Deep AI Enabled Ubiquitous Wireless Sensing: A Survey

CHENNING LI[1], ZHICHAO CAO[1], and YUNHAO LIU[1,2], [1]Michigan State University, USA and [2]Tsinghua University, China

With the development of the Internet of Things (IoT), kinds of wireless signals (e.g., Wi-Fi, LoRa, RFID) are full of our living and working spaces nowadays. Beyond communication, wireless signals can sense the status of surrounding objects, known as wireless sensing, with their reflection, scattering, and refraction while propagating in space. In the last past decade, many sophisticated wireless sensing techniques and systems are widely studied for various applications (e.g., gesture recognition, localization, object imaging). Recently, deep Artificial Intelligence (AI), also known as Deep Learning (DL), has shown great success in computer vision. And some works have initially proved that deep AI can benefit wireless sensing as well, leading to a brand-new step toward ubiquitous sensing. In this survey, we focus on the evolution of wireless sensing enhanced by deep AI techniques. We first present a general workflow of Wireless Sensing Systems (WSSs) which consists of signal pre-processing, high-level feature, and sensing model formulation. For each module, existing deep AI-based techniques are summarized, further compared with traditional approaches. Then, we provide a view of issues and challenges induced by combining deep AI and wireless sensing together. Finally, we discuss the future trends of deep AI to enable ubiquitous wireless sensing.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **General and reference → Surveys and overviews**.

Additional Key Words and Phrases: wireless sensing, AI, deep learning, deep neural network, Wi-Fi, acoustic, LoRa, activity recognition, human localization, pose estimation

ACM Reference Format:
Chenning Li[1], Zhichao Cao[1], and Yunhao Liu[1,2]. 2020. Deep AI Enabled Ubiquitous Wireless Sensing: A Survey. *J. ACM* 37, 4, Article 111 (November 2020), 45 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

With the emergence of IoT, wireless communication technologies are widely used with the explosion of IoT devices. Resembling light in its wave property, wireless signals cannot only be modulated to carry data, but also reflected, scattered or refracted by surrounding objects/humans so that the changes of their phase and attenuation can be further utilized for sensing tasks. The raw signal information fetched from radios are either Received Signal Strength Indicator (RSSI) or Channel State Information (CSI) which can be taken as input of WSSs. To achieve various application goals (e.g., detection, recognition, tracking, imaging), elaborate features and models are required for three general function modules of WSSs, which are signal pre-processing, high-level feature and sensing model formulation as summarized in Figure 1.

In recent years, deep AI (e.g., CNN, RNN, GAN), also known as *deep learning*, has shown its success in many vision and audio applications such as speech recognition, object detection and

Authors' address: Chenning Li[1], lichenni@msu.edu; Zhichao Cao[1], caozc@msu.edu; Yunhao Liu[1,2], yunhaoliu@gmail.com, [1]Michigan State University, 426 Auditorium Road, East Lansing, Michigan, USA, [2]Tsinghua University, 30 Shuangqing Rd, Haidian Qu, Beijing Shi, China.

0004-5411/2020/11-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

face completion. In comparison with traditional Machine Learning (ML) models (e.g., k-NN, SVM, boosting, decision tree, Bayesian networks), DL frameworks take the advantage of a huge amount of complex and heterogeneous input data, and are able to efficiently prohibit noises and extract features so that guarantee the accurate results in general application scenarios. As the similarity of inputs and outputs between wireless sensing and vision sensing, WSSs can absolutely benefit from deep AI. As shown in Figure 1, some works have embedded DL frameworks into each part of WSSs. For signal pre-processing, several works [29, 52, 53, 63, 81, 134, 136, 165, 184, 191] leverage the adversarial architecture [32, 36] and attention module [29, 69, 120, 168, 189] for noise reduction. Moreover, various deep neural networks (e.g., CNN [14, 29, 52, 53, 69, 83, 114, 188–191, 194], RNN [53, 63, 87, 92, 191, 194], GAN [29, 52, 134, 165, 191]) can be adopted to automatically extract spatial, temporal and physical features from sanitized signals. We can further employ DL frameworks [29, 49, 63, 65, 69, 114, 134, 135, 188–191, 194] to formulate complex models of wireless sensing as a whole, resolving the generality by extending the WSSs to various environments.

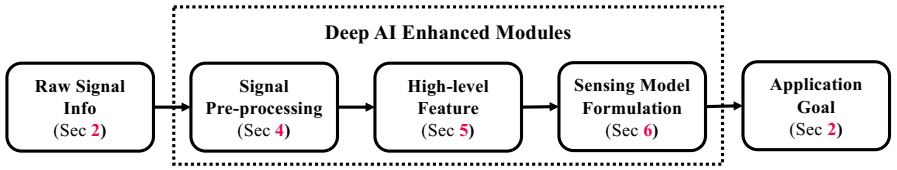


Fig. 1. The general flowchart of WSSs and modules enhanced by deep AI.

Despite the growing interests in deep AI enabled wireless sensing, existing surveys are scattered across different raw signal information and application goals. And a comprehensive survey is lacking for the evolution and principle of systems to combine wireless sensing and deep AI together. This survey fills the gap by summarizing deep AI approaches and comparing them with traditional methods for the middle three function modules in the general workflow, shown in Figure 1. Beyond reviewing relevant literature, we present some unique issues and challenges induced by deep AI. Finally, we wrap up this paper by presenting future trends that use deep AI to enable ubiquitous wireless sensing. Our ultimate prospect is to provide a comprehensive review of deep AI enabled WSSs, which can inspire researchers to adopt DL techniques to resolve intractable issues when deploying WSSs in real life.

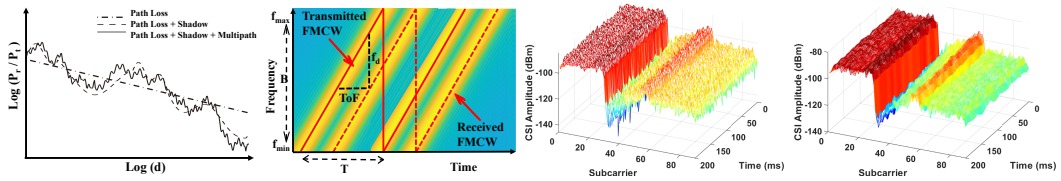
Survey Organization: This article is organized as follows. We first provide the background of DL and wireless sensing in § 2. Existing surveys are discussed in § 3. Then, we present the comparison between deep AI frameworks and traditional methods designed for different function modules in § 4, § 5 and § 6, respectively. Furthermore, issues and challenges induced by deep AI are demonstrated in § 7, followed by § 8 for future trends. We conclude our survey in § 9.

2 BACKGROUND OF WIRELESS SENSING AND DEEP AI

In this section, we first summarize the widely adopted wireless signal inputs and popular application outputs of existing WSSs. Then, we provide a basic overview of deep AI techniques that have the potential to enhance the performance of WSSs.

2.1 Inputs and Outputs of Wireless Sensing

With the development of IoT, multiple wireless technologies are adopted to connect IoT devices in different spectrum bandwidth, communication range, data rate and energy consumption. The widely used and commercialized wireless technologies include Wi-Fi, ZigBee, LoRa, RFID, Ultra-Wideband



(a) Raw RSSI indoors [169]. (b) FMCW with ToF. (c) Raw CSI amplitude[1]. (d) Denoised CSI amplitude.

Fig. 2. Illustration of typical sourcing inputs as sensing mediums, including RSSI, CSI, and FMCW.

(UWB) and Bluetooth. Moreover, due to camera, speaker and microphone are cheap and widely equipped on smartphone and wearable devices, acoustic signal and visible light based wireless communication is also emerged. By taking different types of raw signal information as inputs, we can derive diverse outputs across many sensing purposes which include detection, recognition, identification, positioning, tracking and imaging.

2.1.1 Signal Status as Inputs. Given a transceiver pair, the status of the received signals obtained from either commodity radio or software-defined-radio (SDR) serves as the inputs of a WSSs. Next, we discuss three kinds of signal status which are RSSI, CSI and chirp.

RSSI: RSSI can be directly fetched from most of commodity radios and characterizes the spatial attenuation of signal propagation, covering multiple modalities including Wi-Fi [56, 58], LoRa [60, 60, 72], RFID [78, 91, 132]. Empirically, RSSI follows a propagation fading model (e.g., Log-normal Distance Path Loss (LDPL) model [109]). The LDPL model characterize the variation of received signal power over distance due to path loss and shadowing caused by obstacles through absorption, reflection, scattering, and diffraction. As shown in Figure 2(a), we can leverage the path loss for the power-based ranging assuming that RSSI monotonically decreases with distance. Due to the random shadowing effect, the monotonic trend only holds on relatively large scale. Given the multi-path effect, RSSI can fluctuate on the order of signal wavelength and blurs the monotonic trend, making it almost impossible to distinguish locations in the vicinity [169].

CSI: Along with the success of Orthogonal Frequency-division Multiplexing (OFDM) and Multiple-Input Multiple-Output (MIMO) in Wi-Fi, CSI can be acquired readily and measure the multi-path effect using the amplitude attenuation and phase shift at each carrier frequency [82]. Although CSI is included in Wi-Fi since IEEE 802.11n, it is not reported by all commercial Wi-Fi Network Interface Controllers (NIC). And the 802.11n CSI Tool [40] and the Atheros CSI Tool [159] are most widely used for CSI measurement on commercial Wi-Fi devices. And we illustrate the raw CSI measurements from the commercial Intel 5300 Wi-Fi NIC in Figure 2(c). Given the three antennas from a NIC with 90 subcarriers, each CSI complex value is converted into the amplitude (dBm) representing the intensity and phase information. Due to the narrow-band signal of each subcarrier, raw CSI measurements demonstrate noticeable dynamics across packets and subcarriers, rendering the necessity of the noise reduction in signal pre-processing. For example, multiple algorithms have been proposed to remove the time and phase shift induced by the Cyclic Shift Diversity (CSD), Sampling Time Offset (STO), Sampling Frequency Offset (SFO), and beamforming technique [82, 169], shown in Figure 2(d).

FMCW: Chirp signal refers to signals generated by multiple types of circuits [5, 86, 90] whose carrier frequency changes linearly with time, especially the Frequency Modulated Continuous

[1]Note that the Intel 5300 NIC reports the CSI amplitude in voltage space [40]. We convert the amplitude of CSI $|h|$ into P with the unit of dBm as: $P = 20log(|h|/1000)$ [131].

Wave (FMCW) radio. As shown in Figure 2(b), FMCW transmits a chirp signal (e.g., red solid line) as $S_t(t) = cos(2\pi f_{min} t + \frac{\pi B t^2}{T})$. And $B = f_{max} - f_{min}$ and T denote the frequency bandwidth and sweeping time. Thus, the received FMCW (e.g., red dashed line) becomes $S_r(t) = \alpha S_t(t - ToF)$ with the attenuation factor α and ToF for the reflected signal propagation. We use frequency correlation of these two signals [85, 88] to compute the ToF. And it provides additional resilience to the multi-path effect for measuring Time-of-Flight (ToF) [84] and is employed broadly in Wi-Fi [5, 29, 69, 188–190], acoustic [85, 86, 94, 123] and LoRa [19, 90]. To obtain the required resolution of distance from the corresponding reflector, we can adjust the frequency bandwidth. Large bandwidth enables a fine-grained sensing granularity. For example, RF-Pose, RF-Pose3D [29, 69, 188–190] select the chirp signal with a total bandwidth of $B = 1.78GHz$ from $f_{min} = 5.46GHz$ to $f_{max} = 7.24GHz$, obtaining a distance resolution of 8.4 cm. Then they incorporate the DL techniques with high-quality chirp signals to achieve device-free multi-person pose estimation with the centimeter accuracy [188, 190].

Comparison and Connection: A high-quality signal status and an effective wireless system are two key enablers for sensing. Depending on the sensing granularity, range, deployment cost, and system robustness, RSSI, CSI, and chirp signals can be utilized in various sensing scenarios. Specifically, RSSI suffers more than CSI from the Inter Symbol Interference (ISI) induced by the multi-path effect, while the time-frequency analysis of the latter requires additional computation resources, rendering a high computation complexity in mobile devices [174]. Besides, chirp can provide much finer-grained spatial resolution than CSI, such as dedicated FMCW radios [74] used for device-free human pose estimation [5, 188, 190]. However, CSI can be acquired using low-cost commercial Wi-Fi devices readily, making the ubiquitous deployment much easier.

RSSI is widely used for computation-efficient but short-range indoor acoustic sensing on mobile devices due to the limitation of computation efficiency and hardware deployment. For example, Strata [174] delivers the device-free acoustic tracking system within a 1.0 cm median error but the moving target is however less than 40 cm from the mobile phone. CaField [167] constructs the "fieldprint" with the acoustic biometrics embedded in the RSSI distribution, achieving a detection accuracy of 98.42% for user authentication among 19 human participants, but a displacement of more than 4 cm shall be kept for the smartphone holding in hands. Besides, many researches have adopted RSSI for low-cost but coarse-grained sensing tasks. SpotFi [58] estimates the Angle of Arrival (AoA) with RSSI of Wi-Fi and achieves a median accuracy of 40 cm for device-based human localization. Karanam et al. [56] further construct the 3D Wi-Fi imaging of meter-scale objects by modelling RSSI distribution with the loopy belief propagation model [171] while SateLoc [72] collects RSSI from LoRa nodes and improves the spatial resolution of RSSI using the fingerprinting of the satellite images, achieving a median localization error of 47.1 m in a 227,500 m^2 urban area.

To improve the sensing granularity and system robustness, CSI have been broadly utilized for Wi-Fi sensing. Compared with SpotFi [58], which pinpoints the target with RSSI information available from each of the Access Points (APs), RIM [150] employs the fingerprinting-like localization system with the correlation of consecutive CSI measurements, achieving a median error in moving distance of 2.3 cm and 8.4 cm for short-range and long-distance device-based tracking respectively. Chronos [118] further emulates a wide-band radio by stitching 35 available Wi-Fi bands to alleviate the impact of narrow-band CSI channel on the Synthetic-aperture Radar (SAR) model [31]. More fine-grained WSSs have been developed such as material recognition [124, 181] and decimeter-level device-free localization [70, 71, 97, 99, 160] at room scale.

Given the much finer-grained spatial resolution than CSI, chirp signals can be utilized for short-range acoustic sensing with a narrow bandwidth (10KHz to 22KHz) [86, 87, 123]. To track the tiny motion, BreathJunior [123] utilizes the phase change of the each frequency component of the demodulated chirp signals to monitor the motion and respiration of infants. Furthermore,

LoRa [13] has emerged as a promising technique for Low-Power Wide Area Network (LP-WAN). To support the long-range communication (e.g., $\sim km$) at low power consumption, LoRa employs a variant of CSS modulation, which is an alternative PHY standard for IEEE 802.15.4a. And it has similar principles with FMCW for sensing tasks except for modulation implementation [26]. The PHY technique of LoRa, CSS, is inherently robust to interference and noise in the ISM band (e.g., Wi-Fi, RFID, etc.) [157]. Specifically, WideSee [19] integrates the LoRa technique with the drone and delivers the single-transceiver system within the average localization error of 4.6m in a high-rise building structure with a size of $20 \times 42 \times 85m^3$. μLocate [90] further presents a novel backscatter architecture enabling CSS backscatter at significantly lower power, achieving meter-level localization accuracy at the building scale.

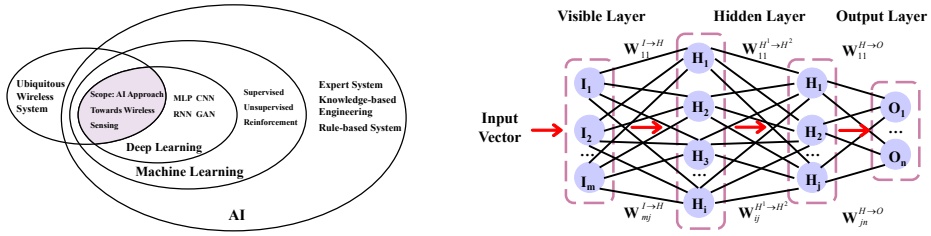
2.1.2 Application Goals as Outputs. WSSs can be employed for multiple applications, covering coarse-grained detection&recognition, and fine-grained image generation, shown in Table 1.

Table 1. Summary of Existing Applications in terms of Applied Output

Detection & Recognition: Binary & Multiple Classification
Event Detection: Presence [8, 98, 196], Fall [41, 128, 182], Smoking [192, 193], Intrusion & Attack [54, 73, 163]; **Activity Recognition:** Gesture [4, 63, 121, 184, 194], Sign [66, 83, 89], Behavior [8, 34, 49, 52, 69, 78, 92, 138, 139, 145, 165, 166, 178]; **Status Recognition:** Material [124, 173, 181], Sleeping [76, 77, 191], Emotion [187], User Identification & Authentication [29, 48, 49, 57, 63, 114, 117, 125, 137, 161, 167, 177, 184–186, 201], Keystroke [11, 12, 67]
Numerical Analysis: Discrete Estimation
Anchor Estimation: Device-based localization & tracking [19, 58, 59, 72, 88, 90, 91, 110, 118, 119, 150], Device-free localization & tracking [7, 70, 71, 85, 87, 97, 99, 131, 195], Handwriting [115, 151, 174], Walking Direction [153, 164, 180]; **Physical Estimation:** Human counting [38, 95, 127], Queue Length [140, 141, 148], Breathing/Respiration Rate [3, 9, 123], Moisture & Salinity [27], Protein [173]
Image Generation: Continuous Estimation
Spatial Image: Imaging [50, 56, 86, 124, 132, 199, 200], Holography [47], Floor Plane [14, 94], Mesh [189]; **Multi-target Sensing:** Multi-person Localization and Tracking [6, 55, 160, 190], Pose estimation [5, 53, 188, 190]

Detection & Recognition: Appendix A shows representatives for Detection & Recognition sensing tasks, corresponding to the binary and multi-class classification. And we further divide them into three groups, including event detection, activity recognition, and status recognition. Given the ubiquitous and untraceable wireless signals, we can achieve the security and surveillance monitoring, such as the detection of human presence, fall, smoking and intrusion. Besides, WSSs can also be deployed for the smart home, smart factory and Virtual Reality (VR), make it more efficient to interact with machines by gestures, activities, signs and some specific behaviors. Further, we can employ the WSSs for the health care, such as the Emerald [1] from MIT, a touchless sensor and ML platform for health analytic. And it can monitor the status of sleeping, vitals, and behavior wirelessly, even for the current COVID-19 monitoring.

Numerical Analysis: Beyond classification, finer-grained sensing tasks require the concrete measurements for the numerical analysis with the discrete estimation. Illustrated in **Appendix B**, on one hand, the anchor based numerical estimation can measure the relative location or the motion of targets with the anchors (e.g., transceivers), such as the localization & tracking, handwriting,



(a) Venn diagram for the evolution of AI.

(b) Structure of multi-layer perceptrons with weights.

Fig. 3. Evolution of AI approaches representing our scope and the initial deep learning model.

and walking direction. Note that localization & tracking can be divided into two categories, device-based and device-free, depending on the need to carry transceivers with the target. On the other hand, WSSs can be adopted to estimate physical parameters, including the queue length, the breathing/respiration rate, the soil moisture & salinity and protein containment in liquids.

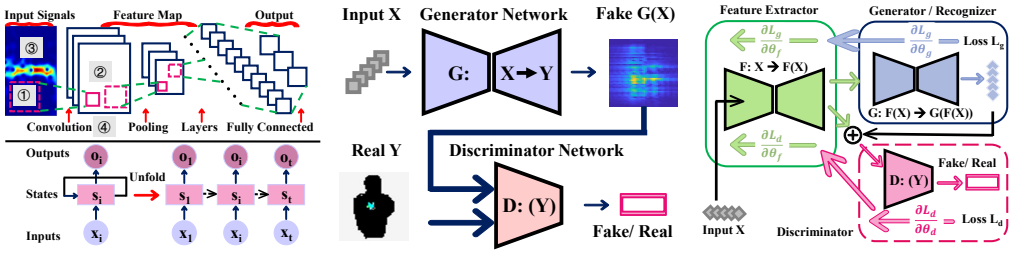
Image Generation: Compared with the discrete estimation of numerical analysis tasks, the image generation is much more challenging for continuous estimations, shown in **Appendix C**. And each measurement has the inter-connected relation across spatial vantages or temporal snapshots. For example, the wireless imaging requires pixel-wise estimations distributed uniformly in space. Moreover, we also have the multi-target sensing, such as multi-person localization & tracking and pose estimation with key joints of the head, shoulder, elbow, chest and so on. In a conceptual sense, image generation is analogous to the concurrent numerical analysis while each of the estimation is correlated in space or time.

2.2 AI and Deep Learning

To demonstrate our scope from the view of the evolution of AI, we first illustrate the relation among AI, ML and DL in Figure 3(a).

In the early days of AI, the field focused on problems that can be described by a list of formal, mathematical rules [35], making it relatively straightforward for computers, such as recognizing spoken words or faces in images. To incorporate the hard-code knowledge about the world in formal languages for computers, several knowledge based approaches were proposed including expert system [106], knowledge-based engineering [104], and rule-based system [175]. However, difficulties faced by systems with hard-code knowledge suggest that AI systems require the ability to acquire their own knowledge, by extracting patterns from raw data. And this capability is known as ML [35], which essentially enables algorithms to make predictions, classifications, or decisions based on raw data, without being explicitly programmed [179]. And we can divide them into three categories including supervised learning, unsupervised learning, and reinforcement learning, depending on their frameworks and data requirements. Note that most existing WSSs resort to ML algorithms for feature extraction, such as k-Nearest Neighbor (k-NN) [91, 124, 132, 132, 161, 184] and Support Vector Machines (SVM) [48, 114, 125, 137, 184, 186, 187] of supervised learning as well as clustering [139, 201] of unsupervised learning.

Given the potential massive noises of wireless signals, the limited ability of pattern representation prevents it from exploring more, especially for the high-level and abstract features of raw data. Thus, DL techniques are proposed for representation learning by introducing hierarchy multi-layer nonlinear processing units, namely feed-forward Artificial Neural Network (ANN). And the basic idea is to render complex concepts out of simpler and predefined operations of units



(a) Spatial CNN & Temporal RNN. (b) Generative adversarial network. (c) Domain-Adversarial network.

Fig. 4. Typical architectures of deep neural network integrated in wireless sensing systems.

(or neurons). Specifically, we leverage a weighted combination of layered hidden units with a non-linear activation function, resembling the perception process in human brain, where a specific set of units are activated given the current environment or task, influencing the output of the neural network model [179]. And the initial ANN is the Multi-layer Perceptrons (MLPs), which consist of more than one hidden layer of operations [23]. As shown in Figure 3(b), the input vector is presented at the **visible layer**, followed by a series of **hidden layers** to extract increasingly abstract features. These layers are called "hidden" because their values are not given in the data, instead the model must determine which concepts are useful for explaining the relationships in the observed data [35]. IntuWition [181] employs MLPs for Wi-Fi based material recognition and achieves the average accuracy of 95% and 92% in classifying five types of materials for Line-of-sight (LoS) and Non-line-of-sight (NLoS) scenarios. Besides, it also demonstrates that MLPs provide the maximum classification accuracy, compared with conventional ML models: REF-SVM [45], k-NN [37], PCA, and Naïve Bayes [103].

Convolutional Neural Network (CNN) employs several techniques to reduce the model complexity significantly while retaining the robust feature extraction ability, shown in Figure 4(a). First, CNN adopts the small convolutional kernel corresponding to the receptive field (e.g., the pink dashed rectangle) of the input layer for **local connectivity** ① with the output layer (e.g., the pink solid rectangle). Note that the **receptive field** can be defined as the region containing any input pixel with a non-negligible impact on the output [80]. Second, the **parameter sharing** ② is adopted to reduce the number of parameters while mitigating the risk of over-fitting by employing the same kernel to scan the whole input feature map. Third, CNN can work with inputs of variable size ③ adaptively by introducing the convolution operation. Fourth, **convolution operations** ④ are invariant in terms of translation, scale, and shape, namely the equivariant representation. And it can enhance the feature extraction since essential features may show up at different locations of the input image, with various affine patterns [179]. Besides, several network structures are designed for data adaption and over-fitting mitigation, including the pooling layer, the fully connected layer, batch normalization, Dropout technique and so on. Owing to properties mentioned above, CNN achieves remarkable performance for WSSs in multiple applied outputs. For example, SignFi [83] feeds the pre-processed CSI measurements to a 9-layer CNN for feature extraction representing sign gestures and the average recognition accuracy is 86.66% for 150 sign gestures performed by 5 different users. And Shi et al. [114] integrate the 3-layer CNN with SVM and achieves over 91% for recognizing 11 activities.

CNN can be adopted for spatial feature extraction while Recurrent Neural Network (RNN) can deal with the sequential data. Since RNN has an internal memory for temporal feature extraction

from sequential input. Illustrated in Figure 4(a), RNN performs the same function for each state with corresponding input while the output of the current states also depends on the past states. By unfolding the recurrent structure of RNN, the output is copied and sent back into the recurrent network from the previous states. RNN can model sequential feature of input data that each sample can be assumed to be dependent on previous ones, enabling it suitable for massive sequential wireless signals. For example, RTrack [87] leverages RNN to estimate the distance and AoA from continuous acoustic signals, achieving the room-scale human tracking with 1.2cm-3.7cm error.

Another promising technique is the Generative Adversarial Network (GAN) [36], which trains a generative model $\mathcal{G} : X \rightarrow Y$ and a discriminative model $\mathcal{D}(Y)$. Illustrated in Figure 4(b), the former model seeks to approximate the target data distribution Y from training data while the latter estimates the probability that a sample comes from the real training data Y rather than the output $\mathcal{G}(X)$ of $\mathcal{G}$ [36]. Give the imaging task of the human holding a phone, we can feed the wireless signals as input to generate a fake image of the target person. Then we can input the ground truth image derived by the computer vision algorithm (the black silhouette with a light blue chunk as the phone) and the fake one into the discriminator network. Thus, we can refined the fake image by inducing the generator network to generate more vivid images to fool the discriminator network.

Existing WSSs have incorporated GANs for sensing procedures, such as the adversarial architecture for domain adaptation of signal processing [32].

3 RELATED WORK

Most of surveys have explored the wireless sensing and DL techniques separately. On one hand, existing contributions are scattered across different sensing modalities (e.g., Wi-Fi, acoustic, LoRa) and focused applications (e.g., recognition, localization, health monitoring). On the other hand, none of existing works demonstrates the evolution from traditional WSSs to DL techniques. And more comparisons and connections are required for further combination from the view of the general workflow for WSSs, covering the signal processing, algorithm design, and model generalization. To make our contributions more clear, we summary existing works from three dimensions, including the modality/input, application/output and topic focus, shown in **Appendix D**.

Given the general workflow in Figure 1, the break-through progress of wireless sensing usually resorts to high-quality sourcing input as the sensing medium, such as Wi-Fi, acoustics, LoRa as well as visible light. For example, Yang et al. [169] present a survey on the evolution from RSSI to fine-grained CSI in terms of network layering, time & frequency resolution, stability, and accessibility. In a conceptual sense, *CSI is to RSSI what a rainbow (color spectrum) is to a sunbeam, where components of different wavelengths are separated.* However, it only focuses on basic principles and research trends of CSI in the field of indoor localization. Given CSI of Wi-Fi as the sourcing input, all the following six surveys [107, 130, 142, 152, 172, 202] focus on the device-free human behavior recognition with respective emphasis, including the comparison between data-driven and model-based Wi-Fi radar [202], performance improvement induced by Fresnel zone model based approaches [152] as well as DL techniques [172]. By focusing on the human-centric applications, Savazzi et al. [107] highlight new radio technologies and unexplored bands for more practical deployment in assisted living applications while Wang et cl. [142] review the human behavior recognition at three granularities of signal, action, and activity. Given the high-level feature of wireless signals, Wang et al [130] further provide a comprehensive overview on the working principle and system architecture of the device-free WSSs. Beyond Wi-Fi sensing, Liu et al. [75] compare acoustic with inertial sensors, vision-based as well as RF signals and further explore the feature design and model mechanisms for range-based localization. Systematically, Cai et al. [17] present a general 3-layer framework to categorize main building blocks of acoustic sensing system, including the physical layer, processing layer, and application layer. It however

only focuses on conventional signal processing systems and regards the DL techniques as the future direction. Sundaram et al. [112] further propose the research problems, current solutions, and open issues for LoRa networking, especially for the link coordination, collision avoidance. Given the highly accurate and noise-resilience visible light, the following three surveys [10, 43, 100] review literatures for visible light positioning (VLP) with various topic focuses. Specifically, Hassan et al. [43] provide a classification of VLP systems on the photodiode and camera based receivers, covering the basic principle and architecture with associated position computation algorithms. To further highlight challenges on real-world deployment from ideal simulations, Afzalan et al. [10] focus on performance-based evaluation of real-world VLP systems as well as different components and fundamental concepts, enabling the VLP system in the wild. Recently, Rahman et al. [100] summarize the current advances in VLP for navigation, tracking and security, which can be divided into solutions with modified and unmodified light source.

Given the growing research interest of wireless sensing, two surveys [158, 176] focus on indoor localization, covering Wi-Fi, Acoustic, RFID, UWB, Bluetooth, etc. And the former [176] emphasizes the comparison of input signals and work principle of localization systems while the latter [158] analyzes smartphone-based methodologies of signal processing and data fusion techniques from the device perspective, including the device based and device free WSSs. The most closely related to our work are [20, 82, 129, 179] (painted blue in **Appendix D**). By covering sufficient applied outputs such as detection & recognition, numerical analysis, and image generation sensing tasks, Ma et al. [82] give a comprehensive survey on modeling and learning based Wi-Fi WSSs but however mainly focus on the workflow of CSI of Wi-Fi and the traditional signal processing techniques. To demystify the promising wireless sensing with multiple input signals, Chen et al. [20] provide a general picture of this cross-disciplinary area, including the wireless communication, ML, and Human-computer Interaction (HCI). While Zhang et al. [179] provide an encyclopedic review of mobile and wireless networking research based on DL, covering exiting DL computing framework and mobile data analysis. However, it emphasizes more on mobile networking (e.g., 5G) and DL driven network problems. To alleviate the intensive training effort induced by DL techniques, Wang et al. [129] mainly focus on deep similarity evaluation networks and deep GANs, delivering a general framework of DL with less training efforts for WSSs.

To the best of our knowledge, none of existing works focuses on the topic of this survey, covering the evolution, comparison and connection of traditional WSSs to DL approaches. Besides, it is different from existing ones in that its scope is not limited to any specific type of sourcing input or applied output. We hope this survey can provide a comprehensive review of the combination of AI approaches toward wireless sensing. And more researchers can be inspired to employ DL techniques to resolve intractable issues for WSSs in real life.

4 SIGNAL PRE-PROCESSING

Upon receiving the raw sourcing inputs, several signal processing algorithms are required to focus on the components of interest, shown in Table 2. Given that wireless environments are increasingly complex, heterogeneous and evolving, the noise reduction, data adaptation, and transform methods are employed for further feature design and model formulation.

4.1 Noise Reduction

Depending on the target components of signals, we have multi-scale noises with various granularity, which can also be beneficial sometimes. Due to the hardware heterogeneity and background interference, raw measurements can be first mixed with multiple types of phase shifts. On one hand, elaborate algorithms are required to mitigate or even remove the estimation error for further feature extraction. On the other hand, those hardware-specific characteristics can also be utilized

Table 2. Summary of Existing Algorithms in terms of Signal Processing

Noise Reduction
Model-driven: Power Model [97, 121, 138, 153], Linear Regression [9, 27, 56, 58, 70, 83, 85, 124, 160, 187], Conjugate Multiplication [53, 63, 71, 99, 194], Coordinate Transform [14, 53, 123, 194]; **Data-driven:** Background Subtraction [5–7, 27, 28, 28, 34, 49, 50, 71, 86, 86, 87, 181, 188, 190], Filters [4, 7, 19, 34, 48, 49, 52, 63, 88, 97, 99, 110, 114, 123, 125, 137, 139, 151, 153, 161, 166, 167, 173, 177, 185, 194], Frequency Response Compensation [86, 88], Mathematical Operation [47, 52, 57, 59, 78, 83, 88, 117, 118, 124, 132, 151, 165, 186, 192].
Data Adaptation
Compression: PCA [48, 57, 63, 97, 121, 137, 138, 161, 184, 194], ICA [92], SVD [8, 33, 58, 71, 178], Graph-based Path Matching [56, 63, 97, 99, 150], Component Selection [7, 70, 86–88, 90, 97, 114, 119, 125, 131, 167, 174, 177, 201]; **Composition:** Multi-devices [5–7, 47, 50, 59, 78, 86, 91, 117, 131, 132, 139, 165, 186, 188, 190, 194], Multi-channels [34, 118, 132, 181], Virtual Samples [8, 55, 58, 69, 121, 150, 184]
AI Approach
Domain-dependent Noise Removal [29, 52, 53, 63, 81, 134, 136, 165, 184, 191], **Attention Module** [29, 69, 189], **ROI Detection** [29, 69, 189, 190]

for device recognition, such as collision avoidance for LoRa communication [28, 157]. Besides, we have to deal with the multi-path interference for the path of interest due to the multi-path effect. Specifically, SpotFi [58] proposes that the radiated signal could reflect off multiple objects and arrive at the APs, typically within 6-8 significant paths [58, 99]. And such a path refinement can be more important for the indoor localization problem [97, 99, 160], which generalizes the target person as a single reflector. To make it worse, the targeting reflected path can sometimes be much weaker than the uninterested ones. For example, those bouncing off the wall are much stronger than the reflections from objects inside the room. Thus, the former can overwhelm the distortions induced by the target object, preventing it from registering the minute variations. This behavior is called the **Flash Effect** [8] since it is analogous to how a mirror in front of a camera reflects the cameras flash and prevents it from capturing objects in the scene. Equivalently, the **near-far problem** [6] indicates that the reflection off the nearest target can obfuscate the signals from distant targets for multi-target sensing. Nevertheless, the multi-path effect can also be useful for fine-grained sensing tasks, enabling more diversities for feature space such as activity recognition [4, 52, 192, 194] and user identification [48, 125, 137, 161, 177, 185]. Finally, some uninterested components can still be contained even if we can extract the target reflected path accurately, say the environmental-dependent components. And it should be removed otherwise it usually requires extra deployment cost such as data collection and model re-training when the system is employed in a new environment, namely the **cross-domain sensing** [194]. Literately, the aforementioned three types of noises with the increasing granularity can be eliminated as well as utilized for specific purposes. And dedicated algorithms can be divided into two parts, including the data-driven and model-driven algorithms.

4.1.1 Model-driven Reduction. To remove phase offsets induced by the hardware heterogeneity, we can first formulate them with raw measurements of the souring input using the model-driven reduction. Given the raw CSI measurement $H_{t,f,s}$ for the packet t at the subcarrier f of the antenna s, one intuitive method to remove the offsets is to compute the **power** of $H_{t,f,s}$ [97, 121, 138, 153], by which CARM [138] associates the power $|H_{t,f,s}|^2$ with the path length change for further feature

extraction. However, we can find that extra valuable information such as ToF and AoA can also be reduced by computing the power of CSI for reflected paths of interest, which hinders some other models (e.g., the trilateration/triangulation model and joint parameter estimation model) for final applied outputs. Note that by focusing on the SFO and STO, the erroneous version of CSI measurements can also be formulated [99] as $\hat{H}_{t,f,s} = H_{t,f,s}e^{2\pi(\Delta f\epsilon_t+\Delta t\epsilon_f)+\xi_s}$, at subcarrier f of antenna s for packet t, where ϵ_t and ϵ_f are the STO and SFO between transceivers. And ξ_s is the initial phase of the receiver sensor, which can be manually calibrated since it keeps constant every time the receiver starts up [162]. Observing the phase offset are linear across subcarriers, SpotFi [58] proposes a linear fitting method for accurate ToF estimation. And multiple following works [9, 27, 56, 58, 70, 83, 85, 124, 160, 187] adopt comparable **linear regression** algorithms for noise reduction. For example, Karanam et al. [56] leverage the linear approximation, namely the WKB approximation [21], to model the interaction interference of the transmitted wave approximation with the area of interest for 3D Wi-Fi imaging. Wang et al. [124] first adopt the phase unwrapping [42] to derive the adjusted phase at each subcarrier of CSI, delivering a Wi-Fi based material detection system. We can also extend the linear regression with multiple phase shifts, namely multiple linear regression [83]. Thus, it extracts the correlated spatial features accurately using CNN to recognize 150 sign gestures.

Nevertheless, the linear fitting can eliminate the absolute ToF of the LoS signal, making it impossible to extract the reflected path of interest using the jointly parameter estimation model [71, 99, 160]. Since the phase contributed by the absolute ToF are also linear across subcarriers [99]. Thus, the **conjugate multiplication** [71, 99] is proposed to filter out irrelevant noises while retaining indispensable channel responses. And the basic idea is that the time-variant random phase offsets are the same across different antennas on a Wi-Fi card [58, 70] as they share the same RF oscillator. Therefore, we can select a antenna as the reference one s_0 and calculate the conjugate multiplication $C(t, f, s) = \hat{H}(t, f, s) * \hat{H}(t, f, s_0) = H(t, f, s) * H(t, f, s_0)$, resolving the offset noise and multi-path interference simultaneously. Beyond that, the **coordinates system transform** [14, 56, 123, 194] can also be utilized for noise reduction. One interesting design is BreathJunior [123], which transforms the white noise into multiple FMCW signals at the receiver while preserving the multi-path reflection information with a negligible SNR loss. Thus, it can demodulate these orthogonal FMCW chirps to sense the minute respiration motion of infants and compute their distance from the smart speaker, without any noise damage to infants. Besides, DLoc [14] transforms the extracted 2D AoA-ToF profile into the global 2D Cartesian plane, which represents the location probability spread out over the X-Y plane and can be further fed into the designed neural network for imaging translation. And it outperforms the state-of-the-art, SpotFi [58], for Wi-Fi based localization by 80% across the 2000 sq.ft. area. Finally, the remaining challenge is the environment-dependent components of signals for cross-domain sensing, which cannot be resolved due to the complex mapping functions in the high-level feature space. To filter out the environment-dependent components while reducing the deployment cost such as data collection, we introduce AI approaches for domain-independent feature extraction in § 4.3.

4.1.2 Data-driven Reduction. Another type of noise reduction algorithms is the data-driven reduction by mining the massive wireless signals. For example, we can remove the background noise by recording it in advance for further calibration. And one intuitive method is to subtract directly the pre-recorded interference from the raw sourcing input. For example, the **background subtraction** widely used for human localization and tracking [5–7, 27, 28, 28, 34, 49, 50, 71, 86, 86, 87, 181, 188, 190]. To eliminate the reflections from static objects (wall, furniture), we can measure the interference initially and subtract it for the following collected signals. Further, we can adapt the measured interference iteratively for continuous calibration, namely the Successive Interference Cancellation

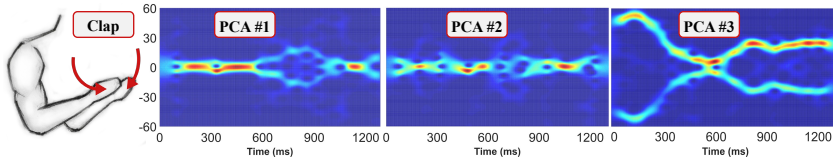
(SSC) [6, 28, 181]. Besides, we can also leverage multiple types of **filters** for data-driven noise reduction, such as the frequency response filter [97, 114, 123, 125, 137, 139, 161, 167, 177, 194], say the Butterworth IIR filter with a flat amplitude response, Kalman Filter [88, 99], Savitzky-Golay Filter [151, 153], Moving Average Filter [110, 185], even the wavelet filter [4, 22]. Specifically, CapCam [173] convolvs the input image with the designed wavelet filter to extract the vertical line of the symmetric neighbor pixels for reflection symmetry detection. To deal with the non-flat frequency response in the acoustic channel of the speaker, we can measure the frequency response and compensate for its effect, namely **frequency response compensation** [86, 88]. For example, AIM [86] develops an two-step phase gradient algorithm for acoustic imaging, namely MPGA. And it first compensates for quantization errors and then estimates phase errors using a stochastic method. Thus, it can remove the assumption of narrow beam signals and capture the impact of closely spaced dominant reflectors [86]. The similar issue also bothers the Wi-Fi sensing. Some works also employ **mathematical operations** to provides additional resilience to noise [47, 52, 57, 59, 78, 83, 88, 117, 118, 124, 132, 151, 165, 186, 192], such as interpolation [7, 52, 118, 192], ratio computation [117, 132, 151], Phase Unwrapping [83, 124, 186], and Gaussian mask [47, 57, 96]. For example, Chronos [118] interpolates the continuous subcarriers of CSI to estimate the offset-free zero-subcarrier, which overlaps with DC offsets in hardware that are extremely difficult to remove [46]. FingerDraw [151] propose a novel CSI-quotient operation which computes the quotient of two CSI signals from various antennas at the same receiver. And it can cancel random phase offsets which are identical among the antennas of the same receiver based on the model-driven algorithms in [71, 99, 128]. In addition, it effectively maximizes SNR so that the resolution of CSI signals can be significantly improved [151] for tiny finger tracking.

The rationale behind the data-driven reduction is to leverage the pre-measured interference for the following calibration, whose performance depends on the granularity and adaptation of the interference measurements. From the view of DL techniques, the calibration can be substituted using the data-hungry network, which can distinguish the signals of interest and noise, such as the phase offsets induced by hardware heterogeneity and background interference, the irrelevant reflected paths induced by the flash effect and near-far problem, and the non-flat frequency response. We however witness few works on this issue and the reason can be the difficult modification of DL techniques for massive dirty wireless noise signals (See § 7).

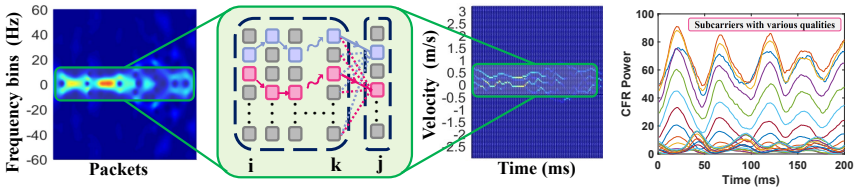
4.2 Data Adaptation

To balance the computation efficiency and signal granularity, data adaptation is required for further transform method and feature extraction. On one hand, we can leverage the compression algorithms to remove the redundant components to improve computation efficiency. On the other hand, composition algorithms can be utilized to acquire multi-dimensional information, rendering it more sensible to fine-grained sensing variance.

4.2.1 Compression Adaptation. Given the massive sanitized sourcing inputs, employing transform methods for further feature extraction directly can be too computation-intensive to be running in real time [63, 194]. For example, the size of CSI measurement [82] can be $3 \times 30 \times 1000 \times 32/8 = 360KB/s$ for a 20 MHz Wi-Fi channel with each packet represented by 32 bits at the sampling rate of 1000 from 30 subcarriers of 3 pairs of transceivers [63, 97, 194]. And traditional compression algorithms include Principal/Independent Component Analysis (PCA/ICA) using the linear transformation [48, 57, 63, 92, 97, 121, 137, 138, 161, 184, 194], Singular Value Decomposition (SVD) [8, 33, 58, 71, 178], graph-based path matching [56, 63, 97, 99, 150], and component selection [7, 70, 86–88, 90, 97, 114, 119, 125, 131, 167, 174, 177, 201].



(a) The first three PCA components of the CSI spectrogram for gesture recognition (e.g., clapping).



(b) Graph-based path matching of CSI for data compression. (c) Subcarriers selection.

Fig. 5. Multiple compression algorithms, including PCA, graph-based path matching and frequency selection.

PCA is widely used for dimensionality reduction of data while retaining most of the most information with selected primary components, which is a set of linearly uncorrelated variables and ordered by the fraction of the total information each contains. Note that the input is assumed to be a set of possibly correlated variables, such as the CSI matrix with multiple subcarriers. And the number of selected primary components depends on the demanding granularity of the specific sensing task. Note that more components contain more complementary information while consuming more time for further processing. And most existing works select only one primary component [97, 111, 122, 194], such as the first one[97, 194] after noise reduction using the filter or the third one without the filter [111, 122]. Illustrated in Figure 5(a), the first three primary components are plotted given the CSI spectrogram of Wi-Fi for gesture recognition [194]. We can observe that different primary components contain diverse scales of information corresponding to the motion of various body parts. Specifically, PCA #1 shows the dominant power distribution (e.g., the torso) while PCA #3 provides finer-grained motion pattern which can represent the symmetrical arm movements for the clapping gesture [64]. Similarly, we also have **ICA** [92] and **SVD (or eigen decomposition)** using the coordinate transformation for data separation and dimensionality reduction, respectively. For example, Ohara et al. [92] employ ICA to separate the events caused by objects since CSI of Wi-Fi consists of the mixed effects of multiple indoor objects. And the eigen decomposition is incorporated in MUSIC algorithm (See § 6.2) for AoA estimation with the covariance or correlation matrix of the input [8, 58, 71, 178].

Given a continuous sequence of data, the **graph-based path matching algorithm** can also be utilized for data compression [56, 63, 97, 99, 150]. For example, the optimal combination of multiple parameters (e.g., AoA, ToF, velocity) should be estimated with continuous packets for human localization and tracking. Illustrated in Figure 5(a). give the CSI spectrogram with i to j packets, we can integrate the physical restrictions such as the velocity continuation for the parameter estimation between each pair of adjacent packets, say the k and j packets. Specifically, Widar [97] properly decimates the velocity change between adjacent packets with the knowledge of the maximum acceleration of walking human, say $3.2m/s^2$. And it then can select the dominant velocity change paths for human tracking. Further, we can also compress the data with the **component selection**, including the frequency response [7, 87, 88, 90, 97, 119, 131] and time delay [86, 125, 174, 177]. The

rationale is that subcarriers of CSI with different frequency are experiencing frequency-selective fading [40] and the target reflection has a specific range for the propagation delay. Figure 5(c) shows multiple waveforms of Channel Frequency Response (CFR) power measurements across subcarriers of Wi-Fi, each of which has various dynamics while the target person is walking. To choose appropriate subcarrier for maximizing the proportion of reflecting power while suppressing the noise, selected subcarriers should have larger dynamics, especially compared with the latest static one [97]. Besides, LiFS [131] adopts a threshold to filter out those "dirty" subcarriers based on whether the power decrease is large enough for human tracking. Specifically, if a target is not located on the LoS path, the threshold δ_{eff} for power decrease is defined as the averaged standard deviation over all subcarriers. Further, the frequency selection can also be employed initially such as ReMix [119] for in-body deep-tissue backscatter, which can achieve optimal transmission efficiency and safety limits by distinguishing the required weak frequency response from the backscatter signals. As for the delay selection, Strata [174] estimates the channel impulse response (CIR) of acoustic, which characterizes signal traversal paths with different delays in time domain. And the delay selection is required to explicitly take into account the multi-path propagation since each channel tap corresponds to the multi-path effects within a certain delay range. Additionally, WiWho [177] extracts the CIR of Wi-Fi within the delay of less than 0.5 microseconds to remove distant multi-path components and then converts the CIR back to CFR using FFT. Thus, we can focus on the reflected path of the target gait within a room which is necessary for user identification.

4.2.2 Composition Adaptation. To provide additional resilience to interference, multi-dimensional and orthogonal measurements are required for composition adaptation from **multiple devices**, channels, and virtual samples. Specifically, Multiple-Input Multiple-Output (MIMO) used for Wi-Fi cannot enhance a high throughput to meet the growing demands of wireless data traffic but also provide multiple spatial vantage views of sourcing input collection for sensing tasks. Besides, all those SAR [31] based WSSs [8, 47, 50, 56, 59, 86] require the mechanical scanning or a massive antenna array as a filled 2D aperture. Ubicarse [59] proposes the circular SAR by moving a single antenna in a circle to emulate a circular antenna array, delivering a transition-resilient SAR system for the complex unknown trajectories. And it however increases the deployment cost significantly. For example, Wision [50] employs a $(8, 8)$ stationary antenna array while Karanam et al. [56] demand on the antenna scanning to formulate a virtual antenna array equivalent to 150×150.

To alleviate the limitation of narrow bandwidth on the spatial resolution, **multi-channel information** is integrated [34, 118, 132, 181], such as the channel stitching for emulation of the wide-band radio [118, 181]. The basic idea is to transmit packets on multiple Wi-Fi bands (35 available Wi-Fi bands) and stitch the information together to give the illusion of a wide-band radio [118]. Furthermore, **virtual samples** can also be utilized with the virtual antenna array [8, 58, 150] or combination of separate samples [69, 121, 184]. Specifically, mainstream APs only have a few antennas, limiting the resolution and robustness of the measurements. Thus, RIM employs a number of virtual antennas emulated by the sequence of channel snapshots recorded by a moving antenna, forming a virtual massive antenna array whose size is the number of channel snapshots [150]. And it can further bootstrap the super-resolution feature extraction. Note that the virtual antenna array requires the moving transceivers for multiple channel snapshots acquisition, say device-based sensing [58, 150]. To recognize multi-user gesture using Wi-Fi, WiMU [121] generates virtual samples for any plausible combination of gestures using the training samples collected from a single user, which are identical to the real samples that would result from real users performing that combination of gestures. Thus, it can reduce the deployment cost significantly while recognizing the simultaneously performed gestures.

4.3 AI Approach toward the Signal Pre-processing

Illustrated in Table 2, most researches leverage DL techniques for **domain-independent feature extraction** [29, 52, 53, 63, 81, 134, 136, 165, 184, 191]. To mitigate the degradation induced by the feature shift in new environments, DFAR [136] develops a maximum-minimum adversarial approach for transferring the feature in cross-scenarios conditions. Then it customizes a multi-layer CNN for the cross-scenario activity recognition via mmWave. Given the domain-dependent CSI of Wi-Fi, DFGR [81] designs a deep feature extraction as well as a deep similarity evaluation network. And the former learns discriminative deep features while the latter evaluates the transferrable similarity from the training set to the new testing conditions. To further reduce the training effort while being scaled up to cross-scenario conditions, CrossSense [184] employs the transfer learning technique to remove the irrelevant components by designing a 7-layer feed-forward ANN roaming model. Further, EI [52] and RF-Sleep [191] integrate the adversarial architecture into the feature extractor, which can learn environmental-independent components by specially designed loss functions for the generator and discriminator as the penalty. Unfortunately, all cannot eliminate the deployment cost including data collection or model re-training. To achieve the zero-effort cross-domain sensing, Widar3.0 [194] designs a Wi-Fi based gesture recognition system to be deployed in cross-scenarios conditions such as new users, positions, orientations and environments out of the training dataset without any deployment cost. It first derives the Body-coordinate velocity Profile (BVP) with CSI measurements, which is a domain-independent physical feature since BVP is inherently irrelevant to the domains such as the deployment of the transceivers or the orientation of performers. It further enhances the domain-specific noise reduction by integrating a CNN-based GRU model and achieves a comparable gesture recognition performance when applying in new domains directly. Importantly, it demonstrates that the potential of incorporating signal processing with DL techniques for noise reduction [29, 53, 63, 165, 194]. Subsequently, WiPose [53] removes the posture-irrelevant noise using the 3D BVP and employs a 7-layer CNN-LSTM model for enhancement while WiHF [63] designs a new domain-independent feature, the motion change pattern, and a collaborative CNN-GRU model to recognize gestures and users simultaneously.

To further improve the computation-efficiency for sensing tasks in complex scenarios, DL techniques can also be utilized for data compression, such as the **attention module** [120] and **the network for Region of Interest (ROI) detection**. And the former is proposed to equip the neural network with the ability to focus on a subset of its inputs while the latter is widely used for object detection and segmentation in the field of computer vision. Specifically, given that the human body is best modeled as a quasi-specular reflector [5], the attention module is employed in the designed neural network to solve the missing body part for individual RF frames, rendering a reliable WSSs for human mesh recovery [189], multi-person activity recognition [69], and fine-grained user identification [29], especially through the wall. The basic idea of the attention module is to emulate the working mechanisms of human brain for content awareness and compute a mask for the input feature. And many related works have been proposed for further exploration, such as bottleneck attention module [93] and convolutional block attention module [149] to be integrated with any feed-forward CNN. The attention module achieves significant progress for computer vision tasks and is supposed to compress massive wireless signals effectively. Further, the network designed for ROI detection is also adopted to focus on signals of interest, such as the Region Proposal Network (RPN) for multi-person detection [69, 189, 190]. Basically, it's analogous to the beamforming technique via directional antennas [6], which is used for multi-target segmentation.

Table 3. Summary of Existing Measurements in terms of Feature Design

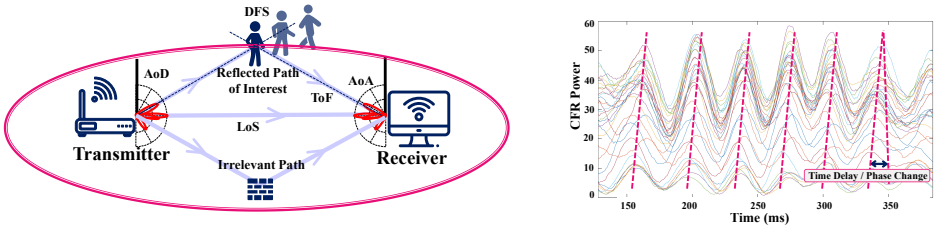
Spatial Vantages : AoA [14, 55, 58, 70, 71, 71, 87, 99, 110, 117, 160], Amplitude [19, 27, 34, 50, 56, 83, 91, 99, 110, 123, 124, 131, 139, 160, 161, 178], Power Distribution [5, 8, 19, 49, 56, 59, 72, 117, 137, 165, 185, 188–190], Wave Front [14, 47, 50, 86, 87], Sound Field [167], TRRS [150, 155]
Temporal Snapshots : ToF [6, 7, 9, 14, 27, 27, 58, 70, 87, 99, 118, 160], Frequency Shift [6, 7, 85, 88, 94], Phase [9, 34, 83, 90, 123, 124, 132, 138, 151, 153, 161, 174, 186], PLCR [71, 97, 121, 138], Circularity [137, 177, 192, 201], Rhythm Motion Pattern [4, 63, 78, 139, 187, 192, 201]
Physical Properties : DFS [85, 88, 99, 160], Hardware Behavior [119, 157], Coherence [70], BVP [53, 194], Polarization [181], Surface Tension [173], Soil Electrical Conductivity [27]
AI Approach: MLP [181, 184], CNN [14, 29, 34, 49, 52, 53, 65, 69, 83, 114, 135, 165, 166, 188–191, 194], RNN [53, 63, 87, 92, 191, 194], Adversarial Architecture [29, 52, 134, 165, 191]

5 HIGH-LEVEL FEATURE

Given the sanitized signals of interest from the signal processing module, the next challenge is to feature extraction for further model formulation. Table 3 presents multiple specially designed features given the final applied output, such as the spatial feature for imaging, temporal feature for tracking, and physical feature for material recognition.

5.1 Spatial Vantages

Intuitively, various devices with corresponding spatial vantages can obtain the complementary information of surroundings. And recording the spatial distribution of transmitted wireless radiations can help pinpoint the target by traversing or reflecting off it. **AoA** has been broadly used as a spatial feature, which indicates the direction of the signal arriving at the receiver and can be estimated using multiple antennas of the same receiver [14, 58, 70, 71, 71, 87, 99, 110, 117, 160]. Thus, we can determine the target with multiple directions derived by corresponding devices. Besides, the concept of the angle can also be extended to the angle of departure from the transmitter, namely Angle of Departure (AoD). For example, md-Track profiles each single reflected path using the AoA, AoD, Doppler Shift, and the attenuation. And it employs the joint parameter estimation model iteratively to handle multiple reflected paths arriving along different propagation directions, illustrated in Figure 6(a) for the device-free tracking task. Another representative is the **amplitude attenuation** for computing the propagating fading such as the reflection, refraction, diffraction, absorption, scattering, and shadowing fading [19, 27, 34, 50, 56, 83, 91, 99, 110, 123, 124, 131, 139, 160, 161, 178], say LDPL model [109] used for RSSI. Specifically, LiFS [131] formulates the amplitude attenuation between the transmitter and the receiver as the summation of the propagation fading, diffraction fading and the target absorption fading. And it achieves a comparable performance with the fingerprinting but requires a dense deployment of devices by employing the radio tomography imaging approach for localization, say 40 antennas [99]. Another example is the LANDMARK [91], which scans the 8 discrete power levels of RFID and estimates the signal strength of tags. Since none of the currently available RFID products provides the signal strength of tags directly at that time. LANDMARK however employs a massive reference tag array and multiple readers as cooperated eyes for the monitoring area. Thus, it can measure the similarity of the received signal strength between the reference tags and the tracking ones, delivering the first sensing RFID system for indoor objects localization. Further, the **power distribution** indicates the power heatmap in 2D plane, such as the Cartesian plane [5, 49, 56, 72, 188–190] or the spectrogram [19, 57, 97, 137, 165]. Thus, it can estimate the distributed probability for localization and tracking. To describe the signal power



(a) AoA/AoD/ToF/DFS profiling the reflected path. (b) Phase change across Wi-Fi subcarriers.

Fig. 6. Joint multi-parameter estimation and the temporal phase change for the time delay.

below the noise floor reliably, SateLoc [72] employs the Expected Signal Power (ESP) measurement by incorporating the SNR information with RSSI, delivering the probabilistic power distribution map for each target node of the monitoring area. And it outperforms the state-of-the-art localization works using LoRa. Besides, XModal-ID [57] captures the gait characteristics of a person based on Wi-Fi CSI spectrogram using STFT. Thus, it can identify users without the need to know the track of the person, even through the wall.

Additionally, exploring the physical phenomenon across modalities and fields also enhance the spatial feature design, such as the **wave front** [14, 47, 50, 86, 87], sound field [167], time reversal focusing effect [150]. Given the propagating electromagnetic waves with a precisely known amplitude and phase, a two-dimensional wave front is formed by radiations while encoding the pixel-wise amplitude and phase distribution of the traversed or reflected plane [47]. Resembling light, wireless signals inherently represent the intensity and direction information of each pixel with a complex value[47] while the light can only record the pixel-wise intensity using the light amplitude. Given the propagation mechanism of a wave field in the microscopy, the angular-spectrum relation models the propagation of a wave field, rendering the light field at an arbitrary depth in space. Thus, it can be utilized for the 3D hologram in the whole space, which is a prospecting problem for wireless sensing. Besides. CaField [167] proposes the biometric feature with the voice propagation stage as the “fieldprint", analogous to “voiceprints". And a fieldprint is extracted from the **sound field** [16], which is a physical field of acoustic energy created by the propagation of sound over the air. The sound field can be used for voice authentication (either a human or a loudspeaker) since it is only affected by the physical structure, such as physiological features of the mouth, head, and torso, or dimensions of the mechanical components. Another example is the **time-reversal focusing effect** applied in the fields of ultrasonic, acoustics, light and electromagnetism [61, 62]. And it profiles the focused energy of the transmitted signal in both space and time domains when combined with its time-reversed and conjugated counterpart [150]. To represent the effect with the Wi-Fi CSI measurements, RIM [150] leverages the Time-reversal Resonating Strength (TRRS) [155] for the multi-path profile which can underpin a high-resolution location map. Thus, it can compute the fingerprinting of the multi-path effect for device-based tracking.

5.2 Temporal Snapshots

Sequential data is to spatial one what the video is to an image. And a temporal feature can capture the temporal correlation among each frame of the sequential signals, analogous to continuous snapshots of videos. One well-known temporal feature is the **ToF**, which records the propagation time the targeting signal takes to travel along a particular path from the transmitter to the receiver. Given the series of ToF estimations as snapshots, we can track the distance of the moving target

continuously [6, 7, 9, 14, 27, 27, 58, 70, 87, 99, 118, 160]. To extract the ToF accurately, Chronos [118] utilizes the multi-channel stitching to emulate a wide-band radio, achieving the sub-nanosecond ToF estimation using Wi-Fi (See § 4.2) while Strobe [27] leverages the relative ToF to provide a high accuracy distance measurements without requiring an ultra-wide bandwidth. Since the resolution of relative ToF is only constrained by carrier frequency, not bandwidth [58]. And the relative ToF error is 0.006 ns at 2.4GHz for Wi-Fi. Besides, we can also estimate the ToF using the **frequency shift** of chirp signals due to the duality characteristic of chirps between time and frequency [28, 157]. Given the FMCW radio, WiTrack [7] first integrates the ToF with the direction information derived by three directional antennas. By recording multiple ToF snapshots corresponding to different body parts, it achieves the 3D human tacking, elderly fall detection and control appliances interaction. WiTrack2.0 [6] further designs a multi-shift FMCW scheme to avoid interference among multiple transceivers and leverages the successive silhouette cancellation based on the background subtraction (See § 4.1) to detect the ToF of weak reflection paths for multi-person localization. To measure the frequency shift of acoustic [85, 88, 94], CAT first calibrates the frequency shift induced by the sampling frequency offset using the least square regression and incorporates the frequency shift estimation with Doppler shift over the time. Since the former feature can give the distance estimation without incurring error accumulation and the latter provides more accurate distance change in a short-term. And droneTrack [88] extracts the frequency shift using the MUltiple SIgnal Classification (MUSIC) model, resolving multi-path and enhancing the distance estimation.

Given the spectrogram of multi-FMCW reflected chirps at the receiver, each frequency bin corresponds to reflections at different distances. To further improve the resolution of ToF impacted by the bandwidth and the multi-path effect, BreathJunior [123] leverages the phase of each frequency component of the demodulated acoustic signals for minute breathing motion detection of infants. Since a tiny 1mm displacement will result in a significant 0.185 radian phase difference for the measured spectrogram. And the **phase change** have been broadly adopted for minute distance estimation [9, 34, 83, 90, 123, 124, 132, 138, 151, 153, 161, 174, 186]. For example, RF-Mehndi [186] employs a RFID tag array as the authentication credential and extracts the unique phase change induced by the fingertip touching for user authentication. Meanwhile, Wang et al. [124] reconstruct the accurate amplitude and phase change from the reflected Wi-Fi signals for dangerous material type detection. Illustrated in Figure 6(b), WiDir [153] further associates the phase changes with the time delay between subcarriers of Wi-Fi CSI. Since the phase change overtime corresponds to the time correlation between instant waveforms for computation. Thus, it can estimate the sign of the time delay based on the Fresnel Zone [101] readily for walking direction detection.

Beyond the temporal snapshots of the phase change, several temporal features are also designed with definite physical characteristics, including the **Path Length Change Rate (PLCR)** [71, 97, 121, 138], **circularity** [137, 177, 192, 201], and **rhythm motion Pattern** [4, 63, 78, 139, 177, 187, 192, 201]. Given the model-driven power model for noise reduction (See § 4.1), CARM [138] first associates the power $|H_{t,f,s}|^2$ with the path length change by dividing the multi-path channel into static path P_s and dynamic path P_d. Then we can further derive the corresponding target moving velocity based on the Doppler Frequency Shift [97, 121, 138]. Since we can find the total CSI power is the sum of a constant offset and a set of sinusoids, where the frequencies of the sinusoids are functions of the speed of path length change [121, 138]. Given the CSI-speed power model, Widar [97] integrates the sanitized CSI power with the moving velocity and pinpoints the target person consecutively using a geometrical tracing estimation (See § 6.1) for the following model formulation while WiMU [121] leverages the frequency difference and combination of various terms in the power for multi-person gesture recognition. To extract the personalized styles of the biometric feature for user identification, AutoID [201] represents the gait circularity using the

designed shapelet for CSI time series, which indicates the local, phase-independent sub-sequences of time series that features the maximal discriminative power for time series classification [170]. Similarly, WifiU [137] correlates the movement speed of different body parts with the spectrogram by extracting 170 user-specific gait features such as gait cycle time, achieving an average accuracy of 93.05% for the top-3 recognition among 50 users. Further, WiHF [63] and Smokey [192] extract the rhythm motion pattern resembling the acceleration information representing unique performing styles, enabling user identification and smoking detection, respectively.

5.3 Physical Characteristics

One common physical characteristic is the **Doppler Frequency Shift (DFS)**, which indicates the introduced frequency shift by movements of the transceivers or the reflector [85, 88, 99, 160]. And droneTrack [88] estimates the target's relative velocity with the drone by measuring the amount of Doppler shift, delivering an automatic following system for drones. Note that the resolution of the Doppler shift depends on the observation interval [160]. To acquire multi-dimensional information, several works leverage the joint parameter estimation model [71, 99, 160] to provide extra resilience to the power fading and the multi-path effect (See § 6.2), which incorporates Doppler shift with ToF, AoA and amplitude attenuation. Besides, the **hardware behavior**, say hardware imperfection or non-linearity which can also be utilized for physical feature extraction. FTrack [157] extracts the frequency shift induced by the hardware imperfection and distinguish various devices for transmission collision avoidance. To isolate the much weaker backscatter signal under deep tissues, ReMix [119] employs the Schottky Diode as the tiny and passive backscatter radio for its non-linearity response, delivering an in-body backscatter system for communication and localization. Exploring physical characteristics can further turn waste into treasure. Given the device-free scenarios, the signal of LoS and signals reflected by static objects are coherent with each other. And coherency is usually considered harmful for the traditional MUSIC algorithm. Thus, different schemes [58, 162] are proposed to remove the coherence among signals for accurate AoA estimation. MaTrack [70] proposes the Dynamic-MUSIC method to utilize this 'bad' **coherence** to distinguish signals reflected from the moving target intelligently. Meanwhile, RIM [150] leverages the time-reversal focusing effect [61, 62] to transform the power distribution induced by the multi-path effect to fingerprints. In other word, *the more impact of the multiple paths on the power distribution, the finer granularity of the fingerprinting map we can derive*. We also have some effective scientific observations to extend the range of wireless systems. IntuWition [181] extracts the **polarization** of Wi-Fi signals bouncing off targets, which can sense the material type and surface texture of the reflector. And CapCam [173] measures the **surface tension** by formulating its quantity relationship with the focusing light of the vibrating water surface, achieving the water contamination and alcohol concentration detection. Additionally, Strobe [27] associates the wave propagation characteristics with soil properties, say the **soil permittivity and electrical conductivity**. Thus, it delivers the Wi-Fi based sensing system for the soil moisture and salinity.

5.4 AI Approach toward the High-level Feature

Due to its high-level learning ability, DL techniques have been broadly utilized for feature extraction. Illustrated in Table 3, multiple convolution based networks are designed for spatial vantages analysis, including **MLP** [181, 184], **CNN** [49, 52, 53, 83, 165, 166, 191, 194] and some derivative networks such as the encoder-decoder network [14, 29, 34, 69, 114, 188–190]. For example, intuWition [181] recognizes materials using the polarization of the signals, a physical phenomenon of electromagnetic waves. And it employs a simple MLP to enhance the polarization extraction from the processed wireless signals. Experiments demonstrate that MLP outperforms the traditional ML based feature extractor such as REF-SVM, k-NN, and Naïve Bayes. Compared with MLP, CNN demonstrates the

powerful capability for automatic spatial feature optimization due to its design on local connectivity, parameter sharing, input adaptation and equivalent representation (See § 2). Thus, WSSs incorporate CNN for finer-grained spatial feature extraction. Specifically, Li et al [65] design a simple but effective 6-layer CNN with the power distribution of CSI across multiple APs, outperforming the SVM based model for human activity recognition. To avoid the handcraft feature design for a specific application, Wang et al. [135] integrate the sparse auto-encoder network for the spatial feature extraction of multiple applications, achieving a higher accuracy for location, activity, and gesture recognition simultaneously. SignFi [83] further employs a 9-layer CNN to capture the spatial correlation from the amplitude and phase of CSI measurements to recognize 276 sign gestures. Given the conventional SVM spoofing scheme, Shi et al. [114] integrate the CNN to extract the 9 manually designed physio-logical (e.g., body shape, height, and weight) and behavioral characteristics (e.g., walking patterns) from daily activities, rendering distinctive spatial features for authentication among 11 users. Further, DLoc [14] employs a CNN based encoder-decoder network composed of a stack of CNNs for sequence-to-sequence translation in the natural language processing. DLoc formulates the tracking problem in the monitoring area as the image translation. The encoder-decoder network takes as input the heatmap in Cartesian plane after 2D FFT and outputs the pixel-wise spatial probability distribution map for indoor navigation. We can see the utilization of DL network can alleviate the cumbersome efforts for manually feature design and further bootstrap applications of WSSs.

DL techniques can also be used to capture the temporal snapshots of sequential signals, such as **RNN**, **LSTM**, and **GRU** [53, 63, 87, 92, 191, 194]. Given the noisy 2D AoA-ToF profiles, RTrack [87] designs a RNN to map the error-prone 2D AoA-distance profile derived by the 2D MUSIC [144] to a fine-grained one of the target reflector, even under low SNR. The rationale is that RNN can exploit the temporal structure among consecutive 2D profile and correct the impact of noise, multi-path and mobility issues [87]. Note that most WSSs [14, 58, 99, 160] incorporate the temporal snapshots with the spatial vantages for distance and direction estimation collaboratively. And it also applies to neural networks, say CNN based GRU network designed by Widar3.0 [194] to extract the spatial and temporal features. Furthermore, WiPose [53] employs a stacked 4-layer CNN followed by a 3-layer LSTM, which can capture relatively long movements of body parts for pose estimation using commercial Wi-Fi. Eventually. comprehensive evaluations are also presented to verify the effectiveness of RNN for temporal feature extraction. RTrack [87] demonstrates the increasing correlation between the output of a deeper intermediate layer of RNN and the ground truth. Meanwhile, Widar3.0 [194] verifies that the simple single-layer GRUs are sufficient for capturing temporal dependencies for short-time human gesture. Note that exploring the design of the neural network cannot only optimize the selection of DL models, but also enhance the interpretability of AI approaches. Thus, we can tailor DL techniques for wireless signals, such as the design of loss functions [52, 53, 63], to further narrow the gap between them.

To extract the physical features efficiently, the **adversarial architecture such as GAN** [29, 52, 165, 191] can be utilized to learn the hidden connection between the sourcing inputs and the applied outputs. For example, RF-Sleep [191] leverages the conditional adversarial architecture to distinguish the fake samples from the ground truth, rendering a CNN-RNN feature extractor to learn the optimal feature for sleep stage detection. EI [52] employs a GAN model to remove the irrelevant components from the CNN feature extractor, reducing the deployment cost for cross-domain scenarios. Integrating GANs to WSSs is a promising direction. While it demands on numerous ad hoc “tricks" to achieve model convergence [179]. Nevertheless, the rapid development of GAN in the field of computer vision, such as the Pix2pix GAN [51] and Cycle-GAN [198], inspires us to exploit the potential combination of GANs and WSSs for the future work (See §6).

6 SENSING MODEL FORMULATION

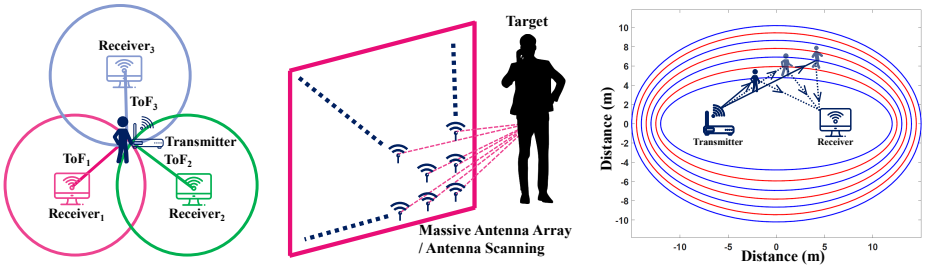
Given the specially designed features, the remaining challenge is to formulate a model to bridge the "illuminated" sourcing input with the final applied output.

Table 4. Summary of Existing Methodologies in terms of Model Formulation

Geometrical Model: Trilateration/Triangulation [7, 7, 58, 90, 118, 151, 174, 190, 190], Tracing Estimation [63, 78, 85, 97, 123, 150, 192], SAR Model [5, 8, 50, 56, 59, 86], Fresnel Zone Model [131, 151, 153], Skeleton Model [5, 53, 188, 190]
Statistical Model: Joint Multi-parameter Estimation [19, 55, 71, 99, 160], Lagrange Multiplier [25, 53, 56, 118, 181, 185, 194, 201], MUSIC [8, 27, 55, 58, 70, 71, 87, 88, 162], Propagation Model [47, 56, 72, 110, 119, 124, 131, 132]
ML-based Model : k-NN [91, 124, 132, 132, 161, 184], SVM [48, 65, 114, 125, 137, 184, 186, 187], Hidden Markov Models [92, 117, 138], Random Forest [72], Decision Tree [177, 178];
AI Approach: Customized Network [29, 49, 63, 65, 69, 114, 134, 135, 188–191, 194], Image Translation [14, 51, 53, 87, 87, 96, 126, 198], Multi-task Learning [29, 63, 126], Transfer Learning [134, 184]

6.1 Geometrical Model

Given the extracted features for distance and direction estimation, the geometric properties of triangles can be utilized to estimate the location with a direct (shortest) path for the transmitted signal, covering **trilateration** and **triangulation**. And the former locates the target with the distance measurement from multiple reference points while the latter leverages the AoA information with respect to the reference point for localization. The basic idea is to pinpoint the intersection of the potential curves corresponding to each pair of transceivers, say the intersection of ellipses for the device-free sensing or circles for the device-based sensing, illustrated in Figure 7(a). For example, given the ellipse whose foci are the transmit antenna and the receive antenna with ToF, WiTrack [7] can uniquely localize the target person using the intersection of two ellipses. It can be further extended into 3D using the intersection of three ellipsoids. To enhance the accuracy and reliability of geometrical trilateration, Chronos [118] first refines the distance measurements by utilizing geometrical constraints, imposed by the physical layouts of transceivers. Then it formulates the trilateration computation as a quadratic optimization problem. Besides, we can also extract the motion of the target using the continuous **tracing estimation** [63, 78, 85, 97, 123, 150, 192]. Specifically, Widar [97] extracts the PLCR of the target representing the radial moving velocity (See § 5.2) and formulates the tracking path as continuous splines of the segmented uniform motion, namely the successive tracking. And LANDMARK2.0 [78] explores the sequential frequency pattern conveyed by the RFID array and estimates the binary frequency trace sequence for the first tag-free RFID system for activity monitoring. Further, several works [5, 8, 50, 56, 59, 86] borrow the idea of the imaging radar system (e.g., SAR [31, 116]) for power distribution of the monitoring area, which can transmit a short narrow pulse and waits for the pulse to hit an object and return back [50]. And SAR model employs a massive antenna array (or equivalently, the antenna scanning) as receiver for the imaging task, illustrated in Figure 7(b). Besides, they develop the modified **SAR model** for wireless sensing. Specifically, Wision [50] integrates the beamforming at the transmitter to extract the depth information beyond the power distribution for Wi-Fi imaging. To provide the required accuracy of the moving path for the antenna scanning, Ubicarse [59] enables handheld devices to



(a) Trilateration via ToF. (b) SAR model via massive antennas. (c) Fresnel zones of Wi-Fi transceivers.

Fig. 7. Illustration of the geometrical model such as trilateration for device-based localization, SAR for imaging, and Fresnel zone model indicating the most reinforced (red) and degraded (blue) locations of reflector.

emulate large antenna arrays by twisting them along coarse-grained paths. Moreover, it combines the SAR based localization system with stereo-vision algorithms for object localization.

Another geometrical model is the **Fresnel Zone model** used for localization and tracking [131, 151, 153], representing one of a series of confocal prolate ellipsoidal regions of space between a transmitter and a receiver [101]. Illustrated in Figure 7(c), the Fresnel Zone model indicates a series of concentric ellipsoidal regions of alternating reinforced strength and weakened strength of the wave superposition. And it is caused by the phase change of the reflected path induced by the moving target, resulting in constructive and destructive interference for the phase of the wave superposition. Since the phase change $2\pi d/\lambda$ corresponds to the path of length d with the radio wave of wavelength λ. Given the multi-path effect for Wi-Fi sensing, the received signal can be expressed as a summation of all the paths, namely the wave superposition $\sum a_i e^{-j2\pi d_i/\lambda}$, where i is path number and a means attenuation coefficient of each path [153]. By focusing on the LoS and reflected paths for the target person, WiDir [153] observes that when the target person is along a radio propagation path, radio waves bouncing off those surfaces may either weaken or reinforce the phase with the signal that travels directly to the receiver, depending on the target's relative location to the pair of transceivers. Thus, it measures phase change dynamics from multiple Wi-Fi subcarriers and infers the walking direction of the target person in a device-free manner. Meanwhile, FingerDraw [151] employs the data-driven noise reduction, the CSI quotient (See § 4.1), for the Fresnel Zone model, achieving the sub-wavelength level finger motion tracking. LiFS [131] extracts the designed spatial power fading from CSI measurement, delivering a complex non-linear Fresnel model for the device-free localization. The last introduced geometrical model is the **skeleton model** [5, 53, 188, 190], widely used for inertial sensor measurements, in which the human body is formulated as individual rigid bodies with fixed length. WiPose [53] incorporates the skeleton knowledge [2] with the DL neural network and designs specific loss functions to represent the model restrictions, such as the connection and fix-length of the rigid bodies.

6.2 Statistical Model

Statistical models formulate the mapping from inputs to outputs for numerical optimization, which relies on empirical measurements or probability functions to characterize wireless channels [82].

To acquire the multi-dimensional and complementary information for sensing tasks, one intuitive idea is to extract multiple features to provide additional resilience to the noise, covering the spatial vantages, temporal snapshots and physical characteristics. Thus, we can leverage the **joint multi-parameter estimation** [19, 71, 99, 160] to integrate various features and derive the final results.

Specifically, IndoTrack advances MaTrack [70] by incorporates the Doppler shift into the existing AoA estimation of MaTrack. Widar2.0 further integrates the ToF, AoA, Doppler shift and amplitude attenuation and employs the sage [24] algorithm for computation efficiency. To improve the accuracy of the joint estimation, especially under the severe multi-path effect, md-Track [160] adds the extra AoD and designs a path separation algorithm to analysis each group of estimated parameters for the corresponding path iteratively, achieving the state-of-the-art in real time.

Another track is the **Lagrange Multiplier** [53, 56, 118, 181, 185, 194, 201], which resolves the under-determined problems by employing the regulation term to control the sparsity of solutions. To estimate the frequency response, Chronos [118] leverages the Non-uniform Discrete Fourier Transform (NDFT) [15] for measurements of multiple channels spaced uniformly, which can be re-formulated using the Lagrange multipliers as $min_p||\hat{h} - \mathcal{F}p||_2^2 + \alpha||p||_k$, where $\hat{h}$ is the measured wireless channels at different zero-subcarriers (See § 4.1) and $\mathcal{F}$ denotes the Fourier Matrix [118]. And $||\cdot||_k$ is the L_k norm while the factor α enforces the level of sparsity. It has been well-studied in convex optimization theory that the L_1 norm of a vector favors sparse solutions [118, 181] while others [53, 194] adopt the idea of compression sensing by applying the L_0 norm.

As an exemplar of the statistical model for numerical estimation, **MUSIC** [108] is proposed to extract the spatial feature accurately such as AoA and has multiple variants accordingly [8, 27, 58, 70, 71, 87, 88, 162]. Mathematically, we compute the eigenvectors of XX^H that correspond to the eigenvalue zero and that they are orthogonal to the steering vectors [58], where X is the measurement matrix corresponds to the CSI matrix itself. To make the derivation more robust and accurate, the number of antennas has to be larger than the number of propagation paths indoors [108]. And SpotFi [58] observes the multiple OFDM subcarriers can be equivalent to antennas for channel measurements. Given the "bad" coherence (See § 5.3), MaTrack [70] further proposes the Dynamic-MUSIC model to detect the subtle reflected signals from the human body while differentiating them from those signals bouncing off static objects (furniture, walls, etc.). Beyond deriving the AoA using multiple antennas, MUSIC can also be utilized for acoustic sensing, which usually lacks of a microphone array. And droneTrack [88] integrates the MUSIC model for FMCW detection to significantly improve the capability of resolving multiple paths and enhance the distance estimation. Additionally, RTrack [87] extends the MUSIC to the 2D AoA-distance profile, which jointly estimates the distance and AoA even under low SNR at the room scale.

Finally, multiple fading effects can also be utilized for applied outputs with the **propagation model** [47, 56, 72, 110, 119, 124, 131, 132]. Specifically, Karanam et al. estimate the binary occupancy state of each voxel in the 3D monitoring area by using loopy belief propagation [171], delivering a drone based 3D imaging system. Depatla et al. [25] achieve the similar 3D binary imaging through the wall by collecting signals of Wi-Fi and UWB via unmanned vehicles. And it associates the power measurement with each voxel with the Lagrange multiplier [21] and compress sensing techniques. And ReMix [119] models the signal refraction propagating through multiple in-body mediums by segmenting the propagation path with splines of piece-wise segments. Thus, it can localize the deep-tissue medical implant with 1.4cm accurately in the surface and depth error.

6.3 Machine Learning based Model

Machine learning, or shallow learning based models are often employed to find the boundaries of sourcing inputs for applied outputs. Illustrated in Table 4, multiple WSSs resort to ML for model formulation, especially the detection & recognition sensing tasks, including k Nearest Neighbor (k-NN) [91, 124, 132, 132, 161, 184], SVM [48, 114, 125, 137, 184, 186, 187], and Self-Organizing Map (SOM) [133]. Specifically, **k-NN** determines the classification with the majority vote of the ground truth labels if its k nearest neighbors. TagScan [132] extracts the designed feature representing

the ratio of RSSI and phase change only dependent on the target material, namely RP-rate. And it employs the k-NN for material recognition using RFID. While Wang et al. adopt k-NN to select good subcarriers of CSI for object risk estimation. And **SVM** separates sampling points of sourcing input by a set of hyper-planes in a high dimensional space to maximize the functional margin, such as the distance to the nearest training data points of any class [82]. For example, EQ-Radio [187] designs the $l_1 - SVM$ [197] for final applied outputs from the inputs while previous works treat the mapping problem as a search problem, which is computation-intensive due to exponentially large searching space. And it extracts the emotion-specific features from FMCW and differentiate among various emotional states automatically. Moreover, RF-Mehndi [186] employs the SVM with the polynomial kernel to analyze the impact of fingertip touching on the RFID tag array, achieving higher than 99% accuracy for user authentication among 15 users.

We also have other WSSs with various ML based models, say Hidden Markov Models (HMM) [92, 117, 138], Random Forest [72], Decision Tree [177, 178], and Naive Bayes [178]. Mathematically, **HMM** formulates the classification problem as a Markov process wherein the true states are hidden. CARM [138] constructs a HMM for each activity to estimate the mean vector and covariance matrix corresponding to each state and the transition probabilities for the HMM while Duet [117] represents the absolute locations of target as symbolic ones and extracts the entry and exit boundaries to such symbolic spaces (e.g., the doors to each room and the side of the couch for sitting). It then leverages a HMM for each symbolic space to reason about entry and exit events in a probabilistic manner. Besides, **Decision Tree** and **Naïve Bayes** can be utilized to find a branching rule to predict the target classes and construct a lightweight classifier based on the Bayes' theorem [82], respectively.

6.4 AI Approach toward the Sensing Model Formulation

Compared with the aforementioned three conventional methodologies for the model formulation, DL models can bridge the "illuminated" sourcing input with the final applied output more effectively, especially for highly fine-grained sensing tasks, say skeleton recovery, multi-person activity recognition and mesh reconstruction of humans. Multiple DL techniques can be transferred from other fields, such as **image translation** [14, 51, 53, 87, 87, 96, 126, 198], **multi-task learning** [29, 63, 126], and **transfer learning** [134, 184]. To demonstrate the effectiveness of tailoring processing for integration, we wrap up this part by presenting several WSSs with **customized networks** [29, 49, 63, 65, 69, 114, 134, 135, 188–191, 194]

Given the noisy RF signals, the spatial resolution keeps low due to the narrow bandwidth and the multi-path effect, such as the decimeter-level indoor human localization [70, 71, 97, 99, 160]. And the centimeter-level pose estimation and skeleton recovery seems impossible with existing signal processing methodologies, especially through the wall. For example, RF-Capture [5], the first work to capture the human skeleton without carrying any devices through the walls, leverages the voxel information of FMCW radar and coarse-to-fine antenna scan for the scratched skeleton reconstruction. Then RF-Pose [188] captures vertical and horizontal two-dimensional heatmaps of human skeleton using the reflected chirp signals and feeds them into the teacher-student DNN for 14 keypoints estimation of human body. Further, RF-Pose3D [190] achieves the multi-person 3D skeleton recovery by integrating the Region Proposal Network (RPN), used for object detection in the computer vision, for multi-target separation and feature extraction. And it achieves an average error of 4.5 cm for the keypoint estimation in the 3D Cartesian coordinate system. Further, the fine-grained pose estimation can bootstrap multiple wireless sensing tasks through occlusions. Resembles RF-Pose3D [190], RF-Action [69] designs the hierarchical co-occurrence network by introducing the attention module [120] and multi-proposal module for multi-person activity recognition. Meanwhile, RF-Avatar [189] integrates the skinned multi-person linear mesh representation model [79] and builds the trajectory-CNN model with the attention module for

the 3D human mesh recovery. To extract more persistent human-identifying features like the body size and shape for user identification, RF-ReID [29] incorporate the hierarchical attention module [168] and an environmental discriminator. Note that we observe that multiple matured DL models can be transferred from other fields for wireless sensing. For example, Proffitt et al. [96] incorporate the vision-based Mask R-CNN [44] on the recovered wave front of the imaging plane, rendering the static imaging objects with the classification. Given the raw wireless signals, Wang et al. [200] customize the multi-task DNN [105] with tailored loss functions of body segmentation and joint estimation for person perception and pose estimation. Based on the similarity of the images and wireless signals, specific wireless sensing tasks [14, 53, 96, 126] confront the similar problem formulation (e.g., image translation [51, 198], pose estimation [44, 105]) and adopt unmodified DL technique in computer vision. And we will explore more in § 8.2.

7 ISSUE & CHALLENGE

While DL techniques demonstrate the promising ability for WSSs, several important issues and challenges are still remained to be addressed with the employment of DL techniques. And we summarize them with the existing works, covering issues of the collaboration between traditional signal processing and DL techniques for scalability and generalization, interpretable and comprehensive systems design for privacy and security as well as robustness and sensibility, respectively.

7.1 Scalability & Generalization

DL techniques rely on massive high-quality data for a scaled and generalized performance. As the architecture increasingly complex and evolved, a larger data volume and a higher quality are required with more parameters to be learned and configured. Compared with other fields such as computer vision and natural language processing, wireless measurements are inherently massive but noisy (e.g., RSSI and CSI) from commercial devices (e.g., 5300 Wi-Fi NIC, Semtech SX1276 LoRa node). Since most wireless data collected by sensors and network equipment are frequently subject to loss, redundancy, mislabelling and class imbalance [179], rendering it much more difficult for the scalability and generalization of training and learning processes. For example, trained wireless systems generally suffer from the interference by multiple targets and transceivers as well as the background noise when transferred into a large-scale scenario [81, 134, 136, 184].

To make the WSSs scalable and generalizable, more efficient integrations between the traditional signal processing and DL techniques are required with the inherently massive but noisy wireless measurements. On one hand, we can extract components of wireless signals with definite physical characteristics by the **traditional signal processing**, such as PLCR, polarization, BVP, motion change pattern [53, 63, 194] (See § 5). Since they are inherently consistent when transferred to large-scale problems and new environments, free from the domain-dependent interference (See § 4.1). On the other hand, we can tailor the **deep learning technique** for the specific sensing task, especially to fuse multi-dimensional information with customized loss functions. In a conceptual sense, we cannot only give a more definite components of wireless measurements via **traditional signal processing** for learning efficiently, but also teach the neural network to learn what should be learnt with specially designed loss functions of **deep learning technique**. For example, Widar3.0 [194] designs the domain-independent BVP to represent the hand motion in the body coordinates while WiHF [63] proposes the motion change pattern to indicate the personalized arm motion and pause for the cross-domain gesture recognition. WiPose [53] incorporates the skeleton model for the cross-domain pose estimation via customized loss functions. It first designs the smooth loss to guarantee the smooth movement of joints across time. Then the rotation loss is added to penalize the deformed estimated joints given the rigid skeleton model. Further, more well-defined DL

techniques can be transferred from other fields, such as the GANs, transfer learning in computer vision [52, 68, 81, 134, 184]. And more details can be found in § 4.3 and § 8.3.

7.2 Privacy & Security

While the untraceable WSSs provides a non-intrusive and non-obtrusive sensing method for the real world, it also brings many privacy and security concerns, such as the multiple supervision applications including daily activities, breathing and respiration rate estimation, and pose estimation of humans, even through the wall. And it can cause serious damages to the victims if these sensing information is leaked and obtained by malicious hackers and attackers. Note that DL techniques can also be utilized to improve the wireless network security since it has been a hot topic for the AI security, which leverages AI to autonomously identify and respond to potential cyber threats based on similar or previous activity. Privacy and security issues can be however exacerbated with the introduced DL techniques. For example, we can fool the WSSs, especially the deep neural network, using faking signals. On one hand, it has been noticed most of the mainstream neural networks can be easily fooled into misclassification by adding only a small amount of noise into the original data. On the other hand, we can also forge the high-level wireless signals on purposes, such as the fake CSI spectrogram generated by XModal-ID [57].

To further alleviate privacy and security concerns, especially introduced by DL techniques, we have to make the processing and function of the deep neural network much more interpretable. For example, RTrack [87] evaluates the correlation between the expected motion parameters with the feature extracted by each layer of the specially designed RNN. And various PCA components corresponds to the target distance, AoA, radial and tangential speed, respectively. Thus, new protocols, policies, architectures, and algorithms can be designed accordingly given the knowledge of each procedure for the WSSs.

7.3 Robustness & Sensibility

The robustness and sensibility of the WSSs can be balanced by adapting the interference resilience and sensing granularity, depending on the sensing task. Specifically, the target person can be generalized as a single point with the moving torso for localization and tracking while the movement of body parts and joints has to be considered for the gesture recognition and pose estimation. And various PCA components [63] can be selected to represent corresponding body parts (See § 4.2). On one hand, robust WSSs require the extraordinary resilience to the interference and noise by sacrificing the spatial resolution, such as the CSS of LoRa resilient to the multi-path effect. On the other hand, sensitive WSSs adopt fine-grained CSI and FMCW for accurate sensing while relying on specially designed procedures for noise reduction, such as imaging [50, 86, 124, 199, 200] and multi-person localization and tracking [6, 55, 160, 190].

To achieve the balanced performance for the specific sensing task, a comprehensive system is in demand to adapt the robustness and sensibility of the sensing system. In a conceptual sense, a trade-off can be optimized between the interference resilience and the sensing granularity.

8 FUTURE TREND

Issues and challenges provide promising research topics with AI approaches toward wireless sensing. And we present the future trends to bootstrap the deep AI enabled ubiquitous wireless sensing across modalities, fields and frameworks, respectively. Specifically, the cross-modality sensing enhances the signal pre-processing with multiple kinds of input signals while cross-field sensing incorporates knowledges of other fields for high-level feature extraction. Further, the cross-framework sensing relies on well-defined deep AI frameworks transferred from vision or audio processing to optimize the sensing modal formulation.

8.1 Cross-modality Sensing

Each kind of input signals provides multi-dimensional and complementary measurements for the ubiquitous sensing. Beyond the focused signals status such as RSSI, CSI and FMCW of Wi-Fi, acoustic, and LoRa in this survey, multi-modality techniques can provide a promising research trend for further incorporation, covering the ultrasound, RFID, visible light, laser, mmWave, Bluetooth, the vision-based images and Inertial Measurement Units (IMU). And the feasibility and potential of the cross-modality sensing have been demonstrated in recent works [18, 52, 57, 59, 85, 96, 117, 126, 150]. Specifically, AmphiLight [18] integrates the laser light with ultrasound to deliver a bidirectional air-water wireless communication systems, achieving up to 5.04 Mbps at 6.5 m in the air and 2.5 m underwater. Since the laser light is capable of adapting to water dynamics for high-throughput communication while the ultrasonic can recover the dynamical wave surface for high-accuracy localization. Besides, CAT [85] incorporates IMU measurements (e.g., accelerometer and gyroscope) with the FMCW radio of acoustic to improve the localization accuracy while Ubicarse [59] combines the RF localization systems with the stereo-vision algorithms for object Geo-tagging. To meet the rigid time deadline for anti-wave generation, MUTE [113] deploys an extra RF device to forward the ambient noise over the wireless radio, rendering a glimpse into the actual sounds at the receiver side for instant noise cancellation. Since wireless signals travel much faster than sound. Systematically, the integration of cross-modality technique can improve existing WSSs in multiple aspects by providing a larger coverage, reducing the deployment cost and improving the robustness and generalization as well.

8.2 Cross-field Sensing

Exploring the cross-field observations cannot only extend the application range of wireless sensing but also provide more signal processing techniques and physical phenomenon for further ubiquitous sensing. For example, Radio Tomographic Imaging (RTI) has been researched broadly for imaging the attenuation caused by physical objects in wireless networks [25, 56, 96, 146, 147]. All however can only achieve the decimeter-level imaging for moving objects due to the low spatial resolution of wireless signals as well as the interfering background noise. A finger-grained wireless sensing has been encouraged by observations from other fields, covering the medical communication, PhotogAcoustic Tomography (PAT), micro-scale material, and coherent light propagation. On one hand, medical communication require to explore the potential for in-body wireless communication at the centimeter level, such as implant localization and charging [30, 119]. On the other hand, principles of PhotogAcoustic Tomography (PAT) and micro-scale material in physics verify the feasibility of finer-grained imaging of at the cellular level. Physically, PAT is an emerging imaging modality for preclinical research and clinical practice. The rationale is the acoustic detection of optical absorption from either endogenous chromophores or exogenous contrast agents, enabling the super-resolution observations of cellular and subcellular structures [156]. Since ultrasound scatters much less than light in tissue. Guo et al. [39] further leverage the property of the scattering medium for single-shot compression PAT, in which it can increase the imaging efficiency greatly with one-time optical illumination for the imaging target. We expect it the principal of PAT can guide the finger-grained wireless sensing for medical research. Besides, Holl et.al [47] explore the principals of the coherent light propagation and microscope imaging, indicating the feasibility of indoor holography via Wi-Fi radiations. And we regard it as a future trend for multi-dimensional and multi-level wireless sensing.

8.3 Cross-framework Sensing

From the perspective of both spatial and temporal dimensions, we point out that wireless signals have important similarities with videos or speech [179]. Specifically, both videos and massive packets of wireless signals are composed of sequences of "snapshots", which suggests that well-established frameworks for CV and NLP can be transferred into existing WSSs. We however observe that current DL techniques employed in the wireless sensing community largely rely on supervised learning. Given the large-scale of unlabelled and mislabeled wireless signals, unsupervised learning can be adopted to further enhance sensing ability of the WSSs, such as auto-encoder, restricted Boltzmann machine and GAN. For example, Cycle-gan [198] leverages the cycle consistency for the unpaired image-to-image translation, in which the image data used for training has no need to be paired for the original input and the ground truth. Thus, it can fully explore the unlabelled images, which is exactly suitable for massive unlabelled wireless signals for the imaging sensing task. Besides, we have multiple variants of GAN, which have been evaluated comprehensively in CV and NLP, such as the StackGAN [183] for the text-to-image synthesis, the 3D-GAN [154] to generate new 3D models of different objects, the Age-cGANs [143] to explore the face variance as aging as well as the AF-DCGAN [68] to accelerate the fingerprint dataset construction for Wi-Fi positioning. And we believe all have the potential to be utilized in a comparable wireless sensing problem. Another matured technique is to employ the transfer learning for the cross-domain scenarios in WSSs. And it can learn a transferable knowledge from a source domain to aid learning in a target domain, enabling the directly deployment of the trained model in new scenarios [134, 184]. For example, DFHGR [134] develops a scenario transferring network by designing a single scenario GAN to generate virtual samples of the recognized gestures, delivering a general gesture recognition system with less training effort, especially in new scenarios.

9 CONCLUSIONS

AI approaches have been increasingly employed in multiple fields and demonstrate significant potentials for applications, making it indispensable in WSSs toward the ubiquitous sensing. In this paper, we focus on the evolution, comparison and connection of AI approaches applied in existing WSSs and show several open and challenging issues. Given a comprehensive workflow for the general processing of WSSs, various modalities of sourcing inputs are discussed about respective advantages and limitations. We further summarize and compare existing researches and AI approaches from aspects of signal processing, feature design and model formulation and delivering multiple applied outputs, separately. To provide an encyclopedic view of AI approaches in the wireless sensing community, we present the remaining issues and challenges induced by DL techniques. Drawing from our experience, we discuss potentials and future trends of DL techniques toward WSSs by pinpointing the crossovers between wireless sensing and AI approaches across modalities of sourcing inputs, fields of applications and frameworks of matured DL techniques.

ACKNOWLEDGMENTS

We sincerely thank the anonymous reviewers for their helpful comments on the paper. This work was partially supported by National Science Foundation (NSF) Awards CNS-1909177, and CNS-1919154.

REFERENCES

[1] Emerald. https://www.emeraldinno.com/clinical.
[2] Karim Abdel-Malek and Jasbir Arora. 2013. *Human Motion Simulation: Predictive Dynamics.*
[3] Heba Abdelnasser, Khaled A. Harras, and Moustafa Youssef. 2015. UbiBreathe: A Ubiquitous Non-Invasive WiFi-Based Breathing Estimator. In *Proceedings of ACM MobiHoc.*

[4] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. 2015. Wigest: A Ubiquitous WiFi-based Gesture Recognition System. In *Proceedings of IEEE INFOCOM*.

[5] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Fredo Durand. 2015. Capturing the Human Figure through a Wall. (2015).

[6] Fadel Adib, Zachary Kabelac, and Dina Katabi. 2015. Multi-Person Localization via RF Body Reflections. In *Proceedings of USENIX NSDI*.

[7] Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2014. 3D Tracking via Body Radio Reflections. In *Proceedings of USENIX NSDI*.

[8] Fadel Adib and Dina Katabi. 2013. See through Walls with WiFi!. In *Proceedings of ACM SIGCOMM*.

[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart Homes That Monitor Breathing and Heart Rate. In *Proceedings of ACM CHI*.

[10] Milad Afzalan and Farrokh Jazizadeh. 2019. Indoor positioning based on visible light communication: A performance-based survey of real-world prototypes. *ACM Computing Surveys (CSUR)* (2019).

[11] Kamran Ali, Alex X. Liu, Wei Wang, and Muhammad Shahzad. 2015. Keystroke Recognition Using WiFi Signals. In *Proceedings of ACM MobiCom*.

[12] K. Ali, A. X. Liu, W. Wang, and M. Shahzad. 2017. Recognizing Keystrokes Using WiFi Devices. *IEEE Journal on Selected Areas in Communications* (2017).

[13] LoRa Alliance. 2019. LoRaWAN for Developer.. In *Retrieved Jan 10, 2019 from https://lora-alliance.org/lorawan-for-developers*.

[14] Roshan Ayyalasomayajula, Aditya Arun, Chenfeng Wu, Sanatan Sharma, Abhishek Sethi Deepak Vasisht, and Dinesh Bharadia. 2020. Deep Learning based Wireless Localization for Indoor Navigation. In *Proceedings of ACM MobiCom*.

[15] S. Bagchi and S. K. Mitra. 1996. The nonuniform discrete Fourier transform and its applications in filter design. I. 1-D. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* (1996).

[16] Leo L. Beranek and Tim J. Mellow (Eds.). 2012. *Acoustics: Sound Fields and Transducers*.

[17] Chao Cai, Rong Zheng, and Menglan Hu. 2019. A survey on acoustic sensing. *ArXiv* (2019).

[18] Charles J. Carver, Zhao Tian, Hongyong Zhang, Kofi M. Odame, Alberto Quattrini Li, and Xia Zhou. 2020. AmphiLight: Direct Air-Water Communication with Laser Light. In *Proceedings of USENIX NSDI*.

[19] Lili Chen, Jie Xiong, Xiaojiang Chen, Sunghoon Ivan Lee, Kai Chen, Dianhe Han, Dingyi Fang, Zhanyong Tang, and Zheng Wang. 2019. WideSee: Towards Wide-Area Contactless Wireless Sensing. In *Proceedings of ACM SenSys*.

[20] Xiaojiang Chen, Lili Chen, Chao Feng, Dingyi Fang, Jie Xiong, Zheng Wang, and Yong Cui. 2019. Sensing Our World Using Wireless Signals. *IEEE Internet Computing* (2019).

[21] Weng Cho Chew. 1995. *Waves and Fields in Inhomogeneous Media*. IEEE Press.

[22] Marcelo Cicconet, Vighnesh Birodkar, Mads Lund, Michael Werman, and Davi Geiger. 2017. A convolutional approach to reflection symmetry. *Pattern Recognition Letters* (2017).

[23] Ronan Collobert and Samy Bengio. 2004. Links between perceptrons, MLPs and SVMs. In *Proceedings of ICML*.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B* (1977).

[25] Saandeep Depatla, Chitra R. Karanam, and Yasamin Mostofi. 2017. Robotic Through-Wall Imaging: Radio-Frequency Imaging Possibilities with Unmanned Vehicles. *IEEE Antennas and Propagation Magazine* (2017).

[26] Yingfei Diao Diao, Minyue Fu, and Huanshui Zhang. 2010. An overview of range detection techniques for wireless sensor networks. In *Proceedings of World Congress on Intelligent Control and Automation*.

[27] Jian Ding and Ranveer Chandra. 2019. Towards Low Cost Soil Sensing Using Wi-Fi. In *Proceedings of ACM MobiCom*.

[28] Rashad Eletreby, Diana Zhang, Swarun Kumar, and Osman Yağan. 2017. Empowering Low-Power Wide Area Networks in Urban Settings. In *Proceedings of ACM SIGCOMM*.

[29] Lijie Fan, Tianhong Li, Rongyao Fang, Rumen Hristov, Yuan Yuan, and Dina Katabi. 2020. Learning Longterm Representations for Person Re-Identification Using Radio Signals. In *Proceedings of the IEEE CVPR*.

[30] Xiaoran Fan, Longfei Shangguan, Richard Howard, Yanyong Zhang, Yao Peng, Jie Xiong, Yunfei Ma, and Xiang-Yang Li. 2020. Towards Flexible Wireless Charging for Medical Implants Using Distributed Antenna System. In *Proceedings of ACM MobiCom*.

[31] J. Patrick Fitch. 1988. *Synthetic Aperture Radar*. Springer-Verlag.

[32] Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised Domain Adaptation by Backpropagation. In *Proceedings of ICML*. JMLR.org.

[33] Chuhan Gao, Yilong Li, and Xinyu Zhang. 2018. LiveTag: Sensing Human-Object Interaction through Passive Chipless WiFi Tags. In *Proceedings of USENIX NSDI*.

[34] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang. 2017. CSI-Based Device-Free Wireless Localization and Activity Recognition Using Radio Image Features. *IEEE Transactions on Vehicular Technology* (2017).

[35] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

[36] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Proceedings of NIPS*.
[37] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. 2003. KNN Model-Based Approach in Classification. (2003).
[38] Xiaonan Guo, Bo Liu, Cong Shi, Hongbo Liu, Yingying Chen, and Mooi Choo Chuah. 2017. WiFi-Enabled Smart Human Dynamics Monitoring. In *Proceedings of ACM SenSys*.
[39] Yuning Guo, Baowen Li, and Xiaobo Yin. 2020. Single-Shot Compressed Photoacoustic Tomographic Imaging with a Single Detector in a Scattering Medium. *Phys. Rev. Applied* (2020).
[40] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Predictable 802.11 packet delivery from wireless channel measurements. In *Proceedings of ACM SIGCOMM*.
[41] C. Han, K. Wu, Y. Wang, and L. M. Ni. 2014. WiFall: Device-free fall detection by wireless networks. In *Proceedings of IEEE INFOCOM*.
[42] Amer A. Hassan, Wayne E. Stark, John E. Hershey, and Sandeep Chennakeshu. 1996. Cryptographic Key Agreement for Mobile Radio. *Digital Signal Processing* (1996).
[43] Naveed Ul Hassan, Aqsa Naeem, Muhammad Adeel Pasha, Tariq Jadoon, and Chau Yuen. 2015. Indoor positioning using visible led lights: A survey. *Comput. Surveys* (2015).
[44] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. In *Proceedings of IEEE/CVF ICCV*.
[45] Marti A. Hearst. 1998. Support Vector Machines. *IEEE Intelligent Systems* (1998).
[46] Juha Heiskala and John Terry. 2001. *OFDM Wireless LANs: A Theoretical and Practical Guide*. Sams.
[47] Philipp M. Holl and Friedemann Reinhard. 2017. Holography of Wi-fi Radiation. *Phys. Rev. Lett.* (2017).
[48] Feng Hong, Xiang Wang, Yanni Yang, Yuan Zong, Yuliang Zhang, and Zhongwen Guo. 2016. WFID: Passive Device-free Human Identification Using WiFi Signal. In *Proceedings of EAI MobiQuitous*.
[49] Chen-Yu Hsu, Rumen Hristov, Guang-He Lee, Mingmin Zhao, and Dina Katabi. 2019. Enabling Identification and Behavioral Sensing in Homes Using Radio Reflections. In *Proceedings of CHI*.
[50] Donny Huang, Rajalakshmi Nandakumar, and Shyamnath Gollakota. 2014. Feasibility and Limits of Wi-fi Imaging. In *Proceedings of ACM SenSys*.
[51] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *Proceedings of IEEE CVPR*.
[52] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, Wenyao Xu, and Lu Su. 2018. Towards Environment Independent Device Free Human Activity Recognition. In *Proceedings of ACM MobiCom*.
[53] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun Sun, and Lu Su. 2020. Towards 3D Human Pose Construction Using WiFi. In *Proceedings of ACM MobiCom*.
[54] Zhiping Jiang, Jizhong Zhao, Xiang-Yang Li, Jinsong Han, and Wei Xi. 2013. Rejecting the Attack: Source Authentication for Wi-Fi Management Frames using CSI Information. (2013).
[55] Chitra R. Karanam, Belal Korany, and Yasamin Mostofi. 2019. Tracking from One Side: Multi-Person Passive Tracking with WiFi Magnitude Measurements. In *Proceedings of ACM IPSN*.
[56] Chitra R. Karanam and Yasamin Mostofi. 2017. 3D Through-wall Imaging with Unmanned Aerial Vehicles Using Wifi. In *Proceedings of ACM IPSN*.
[57] Belal Korany, Chitra R. Karanam, Hong Cai, and Yasamin Mostofi. 2019. XModal-ID: Using WiFi for Through-Wall Person Identification from Candidate Video Footage. In *Proceedings of ACM MobiCom*.
[58] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, Sachin Katti, Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. 2015. *SpotFi: Decimeter Level Localization Using WiFi*. Proceedings of ACM SIGCOMM.
[59] Swarun Kumar, Stephanie Gil, Dina Katabi, and Daniela Rus. 2014. Accurate Indoor Localization with Zero Start-up Cost. In *Proceedings of ACM MobiCom*.
[60] K. Lam, C. Cheung, and W. Lee. 2017. LoRa-based localization systems for noisy outdoor environment. In *Proceedings of IEEE WiMob*.
[61] G. Lerosey, J. d. Rosny, A. Tourin, A. Derode, G. Montaldo, and M. Fink. 2005. Time reversal of electromagnetic waves and telecommunication. *Radio Science* (2005).
[62] Geoffroy Lerosey, Julien de Rosny, Arnaud Tourin, and Mathias Fink. 2007. Focusing beyond the diffraction limit with far-field time reversal. *Science* (2007).
[63] Chenning Li, Manni Liu, and Zhichao Cao. 2020. WiHF: Enable User Identified Gesture Recognition with WiFi. In *Proceedings of IEEE INFOCOM*.
[64] C. L. Li, M. Liu, and Z. Cao. 2020. WiHF: Gesture and User Recognition with WiFi. *IEEE Transactions on Mobile Computing* (2020). https://doi.org/10.1109/TMC.2020.3009561

[65] He Li, Kaoru Ota, Mianxiong Dong, and Minyi Guo. 2018. Learning Human Activities through Wi-Fi Channel State Information with Multiple Access Points. *IEEE Communications Magazine* (2018).
[66] Hong Li, Wei Yang, Jianxin Wang, Yang Xu, and Liusheng Huang. 2016. WiFinger: Talk to Your Smart Devices with Finger-Grained Gesture. In *Proceedings of ACM UbiComp*.
[67] Mengyuan Li, Yan Meng, Junyi Liu, Haojin Zhu, Xiaohui Liang, Yao Liu, and Na Ruan. 2016. When CSI Meets Public WiFi: Inferring Your Mobile Phone Password via WiFi Signals. In *Proceedings of ACM CCS*.
[68] Qiyue Li, Heng Qu, Zhi Liu, Nana Zhou, Wei Sun, Stephan Sigg, and Jie Li. 2020. AF-DCGAN: Amplitude Feature Deep Convolutional GAN for Fingerprint Construction in Indoor Localization Systems. *IEEE Transactions on Emerging Topics in Computational Intelligence* (2020).
[69] Tianhong Li, Lijie Fan, Mingmin Zhao, Yingcheng Liu, and Dina Katabi. 2019. Making the Invisible Visible: Action Recognition Through Walls and Occlusions. In *Proceedings of IEEE/CVF ICCV*.
[70] Xiang Li, Shengjie Li, Daqing Zhang, Jie Xiong, Yasha Wang, and Hong Mei. 2016. *Dynamic-MUSIC: accurate device-free indoor localization*.
[71] Xiang Li, Daqing Zhang, Qin Lv, Jie Xiong, Shengjie Li, Yue Zhang, and Hong Mei. 2017. IndoTrack: Device-Free Indoor Human Tracking with Commodity Wi-Fi. *Proceedings of ACM IMWUT* (2017).
[72] Yuxiang Lin, Wei Dong, Yi Gao, and Tao Gu. 2020. SateLoc: A Virtual Fingerprinting Approach to Outdoor LoRa Localization using Satellite Images. In *Proceedings of IEEE IPSN*.
[73] H. Liu, Y. Wang, J. Liu, J. Yang, Y. Chen, and H. V. Poor. 2018. Authenticating Users Through Fine-Grained Channel Information. *IEEE Transactions on Mobile Computing* (2018).
[74] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang. 2019. Wireless Sensing for Human Activity: A Survey. *IEEE Communications Surveys & Tutorials* (2019).
[75] Manni Liu, Linsong Cheng, Kun Qian, Jiliang Wang, Jiongtao Wang, and Yunhao Liu. 2020. Indoor acoustic localization: a survey. *Human-centric Computing and Information Sciences* (2020).
[76] X. Liu, J. Cao, S. Tang, and J. Wen. 2014. Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals. In *Proceedings of IEEE Real-Time Systems Symposium*.
[77] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo. 2016. Contactless Respiration Monitoring Via Off-the-Shelf WiFi Devices. *IEEE Transactions on Mobile Computing* (2016).
[78] Y. Liu, L. Chen, J. Pei, Q. Chen, and Y. Zhao. 2007. Mining Frequent Trajectory Patterns for Activity Monitoring Using Radio Frequency Tag Arrays. In *Proceedings of IEEE PERCOM*.
[79] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graph.* (2015).
[80] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. 2016. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. In *Proceedings of NIPS*.
[81] Xiaorui Ma, Yunong Zhao, Liang Zhang, Qinghua Gao, Miao Pan, and Jie Wang. 2020. Practical Device-Free Gesture Recognition Using WiFi Signals Based on Metalearning. *IEEE Transactions on Industrial Informatics* (2020).
[82] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi Sensing with Channel State Information: A Survey. (2019).
[83] Yongsen Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. 2018. SignFi: Sign Language Recognition Using WiFi. (2018).
[84] Bassem R. Mahafza. 2000. *Radar Systems Analysis and Design Using MATLAB*. CRC Press, Inc.
[85] Wenguang Mao, Jian He, Huihuang Zheng, Zaiwei Zhang, and Lili Qiu. 2016. High-Precision Acoustic Motion Tracking: Demo. In *Proceedings of ACM MobiCom*.
[86] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. AIM: Acoustic Imaging on a Mobile. In *Proceedings of ACM MobiSys*.
[87] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-Based Room Scale Hand Motion Tracking. In *Proceedings of ACM MobiCom*.
[88] Wenguang Mao, Zaiwei Zhang, Lili Qiu, Jian He, Yuchen Cui, and Sangki Yun. 2017. Indoor Follow Me Drone. In *Proceedings of ACM MobiSys*.
[89] Pedro Melgarejo, Xinyu Zhang, Parameswaran Ramanathan, and David Chu. 2014. Leveraging Directional Antenna Capabilities for Fine-Grained Gesture Recognition. In *Proceedings of ACM UbiComp*.
[90] Rajalakshmi Nandakumar, Vikram Iyer, and Shyamnath Gollakota. 2018. 3D Localization for Sub-Centimeter Sized Devices. In *Proceedings of ACM SenSys*.
[91] L. M. Ni, Yunhao Liu, Yiu Cho Lau, and A. P. Patil. 2003. LANDMARC: indoor location sensing using active RFID. In *Proceedings of IEEE PERCOM*.
[92] Kazuya Ohara, Takuya Maekawa, and Yasuyuki Matsushita. 2017. Detecting State Changes of Indoor Everyday Objects Using Wi-Fi Channel State Information. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (2017).
[93] Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In-So Kweon. 2018. BAM: Bottleneck Attention Module. (2018).

[94] Swadhin Pradhan, Ghufran Baig, Wenguang Mao, Lili Qiu, Guohai Chen, and Bo Yang. 2018. Smartphone-Based Acoustic Indoor Space Mapping. *Proceedings of ACM IMWUT* (2018).
[95] Pichaya Prasertsung and Teerayut Horanont. 2017. How Does Coffee Shop Get Crowded? Using WiFi Footprints to Deliver Insights into the Success of Promotion. In *Proceedings of ACM UbiComp*.
[96] Paul C. Proffitt and Honggang Wang. 2018. Static Object Wi-Fi Imaging and Classifier. In *Proceedings of IEEE International Symposium on Technologies for Homeland Security (HST)*.
[97] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. 2017. Widar: Decimeter-Level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi. In *Proceedings of ACM Mobihoc*.
[98] K. Qian, C. Wu, Z. Yang, Y. Liu, and Z. Zhou. 2014. PADS: Passive detection of moving targets with dynamic speed using PHY layer information. In *Proceedings of ICPADS*.
[99] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2.0: Passive Human Tracking with a Single Wi-Fi Link. In *Proceedings of ACM MobiSys*.
[100] A B M Mohaimenur Rahman, Ting Li, and Yu Wang. 2020. Recent Advances in Indoor Localization via Visible Lights: A Survey. *Sensors* (2020).
[101] Theodore S Rappaport et al. 1996. *Wireless communications: principles and practice*. prentice hall PTR New Jersey.
[102] S. Ren, K. He, R. Girshick, and J. Sun. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).
[103] Irina Rish. 2001. An Empirical Study of the Naïve Bayes Classifier. *IJCAI 2001 Work Empir Methods Artif Intell* (2001).
[104] Gianfranco La Rocca. 2012. Knowledge based engineering: Between AI and CAD. Review of a language based technology to support engineering design. *Advanced Engineering Informatics* (2012).
[105] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of Springer MICCAI*.
[106] S. Sahin, M.R. Tolun, and R. Hassanpour. 2012. Hybrid expert systems: A survey of current approaches and applications. *Expert Systems with Applications* (2012).
[107] Stefano Savazzi, Stephan Sigg, Monica Nicoli, Vittorio Rampa, Sanaz Kianoush, and Umberto Spagnolini. 2016. Device-Free Radio Vision for Assisted Living: Leveraging wireless channel quality information for human sensing. *IEEE Signal Processing Magazine* (2016).
[108] R. Schmidt. 1986. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation* (1986).
[109] S. Y. Seidel and T. S. Rappaport. 1992. 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE Transactions on Antennas and Propagation* (1992).
[110] Souvik Sen, Jeongkeun Lee, Kyu-Han Kim, and Paul Congdon. 2013. Avoiding Multipath to Revive Inbuilding WiFi Localization. In *Proceedings of ACM MobiSys*.
[111] Muhammad Shahzad and Shaohu Zhang. 2018. Augmenting User Identification with WiFi Based Gesture Recognition. *Proceedings on ACM IMWUT* (2018).
[112] J. P. Shanmuga Sundaram, W. Du, and Z. Zhao. 2020. A Survey on LoRa Networking: Research Problems, Current Solutions, and Open Issues. *IEEE Communications Surveys & Tutorials* (2020).
[113] Sheng Shen, Nirupam Roy, Junfeng Guan, Haitham Hassanieh, and Romit Roy Choudhury. 2018. MUTE: Bringing IoT to Noise Cancellation. In *Proceedings of SIGCOMM*.
[114] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 2017. Smart User Authentication Through Actuation of Daily Activities Leveraging WiFi-enabled IoT. In *ACM on ACM MobiHoc*.
[115] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. 2015. WiDraw: Enabling Hands-Free Drawing in the Air on Commodity WiFi Devices. In *Proceedings of ACM MobiCom*.
[116] Paul Thompson, Daniel E. Wahl, Paul H. Eichel, Dennis C. Ghiglia, and Charles V. Jakowatz. 1996. *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Kluwer Academic Publishers.
[117] Deepak Vasisht, Anubhav Jain, Chen-Yu Hsu, Zachary Kabelac, and Dina Katabi. 2018. Duet: Estimating User Position and Identity in Smart Homes Using Intermittent and Incomplete RF-Data. *Proceedings on ACM IMWUT* (2018).
[118] Deepak Vasisht, Swarun Kumar, and Dina Katabi. 2016. Decimeter-Level Localization with a Single WiFi Access Point. In *Proceedings of USENIX NSDI*.
[119] Deepak Vasisht, Guo Zhang, Omid Abari, Hsiao-Ming Lu, Jacob Flanz, and Dina Katabi. 2018. In-Body Backscatter Communication and Localization. In *Proceedings of ACM SIGCOMM*.
[120] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proceedings of NIPS*.
[121] Raghav H Venkatnarayan, Griffin Page, and Muhammad Shahzad. 2018. Multi-User Gesture Recognition Using WiFi.. In *Proceedings of ACM MobiSys*.
[122] Aditya Virmani and Muhammad Shahzad. 2017. *Position and Orientation Agnostic Gesture Recognition Using WiFi*.

[123] Anran Wang, Jacob E. Sunshine, and Shyamnath Gollakota. 2019. Contactless Infant Monitoring Using White Noise. In *Proceedings of ACM MobiCom*.
[124] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang. 2018. Towards In-baggage Suspicious Object Detection Using Commodity WiFi. In *Proceedings of IEEE CNS*.
[125] Fei Wang, Jinsong Han, Ziyi Dai, Han Ding, and Dong Huang. 2019. WiPIN: Operation-free Person Identification Using WiFi Signals. In *Proceedings of IEEE GLOBECOM*.
[126] Fei Wang, Sanping Zhou, Stanislav Panev, Jinsong Han, and Dong Huang. 2019. Person-in-WiFi: Fine-Grained Person Perception Using WiFi. In *Proceedings of IEEE/CVF ICCV)*.
[127] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. 2016. Human Respiration Detection with Commodity Wifi Devices: Do User Location and Body Orientation Matter?. In *Proceedings of ACM UbiComp*.
[128] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li. 2017. RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices. *IEEE Transactions on Mobile Computing* (2017).
[129] Jie Wang, Qinhua Gao, Xiaorui Ma, Yunong Zhao, and Yuguang Fang. 2020. Learning to Sense: Deep Learning for Wireless Sensing with Less Training Efforts. *IEEE Wireless Communications* (2020).
[130] Jie Wang, Qinhua Gao, Miao Pan, and Yuguang Fang. 2018. Device-Free Wireless Sensing: Challenges, Opportunities, and Applications. *IEEE Network* (2018).
[131] Ju Wang, Hongbo Jiang, Jie Xiong, Kyle Jamieson, Xiaojiang Chen, Dingyi Fang, and Binbin Xie. 2016. LiFS: Low Human-Effort, Device-Free Localization with Fine-Grained Subcarrier Information. In *Proceedings of ACM MobiCom*.
[132] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. 2017. TagScan: Simultaneous Target Imaging and Material Identification with Commodity RFID Devices. In *Proceedings of ACM MobiCom*.
[133] J. Wang, L. Zhang, Q. Gao, M. Pan, and H. Wang. 2018. Device-Free Wireless Sensing in Complex Scenarios Using Spatial Structural Information. *IEEE Transactions on Wireless Communications* (2018).
[134] Jie Wang, Liang Zhang, Changcheng Wang, Xiaorui Ma, Qinghua Gao, and Bin Lin. 2020. Device-Free Human Gesture Recognition With Generative Adversarial Networks. *IEEE Internet of Things Journal* (2020).
[135] Jie Wang, Xiao Zhang, Qinhua Gao, Hao Yue, and Hongyu Wang. 2017. Device-Free Wireless Localization and Activity Recognition: A Deep Learning Approach. *IEEE Transactions on Vehicular Technology* (2017).
[136] Jie Wang, Yunong Zhao, Xiaorui Ma, Qinghua Gao, Miao Pan, and Hongyu Wang. 2020. Cross-Scenario Device-Free Activity Recognition Based on Deep Adversarial Networks. *IEEE Transactions on Vehicular Technology* (2020).
[137] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait Recognition Using WiFi Signals. In *Proceedings of ACM UbiComp*.
[138] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In *Proceedings of ACM MobiCom*.
[139] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: Device-free Location-oriented Activity Identification Using Fine-grained WiFi Signatures. In *Proceedings of ACM MobiCom*.
[140] Yan Wang, Jie Yang, Yingying Chen, Hongbo Liu, Marco Gruteser, and Richard P. Martin. 2014. Tracking Human Queues Using Single-Point Signal Monitoring. In *Proceedings of ACM MobiSys*.
[141] Yan Wang, Jie Yang, Hongbo Liu, Yingying Chen, Marco Gruteser, and Richard P. Martin. 2013. Measuring Human Queues Using WiFi Signals. In *Proceedings of ACM MobiCom*.
[142] Zhu Wang, Bin Guo, Zhiwen Yu, and Xingshe Zhou. 2018. Wi-Fi CSI-Based Behavior Recognition: From Signals and Actions to Activities. *IEEE Communications Magazine* (2018).
[143] Z. Wang, W. Luo X. Tang, and S. Gao. 2018. Face Aging with Identity-Preserved Conditional Generative Adversarial Networks. In *Proceedings of IEEE CVPR*.
[144] M. Wax, Tie-Jun Shan, and T. Kailath. 1984. Spatio-temporal spectral analysis by eigenstructure methods. *IEEE Transactions on Acoustics, Speech, and Signal Processing* (1984).
[145] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. 2019. From Real to Complex: Enhancing Radio-Based Activity Recognition Using Complex-Valued CSI. *ACM Transactions on Sensor Networks* 15, 3 (2019).
[146] J. Wilson and N. Patwari. 2010. Radio Tomographic Imaging with Wireless Networks. *IEEE Transactions on Mobile Computing* (2010).
[147] Joey Wilson and Neal Patwari. 2011. See-Through Walls: Motion Tracking Using Variance-Based Radio Tomography Networks. *IEEE Transactions on Mobile Computing* (2011).
[148] M. Won, S. Zhang, and S. H. Son. 2017. WiTraffic: Low-Cost and Non-Intrusive Traffic Monitoring System Using WiFi. In *Proceedings of the 26th International Conference on Computer Communication and Networks*.
[149] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. CBAM: Convolutional Block Attention Module. (2018).

[150] Chenshu Wu, Feng Zhang, Yusen Fan, and K. J. Ray Liu. 2019. RF-Based Inertial Measurement. In *Proceedings of ACM SIGCOMM*.
[151] Dan Wu, Ruiyang Gao, Youwei Zeng, Jinyi Liu, Leye Wang, Tao Gu, and Daqing Zhang. 2020. FingerDraw: Sub-Wavelength Level Finger Motion Tracking with WiFi Signals. (2020).
[152] Dan Wu, Daqing Zhang, Chenren Xu, Hao Wang, and Xiang Li. 2017. Device-Free WiFi Human Sensing: From Pattern-Based to Model-Based Approaches. *IEEE Communications Magazine* (2017).
[153] Dan Wu, Daqing Zhang, Chenren Xu, Yasha Wang, and Hao Wang. 2016. WiDir: walking direction estimation using wireless signals. In *Proceedings of ACM UbiComp*.
[154] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T Freeman, and Joshua B Tenenbaum. 2016. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. In *Proceedings of NIPS*.
[155] Z. Wu, Y. Han, Y. Chen, and K. J. R. Liu. 2015. A Time-Reversal Paradigm for Indoor Positioning System. *IEEE Transactions on Vehicular Technology* (2015).
[156] Jun Xia, Junjie Yao, and Lihong Wang. 2014. Photoacoustic tomography: principles and advances. In *Electromagnetic waves (Cambridge, Mass.)*.
[157] Xianjin Xia, Yuanqing Zheng, and Tao Gu. 2019. FTrack: Parallel Decoding for LoRa Transmissions. In *Proceedings of ACM SenSys*.
[158] Jiang Xiao, Zimu Zhou, Youwen Yi, and Lionel M. Ni. 2016. A Survey on Wireless Indoor Localization from the Device Perspective. *ACM Computing Surveys*. (2016).
[159] Yaxiong Xie, Zhenjiang Li, and Mo Li. 2015. Precise Power Delay Profiling with Commodity WiFi. In *Proceedings of ACM MobiCom*.
[160] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. 2019. mD-Track: Leveraging Multi-Dimensionality for Passive Indoor Wi-Fi Tracking. In *Proceedings of ACM MobiCom*.
[161] Tong Xin, Bin Guo, Zhu Wang, Mingyang Li, Zhiwen Yu, and Xingshe Zhou. 2016. Freesense: Indoor Human Identification with WiFi Signals. In *Proceedings of IEEE GLOBECOM)*.
[162] Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: A Fine-Grained Indoor Location System. In *Proceedings of USENIX NSDI*.
[163] Jie Xiong and Kyle Jamieson. 2013. SecureArray: Improving Wifi Security with Fine-Grained Physical-Layer Information. In *Proceedings of ACM MobiCom*.
[164] Yang Xu, Wei Yang, Jianxin Wang, Xing Zhou, Hong Li, and Liusheng Huang. 2018. WiStep: Device-Free Step Counting with WiFi Signals. In *Proceedings of ACM IMWUT*.
[165] Hongfei Xue, Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shiyang Wang, Ye Yuan, Shuochao Yao, Aidong Zhang, and Lu Su. 2020. DeepMV: Multi-View Deep Learning for Device-Free Human Activity Recognition. *Proceedings of ACM IMWUT*. (2020).
[166] Hongfei Xue, Wenjun Jiang, Chenglin Miao, Ye Yuan, Fenglong Ma, Xin Ma, Yijiang Wang, Shuochao Yao, Wenyao Xu, Aidong Zhang, and Lu Su. 2019. DeepFusion: A Deep Learning Framework for the Fusion of Heterogeneous Sensory Data. In *Proceedings of ACM Mobihoc*.
[167] Chen Yan, Yan Long, Xiaoyu Ji, and Wenyuan Xu. 2019. The Catcher in the Field: A Fieldprint Based Spoofing Detection for Text-Independent Speaker Verification. In *Proceedings of the ACM CCS*.
[168] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical Attention Networks for Document Classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
[169] Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor localization via channel response. *ACM Comput. Surv* (2013).
[170] Lexiang Ye and Eamonn Keogh. 2009. Time Series Shapelets: A New Primitive for Data Mining. In *Proceedings of ACM SIGKDD*.
[171] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. 2000. Generalized Belief Propagation. In *Proceedings of NIPS*.
[172] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee. 2017. A Survey on Behavior Recognition Using WiFi Channel State Information. *IEEE Communications Magazine* (2017).
[173] Shichao Yue and Dina Katabi. 2019. Liquid Testing with Your Smartphone. In *Proceedings of ACM MobiSys*.
[174] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. 2017. Strata: Fine-Grained Acoustic-based Device-Free Tracking. In *Proceedings of ACM MobiSys*.
[175] Valentin Zacharias. 2008. Development and Verification of Rule Based Systems — A Survey of Developers. In *Rule Representation, Interchange and Reasoning on the Web*, Nick Bassiliades, Guido Governatori, and Adrian Paschke (Eds.). Springer Berlin Heidelberg.
[176] F. Zafari, A. Gkelias, and K. K. Leung. 2019. A Survey of Indoor Localization Systems and Technologies. *IEEE Communications Surveys & Tutorials* (2019).

[177] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-based Person Identification in Smart Spaces. In *Proceedings of ACM/IEEE IPSN*.
[178] Yunze Zeng, Parth H. Pathak, Chao Xu, and Prasant Mohapatra. 2014. Your AP Knows How You Move: Fine-Grained Device Motion Recognition through WiFi. In *Proceedings of HotWireless*.
[179] C. Zhang, P. Patras, and H. Haddadi. 2019. Deep Learning in Mobile and Wireless Networking: A Survey. *IEEE Communications Surveys & Tutorials* (2019).
[180] D. Zhang, H. Wang, and D. Wu. 2017. Toward Centimeter-Scale Human Activity Sensing with Wi-Fi Signals. *Computer* (2017).
[181] Diana Zhang, Jingxian Wang, Junsu Jang, Junbo Zhang, and Swarun Kumar. 2019. On the Feasibility of Wi-Fi Based Material Sensing. In *Proceedings of ACM MobiCom*.
[182] F. Zhang, C. Chen, B. Wang, and K. J. R. Liu. 2018. WiSpeed: A Statistical Electromagnetic Approach for Device-Free Indoor Speed Estimation. *IEEE Internet of Things Journal* (2018).
[183] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. 2017. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. In *Proceedings of IEEE/CVF ICCV*.
[184] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, and Zheng Wang. 2018. CrossSense: Towards Cross-Site and Large-Scale WiFi Sensing. In *Proceedings of ACM MobiCom*.
[185] Jin Zhang, Bo Wei, Wen Hu, and Salil S Kanhere. 2016. Wifi-id: Human Identification Using WiFi Signal. In *Proceedings of IEEE DCOSS*.
[186] Cui Zhao, Zhenjiang Li, Han Ding, Jinsong Han, Wei Xi, Ting Liu, and Ruowei Gui. 2019. RF-Mehndi: A Fingertip Profiled RF Identifier. In *Proceedings on IEEE INFOCOM*.
[187] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of ACM MobiCom*.
[188] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi. 2018. Through-Wall Human Pose Estimation Using Radio Signals. In *Proceedings of IEEE CVPR*.
[189] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Tianhong Li, Hang Zhao, Antonio Torralba, and Dina Katabi. 2019. Through-Wall Human Mesh Recovery Using Radio Signals. In *Proceedings of IEEE/CVF ICCV*.
[190] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-based 3D Skeletons. In *Proceedings of ACM SIGCOMM*.
[191] Mingmin Zhao, Shichao Yue, Dina Katabi, Tommi S Jaakkola, and Matt T Bianchi. 2017. Learning Sleep Stages from Radio Signals - A Conditional Adversarial Architecture. In *Proceedings on ICML*.
[192] Xiaolong Zheng, J. Wang, Longfei Shangguan, Zimu Zhou, and Yunhao Liu. 2016. Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures. In *Proceedings of IEEE INFOCOM*.
[193] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu. 2017. Design and Implementation of a CSI-Based Ubiquitous Smoking Detection System. *IEEE/ACM Transactions on Networking* (2017).
[194] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2019. Zero-Effort Cross-domain Gesture Recognition with WiFi. In *Proceedings of ACM MobiSys*.
[195] Z. Zhou, C. Wu, Z. Yang, and Y. Liu. 2015. Sensorless Sensing with WiFi. *Tsinghua Science and Technology* 20, 1 (2015).
[196] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu. 2014. Omnidirectional Coverage for Device-Free Passive Human Detection. *IEEE Transactions on Parallel and Distributed Systems* (2014).
[197] Ji Zhu, Saharon Rosset, Trevor Hastie, and Rob Tibshirani. 2003. 1-Norm Support Vector Machines. In *Proceedings of NIPS*.
[198] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proceedings of IEEE/CVF ICCV*.
[199] Yanzi Zhu, Yuanshun Yao, Ben Y. Zhao, and Haitao Zheng. 2017. Object Recognition and Navigation Using a Single Networking Device. In *Proceedings of ACM MobiSys*.
[200] Yanzi Zhu, Yibo Zhu, Ben Y. Zhao, and Haitao Zheng. 2015. Reusing 60GHz Radios for Mobile Radar Imaging. In *Proceedings of ACM MobiCom*.
[201] Han Zou, Yuxun Zhou, Jianfei Yang, Weixi Gu, Lihua Xie, and Costas J Spanos. 2018. WiFi-based Human Identification via Convex Tensor Shapelet Learning. In *Proceedings of AAAI*.
[202] Yongpan Zou, Weifeng Liu, Kaishun Wu, and Lionel M. Ni. 2017. Wi-Fi Radar: Recognizing Human Behavior with Commodity Wi-Fi. *IEEE Communications Magazine* (2017).

Note that we list the abbreviations used throughout Table 6, 7, 8 in Table 5.

Table 5. List of abbreviations in alphabetical order

SS	Signal Status	**DB**	Device-based	**DF**	Device-free
DR	Detection&Recognition	**NA**	Numerical Analysis	**IG**	Image Generation
NR	Noise Reduction	**DA**	Data Adaptation	**TM**	Transform Method
SF	Spatial Feature	**TF**	Temporal Feature	**PF**	Physical Feature
GM	Geometrical Model	**SM**	Statistical Model	**MM**	Machine Learning Model

A DETECTION & RECOGNITION

Table 6. Summary of Wireless Sensing with AI Approach: Detection & Recognition

Ref.	Input & Output	Signal Pre-processing	High-level Feature	Model Formulation
APsense [178]	**SS**: CSI Wi-Fi; **DR**: Device Motion Recognition	**DA**: Eigen Decomposition; **TM**: Correlation;	**SF**: Multiple Statistics with Amplitude;	**MM**: Naïve Bayes & Decision Tree
AutoID [201]	**SS**: CSI Wi-Fi; **DR**: Gait-based User Identification	**DA**: Component Selection (Regularization Term); **TM**: DWT;	**TF**: Rhythm Motion Pattern (Shapelet [170]);	**MM**: Lagrange Multiplier (Clustered Concurrent Shapelets)
CaField [167]	**SS**: Single-carrier Channel Acoustic; **DR**: Speaker Verification	**NR**: Filterbank; **DA**: Long-time Average Fieldprint; **TM**: STFT, FFT;	**SF**: Directivity; **PF**: Sound Field [16]	**GM**: Gaussian Mixture Model for Likelihood Matching
CARM [138]	**SS**: CSI Wi-Fi; **DR** Activity Recognition:	**NR**: Power; **DA**: PCA; **TM**: DHT, DWT;	**TF**: PLCR; **PF**: Phase	**SM**: Hidden Markov Model for Likelihood Matching
Cross Sense [184]	**SS**: CSI Wi-Fi; **DR**: Gait & Gesture Recognition across Scenarios	**DA**: PCA, Virtual Samples; **TM**: DWT, DTW	**PF**: Statistical Features; **AI**: ANN Roaming Model	**MM**: Expert Mixture Model (k-NN, SVM); **AI**: Transfer Learning
Deep Fusion [166]	**SS**: CSI Wi-Fi, Ultrasound; **DR**: Human Activity Recognition	**NR**: Filter; **DA**: Cross-sensor Correlation; **TM**: FFT;	**AI**: CNN-based Sensor-Representation Module;	**AI**: Weighted-Combination Module & Cross-Sensor Module
DeepMV [165]	**SS**: CSI Wi-Fi, Ultrasound; **DR**: Device-Free Human Activity Recognition	**AI**: Adversarial Network with the Domain Discriminator; **NR**: Normalization, Standardization;	**SF**: Power Distribution (Spectrogram); **AI**: CNN-based Feature Extractor;	**AI**: Hierarchically-Weighted-Combination Module
DFAR [136]	**SS**: 24GHz FMCW mmWave; **DR**: Cross-scenario DF Activity Recognition	**TM**: FFT with Multiple Time Series Signals	**AI**: A Maximum-minimum Adversarial Architecture	**AI**: Multi-layer CNN with A Center Alignment Strategy

DFGR [81]	**SS**: CSI Wi-Fi **DR**: Cross-scenario DF Gesture Recognition	**TM**: Low-pass Filter; **DA**: Radio Image across Subcarriers and Packets	**PF**: Power Distribution; **AI**: Deep Feature Extraction Network	**AI**: Deep Similarity Evaluation Network for Transfer Learning
DFHGR [134]	**SS**: 24GHz FMCW mmWave; **DR**: DF Gesture Recognition	**DA**: Virtual Samples Augmentation	**AI**: Four Conv Blocks, Four Max-pooling Layers, and Two FCN	**AI**: Transfer Learning; SS & ST GAN [134]
DFLAR [34]	**SS**: CSI Wi-Fi; **DR**: DF Localization & Activity Recognition	**NR**: Background Subtraction, Filter; **DA**: Multi-channel Information;	**SF**: Amplitude; **TF**: Phase;	**AI**: Sparse Auto-Encoder Neural Network
Duet [117]	**SS**: CSI Wi-Fi, FMCW Radio; **DR**: DF Tracking with the Identification	**NR**: Ratio of Channels; **DA**: 8-antenna Switched Array	**SF**: AoA, Power Distribution;	**MM**: Hidden Markov Model, Probabilistic Logic Framework
E-eyes [139]	**SS**: CSI Wi-Fi; **DR**: DF Activity Recognition	**DA**: Data Fusion across multiple links; **TM**: Multi-dimensional DTW, EMD	**SF**: Amplitude; **TF**: CSI Time Series Pattern	**MM**: Cluster-based Semi-supervised Approach
EI [52]	**SS**: CSI Wi-Fi, Ultrasound, mmWave, Visible Light; **DR**: Human Activity Recognition	**NR**: Filter, Interpolation; **AI**: Loss Functions for Constraints (Confidence, Smoothing, Balance)	**AI**: Domain-independent Feature Extractor (CNN)	**AI**: GAN (Activity Recognizer, Domain Discriminator)
EQ-Radio [187]	**SS**: FMCW Radio; **DR**: Emotion Recognition	**Resemble** Vital-Radio [9]	**PF**: Statistical Features for the Sequential Pattern	**MM**: l_1-SVM Classifier [197]
intu Wition [181]	**SS**: CSI Wi-Fi; **DR**: Material Recognition	**NR**: Dynamic Background Subtraction; **DA**: Channel Stitching;	**PF**: Polarization	**AI**: Multi-layer Perceptrons; **SM**: 2D NDFT
LAND MARK 2.0 [78]	**SS**: RSSI RFID; **DR**: Activity Monitoring	**NR**: Detection Threshold; **DA**: Reference Tag Array;	**TF**: Frequent Trajectory Finding for Motion Pattern	**GM**: Fault-tolerant Tracing Mining using Depth-first Search
Marko [49]	**SS**: FMCW Radio; **DR**: Identification & Behavioral Sensing	**NR**: Background Subtraction, Filtering Mask;	**SF**: Power Distribution (Vertical & Horizontal);	**AI**: 2-branch 10-layer CNN Architecture
RF-Action [69]	**SS**: FMCW Radio; **DR**: Activity Recognition	**Resemble** RF-Pose3D [190]; **AI**: Attention Module [120];	**AI**: Multi-proposal Module; Hierarchical co-occurrence Network	**AI**: Modality-independent Action Detection Network

RF-Mehndi [186]	**SS**: Backscatter RFID; **DR**: DB User Identification	**NR**: Phase Unwrapping, Phase Shifting Calibration **DA**: 3×3 Tag Array;	**TF**: Phase Difference of Tags (PDoT)	**MM**: SVM Classifier using Sequential Minimal Optimization
RF-ReID [29]	**SS**: FMCW Radio; **DR**: User Identification	**Resemble** WiTrack [7] **AI**: Hierarchical Attention Module [168];	**AI**: Spatial-temporal Convolution	**AI**: Multi-task Learning & Environment Discriminator
RF-Sleep [191]	**SS**: FMCW Radio; **DR**: Sleep Stages	**AI**: Adversarial Discriminator	**AI**: CNN-RNN Feature Extractor	**AI**: Conditional Adversarial Architecture [32]
SignFi [83]	**SS**: CSI Wi-Fi; **DR**: Sign Gesture Recognition	**NR**: Phase Unwrapping, Multiple Linear Fitting	**SF**: Amplitude & Phase	**AI**: 9-layer CNN Classifier
Smokey [192]	**SS**: CSI Wi-Fi; **DR**: Smoking Detection	**NR**: Foreground Detection, Linear Interpolation; **TM**: Correlation;	**TF**: Periodicity, Rhythmic Composite Smoking Patterns	**GM**: Composite Motion & Threshold Smoking Detection
Widar 3.0 [194]	**SS**: CSI Wi-Fi; **DR**: Gesture Recognition Across domains	**NR**: Filter, Coordinate Transform; **DA**: PCA; **TM**: STFT, EMD Similarity;	**PF**: 2D BVP	**AI**: the CNN-based GRU Network
WiFi-ID [185]	**SS**: CSI Wi-Fi; **DR**: Gait-based User Identification	**NR**: Silence Removal using Median Filter; **TM**: FFT, CWT;	**PF**: Power Distribution, 10 Statistical Features	**SM**: Lagrange Multiplier (Sparse Approximation)
WifiU [137]	**SS**: CSI Wi-Fi; **DR**: Gait-based User Identification	**NR**: 2D Gaussian Low-pass Filter; **DA**: PCA; **TM**: STFT;	**SF**: Power Distribution (Spectrogram Signatures)	**GM**: LibSVM with Radial Basis Function
WiGest [4]	**SS**: RSSI Wi-Fi; **DR**: Gesture Recognition	**NR**: Wavelet Filter; **TM**: FFT, DWT;	**SF**: Gesture Moving Pattern	**SM**: Action Mapping using Count and Frequency
WiHF [63]	**SS**: CSI Wi-Fi; **DR**: Gesture & Identify Recognition	**NR**: Conjugate Multiplication, Filter; **DA**: PCA, Graph-based Path Matching; **TM**: STFT;	**PF**: Arm Motion Change Pattern	**AI**: Dual-task CNN-GRU Network
WiMU [121]	**SS**: CSI Wi-Fi; **DR**: Multi-user Gesture Recognition	**NR**: Power Model; **DA**: PCA, Virtual Samples; **TM**: STFT;	**PF**: Primary & Secondary Frequency	**SM**: Threshold with Jaccard Similarity
Wi-Vi [8]	**SS**: CSI Wi-Fi; **DR**: Through-wall Motion Recognition	**NR**: MIMO Nulling; **DA**: Eigen Decomposition, Virtual Time Samples;	**SF**: Power Distribution;	**GM**: Inverse SAR, MUSIC [108]
WiWho [177]	**SS**: CSI Wi-Fi; **DR**: Gait-based User Identification	**NR**: Bandpass Filter; **DA**: Delay Selection; **TM**: FFT, DTW;	**TF**: Periodicity, Walking Pattern of Gait	**GM**: Decision-tree based Classifier

XModal-ID [57]	**SS**: CSI Wi-Fi; **DR**: Through-wall User Identification	**NR**: Gaussian Mask; **DA**: PCA; **TM**: STFT;	**SF**: Power Distribution (12 Spectrogram Features)	**GM**: 1-layer Neural Network with 30 Units
C. Shi '17 [114]	**SS**: CSI Wi-Fi; **DR**: User Identification	**NR**: Filter; **DA**: Subcarrier Selection; **TM**: FFT;	**PF**: 9 Physiological & Behavioral Characteristics	**MM**: SVM Spoofing Scheme **AI**: 3-layer Stacked Autoencoder
H. Li '17 [65]	**SS**: CSI Wi-Fi; **DR**: Human Activity	Raw CSI Measurements from Multiple APs	**SF**: Power Distribution across Packets and Subcarriers	**MM**: SVM; **AI**: Multilayer CNN
K. Ohara '17 [92]	**SS**: CSI Wi-Fi; **DR**: Indoor Object Event	**NR**: Background Subtraction, Filtering Mask; **DA**: ICA for Signal Separation;	**AI**: Convolutional LSTM Network;	**AI**: Knowledge-based HMM for Error Correction

B NUMERICAL ANALYSIS

Table 7. Summary of Wireless Sensing with AI Approach: Numerical Analysis

Ref.	Input & Output	Signal Pre-processing	High-level Feature	Model Formulation
Breath Junior [123]	**SS**: FMCW Acoustic; **NA**: Infant Localization & Respiratory Rate	**NR**: FIR Filter; **DA**: Circular Microphone Array, ; **TM**: FFT;	**SF**: Amplitude; **TF**: Phase;	**SM**: Progressive Ternary Search for Beamforming
CapCam [173]	**SS**: Capillary Wave; **NA**: Liquid Testing	**NR**: Wavelet Filter; **TM**: Reflection Symmetry Detection;	**PF**: Surface Tension	**GM**: Capillary Wavelength Inference using Virtual Lenses
CAT [85]	**SS**: FMCW Acoustic; **NA**: Motion Tracking	**NR**: Linear Fitting; **TM**: FFT;	**PF**: Peak Frequency Shift, DFS, IMU	**SM**: Tracing Estimation by Combining Velocity and Distance
Chronos [118]	**SS**: CSI Wi-Fi; **NA**: DB Localization	**NR**: Interpolation for Zero-subcarrier; **DA**: Channel Stitching; **TM**: l_1 Norm	**TF**: Sub-nanosecond ToF;	**SM**: Inverse NDFT & Quadratic-constrained Trilateration
CUPID [110]	**SS**: CSI Wi-Fi; **NA**: DB Localization	**NR**: Moving Average Filter; **TM**: IFFT;	**SF**: AoA, Energy of the Direct Path	**SM**: Propagation Model (LoS Path)
DLoc [14]	**SS**: CSI Wi-Fi; **NA**: Indoor Mapping & Localization	**TM**: 2D FFT, Coordinates Transformation;	**SF**: AoA; **TF**: ToF; **PF**: Wave Front	**AI**: Encoder-decoder DNN
drone Track [88]	**SS**: Sine Waves, FMCW Acoustic; **NA**: Following drone	**NR**: Response Compensation, Filter, Subsampling; **DA**: Chopping, Peak Frequency Selection; **TM**: FFT;	**PF**: DFS, Peak Frequency Shift	**SM**:Root-MUSIC

Finger Draw [151]	**SS**: CSI Wi-Fi; **NA**: Finger Drawing Tracking	**NR**: Differential, Savitzky-Golay Filter; **TM**: FFT, DTW;	**TF**: CSI Quotient Between Antennas, Phase Changes	**GM**: Triangulation, Fresnel Zone Model
Indo Track [71]	**SS**: CSI Wi-Fi; **NA**: DF Tracking	**NR**: Conjugate Multiplication; **DA**: Eigenvalue Decomposition;	**SF**: AoA; **TF**: PLCR	**SM**: Doppler-AoA Joint Parameter Estimation, MUSIC
Ma Track [70]	**SS**: CSI Wi-Fi; **NA**: DF Localization	**NR**: Linear Fitting; **DA**: Coherence Merging	**SF**: AoA; **TF**: ToF; **PF**: Coherency	**SM**: Dynamic-MUSIC
LAND MARC [91]	**SS**: Signal Strength RFID; **NA**: DB Localization	**DA**: Reference Tag Array; **TM**: Euclidean Distance for Spatial Similarity;	**SF**: Amplitude;	**MM**: k-Nearest Neighbor Model with Reference Tags
LiFS [131]	**SS**: CSI Wi-Fi; **NA**: DF Localization	**DA**: Frequency Selection, Reference Transceivers (40 Links); **TM**: DTW;	**PF**: Amplitude with Multiple Fading	**GM**: Complex Nonlinear Fresnel Model
ReMix [119]	**SS**: Backscatter RFID ; **NA**: In-body DB Localization	**DA**: Frequency Selection;	**PF**: Non-linearity of Schottky Detector Diode	**SM**: Spline-wise Propagation Model with Multiple Constraints
RIM [150]	**SS**: CSI Wi-Fi; **NA**: Inertial Measurements	**DA**: Virtual Antenna Array; **TM**: Cosine Similarity;	**PF**: Time-reversal Resonating Strength [155]	**GM**: Dynamic Programming Tracing Model
RTrack [87]	**SS**: FMCW Acoustic; **NA**: Room-scale Hand Tracking	**NR**: Interference Cancellation, Filter; **DA**: MIC Array, Frequency Selection; **TM**: FFT, DTW;	**SF**: AoA; **TF**: ToF; **PF**: Wave Front;	**SM**: 2D MUSIC; **AI**: RNN with the Specially Designed State Unit;
SateLoc [72]	**SS**: RSSI LoRa; **NA**: DB Localization	**NR**: Weighted Average Method;	**SF**: Power Distribution;	**GM**: Propagation Model (Expected Signal Power Map); **MM**: Random Forest
SpotFi [58]	**SS**: RSSI Wi-Fi, CSI Wi-Fi; **NA**: DB Localization	**NR**: Linear Fitting; **DA**: Virtual Antenna Array, Eigenvalue Decomposition	**SF**: AoA; **TF**: ToF;	**GM**: Triangulation with AoA and RSSI; **SM**: MUSIC [108]
Strata [174]	**SS**: Single-carrier Wave Acoustic; **NA**: Finger Tracking	**NR**: Least-square Channel Estimation; **DA**: Delay Selection;	**TF**: Phase Change	**SM**: Triangulation with Two Microphones
Strobe [27]	**SS**: CSI Wi-Fi; **NA**: Soil Moisture & Salinity	**NR**: Linear Fitting; **DA**: Frequency Band Selection;	**SF**: Relative Amplitude & **TF**: ToF across Antennas;	**SM**: MUSIC for Resolving Multi-path [108]

Ubicarse [59]	**SS**: CSI Wi-Fi; **NA**: DB Localization	**NR**: Relative Channels across antennas ; **DA**: Circular Antenna Array;	**PF**: Power Distribution	**SM**: Translation-resilient SAR [31]
Vital-Radio [9]	**SS**: FMCW Radio; **NA**: Breathing & Heart Rate Monitoring	**NR**: Linear Regression; **TM**: FFT;	**TF**: ToF with Buckets; **PF**: Phase Variation	**SM**: Tracing Estimation with IFFT of the corresponding FFT bins
Widar [97]	**SS**: CSI Wi-Fi; **NA**: Human Tracking	**NR**: Power, Filter; **DA**: Graph-based Matching, Frequency Selection, PCA; **TM**: STFT;	**TF**: Path Length Change Rate	**GM**: Tracing Model (Velocity Composition)
Widar2.0 [99]	**SS**: CSI Wi-Fi; **NA**: Human Tracking	**NR**: Conjugate Multiplication, Filter; **DA**:Graph-based Path Matching;	**SF**: AoA, Amplitude; **TF**: ToF; **PF**: DFS	**SM**: Joint Parameter Estimation (Sage Algorithm [24])
WiDir [153]	**SS**: CSI Wi-Fi; **NA**: Walking Direction	**NR**: Power, Savitzky-Golay Filter; **TM**: Cross-correlation;	**PF**: Fresnel Direction (Phase Shift)	**GM**: Fresnel Zone Model [101]
WideSee [19]	**SS**: CSS LoRa; **NA**: Building-scale Localization	**NR**: Filter; **DA**: Reconfigurable Directed Antenna System;	**SF**: Amplitude, Power Distribution (Power Spectral Density);	**SM**: Jointly Direction-related Parameter Estimation, Global Optimum Search
WiTrack [7]	**SS**: FMCW Radio; **NA**: 3D Human Tracking	**NR**: Background Subtraction, Filter, Interpolation; **DA**: Frequency Selection;	**TF**: ToF;	**GM**: 3D Triangulation
μ Locate [90]	**SS**: CSS LoRa; **NA**: 3D localization for sub-cm Objects	**DA**: Dynamic Frequency Selection;**TM**: FFT	**TF**: Below-Noise Backscatter Phase	**SM**: Triangulation using the Least Squares Estimation
J. Wang '17 [135]	**SS**: RSSI Wi-Fi; **DR**: DF Location and Gesture Recognition	**NR**: Wavelet Filter	**AI**: Sparse Autoencoder Network with Layer-by-layer Initialization	**MM**: Softmax-regression-based Machine Learning Framework

C IMAGE GENERATION

Table 8. Summary of Wireless Sensing with AI Approach: Image Generation

Ref.	Input & Output	Signal Pre-processing	High-level Feature	Model Formulation
AIM [86]	**SS**: FMCW Acoustic; **IG**: 2D imaging	**NR**: Background Subtraction, Response Compensation, Filter; **DA**: Delay Selection; **TM**: 2D FFT	**SF**: Wave front (Intermediate-frequency Signal)	**SM**: SAR with the Range Mitigation Algorithm

mD-Track [160]	**SS**: CSI Wi-Fi; **NA**: DF Multi-target Tracking	**NR**: Linear Fitting; **DA**: Iterative Path Parameter Refinement;	**SF**: AoA,AoD; **TF**: ToF; **PF**: DFS	**SM**: Joint Multi-dimensional Estimator
RF-Avatar [189]	**SS**: FMCW Radio; **IG**: 3D Human Mesh Recovery	**Resemble** RF-Pose3D [5]; **AI**: Attention Module [120], Recurrent RPN [102],	**SF**: Power Distribution; **AI**: Trajectory-CNN	**GM**: Skinned Multi-person Linear Model [79]; **AI**: Trajectory Proposal Network
RF-Capture [5]	**SS**: FMCW Radio; **IG**: 3D Position of Body Parts	**NR**: Background Subtraction; **DA**: Coarse-to-fine Antenna Scanning, Snapshot Synthesis; **TM**: FFT;	**SF**: Power Distribution;	**GM**: Skeletal Stitching with the Skeleton Knowledge
RF-Pose [188]	**SS**: RF Radio; **IG**: 2D Pose Estimation	**Resemble** RF-Capture [5];	**AI**: Spatial-temporal Convolution	**AI**: Encoder-Decoder Network with skeleton Association
RF-Pose3D [190]	**SS**: RF Radio; **IG**: 3D Multi-person Pose Estimation	**Resemble** RF-Capture [5]; **AI**: Region Proposal Network [102]	**SF**: Power Distribution; **AI**: ResNet-based Network;	**SM**: 3D Skeleton Triangulation
SAMS [94]	**SS**: FMCW Acoustic; **IG**: Indoor Space Mapping	**NR**: Geometric Constraints & Polynomial Fitting; **TM**: FFT;	**SF**: Frequency Shift; **PF**: Surrounding Depth,	**SM**: Contour Construction Model, Dead Reckoning with Inertial Measurements
TagScan [132]	**SS**: RSSI RFID; **IG**: Imaging & Material Identification	**NR**: RP-Rate (Ratio of RSS and Phase Change); **DA**: Weighted Tags and Channels;	**SF**: RSS; **TF**: Phase Change	**SM**:Propagation Model; **MM**: k-NN Classifier [37]
WiPose [53]	**SS**: CSI Wi-Fi; **IG**: Pose Estimation on the Spot	**NR**: Conjugate Multiplication, Coordinate Transform; **TM**: STFT;	**PF**: 3D BVP [194]	**GM**: Forward Kinematics [2] (Skeleton Model); **AI**: 4-layer CNN & 3-layer LSTM
Wision [50]	**SS**: CSI Wi-Fi; **IG**: 2D Imaging	**NR**: Interference Cancellation; **DA**: Antenna Scanning; **TM**: 2D FFT;	**SF**: Wave Front;	**MM**: SAR& Beamforming
C. Karanam '17 [56]	**SS**: RSSI Wi-Fi; **IG**: 3D Binary Imaging via Unmanned Drones	**NR**: Linear Fitting [21]; **DA**: Graph-based Markov Random Field; **TM**: l_2 Norm	**SF**: Attenuation, Power Distribution	**SM**: Loopy Belief Propagation Model [171], Lagrange Multiplier

S. Depatla '17 [25]	**SS**: UWB, Wi-Fi; **IG**: 3D Binary Imaging via Unmanned Vehicles	**NR**: Linear Fitting [21]; **DA**: Antenna Scanning	**SF**: the first-path power; **TF**: ToF	**SM**: Sparse Signal Reconstruction via Lagrange Multiplier
WiTrack 2.0 [6]	**SS**: FMCW Radio; **NA**: Multi-person Tracking	**NR**: Background Subtraction, Successive Silhouette Cancellation	**TF**: ToF Profile;	**SM**: Multi-shift FMCW for Propagation
C. Karanam '19 [55]	**SS**: CSI Wi-Fi; **IG**: Multi-person DF Tracking	**NR**: Particle Filter with a Joint Probabilistic Data Association Filter; **DA**: Virtual Arrays	**SF**: Amplitude, 2D AoA	**SM**: Joint parameter estimation, 2D MUSIC
C. Wang '18 [124]	**SS**: CSI Wi-Fi; **NA**: Volume Estimation & Tomography	**NR**: Phase Unwrapping [42], Linear Fitting	**PF**: Waveform (Amplitude & Phase)	**SM**: Propagation Model; **MM**: k-NN Classifier
F. Wang '19 [126]	**SS**: CSI Wi-Fi; **IG**: Person Perception and Pose Estimation	Raw CSI Signals	**SF**: Multi-task DNN including CNN, U-Net [105];	**AI**: Mask R-CNN [44]
P. Holl '17 [47]	**SS**: CSI Wi-Fi; **IG**: Indoor Holography	**NR**: Digital Darkfiled Mask; **DA**: Antenna Scanning; **TM**: 2D FFT	**SF**: Wave Front;	**SM**: Angular-spectrum Propagation
P. Proffitt '18 [96]	**SS**: CSI Wi-Fi; **IG**: Static Object Imaging	**NR**: Gaussian Filter; **DA**: Automatic Antenna Scanning	**SF**: Wave Front;	**AI**: Mask R-CNN [44]

D COMPARISON STUDY

Table 9. Summary of Related Surveys on Wireless Sensing

Reference	Modality/Input	Application/Output	Topic Focus
Z. Yang et al. [169]	Wi-Fi (RSSI & CSI)	Numerical Analysis: device-free and device-based indoor localization	Evolution from RSSI to CSI for basic principles, resolution, accessibility, and future trends
Y. Zou et al. [202]	Wi-Fi (CSI)	Detection & Recognition: device-free human behavior recognition	Wi-Fi radar with data-driven and model-based system design
D. Wu et al. [152]			From pattern-based applying machine learning techniques to Fresnel zone model based approaches
S. Yousefi et al. [172]			Evolution and performance improvement from machine learning to deep learning techniques

S. Savazzi et al. [130]		Detection & Recognition: activity and gesture recognition, Numerical Analysis: device-free localization and motion tracking	Leading-edge research and developments with a special focus on assisted living applications
Z. Wang et al. [142]		Detection & Recognition: action & activity recognition, Numerical Analysis: signal recognition	Basics and insights on CSI based behavior recognition applications classified by three granularities
J. Wang et al. [107]	Wireless Signals (RSS, phase, AOA, TOF, and CSI)	Detection & Recognition: security monitoring, emergency rescue, intelligent interaction and monitoring.	A comprehensive overview on working principle and system architecture of the device-free WSSs
M. Liu et al. [75]	Acoustic (Sine wave & FMCW)	Numerical Analysis: absolute and relative range-based indoor localization	Feature design and model mechanisms for range-based localization.
C. Cai et al. [17]		Detection & Recognition: human-computer interface, Numerical Analysis: context-aware application	A general categorization framework from the physical layer, processing layer, and application layer
J. Sundaram et al. [112]	LoRa (chirp spread spectrum)	Numerical Analysis: LoRa networking such as collision avoidance, channel estimation	Research problems, current solutions, and open issues on LoRa networking and sensing
N. Hassan et al. [43]	Visible Light (RSSI)	Numerical Analysis: photodiode and camera based indoor localization	Basic principle and architecture via various receivers along with various position determination algorithms
M. Afzalan et al. [10]		Numerical Analysis: range-based, range-free, and fingerprinting indoor localization	Performance-based evaluation of the deployed VLP systems in real-world, covering different components and fundamental concepts of systems
A. Rahman et al. [100]		Numerical Analysis: indoor localization with modified and unmodified light source	reviews on recent advances of VLP, covering common architecture as well as potential directions
F. Zafari et al. [176]	Wi-Fi, Acoustic, RFID, UWB, Bluetooth, etc	Numerical Analysis: Localization and positioning of human users and devices	Evaluation and comparison of input techniques covering range, scalability and tracking accuracy. Working principles of localization systems
J. Xiao et al. [158]	Wi-Fi, Acoustic, Bluetooth, UWB, RFID, etc	Numerical Analysis: device-free and device-based indoor localization	Conventional methodologies for basic principles of signal processing, and data fusion techniques
Y. Ma et al. [82]	Wi-Fi (CSI)	Detection & Recognition, Numerical Analysis, Image Generation	Signal processing, modeling-based and learning-based algorithms, applications, performance results

X. Chen et al. [20]	Wi-Fi, RFID, UWB, acoustic, LoRa, Visible light, etc	Detection & Recognition, Numerical Analysis, Image Generation	A general picture of the across-disciplinary wireless sensing, covering wireless communication, signal processing, machine learning, ubiquitous computing, and HCI
C. Zhang et al. [179]	Mobile big data of cellular network, Wi-Fi, etc	DL driven mobile networking, like mobile data analysis, wireless sensor network, network control, and security	DL backgrounds, framework and approaches toward mobile networking problems, tailoring DL to mobile networks
J. Wang et al. [129]	Wireless Signals (RSS, Phase, ToF)	Image Generation: Radio Image Construction	Deep Learning for wireless sensing with less training efforts via a deep similarity evaluation networks and deep GANs
This survey	Modalities covering Wi-Fi, acoustic, LoRa	Detection & Recognition, Numerical Analysis, Image Generation	Evolution of model-based methods to deep learning approaches and comparison of multiple sourcing inputs for various applied outputs