micro Doppler | Sensing Uncertainty |
|------|------------|-------------|---------------------|-----------------------------|-------------------|---------------|-------------------|
| S    | [22]       | cluster random process | commun. | L1 | X | X | X |
| D    | [23]       | ray tracing | both | L2 | ✓ | ✓ | X |
|      | [24]       | primitive based | sensing | L2 | ✓ | ✓ | X |
| Q-D  | 3GPP TR 38.901 [25] | GBSM based | commun. | L2 | X | X | X |
|      | QuaDRiGa [26] | GBSM based | commun. | L2 | X | X | X |
|      | IEEE 802.11 ay [19] | Q-D based | commun. | L2 | X | X | X |
|      | METIS [27] | GBSM, map based hybrid | commun. | L2 | X | X | X |
|      | Proposed   | primitive based hybrid | both | L2 | ✓ | ✓ | ✓ |

GBSM means geometry-based stochastic channel model. “S” means statistical, “D” means deterministic, “Q-D” means quasi-deterministic. “L1” means large-scale spatial consistency, “L2” means large-scale and small-scale spatial consistency. “✓” means functionality supported, “X” means functionality not supported.
and shadow fading, were computed via ray tracing and the small-scale channel parameters, including time delays, cluster powers, and arrival/departure angles, were moving, obtained using the statistical method, where the small-scale consistency is maintained via the distribution correlation. For example, the delays and angles of rays were generated based on uniform distributions in [25] where the distribution parameters linearly depend on the correlation distance. In [28], a method to calculate the correlation distance according to the large-scale parameters was proposed. In [19], a channel model consisting of both deterministic and random rays was proposed, where the deterministic rays were generated via ray tracing to maintain the small-scale consistency. However, the modeling of human motions is not considered in the above channel models. For example, the human body was simply treated as a blocker of rays in [19] and [25]. Thus, the micro-Doppler effect of human motion are not characterized in the above models.

The deterministic channel models, such as ray-tracing [23], have potential to capture information of both environment and human target. However, the computation complexity could be huge if the ray tracing method is directly applied on nonrigid human motions. On the other hand, primitive-based channel model [24] is a computationally efficient approximation method for ray tracing, where each dynamic object is modeled as an extended target made of multiple point scatters (primitives) distributed along its body (e.g., a human body usually consists of 16 primitives). The received signal from each primitive is computed using the electromagnetic field method in [29, Sec. 5.8], where the radar cross section (RCS) for simple shapes could be obtained from [30].

As a summary, the quasi-deterministic channel models are able to simulate the rays’ propagation in the sensing environment, and the primitive-based model is able to efficiently simulate the rays from human target. The integration of both models could be used for sensing data set generation. Furthermore, due to the complexity and uncertainty of the real sensing scenario, it is necessary to keep randomness in the channel impulse response, which is referred to as the sensing uncertainty in this article. Specifically, sensing uncertainty could be raised at least by the following three factors.

1) Unpredictable reflections from the wall and random scatterers.
2) Undesired movements of the nontarget objects.
3) Noise.

Examples of spectrograms generated from the real experiment are illustrated in Fig. 3. It can be observed that there are significant blurs aroused by the sensing uncertainty. To our best knowledge, the sensing uncertainty has not been not explicitly considered in the existing channel models.

**B. Channel Modeling**

In order to keep the consistency with the sensing scenario and model the sensing uncertainty, the autoregressive model and quasi-deterministic model are integrated in the proposed DAHC model as illustrated in Fig. 4. Specifically, the channel impulse response from the BS to the kth receiver (the 0-th receiver refers to the co-located BS radar receiver) in the jth time slot can be represented by

\[ h_{k,j}(t) = u_{k,j}(t) + v_{k,j}(t) \]  \( (6) \)

where

1) \( u_{k,j}(t) \) is the target-related channel component consisting of the rays reflecting from the human target.
2) \( v_{k,j}(t) \) is the target-unrelated channel component, consisting of the rays via the LoS path and the other reflection paths. For example, the reflection paths via walls. Due to the self-interference cancelation at the BS radar receiver, we neglect the LoS path in \( v_{0,j}(t) \).

Similar to [24], the target related channel \( u_{k,j}(t) \) is modeled using the following primitive-based method

\[ u_{k,j}(t) = \frac{A}{\sqrt{4\pi}} \sum_{b=1}^{B} \sqrt{G_{b,j}} \exp\left(-\frac{2\pi f_c}{c} 2D_{b,j}\right) \times \exp(j\varphi_{b,j}) \delta(t - \frac{2D_{b,j}}{c}). \] \( (7) \)

In the above equation, \( A \) is the constant related to wave length \( \lambda \) and antenna gain \( P_1 \), which is \( \lambda^2/\sqrt{P_1} \) in the LoS case; \( B \) is the number of primitives; \( G_{b,j} \) is the amplitude accounting for RCS of the bth primitive in the jth time slot; \( D_{b,j} \) is the distance from the bth primitive to the radar in the jth time slot, \( f_c \) is the carrier frequency; \( c \) is the speed of light; and \( \varphi_{b} \) is the initial phase of the bth ray, which follows uniform
distribution in $[−\pi, \pi]$, $\delta(a)$ is the indicator function, whose value is 1 for $a = 0$ and 0 otherwise.

On the other hand, it is not necessary to capture the motions via the primitive-based approach in the target-unrelated channel $v_{k,j}(t)$ as that in target-related channel. Hence, the following autoregressive method is used to model the sensing uncertainty statistically

$$v_{k,j}(t) = \begin{cases} \gamma_1(t), & \text{if } j = 1 \\ \rho v_{k,j-1}(t − T_0) + (1 − \rho)\gamma_j(t), & \text{if } j > 1 \end{cases}$$

where $\gamma_1(t)$ is the target-unrelated channel impulse response of the first slot, and $\rho$ is a hyperparameter controlling the intensity of sensing uncertainty. A larger (or smaller) $\rho$ leads to a weaker (or stronger) sensing uncertainty. In this article, we adopt the quasi-deterministic channel model to generate $\gamma_j(t) \forall j$, thus

$$\gamma(t) = \frac{N}{\lambda} \sum_{n=1}^{N} \sqrt{H_n} \left(\frac{\lambda}{4\pi(D_0 + \tau_n^{\text{cluster}})}\right) \times \sum_{m=1}^{M} a_{n,m} \exp\left(i\phi_{n,m}\right) \delta(t − \tau_{n,m}^{\text{ray}}).$$

In the above expression, $N$ is the number of scattering clusters, $H_n$ is the reflection factor for both first-order and second-order reflections; $\lambda$ is the wave length; $\tau_n^{\text{cluster}}$ is the $n$th cluster’s time delay (in seconds) obtained from ray-tracing; $\tau_{n,m}^{\text{ray}}$, $a_{n,m}$, and $\phi_{n,m}$ are the time delay, amplitude and initial phase of the $m$th ray via the $n$th cluster, which are obtained from Poisson distribution, Rayleigh distribution and uniform distribution, respectively.

C. Data-Assisted Model Calibration

In order to fit the hyperparameter $\rho$ of the DAHC to the real uncertainty level in a particular sensing scenario, it is necessary to make a measurement in the target environment. Denote the received signal in the real scenario and DAHC model simulation as $\{r_j(t)\mod(N_m + 1) = 1, 1 \leq j \leq N_f(N_m + 1)\}$ and $\{\hat{r}_j(t)\mod(N_m + 1) = 1, 1 \leq j \leq N_f(N_m + 1)\}$, respectively. The adaptation of hyperparameter to real uncertainty level is elaborated below.

The spectrograms of both measured and simulated signals are first obtained according to Section IV-A. The Doppler frequency strength versus time and frequency of both spectrograms are quantized into $F$ levels. The probability mass function (PMF) of Doppler frequency strength for both spectrograms can then be obtained, denoted as $\Psi$ and $\Xi$, respectively. Finally, the proper hyperparameter $\rho$ is the one minimizing the Kullback–Leibler (KL) divergence as follows:

$$\min_{\rho} \sum_{j=1}^{F} \Psi(z_j) \left[ \log\left( \frac{\Psi(z_j)}{\Xi(z_j|\rho)} \right) \right]$$

s.t. $0 \leq \rho \leq 1$

(10)

where $z_j$ denotes the signal strength of the $F$th level. The above problem can be solved by 1-D search. In practice, $\rho$ varies in different scenarios and we can store the values of $\rho$ for typical scenarios in a look-up table.

D. Human Kinematic Modeling

The human target is represented by 16 primitives in the proposed DAHC model. In order to model the motion of the 16 primitives, a global human model, namely, the Thalmann model [31], is adopted. In the Thalmann model, the human body is represented as a series of 16 segments, which corresponds to the primitives of the proposed channel model. Based on the biomechanical experimental data, the motions of all the segments, including their positions and orientations of all time slots, are obtained for different categories of human motions.

However, this global human model averages out the personification in the same category of motion, losing the diversity in motion’s data set generation. In the future work, the
motion capture methods from the areas of Graphics [32] could be exploited to simulate the primitives’ motion with personification.

IV. SDP3 PERFORMANCE PREDICTOR AND OPTIMIZER

In this section, we present the mechanisms of the proposed SDP3 performance predictor and optimizer.

A. SDP3 Performance Predictor

The SDP3 performance predictor is trained to predict the sensing performance versus the sensing resource. Specifically, a DSN is trained with the SDP3 data simulator generated data set for motion recognition. Afterward, the approximate expression formula of the motion recognition accuracy versus the number of sensing slots \( N_s \) is derived. The detailed procedures are elaborated as follows.

1) Deep Spectrogram Network: Let \( M \) and \( C = \{1, \ldots, C\} \) be the number of human motion categories and the set of human motion categories, respectively, the input of DSN is the signals \( \{r_j(t)\} \) mod \( (N_m + 1) = 1, 1 \leq j \leq N_s(N_m + 1) \) and the output is the index of human motion category \( c \in C \). In DSN, the spectrogram of input signals is first generated via data cleaning and transformation, followed by model training (in the training phase) or inference (in the inference phase), as shown in Fig. 2.

Data Cleaning: In this step, the received signal \( r_j(t) \) within the sweep duration \( t \in [0, T_w] \) is sampled with a frequency \( f_s \). Let \( x_j = [r_1(f_s), r_2(f_s), \ldots, r_{T_w}(f_s)]^T \in \mathbb{C}^L \) be the vector of samples in the \( j \)th slot, where \( L = T_w f_s \), and \( X = [x_1, x_2, \ldots, x_{N_s}] \in \mathbb{C}^{L \times N_s} \) be the aggregation of sample vectors of all sensing slots. The matrix \( \tilde{X} \) is a superposed signal consisting of the desired signals reflected from the target and undesired signals reflected from walls (or other objects). To extract the desired information, we first decipher the sampled signal \( \tilde{X} \) as follows:

\[
\tilde{X} = \left( s \circ x_1 \circ \hat{s} \circ x_2 \circ \cdots \circ \hat{s} \circ x_{N_s} \right)^* \tag{11}
\]

where \( \hat{s} = [s(1/f_s), s(2/f_s), \ldots, s(T_w)]^T \), \( x_i^* \) denotes the conjugate of \( x_i \), and \( \circ \) denotes the Hadamard product. Then the singular value decomposition (SVD) is applied to \( \tilde{X} \), yielding \( \tilde{X} = \sum_{i=1}^{D} a_i u_i v_i^H \), where \( D \) is the rank of \( \tilde{X} \), \( a_i \) is the \( i \)th largest singular value, \( u_i \) and \( v_i \) are the \( i \)th left-singular vector and right-singular vector, respectively. Removing the first \( d - 1 \) components, which represent the undesired signal paths, the denoised signal matrix is

\[
Y = \sum_{i=d}^{D} a_i u_i v_i^H. \tag{12}
\]

Data Transformation: In this step, the STFT [33] is applied on \( Y \). We first define the sliding window function as

\[
w_b[n] = \frac{i_0(\beta \sqrt{1 - \left(\frac{n - (N_w - 1)/2}{(N_w - 1)/2}\right)^2})}{i_0(\beta)}, 0 \leq n \leq N_w - 1
\]

where \( i_0 \) is the zeroth-order modified Bessel function of the first kind and \( N_w \) is the length of sliding window. Let \( y = 1^T Y \),

\[
\begin{align*}
z[k, f] &= \sum_{k' = -\infty}^{+\infty} y[k'] w_b[k' - k] \exp(-j 2\pi k' f / N) \\
k &\in \left\{ 0, \ldots, \left\lfloor \frac{N_s - N_w}{N_w - N_{ol}} \right\rfloor \right\} \ast (N_w - N_{ol}) \\
f &\in \left\{ \frac{N_{ff}}{2} + 1, \ldots, \frac{N_{ff}}{2} \right\} \tag{13}
\end{align*}
\]

where \( N_{ol} \) and \( N_{ff} \) specify the number of overlap samples between adjoining STFT windows and the number of frequency points, respectively. Then the spectrogram of sensing signal \( x_{i,j} \) mod \( (N_m + 1) = 1, 1 \leq j \leq N \) could be illustrated via \( z[k, f] \).

Model Training and Inference: To classify the motions, the ResNet-32 [34] is adopted as the backbone for the feature extraction. It consists of one input layer, five identical residual blocks and one output layer as illustrated in Fig. 2. The input layer consists of a convolution layer and a pooling layer, and the output layer consists of a global average pooling layer and a fully connected layer with softmax as the activation function. Each residual block consists of six layers: 1) a batch normalization layer; 2) a ReLu activation layer; 3) a convolution layer; 4) a batch normalization layer; 5) a ReLu activation layer; and 6) a convolution layer. The input of the ResNet is the spectrogram, and the output is the index of the estimated human motion category \( \hat{c} \).

2) Motion Recognition Accuracy Model: It is difficult to directly derive the analytical relationship between the motion recognition accuracy and the number of sensing slots, which is denoted as \( A = \Theta(N_s) \). This is because there is no analytical expression to quantify the learning performance of ResNet–32.

To address the above challenge, a promising solution is the curve fitting approach proposed in [35], [36], [37], and [38]. Specifically, it is observed that the recognition accuracy \( A = \Theta(N_s) \) is a nonlinear function of \( N_s \) satisfying the following properties.

1) \( \Theta(N_s) \) is a monotonically increasing function of \( N_s \).
2) As \( N_s \) increases, the magnitude of the partial derivative \( |\partial \Theta / \partial N_s| \) would gradually decrease and become zero when \( N_s \) is sufficiently large, meaning that increasing the number of sensing time slots will not help wireless sensing performance at large \( N_s \).

Hence, the candidate parametric learning curve models to approximate the \( \Theta(N_s) \) are shown in Table II (as proposed in [37]), where \( \alpha, \beta, \gamma, \epsilon \) are tuning parameters.

In order to determine the tuning parameters, we first generate the data set of sensing signals \( \{r_j(t)\} \) mod \( (N_m + 1) = 1, 1 \leq j \leq N_s(N_m + 1) \) for \( Q \) different numbers of sensing slots, denoted as \( \{N_s^{(i)}\} 1 \leq i \leq Q \). After the training and inferencing with DSN, the corresponding recognition accuracies are denoted as \( A^{(i)} 1 \leq i \leq Q \). Then the parameters \( \alpha, \beta, \gamma, \epsilon \) can be calculated via the following least squares fitting

\[
\min_{\alpha, \beta, \gamma, \epsilon} \frac{1}{Q} \sum_{i=1}^{Q} \left| \Theta(N_s^{(i)}) - A^{(i)} \right|^2. \tag{14}
\]
The above problem can be solved by brute-force search, or gradient descent method. Notice that the parameters $\alpha, \beta, \gamma,$ and $\epsilon$ can be fine-tuned in unseen scenarios via online learning methods [39], [40].

**Remark 1:** Apart from the sensing duration (the number of sensing slots $Ns$), the recognition accuracy $A$ is also influenced by the sensing quality (e.g., the sensing SNR, MSE, or SINR measuring the intensity of interference and noise). Nevertheless, in the high-SINR regime (the focus of this article), such influence is negligible as shown in [41], and [42]. On the other hand, in the low-SINR regime, the impact of the sensing quality can be readily integrated into the proposed model by leveraging the experimental data in [41], and [42] and multiplying a discounted weight.

**B. SDP3 Performance Optimizer**

The SDP3 performance optimizer is adopted to investigate the accuracy-throughput (A-T) region of the ISAC system. Denote $N_{c,k}$ as the number of time slots assigned for the $k$th receiver, the throughput of $k$th receiver in the whole scheduling period $T_k$ can be approximated as

$$T_k \approx \frac{N_{c,k} T_s}{T_o} \sum_{m=1}^{M} \log_2(1 + y_{k,m})$$

(15)

where $y_{k,m} = E[y_{k,j,m}] = [E(|H_{k,m}|^2 p_{k,j,m})]/\sigma^2$. In (15), we use $y_{k,m}$ to approximate the instantaneous SNR. This is because the Doppler offset raised by human motion does not dominate the channel gain. Hence, the worst communication throughput among all receivers is

$$T = \min_{k=1,...,K} \frac{N_{c,k} T_s}{T_o} \sum_{m=1}^{M} \log_2(1 + y_{k,m}).$$

(16)

To characterize the performance region, the A-T region $\mathcal{R}(N_s, N_c)$ is defined first as follows:

$$\mathcal{R}(N_s, N_c) = \{(A, T) \in \mathbb{R}_+^2 : N_s + \sum_{k=1}^{K} N_{c,k} = \frac{N}{K} \}$$

(17)

since $\mathcal{R}(N_s, N_c)$ is a solid region which could be determined by its boundary points. That is, seeking the Pareto optimal solutions that maximize the performance pair $\mathcal{R}(A, T)$. To achieve this, a commonly used method is combining the multiple objectives through the weighted sum method with a performance weight vector $[w_A, w_T]$ ($w_A + w_T = 1$) as follows:

$$f(w_A, w_T) = w_A A + w_T T$$

(18)

where $w_A$ and $w_T$ are the weights. Given a pair of weight $(w_A, w_T)$, the optimal $A^*$ and $T^*$ maximizing the objective $f(w_A, w_T)$ could be obtained via the following optimization problem

$$P: \begin{align*}
A^* \in \mathcal{R}(N_s, N_c), & \quad f(w_A, w_T) \\
A = \Theta(N_s), & \quad \sum_{m=1}^{M} P_{k,m} \leq P \forall k \\
N_s + \sum_{k=1}^{K} N_{c,k} = N. & \quad (19)
\end{align*}$$

Hence, the achievable A-T region can be characterized by $\{A, T) = \arg \max f(w_A, w_T)|w_A \geq 0, w_T \geq 0\}$. Moreover, the closed-form relationship between the recognition accuracy and communication throughput on the Pareto boundary is elaborated in the following theorem.

**Theorem 1:** Denote $\Theta^{-1}$ as the inverse function of $\Theta$, let $(A^*, R^*)$ be a point on the Pareto boundary, then

$$T_s \Theta^{-1}(A^*) + \sum_{k=1, k \neq 0}^{K} \frac{T_o}{\log_2 \left( \frac{y_{k,m}}{\gamma_{k,0}} \right)} \leq N T_s.$$

(22)

**Proof:** Please refer to Appendix A.

**V. SIMULATION AND EXPERIMENT**

In this section, we verify the proposed SDP3 framework via both simulation and experiment. Specifically, a conference room with size (3, 4.5, 3 m) (i.e., length, width, and
Fig. 5. Human motion data sets obtained by sensing an adult target in a conference room, where the carrier frequencies are 3.5 GHz (upper side) and 60 GHz (lower side), respectively. The scenarios with carrier frequencies 3.5 and 60 GHz are referred to as the scenario 1 and 2, respectively. The height of the adult is 1.75 m. The following three motions are tested: 1) standing; 2) pacing; and 3) walking. The FMCW is with 10 MHz bandwidth and 100 μs sweep time.

The Scenario 1 is sensed on both 3.5 and 60-GHz band, and the Scenario 2 is sensed on 60-GHz band only. In fact, the Scenario 2 is common in practice. It refers to the situation with stronger sensing uncertainty. In the simulation data set, we generate the samples of received sensing signal via both SDP3 data simulator and ray-tracing model, such that their performances can be compared. In both approaches, the adult and child have \( B = 16 \) primitives, respectively. In the SDP3 data simulator, the IEEE 802.11ax/ay channels \([19], [43]\) are adopted to generate the component \( v_{k,j}(t) \) in (6). In both simulation and experiment data sets, 200 samples are generated for each motion.

A. KL Divergence

In this part, the calibration of hyperparameter \( \rho \) for the motion of adult walking, as elaborated in Section III, is first demonstrated, which verifies the existence of sensing uncertainty.

Specifically, the KL divergence between the experimental and SDP3-simulated data samples versus the hyperparameter \( \rho \) is shown in Fig. 6(a)–(c) for different carrier frequencies and scenarios. It can be seen that the KL divergence is sensitive to the value of \( \rho \) of the proposed DAHC model, and the optimal values minimizing the KL divergence are \( \rho^* = 0.997, 0.997 \) and 0.996 for Scenario 1 on 3.5-GHz band, Scenario 1 on 60-GHz band and Scenario 2 on 60-GHz band, respectively. Note that the KL divergence measures the
distance between two distributions, the minimum KL divergence means the best match between the experiment and the SDP3 data simulator. Note that the samples in Fig. 6(b) and (c) are obtained with the same carrier frequency but different scenarios. This demonstrates the interference from nontarget person: it leads to smaller value of calibrated hyperparameter \( \rho \), thus stronger level of sensing uncertainty. In all the three figures Fig. 6(a)–(c), it can be observed that the KL divergence between the experiment and the ray-tracing model is constant and significantly larger than the calibrated DAHC model. This demonstrates that it is necessary to include the sensing uncertainty in the channel model to capture the potential interference from nontarget motions.

In order to further justify the superior performance of the proposed DAHC model in sensing data generation, the grayscale distribution of spectrogram is compared in Fig. 6(d)–(f). Specifically, the spectrograms from experiment, ray-tracing model and DAHC model are generated for different carrier frequencies and scenarios. Then, the PMF of grayscale levels is calculated for each spectrogram. It can be observed that the grayscale PMFs from the experiment and proposed DHAC model match very well, and ray-tracing model fails to match the real experiment. This coincides with the comparison of KL divergence in Fig. 6(a)–(c).

**B. Recognition Accuracy**

In this part, we continue to show that the sensing data set generated via SDP3 data simulator with the DAHC model could achieve the recognition accuracy close to the experiment data set. The hyperparameter \( \rho \) calibrated in the above part is adopted. The simulation data set via ray-tracing model is also investigated as a comparison.

Particularly, 600 samples of sensing signals are picked up from experiment, proposed SDP3 data simulator, ray-tracing-based simulator, respectively. Their spectrograms are used to train the DSN in Section IV-A1, respectively. The training of DSN is implemented via Momentum optimizer with a learning rate of 0.06 and a mini-batch size of 300. The recognition accuracies of the trained DSNs are tested by other samples of the data sets. As a result, the recognition accuracy versus training epoch is illustrated in Fig. 7. It is observed that compared with the ray-tracing-based simulator, the curve of the SDP3 data simulator is closer to that of experiment. For example, it is shown that the accuracy gap are 2% versus 5.33% at 3.5 GHz and 4% versus 6.7% at 60 GHz, respectively. This demonstrates that the DAHC model can simulate the motion recognition performance in real experiment better than the ray-tracing model. Hence, the data set generated by the SDP3 data simulator could be used for the analysis of sensing-communication performance tradeoff. Then the significant effort of extensive real scenario experiment can be saved.

**C. Verification of SDP3 Performance Predictor**

The performance of the SDP3 performance predictor, as proposed in Section IV-A2, is demonstrated in this part, where the motion recognition accuracy versus the number of sensing slots \( N_s \) is approximated analytically. We adopt the samples of received sensing signals generated by SDP3 data simulator. Both 3.5 and 60-GHz bands are considered and five different motions, including child/adult standing, child walking,
child pacing, adult walking and adult pacing, are classified. The samples in the data set are generated with $N_s = 1000$. The received signals of the first 200, 300, 400, 500, 600, and 1000 sensing time slots are used for DSN training and accuracy testing, respectively, yielding the motion recognition accuracies for different values of $N_s$. Specifically, the training procedure is implemented via Adam optimizer with a learning rate of 0.01 and a mini-batch size of 500. The convergence of the training loss for all the values of $N_s$ is shown in Fig. 8(a). Then the trained models are tested by 500 new samples, respectively, and the corresponding recognition accuracies can be obtained as shown in Fig. 8(b). It can be observed that larger $N_s$ leads to better recognition accuracy. Hence, the curving fitting elaborated in Section IV-A can be proceeded. The optimized tuning parameters of all the curve models are listed in Table III, and their comparison with the simulated accuracies is illustrated in Fig. 8(c). It can be observed that the models of $\text{pow}_3$, $\text{log power}$ and $\text{exp}_4$ match the simulated accuracy very well. This is because they possess higher curvature compared with other models. With the best fitting model $\text{pow}_3$, the tuning parameters $\alpha$, $\beta$, and $\gamma$ are obtained by minimizing the MSE as in (14), yielding $(\alpha, \beta, \gamma) = (61906, 2.4297, 0.9499)$. Thus, the relation between the recognition accuracy and number of sensing slots is approximated as

$$A = 0.9499 - 61906 \times N_s^{-2.4297}. \quad (23)$$
D. Sensing-Communication Tradeoff

Finally, the SDP3 performance optimizer is utilized to illustrate the tradeoff between sensing and communication performance. The result with one sensing target and $K=3$ communication receivers is shown in Fig. 9, where the carrier frequency is 3.5 GHz. Two distributions of receivers are considered: the communication receivers are at the locations (1.5, 3, 1), (2.5, 1, 1), and (0.5, 1, 1) for the case 1, and at (1.5, 2, 1), (2.3, 1, 1), and (0.7, 1, 1) for the case 2.

It can be observed that the accuracy-throughput (A-T) region of case 1 is a subset of that of case 2. This is because of the better communication channel in case 2. Both of two regions consist of three zones: 1) sensing saturation zone (bottom); 2) communication saturation zone (left); and 3) sensing-communication adversarial zone (upper right). In the sensing saturation zone, reducing the communication performance can hardly improve the sensing performance, but a slight decrease of the sensing performance can significantly improve the communication throughput. This is because excessive wireless resources have been allocated to sensing task. The situation is the opposite in the communication saturation zone. Therefore, for a practical ISAC system, it is not desirable to enter the sensing or communication saturation zone. On the other hand, both sensing and communication performance varies sensitively with respect to each other in the sensing-communication adversarial zone. This is because the wireless resource allocation between the two functionalities is balanced. The proposed accuracy-throughput region analysis facilitates the search of the best Pareto point with heterogeneous quality-of-service requirements.

VI. Conclusion

This article proposes an SDP3 framework for the study of sensing-communication tradeoff with the particular application of human motion recognition. Specifically, the SDP3 data simulator with a data-driven hybrid channel model is proposed to generate the received sensing signals in a virtual environment. The SDP3 performance predictor is then introduced to approximate the motion recognition accuracy via analytical expression with the simulated data set of sensing signals. Finally, the recognition accuracy and communication throughput tradeoff is characterized by the SDP3 performance optimizer. It is demonstrated that the data set generated by the SDP3 data simulator matches the experiment data set in KL divergence, grayscale PMF and motion recognition accuracy. Hence, the sensing-communication tradeoff can be investigated without extensive experiments. It is also shown that the sensing and communication performance is balanced in the sensing-communication adversarial zone of the A-T region, where both performance varies sensitively with respect to each other.

APPENDIX A

PROOF OF THEOREM 1

Given that $A = A^*$ with $N_s = N_s^*$, the multiobjective optimization problem $P$ could be converted as follows:

$$\mathcal{P}_1 : \max_{N_s} \quad T $$

s.t. \hfill (16), (19), (20), (21) \hfill (24)$$

moreover, applying the water-filling policy to the power constraints, the optimal worst communication throughput among all receivers is

$$T = \min_{k=1,...,K} \frac{N_{c,k}T_s}{T_0} \sum_{m=1}^{N} \log_2 \left( \frac{\gamma_{k,m}}{\bar{\gamma}_{k,0}} \right)$$

where

$$\sum_{m=1}^{M} \left( \frac{1}{\bar{\gamma}_{k,0}} - \frac{1}{\gamma_{k,m}} \right) = 1$$

and the $\mathcal{P}_1$ could be

$$\mathcal{P}_1 : \max_{N_s} \quad T $$

s.t. \hfill (19), (20), (21), (25). \hfill (27)$$

Then the optimal $t$ is derived using KKT optimality conditions. Specifically, the Lagrangian of problem $\mathcal{P}_1$ is given by

$$L = -T^* + \sum_{k=1}^{K} \eta_k \left( T^* - \frac{N_{s,k}T_s}{T_0} \sum_{m=1}^{N} \log_2 \left( \frac{\gamma_{k,m}}{\bar{\gamma}_{k,0}} \right) \right) + \varrho \left( N_s^* + \sum_{k=1}^{K} N_{c,k} - N \right)$$

where $\{\eta_k \geq 0, \varrho\}$ are Lagrangian multipliers. By letting $\partial L/\partial R^* = 0$ and $\partial L/\partial N_{c,k}^* = 0$, we have

$$\sum_{k=1}^{K} \eta_k = 1, \quad \eta_k = 0$$

Now we will prove that $\eta_k \neq 0$ for any $k$ by contradiction. In particular, assume that $\eta_j = 0$ for some $j$. Putting $\eta_j = 0$ into the second equation of (29) yields $\varrho = 0$. Putting $\varrho = 0$ into the second equation of (29) with $k \neq j$ yields $\eta_k = 0$ for
any \( k \neq j \). Finally, based on \( \eta_j = 0 \) and \( \eta_k = 0 \) for \( k \neq j \), we have \( \sum_{k=1}^{K} \eta_k = 0 \). This contradicts to \( \sum_{k=1}^{K} \eta_k = 1 \) of the first equation of (29). Therefore, \( \eta_k \neq 0 \) for any \( k \). Using the above result and the complementary slackness condition

\[
\eta_k \left( T^* - \frac{N^*_{c,k} T_k}{T_0} \right) = \sum_{m=1}^{N} \log_2 \left( \frac{\gamma_{k,m}}{\gamma_{k,0}} \right) = 0 \quad (30)
\]

the following equality is obtained

\[
N^*_{c,1} = \sum_{m=1}^{N} \log_2 \left( \frac{\gamma_{1,m}}{\gamma_{1,0}} \right) = \cdots
\]

\[
N^*_{c,K} = \sum_{m=1}^{N} \log_2 \left( \frac{\gamma_{K,m}}{\gamma_{K,0}} \right) . \quad (31)
\]

Combining the above result with the constraint \( N^*_s = \sum_{k=1}^{K} N^*_{c,k} = N \) yields

\[
N^*_k = \left( \sum_{k=1}^{K} \sum_{m=1}^{N} \log_2 \left( \frac{\gamma_{k,m}}{\gamma_{k,0}} \right) \right) - N^*_s T_k / T_0 = \left( N - N^*_s \right) T_k / T_0 . \quad (32)
\]

According to (25)

\[
A^* = \Theta(N^*_s) \quad (33)
\]

\[
T^* = \frac{\left( N - N^*_s \right) T_k}{\sum_{k=1}^{K} T_k} = \frac{\left( N - N^*_s \right) T_k}{\sum_{k=1}^{K} \sum_{m=1}^{N} \log_2 \left( \frac{\gamma_{k,m}}{\gamma_{k,0}} \right)} . \quad (34)
\]

Rearranging (33)–(34), the proof is completed.

REFERENCES

[1] G. Li, S. Wang, J. Li, R. Wang, X. Peng, and T. X. Han, “Wireless sensing with deep spectrogram network and primitive based autoregressive hybrid channel model,” in Proc. SPAWC, Lucca, Italy, Sep. 2021, pp. 481–485.

[2] X. Li, Y. Cui, J. A. Zhang, F. Liu, D. Zhang, and L. Hanzo, “Integrated human activity sensing and communications,” IEEE Commun. Mag., vol. 61, no. 5, pp. 90–96, May 2023.

[3] F. Liu et al., “Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond,” IEEE J. Sel. Areas Commun., vol. 40, no. 6, pp. 1728–1767, Jun. 2022.

[4] N. Huang, C. Dou, Y. Wu, L. Qian, B. Lin, and H. Zhou, “Unmanned aerial vehicle aided integrated sensing and computation with mobile edge computing,” IEEE Internet Things J., early access, Apr. 25, 2023, doi: 10.1109/JIOT.2023.3270332.

[5] B. K. Chalise, M. G. Amin, and B. Himed, “Performance tradeoff in a unified passive radar and communications system,” IEEE Signal Process. Lett., vol. 24, no. 9, pp. 1275–1279, Sep. 2017.

[6] B. K. Chalise and B. Himed, “Performance tradeoff in a unified multi-static passive radar and communication system,” in Proc. (RadarConf), Oklahoma City, OK, USA, Apr. 2018, pp. 653–658.

[7] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, “Joint transmit beamforming for multilaterer MIMO communications and MIMO radar,” IEEE Trans. Signal Process., vol. 68, pp. 3929–3944, 2020.

[8] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, “Cramér-Rao bound optimization for joint radar-communication beamforming,” IEEE Trans. Signal Process., vol. 70, pp. 240–253, Dec. 2021.

[9] D. W. Bliss, “Cognitive radar and communications signaling: The estimation and information theory odd couple,” in Proc. (RadarConf), Cincinnati, OH, USA, May 2014, pp. 50–55.

[10] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, “Inner bounds on performance of radar and communications co-existence,” IEEE Trans. Signal Process., vol. 64, no. 2, pp. 464–474, Jan. 2016.

[11] F. Wang, W. Gong, and J. Liu, “On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach,” IEEE Internet Things J., vol. 6, no. 2, pp. 2035–2047, Apr. 2019.

[12] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, “Multimodal CSI-based human activity recognition using GANs,” IEEE Internet Things J., vol. 8, no. 24, pp. 17345–17355, Dec. 2021.

[13] K. Niu, F. Zhang, X. Wang, Q. Lv, H. Luo, and D. Zhang, “Understanding WiFi signal frequency features for position-independent gesture recognition,” IEEE Trans. Mobile Comput., vol. 21, no. 11, pp. 4156–4171, Nov. 2022, doi: 10.1109/TMC.2021.3063135.

[14] Y. Zhang et al., “Wiadar:0: Zero-effort cross-domain gesture recognition with Wi-Fi,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 11, pp. 8671–8688, Nov. 2022, doi: 10.1109/TPAMI.2021.3105387.

[15] C. Lin, J. Hu, Y. Sun, F. Ma, L. Wang, and G. Wu, “WiAU: An accurate device-free authentication system with ResNet,” in Proc. 15th Annu. IEEE Int. Conf. Sens. Commun. Netw., Hong Kong, China, 2018, pp. 1–9.

[16] L. Zhang, C. Wang, and D. Zhang, “Wi-PIGR: Path independent gait recognition with commodity Wi-Fi,” IEEE Trans. Mobile Comput., vol. 21, no. 9, pp. 3414–3427, Sep. 2022.

[17] Y. Kim and H. Ling, “Human activity classification based on micro-Doppler signatures using a support vector machine,” IEEE Trans. Geosci. Remote Sens., vol. 47, no. 5, pp. 1328–1337, May 2009.

[18] Y. Kim and T. Moon, “Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 1, pp. 8–12, Jan. 2016.

[19] A. Maltsi et al., “Channel models for IEEE 802.11ay,” document 802.11-15/1150/9, IEEE, New York, NY, USA, 2016.

[20] A. Rosati, G. E. Corazza, and A. Vanelli-Coralli, “OFDM channel estimation based on impulse response decimation: Analysis and novel algorithms,” IEEE Trans. Commun., vol. 60, no. 7, pp. 996–2008, Jul. 2012.

[21] M. Zhang et al., “Channel models for WLAN sensing systems,” documents IEEE 802.11, IEEE Piscataway, NJ, USA, Sep. 2021. [Online]. Available: https://mentor.ieee.org/802.11/documents?is_dcn=Meihong

[22] S. Wang, Y.-C. Wu, M. Xia, R. Wang, and H. V. Poor, “Machine intelligence based on impulse response decay: Analysis and novel algorithms,” IEEE Trans. Commun., vol. 50, no. 2, pp. 18–27, Feb. 1997.

[23] S. Y. Seidel and T. S. Rappaport, “Site-specific propagation prediction for wireless in-building personal communication system design,” IEEE Trans. Veh. Technol., vol. 43, no. 4, pp. 879–891, Nov. 1994.

[24] P. van Dorp and F. C. A. Groen, “Human walking estimation with radar,” IEEE Proc. Radar Sonar Navig., vol. 150, no. 5, pp. 356–365, Oct. 2003.

[25] “Study on channel model for frequencies from 0.5 to 100 GHz,” JGPP, Sophia Antipolis, France, Rep. 38.901, V16.1.0, Dec. 2019.

[26] S. J. Orfanidis, “Introduction to Signal Processing,” 2nd ed., New Brunswick, NJ, USA: Prentice Hall Press, 1989.
[36] L. Zhou et al., “Learning centric wireless resource allocation for edge computing: Algorithm and experiment,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1035–1040, Jan. 2021.

[37] T. Domhan, J. T. Springenberg, and F. Hutter, “Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves,” in *Proc. IJCAI*, Buenos Aires, Argentina, Jul. 2015, pp. 3460–3468.

[38] M. Johnson, P. Anderson, M. Dras, and M. Steedman, “Predicting classification error on large datasets from smaller pilot data,” in *Proc. ACL*, Melbourne, VIC, Australia, Jul. 2018, pp. 450–455.

[39] Y. Li, B. Liang, and A. Tizghadam, “Robust online learning against malicious manipulation with application to network flow classification,” in *Proc. INFOCOM*, Vancouver, BC, Canada, 2021, pp. 1–10.

[40] Y. Li, B. Liang, and A. Tizghadam, “Robust online learning against malicious manipulation and feedback delay with application to network flow classification,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2648–2663, Aug. 2021.

[41] G. Li, S. Wang, K. Ye, M. Wen, D. W. K. Ng, and M. D. Renzo, “Multi-point integrated sensing and communication: Fusion model and functionality selection,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 12, pp. 2660–2664, Dec. 2022.

[42] P. Liu et al., “Toward ambient intelligence: Federated edge learning with task-oriented sensing, computation, and communication integration,” *IEEE J. Sel. Signal. Process.*, vol. 17, no. 1, pp. 158–172, Jan. 2023.

[43] J. Liu et al., *IEEE 802.11ax Channel Model Document*, IEEE Standard 802.11-14/0882r4, Sep. 2014.

Guoliang Li (Graduate Student Member, IEEE) received the B.Eng. and M.Phil. degree (summa cum laude) from the Southern University of Science and Technology, Shenzhen, China, in 2020 and 2023, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer and Information Science, University of Macau, Macau, China.

His research interests include autonomous driving, mobile robotics, and sensor simulation.

Mr. Li was the recipient of the Best Student Paper Award from the IEEE ICDCS’2023 Workshop on SocialMeta.

Shuai Wang (Member, IEEE) received the Ph.D. degree in electrical and electronic engineering from The University of Hong Kong (HKU), Hong Kong, in 2018.

From 2018 to 2021, he was a Postdoctoral Fellow with HKU and then a Research Assistant Professor with the Southern University of Science and Technology, Shenzhen, China. He is currently an Associate Professor with Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen. He has published more than 40 journal papers and more than 20 conference papers. His research interests include autonomous driving, mobile robotics, and wireless communications.

Dr. Wang has received three Best Paper Awards from IEEE ICC/SPCC/ICDCS, one Excellent Young Scientist Award from IEEE ICCC, and two Exemplary Reviewer Awards from IEEE TWC/WCL.

Jie Li received the B.Eng. and M.Eng. degrees from the Southern University of Science and Technology, Shenzhen, China, in 2019 and 2022, respectively.

His research interests include passive sensing, and integrated sensing and communication.

Rui Wang (Member, IEEE) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree in wireless communications from The Hong Kong University of Science and Technology, Hong Kong, in 2008.

From 2009 to 2012, he was a Senior Research Engineer with Huawei Technologies Company Ltd., Shenzhen, China. Since 2012, he has joined the Southern University of Science and Technology of China, Shenzhen, as an Associate Professor. He has research experience in both academia and industry. He also involved in the development of interference mitigation technology for 5G systems. He has contributed over 20 U.S. patent applications and over 30 Chinese patent applications (20 of them have been granted). He has authored over 70 papers in top-level IEEE journals and flagship international conferences, especially in the area of wireless radio resource optimization and interference management.

Fan Liu (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees from Beijing Institute of Technology, Beijing, China, in 2013 and 2018, respectively.

He is currently an Assistant Professor with the Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen, China. He has previously held academic positions with University College London, London, U.K., first as a Visiting Researcher from 2016 to 2018, and then as a Marie Curie Research Fellow from 2018 to 2020. His research interests lie in the general area of signal processing and wireless communications, and in particular in the area of integrated sensing and communications.

Dr. Liu has 10 papers selected as IEEE ComSoc Best Readings. He was listed in the World’s Top 2% Scientists by Stanford University for citation impact in 2021 and 2022. He was the recipient of the 2023 IEEE Communications Society Stephan O. Rice Prize, the 2023 IEEE ICC Best Paper Award, the 2023 IEEE/CIC ICC 2023 Best Paper Award, the 2022 First Prize of Science and Technology Progress of China Institute of Communications, the 2021 IEEE Signal Processing Society Young Author Best Paper Award, the 2019 Best Ph.D. Thesis Award of Chinese Institute of Electronics, the 2018 EU Marie Curie Individual Fellowship, and has been named as an Exemplary Reviewer for IEEE TWC/TCOM/COMML for five times. He is the Founding Academic Chair of the IEEE ComSoc ISAC Emerging Technology Initiative; an Associate Editor of the IEEE OPEN JOURNAL OF SIGNAL PROCESSING and the IEEE COMMUNICATIONS LETTERS; the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS; and a Guest Editor of the IEEE WIRELESS COMMUNICATIONS and IEEE Vehicular Technology Magazine. He has served as the Organizer and Co-Chair for several workshops, special sessions, and tutorials in flagship IEEE conferences, including ICC, GLOBECOM, ICASSP, and MobiCom. He is a TPC Co-Chair of the 2nd–4th IEEE Joint Communication and Sensing Symposium, and a Track Co-Chair for the IEEE WCNC 2024. He is a member of the IMT-2030 (6G) ISAC Task Group.
Xiaohui Peng (Member, IEEE) received the Ph.D. degree in electronic science and technology from Huazhong University of Science and Technology, Wuhan, China, in 2018. He is a Principal Engineer with Huawei Technologies Company Ltd., Shenzhen, China. His research interests include array signal processing and wireless communication, with special emphasis on integrated sensing and communication.

Tony Xiao Han (Senior Member, IEEE) received the B.E. degree in electrical engineering from Sichuan University, Chengdu, China, in 2008, and the Ph.D. degree in communications engineering from Zhejiang University, Hangzhou, China, in 2013. He is currently a Research Expert and Project Manager with Huawei Technologies Company Ltd., Shenzhen, China. He was a Postdoctoral Research Fellow with the National University of Singapore, Singapore. His research interests include wireless communications, signal processing, integrated sensing and communication, and standardization of wireless communication.

Chengzhong Xu (Fellow, IEEE) received the Ph.D. degree from The University of Hong Kong, Hong Kong, in 1993. He is currently a Chair Professor of Computer Science with the University of Macau, Macau, China. Prior to that, he was on the faculty of Wayne State University, Detroit, MI, USA, and Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. He published two research monographs and more than 14,000 citations. He was also a co-inventor of more than 120 patents and a Co-Founder of Shenzhen Institute of Baidou Applied Technology. His recent research interests are in cloud and distributed computing, systems support for AI, smart city, and autonomous driving.

Dr. Xu was a Best Paper Awardee or Nominee of conferences, including HPCA’2013, HPDC’2013, Cluster’2015, ICPP’2015, GPC’2018, UIC’2018, and SoCC’2021. He was a recipient of the Faculty Research Award, the Career Development Chair Award, and the President’s Award for Excellence in Teaching of WSU. He was also a recipient of the “Outstanding Oversea Scholar” Award of NSFC. He serves or served on a number of journal editorial boards, including IEEE TRANSACTIONS ON COMPUTERS, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, Journal of Parallel and Distributed Computing, Science China Information Science, and ZTE Communication. He was the Chair of IEEE Technical Committee on Distributed Processing from 2015 to 2020.