Cross Domain WiFi Sensing: A Survey

***Abstract*—WiFi sensing is of great benefit to our daily lives. Due to pervasiveness, low cost, no-invasion and no-burden, WiFi sensing is advantageous over video surveillance and wearables in many aspects. By analyzing the influence of human bodies on wireless propagation, indoor human situations can be sensed accurately, such as location, activity, gesture and vital signs. However, wireless signals are sensitive to the surroundings. Different domains, such as different environments, environmental dynamics, humans and devices, may cause variations of wireless propagation. As most WiFi sensing models are constructed via supervised learning or dependent on domain-related parameters, domain change will cause significant performance degradation of WiFi sensing due to data distribution discrepancy between the source and the target domain. This is a major challenge that hinders the real-world application of WiFi sensing. To mitigate this problem, researchers either adapt the source sensing model to the target domain or construct the sensing model with domain-independent features. In this survey, we review the works that address the issue of WiFi sensing across different domains utilizing commodity WiFi devices, including different environments, environmental dynamics, human, time, location, orientation and device. We review and category the works by the domains they address, the measurements they use, the methods they leverage and the applications they realize. In the end of the survey, we discuss the limitations of the current works and point out the future research directions. This survey can help the readers have an overall understanding of cross domain WiFi sensing with its challenges, advances and methods.**

***Index Terms*—WiFi sensing, Received Signal Strength (RSS), Channel State Information (CSI), Transfer learning, Domain adaptation.**

I. Introduction

WiFi sensing can bring great benefits to our daily lives. Sensing human situations has become an inevitable part in various fields like surveillance, entertainment, healthcare, etc. The ubiquitous availability of WiFi brings new opportunities for device-free human sensing, which is attracting increasing attentions in recent years. Conventional approaches to human sensing are based on vision or wearables. Vision-based solutions deploy cameras to collect images and realize sensing through image processing. Vision-based solutions can achieve high accuracy, but require Line of Sight (LOS) and luminous lighting to work properly. Cameras impose privacy concerns, hence they are not appropriate to deploy in private spaces such as bedrooms and bathrooms. Sensing based on wearables eliminates the limitations of LOS and does not impose privacy concerns. These solutions exploit the embedded sensors in wearables to realize sensing by analyzing the sensor data. However, wearable-based solutions require the targets to wear dedicated devices, which is an extra burden on human body, causing inconvenience and reluctance.

The authors are with the University of Electronic Science and Technology of China. e-mail: @uestc.edu.cn.

Manuscript received; revised.

To overcome the shortcomings of video surveillance and wearables, human sensing has shed on wireless signals, including radio frequency (RF) [1], [2], radar [3], [4], Long Range (LoRa) [5], [6] and WiFi [7]–[16]. In an indoor environment, human bodies alter the propagation of wireless signals, causing them to carry rich information of humans, which can be utilized to infer human situations. Due to the ubiquity of wireless signals, wireless sensing avoids the inconvenience brought by wearables and eliminates privacy invasion from cameras. Among them, WiFi sensing is the most promosing. WiFi sensing reuses the infrastructure of wireless communication, hence it is easy to deploy and has almost no additional cost. WiFi provides a feasible solution for both device-based and device-free human sensing.

There have been a plethora of applications of different kinds of WiFi sensing, such as localization, tracking, activity recognition, gesture recognition, vital sign monitoring. However, the all face the issue of domain diversity, for example, different environments, environmental dynamics, human diversity, different time. To deploy WiFi sensing in real-world applications, the cross domain issue must be solved. We list the domains that have been addressed mostly by researchers.

- **Environment.** We use an environment to refer to a different room, building, site, etc. We expect a sensing model trained in one environment can work in a different environment with little or even no recalibration effort. This is the most challenging issue, because different environments have different layouts and quite different wireless propagation paths.
- **Environmental dynamics.** An environment may have various dynamics due to various reasons. Common environmental dynamics include opening or closing windows or doors, temperature or humidity varying, rearranging furniture, reconfiguring transceivers, etc. Environmental dynamics will alter the wireless propagation paths and cause the sensing model to degrade.
- **Human.** In real-world applications, the targets of sensing often involve different people. Different people have different impact on wireless propagation due to different physical features. The WiFi sensing model should work with all the possible targets without training for everyone.
- **Time.** Wireless signals are sensitive to the surroundings and have fluctuations over time even if in a static situation. We expect the sensing model to keep high sensing accuracy over time.
- **Location.** For some sensing applications such as activity recognition and gesture recognition, the targets may be at different locations. The WiFi signals are influenced not only by the activity or gesture but also by the location. We expect the sensing model to work across locations for large scale deployment.

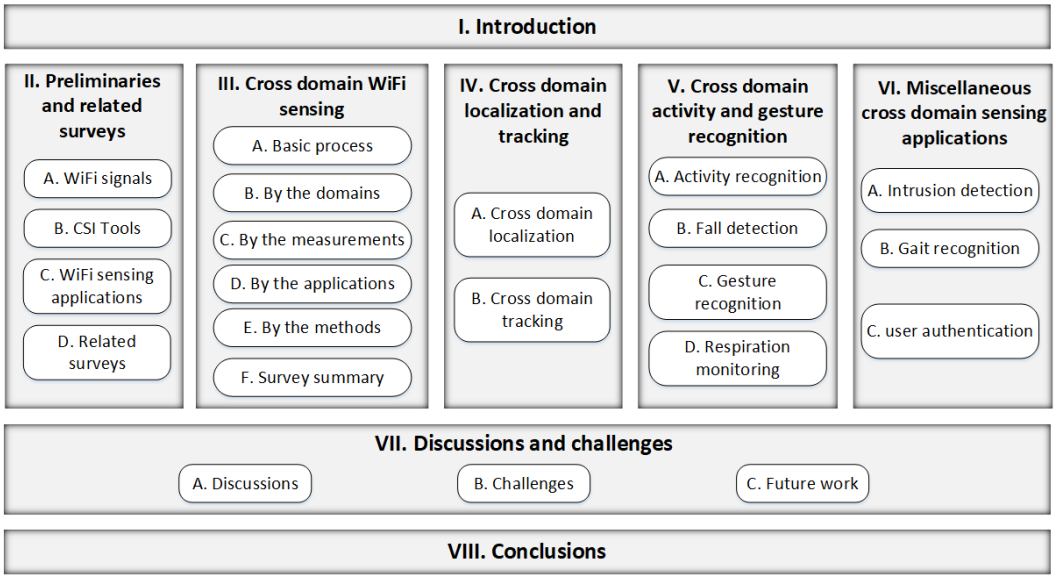


Fig. 1. Organization of the survey

- **Orientation.** Due to the blockage of human body, the WiFi signals are affected differently when the target faces different orientations. Sensing should be orientation independent.
- **Target device.** Some WiFi sensing applications may be device-based, for example, localization and tracking. The target devices, i.e. the devices carried by the targets, have different impact on the received signals due to different capacity and consequently have different sensing results. We expect the sensing model to work across device heterogeneity.
- **Access point.** Modern access points (AP) may adjust their transmitting power automatically. They may appear or disappear from time to time for a target device. The sensing model should work with it.

Most sensing models are constructed through learning or dependent on domain-related parameters. They require the domains to keep unchanged. However, in reality, the domains may change from time to time, causing the sensing model to perform improperly in the changed domains. We may recalibrate the whole area of interest for each domain, but it takes significant recalibration effort. We need a strategy to transfer the knowledge from the source domain to each target domain or extract the domain independent features. If we can achieve this, we can build the universal sensing model for real-world applications and meanwhile keep high accuracy. More and more researchers are noticing the issue and proposing various solutions. In this survey, we review the works that address the issue of WiFi sensing across domains using commodity WiFi devices. We analyze the influence of different domains on WiFi sensing, investigate the methods of cross domain WiFi sensing, and point out the limitations and future research directions. This survey can help the readers have an overall understanding of cross domain WiFi sensing with its challenges, advances and methods. The survey is organized as Fig. 1 illustrates.

II. Preliminaries and Related Surveys

WiFi sensing leverage WiFi signals to sense the human situations in an environment. When a person is in the area covered with WiFi, he or she will exert influence on the surrounding WiFi signals. By analyzing the distortion and fluctuation of the WiFi signals, the sensing system can estimate the whereabouts and doings of the person. There have been a variety of WiFi sensing applications leveraging various technologies. The key is to understand the relationship between human situations and WiFi signals, i.e., how human affects the propagation of WiFi signals. The basic process of WiFi sensing is illustrated in Fig. 2. With a few pairs of WiFi transceivers, the sensing system first collects the WiFi signal data in the form of Received Signal Strength (RSS) or Channel State Information (CSI). The RSS or CSI data are proprocessed and input to the sensing method. Main-stream sensing methods are learning-based or model-based. The learning-based methods construct the learning models to predict the sensing results, while the model-based methods investigate the wireless propagation to infer the sensing results. Leveraging the sensing methods, the sensing systems can realize the sensing tasks, such as localization, tracking, activity recognition, gesture recognition, vital sign monitoring, people counting, intrusion detection, human identification and object detection.

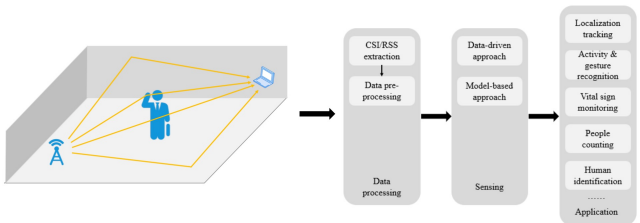


Fig. 2. Basic process of WiFi sensing.

A. WiFi signals

We focus on sensing with WiFi, which is based on the family of IEEE 802.11 standard protocols. The two common measurements of WiFi are RSS and CSI.

1) WiFi propagation: When transmitting in free space, the propagation of wireless signals can be modeled according to the Friis model [17]

$$P^r(d) = \frac{P^t G^t G^r \lambda^2}{(4\pi)^2 d^2} \tag{1}$$

where, P^t represents the transmitting power, $P^r(d)$ represents the receiving power, d is the distance between the transmitter and the receiver, G^t is the gain of the transmitting antenna, G^r is the gain of the receiving antenna, λ is the wavelength of wireless signals.

When transmitting indoors, the wireless signals may be reflected, refracted, scattered and absorbed by various obstacles including the ceiling, the floor, the furniture as well as the human. Considering the multipath effect, the propagation of wireless signals indoors can be modeled as

$$P^r(d) = \frac{P^t G^t G^r \lambda^2}{(4\pi)^2 (4h^2 + \Delta^2 + d^2)} \tag{2}$$

where, h represents the distance from the static reflection object to the LOS path, Δ is the length of the reflection path of the person. The overall power is the sum of the power from all the paths

$$P \propto |P_{LOS} + P_{Static} + P_{Dynamic}|^2 \tag{3}$$

where, P_{LOS} represents the power of the LOS path, P_{Static} represents the power of the static reflection paths, and $P_{Dynamic}$ represents the power of the dynamic reflection paths, usually caused by the targets.

2) WiFi RSS: WiFi RSS has been the most common measurement of WiFi sensing, representing the received signal strength of a wireless link. Due to multipath effect, the received signal is a combination of multiple paths, modeled as

$$s = \sum_{i=1}^{N} \|s_i\| \, e^{-j\theta_i} \tag{4}$$

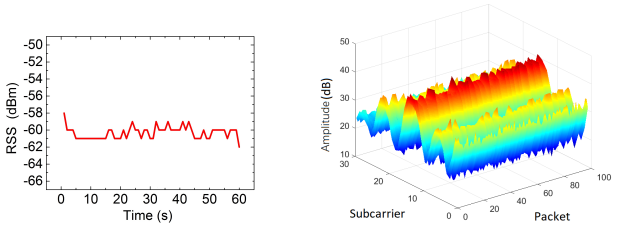
where, s_i represents the amplitude of path i, θ_i represents the phase of path i, N is the number of paths. The received signal strength RSS can be calculated as

$$rss = 10log_2\|s\|^2 \tag{5}$$

RSS can be readily extracted from almost all commercial off-the-shelf (COTS) WiFi devices. Fig. 3(a) shows a time series of RSS values of a wireless link.

3) WiFi CSI: WiFi CSI [7] launches from the IEEE 802.11n standard. Its core technologies include Multiple-input Multiple-output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM), which enable multiple subcarriers to transmit data. In a narrow band flat-fading channel, the OFDM system in the frequency domain is modeled as:

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n} \tag{6}$$



(a) RSS (b) CSI

Fig. 3. The WiFi measurements.

where $\boldsymbol{y}$ represents the received vector, $\boldsymbol{x}$ represents the transmitted vector, $\mathbf{H}$ denotes the channel matrix, and $\boldsymbol{n}$ is the additive white Gaussian noise vector. The estimated value of $\mathbf{H}$ can be expressed as:

$$\bar{\mathbf{H}} = \frac{\boldsymbol{y}}{\boldsymbol{x}} \tag{7}$$

It is a fine-grained value from the physical layer that describes the channel gain from transmitting baseband to receiving baseband. A multipath wireless channel is often portrayed as Channel Impulse Response (CIR), expressed as:

$$h(\tau) = \sum_{i=0}^{N-1} a_i e^{-j\theta_i} \delta(\tau - \tau_i) \tag{8}$$

where N is the number of paths, a_i denotes the amplitude, θ_i denotes the phase, τ_i represents the delay, of the i-th path, $\delta(\tau)$ represents the Dirac delta function. Channel Frequency Response (CFR) $H(f)$ can be acquired by Fourier Transform of CIR:

$$H(f) = F(h(\tau)) \tag{9}$$

Leveraging commodity Network Interface Card (NIC) with a modified driver [18], [19], a discrete version of CFR

$$\mathbf{H} = \{H(f_k)\} \tag{10}$$

can be revealed to the upper layers in the format of CSI, where each $H(f_k)$ is a complex number depicting the amplitude and the phase of subcarrier f_k. Hence the CSI matrix $\mathbf{H}$ is a 3D matrix of complex values representing the amplitude and phase of the multipath WiFi channel:

$$\mathbf{H} = (\boldsymbol{H}_{ij})_{N_{tx} \times N_{rx}} \tag{11}$$

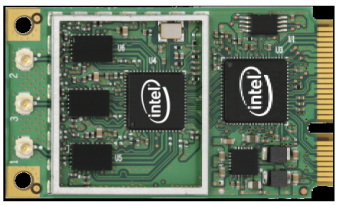
$\boldsymbol{H}_{ij}$ is the CSI of the channel formed by transmitting antenna i and receiving antenna j, containing the information of N_s subcarriers, expressed as

$$\boldsymbol{H}_{ij} = (h_1, h_2, \cdots, h_{N_s}), \ i \in [1, N_{tx}], \ j \in [1, N_{rx}] \tag{12}$$

The k-th subcarrier in $\boldsymbol{H}_{ij}$ can be expressed as

$$h_k = |h_k| e^{j\angle h_k}, \ k \in [1, N_s] \tag{13}$$

here, $|h_k|$ is the amplitude and $\angle h_k$ is the phase of subcarrier k. Fig. 3(b) shows the time series of CSI values of multiple subcarriers.



(a) Intel 5300 NIC

(b) Atheros NIC

Fig. 4. Tools of extracting CSI.

B. CSI tools

The availability of CSI extraction tools is the key enabler of WiFi sensing. Although access to it is not common on COTS WiFi devices, there are some options.

1) Linux 802.11n CSI Tool: University of Washington provides the *Linux 802.11n CSI Tool*[1], enabling CSI extraction on Intel WiFi Link 5300 NICs, using a modified firmware and open source Linux wireless drivers, as shown in Fig. 4(a). IWL5300 provides 802.11n CSI in a format that reports the channel matrices of 30 subcarrier groups, which is about one group for every 2 subcarriers at 20MHz, or one in 4 at 40MHz. Each matrix entry is a complex, specifying the gain and phase of the signal path between a transmiter-receiver antenna pair.

2) Atheros CSI Tool: Non-grouped CSI can be obtained using the *Atheros CSI Tool*[2] maintained by the Wireless And Networked Distributed Sensing (WANDS) research group. Atheros CSI Tool enables extraction of detailed PHY wireless communication information from the Atheros WiFi NICs, as shown in Fig. 4(b). Atheros CSI Tool is built on top of ath9k, which is an open source Linux kernel driver supporting Atheros 802.11n PCI/PCI-E chips. The tool has been tested on Atheros AR9580, AR9590, AR9344 and QCA9558.

3) Nexmon CSI Extractor: Nexmon CSI Extractor[3], maintained by Secure Mobile Networking Lab (SEEMOO) et al., allows per-frame CSI extraction on modern Broadcom and Cypress WiFi chips with up to 80MHz bandwidth in both 2.4 and 5GHz bands. The tool supports BCM4339 on Nexus5, BCM4358 on Nexus6P, BCM43455C0 on Raspberry Pi 3B+ and 4B, and BCM4366C0 on Asus RT-AC86U AP. The tools are open source on github[4].

4) Software Defined Radios: As an alternative to COTS WiFi devices, Software Defined Radios (SDRs) can be used for CSI measurements, such as *Wireless Open Access Research Platform (WARP)*[5] and the *Universal Software Radio Peripheral (USRP)*[6].

C. WiFi sensing applications

There has been a vast variety of applications of WiFi sensing, ranging from localization, tracking, activity recognition, gesture recognition, vital sign monitoring, people counting, intrusion/presence detection, human identification, to object detection, as illustrated in Fig. 5. As our focus is on cross domain WiFi sensing, we introduce the applications and the related works very briefly. For detailed review on WiFi sensing applications, please refer to the related surveys in Table I.

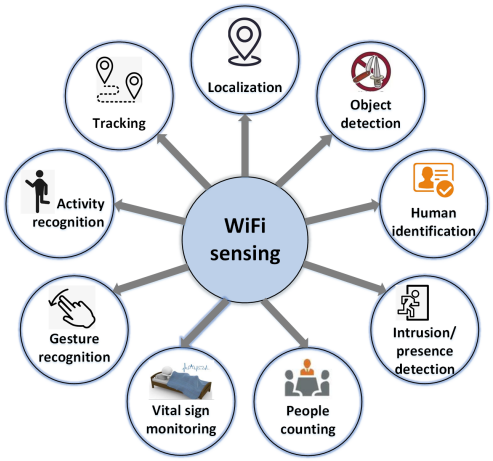


Fig. 5. WiFi sensing applications.

1) Localization: Localization is one of the most common applications of WiFi sensing, making use of RSS or CSI. FILA [17] developed a propagation model representing the relationship between CSI and distances and applied trilateration to accomplish localization. Wu et al. [20] improved FILA with CSI fingerprinting using maximum likelihood algorithm. MonoPHY [21] realized localization with cluster-based CSI fingerprinting using maximum likelihood algorithm. MonoStream [22] enhanced MonoPHY and modeled localization as an object recognition problem. LiFS [23] modeled CSI measurements as a set of power fading based equations and utilized the subcarriers affected by multipath for localization. S-Phaser [24] calibrated CSI phase to compute the direct path length and used a geometric algorithm to determine the target location. Wang et al. proposed DeepFi [25] using CSI amplitude and PhaseFi [26] using calibrated CSI phase to realize localization based on deep learning and the Bayesian algorithm. ConFi [27] organized CSI amplitude into a time-frequency matrix that resembled an image as the feature and employed convolutional neural network (CNN) for localization. CiFi [28] estimated Angle of Arrival (AoA) using phase differences and created AoA images as input to CNN for localization. Gao et al. [29] transformed CSI amplitude and phase into a radio image, extracted color and texture features, and estimated location and activity through deep learning. Zhou et al. [30] built a DNN regression model to represent the relationship between CSI fingerprints and location coordinates and achieved localization by DNN regression. Fei et al. [31] proposed a localization system based on the Fresnel zone model of radio communication that could determine which elliptic region the target located in according to the CSI phase.

2) Tracking: Tracking is to estimate the trajectories of moving target. By detecting the subtle reflection signals from the human body and differentiating them from the reflected

[1]http://dhalperi.github.io/linux-80211n-csitool
[2]https://wands.sg/research/wifi/AtherosCSI/#Overview
[3]https://dl.acm.org/citation.cfm?doid=3349623.3355477
[4]https://github.com/seemoo-lab/nexmon_csi
[5]Rice University. Wireless open access research platform, 2019.
[6]Ettus Research. USRP Software Defined Radio Device, 2018.

signals from static objects, MaTrack [32] identified the target's angle to track the walking human. Widar [33], [34] simultaneously estimated human moving velocities and locations by modeling the relationships between CSI dynamics and locations/velocities exploiting Doppler shifts. Widar2.0 [35] leveraged AoA, Time of Flight (ToF), Doppler shifts, and attenuation together to enable human tracking. IndoTrack [36] proposed Doppler-MUSIC to extract Doppler velocity information from CSI and proposed Doppler-AoA to estimate target velocity and location via probabilistic co-modeling of spatial-temporal Doppler and AoA information. Karanam et al. [37] estimated the AoA of all the signal paths arriving at a receiver array using the corresponding received signal magnitude and utilized a particle filter to track the moving target. Shi et al. [38] leveraged CSI amplitude fingerprint to achieve tracking using Bayes classifier and Bayesian filtering.

3) Activity recognition: Human activity recognition (HAR) is one of the most common applications of WiFi sensing. E-eyes [39] used the distribution of CSI amplitude as the feature to recognize activities. WiseFi [40] harnessed the amplitude and the phase of CSI, and the AOA of blocked signals to localize human and recognize human activities. Wi-Chase [41] utilized variations in phase and magnitude of all subcarriers to recognize daily activities. Zhang et al. [42] proposed a diffraction-based sensing model to quantitatively determine the signal change wrt. a target's motions and linked signal variation patterns with human activities. Chen et al. [43] employed Bidirectional Long Short Term Memory (BLSTM) to learn representative features of activities and leveraged attention mechanism to assign different weights to the learned features. WiAct [44] used the Doppler shift correlation value as the feature to classify activities via Extreme Learning Machine (ELM). WiFall [45], [46] employed the time variability and special diversity of CSI as the indicator to detect human activities and falls through Support Vector Machines (SVM). RT-Fall [47] exploited CSI phase and amplitude to segment and detect the falls in real-time. MAIS [48] identified multiple activities performed by multiple people in the same environment.

4) Gesture recognition: Researchers also exploited WiFi to recognize gestures. WiSee [49] enabled recognition of human gestures and motions by making use of Doppler effect caused by human gestures and motions. WiG [50] accomplished CSI-based gesture recognition by SVM classification. WiFinger [51] recognized finger-grained gestures to realize continuous text input in WiFi devices, based on the patterns in the time series of CSI values. WiHear [52] introduced mouth motion profile that leveraged partial multipath effects and wavelet packet transformation to recognize people talks. DeNum [53] recognized number gestures based on CSI amplitude and phase. WiMU [54] addressed gesture tracking for multiple users performing gestures simultaneously.

5) Vital sign monitoring: Vital sign monitoring is to monitor the vital signs of the targets, such as respiration and heartbeat, which are important for healthcare. UbiBreathe [55] used RSS to realize non-invasive respiratory rate estimation. Wi-Sleep [56] monitored sleep through breath detection by extracting rhythmic patterns associated with respiration and abrupt changes due to body movement. WiHealth [57] could recognize and count breath and heartbeat with different postures. Wang et al. [58] leveraged the Fresnel model to develop the theory relating one's breathing depth, location and orientation to the detectability of respiration. PhaseBeat [59] exploited CSI phase difference to monitor breath and heartbeat. TensorBeat [60] employed CSI phase difference to estimate breath rates for multiple persons. Khan et al. [61] used CSI to monitor the respiration rate of a patient during sleep.

6) People counting: Researchers also put forward solutions to count the number of people in a space or passing by. Frog eye [62] proposed a metric PEM to represent the variation of wireless channels and formulated the relationship between PEM and people count by the Grey Verhulst Model (GVM). Depatla et al. [63] developed a mathematical expression for the RSS probability distribution as a function of the people count, which was the base for the estimation using Kullback-Leibler divergence. Cianca et al. [64] realized crowd counting with a linear discriminant classifier, based on the received raw signal samples (RRSS) extracted from beacon messages. Zhou et al. [65] inferred crowd count by establishing the relationship between the variation of CSI and the number of people with DNN. Doong [66] used mean, standard deviation and five-number summary from windowed CSI data as raw inputs and used stacked denoising autoencoders to extract features and used softmax regression for flow counting. Wi-Count [67] identified the number of people passing by concurrently via modeling the effect of the passing behavior on the phase difference of WiFi signals. Xiao et al. [68] extracted skewness and kurtosis of CSI as the features and applied SVM to estimate the number of people in the queue. WiFlowCount [69] leveraged the Doppler effect induced by human passing to count the number of people in a flow by constructing the spectrogram of Doppler shifts from CSI.

7) Intrusion/presence detection: Intrusion or presence detection aims to determine the presence or intrusion of human. Zhou et al. [70] used CSI fingerprinting and a threshold-based method to detect human presence in an omnidirectional manner. DeMan [71] made use of CSI amplitude and phase to detect moving human and considered human breathing as an intrinsic indicator of stationary human presence. SIED [72] realized intrusion detection of different moving speeds with Hidden Markov Model (HMM), by capturing the variance of variances of amplitudes of each CSI subcarrier. R-TTWD [73] achieved through-the-wall detection of moving human by leveraging the correlated changes over different subcarriers and extracting the first-order difference of eigenvector of CSI amplitudes across different subcarriers.

8) Human identification: Analyzing WiFi signals can associate a person with his or her identity. WFID [74] realized human identification through a linear-kernel SVM based on the feature of CSI subcarrier-amplitude frequency (SAF). WiWho [75] identified a person from a small group of people using CSI by the person's steps and walking gait. WifiU [76] generated spectrograms from CSI to capture gait patterns to recognize humans. Rapid [77] performed human identification using CSI with acoustic information from footstep sound.

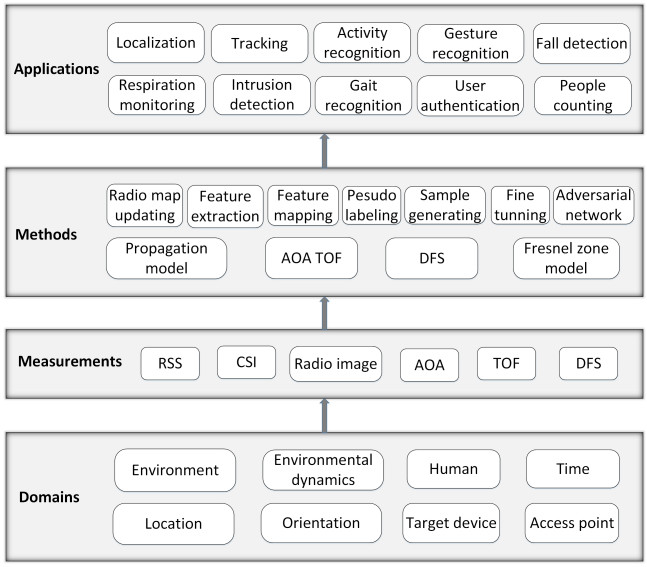


Fig. 6. Basic process of cross domain WiFi sensing.

9) Object detection: Detecting the objects and identifying their types are also possible with WiFi. Wi-Metal [78] could detect metal carried by a standing person and calculate the distance between the metal and the detector. TagFree [79] could distinguish multiple exposed objects. Wang et al. [80] proposed a method to detect metal and liquid objects in baggage and estimate their shape or volume. Hanif et al. [81] proposed a method to identify a large metal sheet carried by a walking pedestrian. Zhou et al. [82] detected metal and liquid concealed by walking pedestrians by analyzing the variations of amplitude caused by different materials.

D. Cross domain WiFi sensing

The surroundings have significant impact on WiFi sensing. This issue has been attracting more and more attention, as it is the prerequisite to large scale application of WiFi sensing. However, the research on domain independent or cross domain WiFi sensing is still in its exploration stage. The basic process of cross domain WiFi sensing is illustrated in Fig. 6. In different domains, the raw measurements RSS and CSI are collected, from which the upper layer measurements AoA, ToF, radio image and Doppler Frequency Shift (DFS) can be extracted. The measurements are then processed by different methods to achieve cross domain sensing and realize the applications working across different domains. The domains may be different environments, environmental dynamics, different people, different time, different locations and orientations, and different devices. There have been many methods proposed over the years to deal with the cross domain issue. The methods are learning-based or model based. Learning based methods include radio map updating, feature extraction, feature mapping, pseudo labeling, sample generating, fine-tuning, and adversarial network. The model-based methods include propagation model, AoA, ToF, DFS and Fresnel zone model. Finally the WiFi sensing applications, such as localization, tracking, activity recognition, gesture recognition and vital sign monitoring, can work properly across different domains. In the next sections of the survey, we categorize the works by the domains they dealt with, the measurements they used, the methods they proposed and the applications they realized, respectively, and review them in each category.

E. Related surveys

There have been some surveys on wireless sensing, as listed in Table I. Yang et al. [7] provided the first survey to review the works using WiFi CSI and focused on localization based on RSS or CSI. Yousefi et al. [8] reviewed the works on behavior recognition using WiFi CSI and discussed the environmental factors that influenced recognition accuracy as a challenge. Ma et al. [9] provided a review on signal processing techniques, algorithms, applications, and performance results of WiFi sensing with WiFi CSI, and highlighted the challenges. Shit et al. [10] reviewed and categorized the works of device-free localization using various wireless signals, and provided a thorough description of localization methods in connection to smart world scenarios. W. Liu et al. [11] reviewed the localization technologies based on WiFi CSI from the aspects of geometry-based methods and fingerprint-based methods. J. Liu et al. [12] surveyed the WiFi sensing systems in terms of basic principles, techniques and structures, based on RSS, CSI, Frequency Modulated Carrier Wave (FMCW) and Doppler shifts, including intrusion detection, room occupancy monitoring, activity recognition, gesture recognition, vital sign monitoring, human identification and localization. Ahmed et al. [13] reviewed the works on device-free gesture recognition using WiFi CSI and summarized the performance of model-based and learning-based approaches. The environmental factors that influenced the recognition accuracy were mentioned. He et al. [14] presented a survey on WiFi vision utilizing CSI, focusing on WiFi imaging, vital sign monitoring, human identification, gesture recognition, gait recognition, activity recognition, fall detection, human detection, and localization. Ngamakeur et al. [15] surveyed device-free localization and tracking in multi-resident environments, using Radio Frequency Identification (RFID), WiFi, ZigBee, Ultra Wideband (UWB), Infrared, acoustics, vibration, electric field. After discussing the steps involved including target detection, target counting, target identification, localization and tracking, the techniques related to each step were discussed. Zhu et al. [16] studied the localization technologies based on various types of wireless fingerprint signals, including RFID, WiFi, Zigbee, UWB, Infrared, Ultrasonic, Bluetooth, visible light, magnetic field, using machine learning and intelligent algorithms, and discussed their advantages and disadvantages, working mechanisms, and principles of localization.

Different from the previous surveys, our survey focuses on the cross domain issue of WiFi sensing, based on RSS or CSI, for various applications. To the best of our knowledge, this is the first survey to review the works on cross domain WiFi sensing. We category the works by the domains they addressed, the measurements they used which may be WiFi RSS or WiFi CSI, the methods they proposed which may be leaning-based or model-based, and the applications they

TABLE I
SUMMARY OF RELATED SURVEYS ON WIRELESS SENSING.

Survey	Year	Title	Measurement	Survey scope
Yang et al. [7]	2013	From RSSI to CSI: Indoor localization via channel response	WiFi RSS, WiFi CSI	Localization
Yousefi et al. [8]	2017	A survey on behavior recognition using WiFi Channel State Information	WiFi CSI	Activity recognition
Ma et al. [9]	2019	WiFi sensing with Channel State Information: A survey	WiFi CSI	Sensing applications
Shit et al. [10]	2019	Ubiquitous localization (UbiLoc): A survey and taxonomy on device free localization for smart world	Wireless signals	Device-free localization
W. Liu et al. [11]	2019	Survey on CSI-based indoor positioning systems and recent advances	WiFi CSI	Localization
J. Liu et al. [12]	2020	Wireless sensing for human activity: A survey	WiFi RSS, WiFi CSI, FMCW, Doppler shift	Intrusion detection, room occupancy monitoring, activity recognition, gesture recognition, vital sign monitoring, human identification, and localization
Ahmed et al. [13]	2020	Device free human gesture recognition using Wi-Fi CSI: A survey	WiFi CSI	Device-free gesture recognition
He et al. [14]	2020	WiFi vision: Sensing, recognition, and detection with commodity MIMO-OFDM WiFi	WiFi CSI	WiFi imaging, vital sign monitoring, human identification, gesture recognition, gait recognition, activity recognition, fall detection, human detection, localization
Ngamakeur et al. [15]	2020	A Survey on device-free indoor localization and tracking in the multi-resident environment	RFID, WiFi, ZigBee, UWB, Infrared, acoustics, vibration, electric field	Device-free localization and tracking in multi-resident environments
Zhu et al. [16]	2020	Indoor intelligent fingerprint-based localization: Principles, approaches and challenges	RFID, WiFi, ZigBee, UWB, Infrared, Ultrasonic, Bluetooth, visible light, magnetic field	Fingerprint-based localization
This survey	2021	Cross domain wireless sensing: A survey	WiFi RSS, WiFi CSI	Cross domain sensing applications

realized, respectively. We analyze and compare the works in each category. Based on the survey, we point out the trend and the remaining challenges, which serve as the basis of future research. The next sections of the survey are the detailed review of them by the domain, by the measurements, by the methods and by the applications.

III. BY THE DOMAINS

We first review and categorize the works by the domains they addess, including environmental dynamics, environment, human, time, location, orientation, target device and AP as well as domain independent solutions. Table II lists the works according to the domains.

A. Environmental dynamics

The wireless environment has dynamics due to various factors. Typical environmental dynamics include furniture rearrangement, temperature and humidity change, opening or closing windows or doors, minor infrastructure reconfiguration. Environmental dynamics have influence on WiFi signal propagation and can cause mismatch between the testing measurements and the training measurements, which is the main reason of performance degradation of WiFi sensing in changed environments.

1) Localization: CRIL [83] integrated RSS-based channel modeling localization with inertial navigation to adapt to dynamic environments. AutoFi [84] applied contaminant removal and autoencoder to calibrate the CSI fingerprint profiles, to solve the inconsistency caused by environmental changes. Liu et al. [85] reshaped the distribution of RSS fingerprints in the target domain based on the transferred knowledge from the source domain to achieve robustness to environmental dynamics and different environments. MoLoc [86] leveraged a CSI fingerprint roaming model to adapt to dynamic ship motions for localization in mobile ship environments. AdapLoc [87] aligned the fingerprints of the source domain and the target domain in a shared space, so as to adapt the source localization model to environmental dynamics.

2) Activity & gesture recognition: WiCAR [88] removed domain specific information while retaining activity related information by deep multi-adversarial domain adaptation, to achieve in-car activity recognition adaptive to driving conditions, car models and human subjects. WiAG [89] achieved gesture recognition agnostic to environmental dynamics as well as positions and orientations, by generating virtual samples for all gestures in the target domains via a translation function. CrossSense [90] enabled gesture training samples to be collected once and used across dynamics and sites, by employing a roaming model to generate synthetic samples for each target domain.

3) Miscellaneous applications: The CSI-based fall detection system FallDeFi [91] used Short-Time Fourier Transform (STFT) to extract the time-frequency features and leveraged a sequential forward selection algorithm to single out the features that were resilient to environmental dynamics, different environments, time and human. RespiRadio [92] could detect

TABLE II
THE DOMAINS AND THE RELATED WORKS.

Domain	Application	Related work
Environmental dynamics	Localization	[83]–[87]
	Activity recognition	[88]
	Gesture recognition	[89], [90]
	Fall detection	[91]
	Respiration monitoring	[92]
	Intrusion detection	[93]
	Gait recognition	[90]
	User authentication	[94], [95]
Cross environment	Localization	[85], [96]–[99]
	Activity recognition	[88], [100]–[108]
	Gesture recognition	[90], [109]–[118]
	Fall detection	[91], [119]
	Intrusion detection	[93]
	Gait recognition	[90]
	People counting	[120], [121]
Human	Localization	[122]–[124]
	Tracking	[125]
	Activity recognition	[88], [101]–[105], [126]–[128]
	Gesture recognition	[111]–[113], [115], [117], [118], [129]
	Fall detection	[91]
Time	Localization	[130]–[147]
	Tracking	[148]
	Fall detection	[91]
	User authentication	[94]
Location	Activity recognition	[149]–[151]
	Gesture recognition	[89], [111], [117], [118], [152]
	User Authentication	[95], [153]
Orientation	Localization	[154]
	Tracking	[148]
	Gesture recognition	[89], [111], [117], [118]
	Gait recognition	[155]
Target device	Localization	[135], [137], [143], [147], [154], [156], [157]
	Tracking	[148]
Access point	Localization	[87], [140], [158], [159]
Domain independent	Localization	[23], [24], [160]–[165]
	Tracking	[32], [33], [35]–[37], [166]
	Activity recognition	[42]
	Gesture recognition	[167]–[169]

a person's respiration rate in dynamic ambient environments by synthesizing a wider-bandwidth WiFi radio. AR-Alarm [93] achieved intrusion detection adaptive to environmental changes and different environments, by utilizing the ratio between the dynamic and the static CSI profiles of the environment and the self-adaptive learning mechanism. CrossSense [90] enabled gait training samples to be collected once and used across dynamics and sites, by employing a roaming model to generate synthetic samples for each target domain. Regani et al. [94] built a long-term driver radio biometric database to make in-car driver authentication adaptive to changing in-car environments. Shi et al. [95] authenticated users by recognizing the activities, leveraging an unsupervised domain discriminator to mitigate the impact of environmental changes and varying location.

B. Cross environment

We use an environment to refer to a room, a building, a site, etc. Different environments have quite different signal propagation paths, due to different layout and WiFi setup. A WiFi sensing model trained in one environment often fails in a different environment. How to build a cross environment sensing model is the most challenging issue for cross domain WiFi sensing.

1) Localization: TransMapping [96] transferred the localization model from one area of a building to another, by learning a mapping function between signal space and location space based on manifold learning and sharing a low-dimensional manifold between data collected in different areas across the building. Wang et al. [97] built a localization model in a multi-floor building, collecting labeled data on one floor and unlabeled data on other floors. They co-embedded these different floors' data in a common low-dimensional manifold and aligned the unlabeled data with the labeled data to propagate the labels. Ohara et al. [98] transferred the localization model in the source environments to the target environment, using the variance of signal strength values within a sliding time window, requiring only unlabeled data and a floor plan of the target environment. FitLoc [99] localized the target over various areas in outdoor environments and similar furnitured indoor environments, by unifying the radio maps over various areas through a transfer scheme that combined Fisher Linear Discriminant Analysis (FLDA) and Bregman Divergence. Liu et al. [85] reshaped the data distribution of RSS fingerprints in the target domain based on the transferred knowledge from the source domain to achieve localization robust to different environments and environmental dynamics.

2) Activity recognition: Chang et al. [100] achieved environment independent activity recognition by leveraging an environment dependency removal method based on Singular Value Decomposition (SVD) to eliminate the background CSI and extract the channel information of signals reflected by human bodies. CARM [101] realized activity recognition robust to different environments and users based on a CSI-speed model that quantified the relation between CSI dynamics and human movement speeds and a CSI-activity model that quantified the relation between human movement speeds and human activities. Based on adversarial networks, EI [102] extracted environment and subject-independent features and achieved activity recognition across human subjects under different environments. WiFit [103] achieved bodyweight exercise counting and classification robust to environments and human using features extracted from CSI amplitude, CSI phase difference and Doppler velocity spectrum. WiSDAR [104] extended the multiple antennas of WiFi devices to construct multiple separated antenna pairs and integrated the features extracted by CNN and LSTM to achieve robustness to environments and human. WiCAR [88] removed domain specific information while retaining the activity related information through deep multi-adversarial domain adaptation and realized in-car activity recognition adaptive to car models, driving

conditions and human subjects. WiDrive [105] achieved in-car activity recognition based on CSI DFS, employing Hidden Markov Models with Gaussian Mixture emissions Model (HMM-GMM) as the classifier, whose parameters were updated by an online adaptation algorithm to adapt for different vehicles and drivers. Sheng et al. [106] achieved CSI-based cross scene activity recognition by integrating the spatial features from CNN into BLSTM and fine-tuning the pre-trained model in the new environment. MatNet-eCSI [107] and HAR-MN-EF [108] employed Matching Network with enhanced CSI, which enhanced the activity dependent features whilst mitigating the behavior unrelated information, to accomplish environment independent activity recognition.

3) Gesture recognition: JADA [109] and WiADG [110] leveraged unsupervised joint adversarial domain adaptation to realize CSI gesture recognition without labeling the training data in new environments. CrossSense [90] scaled up WiFi gesture recognition to new environments by training a roaming model that generated synthetic samples for each target environment. Widar3.0 [111] achieved gesture recognition by deriving the domain independent feature body-coordinate velocity profile (BVP), adaptive to different environments, locations, orientations and persons. Yang et al. [112] used a deep Siamese representation learning architecture incorporating CNN and Bidirectional Recurrent Neural Network (Bi-RNN) for one-shot gesture recognition, and helped to remove structured noise such as various measurement conditions and individual heterogeneity. Wang et al. [113] alleviated the efforts of retraining the gesture recognition system in a new scenario by learning a universal similarity evaluation ability utilizing a deep similarity evaluation networks. To shrink domain discrepancy in different environments, the CSI gesture recognition scheme DANGR [114] adopted domain adaptation based on multi-kernel Maximum Mean Discrepancy (MMD), which matched the mean-embeddings of abstract representations across domains in a Reproducing Kernel Hilbert Space (RKHS). The CSI gesture recognition system presented by Ma et al. [115] based on similarity evaluation could recognize new type of gestures, or gestures performed in a new scenario or by a new user, using very few number of or even one sample of the new gesture. By cross domain distillation, MobileDA [116] allowed a teacher network trained in the server to distill the knowledge for a student network that adapted to the new environment by diminishing domain disparity. Kang et al. [117] realized gesture recognition based on DFS through multi-source unsupervised domain adaptation, which consisted of an adversarial learning scheme together with feature disentanglement and an attention scheme, considering environment, gesture performer, location and orientation. Niu et al. [118] developed a model to quantify the relation between signal frequency and target position, motion direction and speed. They proposed movement fragments and relative motion direction changes as two position independent features for gesture recognition under various conditions, such as different environments, locations, orientations, and persons.

4) Miscellaneous applications: The CSI-based fall detection system FallDeFi [91] used STFT to extract the time-frequency features and leveraged a sequential forward selection algorithm to single out the features that were resilient to different environments, environmental dynamics, time and human. TL-Fall [119] achieved cross environment fall detection based on transfer learning, which was trained with labeled data in the old environment, and trained with only a few labeled data in the new environment. AR-Alarm [93] achieved intrusion detection adaptive to different environments and environmental changes, by utilizing the ratio between the dynamic and the static CSI profiles of the environment and the self-adaptive learning mechanism. CrossSense [90] enabled gait training samples to be collected once and used across sites and dynamics, by employing a roaming model to generate synthetic samples for each target domain. Di Domenico et al. [120] focused on crowd counting without retraining in new environments, by identifying a set of differential CSI feature candidates and selecting the most effective ones via minimization of the summation of the Davies-Bouldin indexes. They [121] presented another crowd counting system without retraining in new environments, by analyzing the shape of the Doppler spectrum of the received WiFi signals, which was correlated to the number of people.

C. Human

Due to different physical features, different people have different influence on the surrounding wireless signals. The sensing model should work well for all targets.

1) Localization & tracking: Wang et al. [122], [123] transferred the distorted RSS across different categories of targets into a latent feature space, where the distributions of the distorted RSS from different categories of targets were unified for localization. FiDo [124] localized different users with labeled data from only one or two example users, leveraging a data augmenter that introduced data diversity with a variational autoencoder, and a domain adaptive classifier that adjusted itself to newly collected unlabeled data with a joint classification-reconstruction structure. FreeTrack [125] was a CSI tracking system based on DNN and Particle Filtering. To combat with environmental variants such as human diversity, fine-tuning was leveraged to enhance the adaptivity.

2) Activity recognition: CARM [101] realized activity recognition robust to different users and environments based on a CSI-speed model that quantified the relation between CSI dynamics and human movement speeds and a CSI-activity model that quantified the relation between human movement speeds and human activities. Based on adversarial networks, EI [102] extracted environment and subject independent features and achieved activity recognition across human subjects under different environments. WiFit [103] achieved bodyweight exercise counting and classification robust to environments and human using features extracted from CSI amplitude, CSI phase difference and Doppler velocity spectrum. WiSDAR [104] extended the multiple antennas of WiFi devices to construct multiple separated antenna pairs and integrated the features extracted by CNN and LSTM to achieve robustness to environments and human. WiCAR [88] removed domain specific information while retaining activity related information by deep multi-adversarial domain adaptation, to

achieve in-car activity recognition adaptive to human subjects, driving conditions and car models. WiDrive [105] achieved in-car activity recognition based on CSI DFS, employing HMM-GMM as the classifier, whose parameters were updated by an online adaptation algorithm to adapt for different drivers and vehicles. CsiGAN [126] enabled CSI-based activity recognition adaptive to user changes based on semi-supervised Generative Adversarial Networks (GAN) to meet the scenarios that unlabeled data from left out users were very limited. TL-HAR [127] transformed CSI amplitude and phase to images for multiple human activity recognition. Zhang et al. [128] synthesized variant activity data through CSI transformation methods to mitigate the impact of activity inconsistency and subject specific issues for activity recognition.

3) Gesture recognition: Based on transfer learning and semi-supervised learning, WiSign [129] achieved CSI-based sign language recognition adaptive to new users, requiring only sparsely labeled training samples. Widar3.0 [111] achieved gesture recognition by deriving the domain independent feature BVP, adaptive to different persons, environments, locations and orientations. Yang et al. [112] used a deep Siamese representation learning architecture incorporating CNN and Bi-RNN for one-shot gesture recognition, and helped to remove structured noise such as individual heterogeneity and various measurement conditions. Wang et al. [113] alleviated the efforts of retraining the gesture recognition system in a new scenario by learning a universal similarity evaluation ability utilizing a deep similarity evaluation networks. The CSI gesture recognition system presented by Ma et al. [115] based on similarity evaluation could recognize new type of gestures, or gestures performed by a new user or in a new scenario, using very few number of or even one sample of the new gesture. Kang et al. [117] realized gesture recognition based on DFS through multi-source unsupervised domain adaptation, which consisted of an adversarial learning scheme together with feature disentanglement and an attention scheme, considering gesture performer, environment, location and orientation. Niu et al. [118] developed a model to quantify the relation between signal frequency and target position, motion direction and speed. They proposed movement fragments and relative motion direction changes as two position independent features for gesture recognition under various conditions, such as different persons, locations, orientations and environments.

4) Fall detection: The CSI-based fall detection system FallDeFi [91] used STFT to extract the time-frequency features and leveraged a sequential forward selection algorithm to single out the features that were resilient to different human, environments, environmental dynamics and time.

D. Time

WiFi signals fluctuate over time, even if nothing really happens, causing distribution discrepancy between the training and the testing samples. The sensing model should work well over time.

1) Localization: Most of the works dealing with time are for localization. Researches on localization over time dates back to more than ten years ago. Yin et al. [130], [131] adapted the temporal radio maps for localization by regression analysis to learn the temporal relationship between the RSS received by sparsely located reference points and that received by the mobile device. Pan et al. [132] proposed LeManCoR to adapt the mapping between RSS and locations over different time based on Manifold Co-Regularization. They [133] then proposed a transfer learning method via dimensionality reduction to find a latent space, in which the distance between RSS distributions in different domains was minimized. They [134] later proposed Transfer Component Analysis (TCA) for domain adaptation, which learned the transfer components across domains in RKHS using MMD. LuMA [135] handled the transfer learning problem caused by time and device in RSS localization by learning a mapping between a source domain and a target domain in a low-dimensional space. Zheng et al. [136] applied a semi-supervised Hidden Markov Model (HMM), called transferred HMM (TrHMM), to transfer the learned localization model from one time to another. WiGEM [137] adjusted the localization model parameters dynamically from the packets captured by the sniffers based on Gaussian Mixture Model (GMM) and Expectation Maximization (EM), thus made the location estimates robust for time and device heterogeneity. EIL [138] achieved time independent localization by eliminating the interference of environment on RSS over time. AcMu [139] realized automatic radio map update for localization, by pinpointing mobile devices with a trajectory matching algorithm with the aid of inertial sensors and using them as mobile reference points. HED [140] combated AP sequence disorders and signal variance due to time for localization, using the relative values of RSS and the order number of a given AP in a sorted fingerprint. Xu et al. [141] proposed a Metric Transfer Learning Framework (MTLF) and verified it on localization over time. Instance weights were learned to bridge the distributions of different domains, while Mahalanobis distance was learned to minimize the intra-class distances and maximize the inter-class distances for the target domain. TKL-WinSMS [142] realized adaptive localization over time by using transfer kernel learning, which learned a domain invariant kernel by matching the source and the target distributions in RKHS. AP-Sequence [143] divided the area into small regions and assigned an AP sequence to each region as a fingerprint. To achieve low overhead in fingerprint map maintenance over time and heterogeneous devices, AP-Sequence leveraged dynamic region partitioning and generated fingerprints based on relative RSS differences among various APs. HAIL [147] leveraged both relative and absolute RSS to achieve adaptive localization over time and heterogeneous devices, using a neural network to measure the fingerprint similarity based on RSS. Rao et al. [144] proposed MSDFL which converted the fingerprints of the changed environments to the original environment via a conversion model based on Multilayer Perceptron (MLP) to solve the problem of environmental change over time in localization. They [145] later utilized transfer learning to learn new feature representations from the CSI samples as fingerprints, which could simultaneously minimize the intra-class differences, maximize the inter-class differences, and minimize the distribution differences between training and testing samples. They [146] also used CSI phase

for localization. In view of the unpredictable nature of calibrated CSI phase over time, they proposed a transfer deep supervised neural network method to solve it.

2) Tracking: Kim et al. [148] exploited maximal RSS combined with PDR to solve the RSS variance problem of tracking over time, device type, device placement and user orientation.

3) Miscellaneous applications: The CSI-based fall detection system FallDeFi [91] used STFT to extract the time-frequency features and leveraged a sequential forward selection algorithm to single out the features that were resilient to different time, environments, environmental dynamics and human. To deal with the in-car changing environment over time for driver authentication, Regani et al. [94] built a long-term driver radio biometric database to made the system adaptive to new in-car environments.

E. Location

Some WiFi sensing applications such as activity recognition and gesture recognition depend on the locations of the targets. The WiFi signals are influenced not only by the activity or the gesture but also by the location.

1) Activity recognition: FALAR [149] reconstructed CSI data by Class Estimated Basis Space Singular Value Decomposition (CSVD) and achieved location independent activity recognition. WiLISensing [150] leveraged CNN and fine-tuning to recognize activities in a position without training or with very few training samples. Li et al. [151] leveraged Angle Difference of Arriva (ADOA) calculated from CSI to realize cross location activity recognition.

2) Gesture recognition: WiAG [89] achieved gesture recognition agnostic to positions and orientations as well as environmental dynamics, by generating virtual samples for all gestures in the target domains via a translation function. Widar3.0 [111] achieved gesture recognition by deriving the domain independent feature BVP, adaptive to different locations, orientations, persons and environments. WiHand [152] realized location independent gesture recognition, by leveraging the low rank and sparse decomposition to separate the gesture signals from the background information. Kang et al. [117] realized gesture recognition based on DFS through multi-source unsupervised domain adaptation, which consisted of an adversarial learning scheme together with feature disentanglement and an attention scheme, considering location, orientation, gesture performer and environment. Niu et al. [118] developed a model to quantify the relation between signal frequency and target position, motion direction and speed. They proposed movement fragments and relative motion direction changes as two position-independent features for gesture recognition under various conditions, such as different locations, orientations, environments, and persons.

3) User authentication: Jung et al. [153] realized identity verification based on the gesture signals of handwritten signature captured by CSI. A ConvNet was pre-trained from one position and transferred to another position via the kernel and the range space projection learning. Shi et al. [95] authenticated users by recognizing the activities, leveraging an unsupervised domain discriminator to mitigate the impact of varying location and environmental changes.

F. Orientation

Orientation of the target also has impact on WiFi sensing. KARMA [154] utilized a one-time fingerprint for localization and applied a set of causality calibration functions in real-time to compensate for the change in the factors such as orientation and device. Kim et al. [148] exploited maximal RSS combined with PDR to solve the RSS variance problem of tracking over user orientation, device type, device placement and time. WiAG [89] achieved gesture recognition agnostic to positions and orientations as well as environmental dynamics, by generating virtual samples for all gestures in the target domains via a translation function. Widar3.0 [111] achieved gesture recognition by deriving the domain independent feature BVP, adaptive to different locations, orientations, persons and environments. Kang et al. [117] realized gesture recognition based on DFS through multi-source unsupervised domain adaptation, which consisted of an adversarial learning scheme together with feature disentanglement and an attention scheme, considering location, orientation, gesture performer and environment. Niu et al. [118] developed a model to quantify the relation between signal frequency and target position, motion direction and speed. They proposed movement fragments and relative motion direction changes as two position-independent features for gesture recognition under various conditions, such as different locations, orientations, environments, and persons. Based on the Fresnel model, WiDIGR [155] generated direction independent signal spectrograms and achieved direction independent gait recognition through signal processing techniques to eliminate the differences among induced signals caused by walking in different directions.

G. Target device

For device-based WiFi sensing, the diversity of target device may affect the performance. Almost all the works addressing this issue are for localization. LuMA [135] handled the transfer learning problem caused by device and time in RSS localization by learning a mapping between a source domain and a target domain in a low-dimensional space. Zheng et al. [156] leveraged a latent multi-task learning algorithm to solve the multi-device localization problem. They employed an alternating optimization approach to iteratively learn feature mappings and multi-task regression models for the devices. WiGEM [137] adjusted the localization model parameters dynamically from the packets captured by the sniffers based on GMM and EM, thus made the location estimates robust for device heterogeneity and time. KARMA [154] utilized a one-time fingerprint for localization and applied a set of causality calibration functions in real-time to compensate for the change in the factors such as device and orientation. AP-Sequence [143] divided the area into small regions and assigned an AP sequence to each region as a fingerprint. To achieve low overhead in fingerprint map maintenance over heterogeneous devices and time, AP-Sequence leveraged dynamic region partitioning and generated fingerprints based

on relative RSS differences among various APs. WiDeep [157] combined stacked denoising autoencoders and a probabilistic framework to handle the noise and the device heterogeneity and captured the relationship between the RSS heard by the mobile phone and its location. HAIL [147] leveraged both relative and absolute RSS to achieve adaptive localization over heterogeneous devices and time, using a neural network to measure the fingerprint similarity based on RSS. Kim et al. [148] exploited maximal RSS combined with PDR to solve the RSS variance problem of tracking over device type, device placement, user orientation and time.

H. Access point

For WiFi sensing, the APs may change and affect the performance. Almost all the works dealing with AP change are for localization. HED [140] combated AP sequence disorders and signal variance due to time for localization, using the relative values of RSS and the order number of a given AP in a sorted fingerprint. By employing implicit crowdsourced signals, LAAFU [158] identified altered APs and filtered them out and employed non-parametric Gaussian process regression to update the radio map. CRISLoc [159] detected altered APs with a joint clustering and outlier detection method, and used a transfer learning approach to reconstruct the CSI fingerprint database on the basis of the outdated fingerprints and a few fresh measurements. AdapLoc [87] aligned the fingerprints of the source domain and the target domain in a shared space, so as to adapt the source localization model to the changed environment, such as transceiver type change or transceiver location change.

I. Domain independent

The above works aim to adapt to different domains, while some other works aim to achieve domain independence. These methods do not require offline training, thus are agnostic to domains.

1) Localization: ArrayTrack [160] calculated the target location by computing the AoA from the frames transmitted by the target and overheard by multiple APs. SpotFi [161] achieved decimeter level localization using a super-resolution AoA estimation algorithm. Chronos [162] computed ToF to localize the targets to within tens of centimeter. SiFi [163] achieved single AP indoor localization based on ToA. LiFS [23] estimated the target location by modeling the CSI measurements of multiple wireless links as a set of power fading based equations using only the subcarriers affected by multipath. S-Phaser [24] utilized calibrated CSI phase to compute the direct path length between an AP and terminals and used a geometric positioning algorithm to determine the target location. Wan et al. [164] and Tong et al. [165] presented a calibration-free CSI phase fingerprinting localization system via hybrid inter-antenna distance and context-aware fingerprinting recognition, in which the fingerprinting database was constructed theoretically without site survey.

2) Tracking: MaTrack [32] proposed the Dynamic-MUSIC method to detect the subtle reflection signals from human body and further differentiate them from those reflected signals

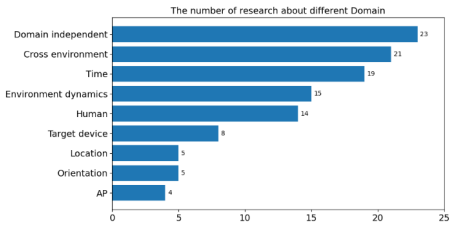


Fig. 7. Number of papers by the domains.

from static objects to identify the target's angle for tracking. IndoTrack [36] used Doppler-MUSIC to extract Doppler velocity information from CSI samples and used Doppler-AoA to determine the trajectory of the target by jointly estimating target velocity and location via probabilistic co-modeling of spatial-temporal Doppler and AoA information. Karanam et al. [37] estimated the AoA of all the signal paths arriving at a receiver array using the corresponding received signal magnitude measurements. Widar [33] simultaneously estimated human moving velocities and locations by modeling the relationships between CSI dynamics and locations and velocities exploiting Doppler shifts. Widar2.0 [35] achieved localization and tracking using a single WiFi link, by building a unified model accounting for AoA, ToF and Doppler shifts together and devising an efficient algorithm for their joint estimation. DeFi [166] estimated the target location by refining the AoA of the target reflection path based on CSI. They eliminated the static paths using a background elimination algorithm, and separated out the target reflection path from the motion paths according to the AoA and equivalent ToF.

3) Activity & gesture recognition: Zhang et al. [42] leveraged the Fresnel diffraction model and proposed a diffraction-based sensing model to quantitatively determine the signal propagation with respect to a target's motions in the first Fresnel zone. WiDeo [167] achieved domain independent motion tracing using RF Backscatter, by calculating each reflection's amplitude, ToF and AoA. AirDraw [168] was a learning-free in-air handwriting system by passive gesture tracking using CSI phase. They denoised CSI measurements by the ratio between two close-by antennas, separated the reflected signal from noise by PCA, and proposed a signal calibration algorithm for tracking correction by eliminating the static components unrelated to hand motion. FingerDraw [169] tracked centimeter-level finger drawings by leveraging the CSI-quotient model, which used the channel quotient between two antennas of the receiver to cancel out the noise in CSI amplitude and the random offsets in CSI phase, and quantified the correlation between CSI value dynamics and object displacement.

J. Remarks

Some works try to achieve domain independence, while others try to adapt to different domains. Among xx works, xx works are on domain independence, xx works are on cross environments, xx works are on time, xx works are on environmental dynamics and xx works are on human diversity.

TABLE III
THE MEASUREMENTS AND THE RELATED WORKS.

Measurement	Application	Related work
RSS fingerprint	Localization	[85], [96], [97], [99], [122], [123], [130]–[136], [138]–[143], [147], [154], [156]–[158]
	Tracking	[148]
RSS propagation	Localization	[83], [98], [137]
RSS AoA	Localization	[160]
CSI amplitude	Localization	[84], [86], [87], [124], [144], [145], [159]
	Tracking	[125]
	Activity recognition	[42], [88], [100]–[104], [106]–[108], [126]–[128], [149], [150]
	Gesture recognition	[89], [90], [113]–[115], [129], [152], [169]
	Fall detection	[91], [119]
	Respiration monitoring	[92]
	Gait recognition	[90], [155]
	User authentication	[95], [153]
	People counting	[120]
CSI phase	Localization	[31], [86], [146], [164], [165], [170]
	Activity recognition	[101], [103], [107], [127]
	Gesture recognition	[109], [110], [112], [114], [116], [168], [169]
	Intrusion detection	[93]
	User authentication	[94]
CSI propagation	Localization	[23], [24]
CSI AoA/ToF	Localization	[161]–[163]
	Tracking	[32], [35]–[37], [166]
	Activity recognition	[151]
	Gesture recognition	[167]
CSI DFS	Tracking	[33], [35], [36]
	Activity recognition	[103], [105]
	Gesture recognition	[111], [117], [118]
	People counting	[121]

IV. BY THE MEASUREMENTS

Researchers used different measurements from WiFi signals to realize sensing, including RSS, CSI and the derived measurements from them. We review the works from the perspective of measurements, as listed in Table III.

A. RSS fingerprint

A RSS fingerprint is usually defined as the vector containing the RSS values from the visible APs. Suppose there are m visible APs in the environment. A RSS fingerprint is defined as

$$\boldsymbol{s} = (s_1, s_2, \cdots, s_i, \cdots, s_m) \tag{14}$$

where s_i represents the RSS value from the i-th AP.

Almost all the works based on RSS fingerprint that dressed the cross domain issue were on localization. Yin et al. [130], [131], LeManCoR [132], Pan et al. [133], [134], TrHMM [136], EIL [138], AcMu [139], Xu et al. [141] and TKL-WinSMS [142] used RSS fingerprints to achieve indoor localization over time. TransMapping [96], Wang et al. [97] and FitLoc [99] used RSS fingerprints to achieve indoor localization across different environments. Zheng et al. [156], WiDeep [157] and HAIL [147] realized RSS fingerprinting localization over heterogeneous target devices. LuMA [135] realized RSS fingerprinting localization over time and target devices. KARMA [154] achieved RSS fingerprinting localization over orientations and target devices. TLCS [122], [123] used RSS fingerprints for indoor localization over different target categories. HED [140] realized RSS fingerprinting localization over time and AP changes. LAAFU [158] realized RSS fingerprinting localization over AP changes. Liu et al. [85] achieved RSS fingerprinting localization adaptive to different environments and environmental dynamics. AP-Sequence [143] divided the area into small regions and assigned an AP sequence to each region as a fingerprint and achieved low overhead fingerprint map maintenance over time and target devices. Kim et al. [148] exploited maximum RSS combined with PDR to achieve tracking over time, orientation and target devices.

B. RSS propagation

Some works modeled radio propagation to realize cross domain localization. WiGEM [137] leveraged RSS propagation model to achieve indoor localization over time and target devices. Ohara et al. [98] leveraged RSS variance model to achieve localization across different environments. CRIL [83] leveraged RSS channel model and INS to achieve localization robust to environmental dynamics.

C. RSS AoA

To achieve domain independent localization, ArrayTrack [160] utilized the AoA information calculated from RSS by multiple APs using the frames sent out by the mobile device to locate the mobile device.

D. CSI amplitude

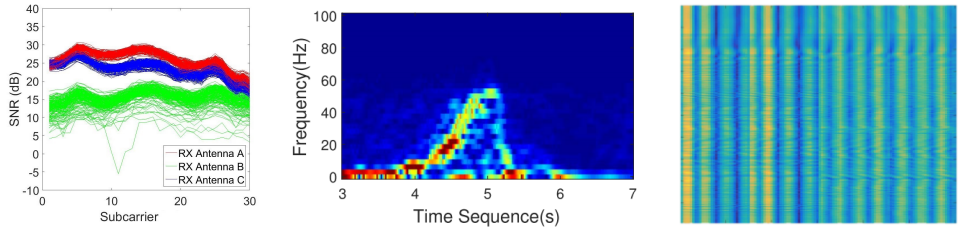
Since the extraction of CSI became possible, WiFi sensing based on CSI has exploded. CSI contains finer-grained information than RSS and outperforms RSS with respect to WiFi sensing significantly. The most common use of CSI is the amplitude values of the subcarriers. There are different ways of using CSI amplitude, including CSI amplitude fingerprint, CSI amplitude spectrogram and CSI amplitude image.

A CSI amplitude fingerprint is defined as the vector of amplitude values from all the subcarriers. Suppose there are N subcarriers in the communication link. A CSI amplitude fingerprint is defined as

$$\boldsymbol{h} = (|h_1|, |h_2|, \cdots, |h_i|, \cdots, |h_N|) \tag{15}$$

where $|h_i|$ represents the amplitude value of the i-th subcarrier. In Fig. 8(a), each waveform represents a CSI amplitude fingerprint.

Some researchers use CSI amplitude spectrogram to realize WiFi sensing. A CSI amplitude spectrogram is generated by using Short Time Fourier Transform (STFT) on CSI amplitude



(a) CSI amplitude (b) CSI spectrogram (c) Amplitude image

Fig. 8. CSI amplitude and its derivations.

to extract features on both the time domain and the frequency domain. Fig. 8(b) is an example of CSI amplitude spectrogram.

Some researchers transformed CSI amplitude to an image to realize WiFi sensing, so that image processing technology can be applied. Fig. 8(c) is an example of CSI amplitude image, on which the x-axis represents the subcarriers, the y-axis represents the packet/time, and the color represents the amplitude value.

Localization & tracking. Using CSI amplitude fingerprint, AutoFi [84] realized localization adaptive to environmental dynamics, MSDFL [144], [145] realized localization over time, FiDo [124] realized localization across human, CRISLoc [159] realized smartphone-based localization adaptive to AP changes, FreeTrack [125] realized tracking across human, AdapLoc [87] realized localization adaptive to environmental dynamics and AP changes. Using CSI amplitude and phase fingerprint, MoLoc [86] realized robust localization in the mobile ship environment.

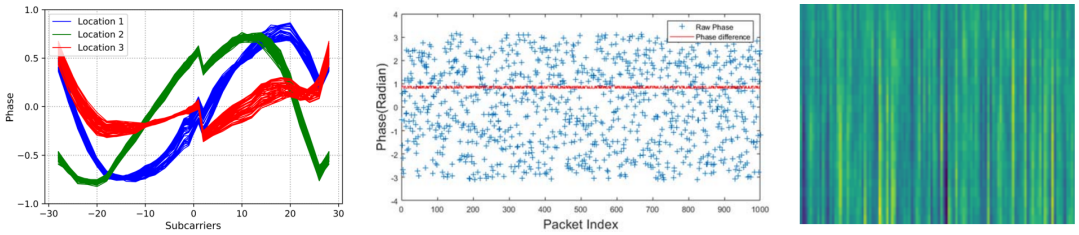
Activity recognition. Using CSI amplitude, FALAR [149] realized location independent activity recognition, WiLISensing [150] leveraged CNN and fine-tuning to recognize activities across locations, EI [102] achieved cross environment and cross human activity recognition, CsiGAN [126] realized human independent activity recognition, Sheng et al. [106] achieved cross environment activity recognition, HAR-MN-EF [108] accomplished environment independent activity recognition. Zhang et al. [42] leveraged the Fresnel diffraction model to quantitatively determine the signal propagation with respect to the target motions in the first Fresnel zone, hence to detect human activities using CSI amplitude. WiSDAR [104] used CSI amplitude and spectrogram for activity recognition robust to environments and human. WiCAR [88] used CSI amplitude spectrogram to realize in-car activity recognition adaptive to new driving conditions, car models and human subjects. Zhang et al. [128] achieved activity recognition based on CSI amplitude spectrogram by synthesizing variant activity data through CSI transformation to mitigate the impact of activity inconsistency and subject-specific issues. Using CSI amplitude and phase, Chang et al. [100] transformed CSI to images to perform cross environment activity recognition, TL-HAR [127] transformed CSI to images to realize multiple human activity recognition, MatNet-eCSI [107] achieved activity recognition across environments. CARM [101] used CSI amplitude, CSI phase and spectrogram for human activity recognition robust to environments and human. WiFit [103] achieved bodyweight exercises monitoring robust to environments and human, using the features extracted from CSI amplitude, CSI phase difference and Doppler velocity spectrum.

Gesture recognition. Using CSI amplitude, WiAG [89] realized gesture recognition agnostic to location and orientation as well as environmental changes, WiSign [129] realized sign language recognition adaptive to new users, WiHand [152] achieved gesture recognition independent on locations. CrossSense [90] used both CSI amplitude fingerprint and RSS fingerprint for gesture recognition and localization across environments and human. Wang et al. [113] used CSI amplitude image for gesture recognition across environments and human. Ma et al. [115] transformed CSI amplitude to image for gesture recognition adaptive to new environments and users. DANGR [114] used both CSI amplitude and phase for cross environment gesture recognition. FingerDraw [169] tracked centimeter-level finger drawings by leveraging the CSI-quotient model and the Fresnel model using CSI amplitude and phase.

Miscellaneous. FallDeFi [91] used CSI amplitude and spectrogram for fall detection across environments and human, TL-Fall [119] used CSI amplitude for fall detection across environments. RespiRadio [92] used CSI amplitude for respiration rate monitoring in dynamic environments based on the Fresnel model. CrossSense [90] used both CSI amplitude fingerprint and RSS fingerprint for gait recognition across environments and environmental dynamics. WiDIGR [155] achieved direction independent gait recognition based on the Fresnel model and a series of signal processing techniques to generate a direction independent signal spectrogram. Jung et al. [153] verified identity using CSI amplitude based on the gesture signals of handwritten signature adaptive to new locations. Shi et al. [95] presented user authentication by recognizing activities, using CSI amplitude, relative amplitude and STFT holograms adaptive to locations and environmental changes. Di Domenico et al. [120] focused on crowd counting without retraining in new environments, by identifying a set of differential CSI feature based on CSI amplitude.

E. CSI phase

Human and their movements may change the length of some signal propagation paths, resulting in phase offset, which can be used to deduct human motions. As the phase offset may be caused by various factors, the raw phase extracted from commodity WiFi can not be used directly. The researchers



(a) CSI phase (b) Phase difference (c) Phase image

Fig. 9. CSI phase and its derivations.

proposed various ways to remove random phase shifts to use CSI phase, including CSI phase fingerprint, CSI phase difference and CSI phase image.

A CSI phase fingerprint is defined as the vector of phase values from all the subcarriers. Suppose there are N subcarriers in the communication link. A CSI phase fingerprint is defined as

$$\boldsymbol{a} = (\angle h_1, \angle h_2, \cdots, \angle h_i, \cdots, \angle h_N) \tag{16}$$

where $\angle h_i$ represents the phase value of the i-th subcarrier. Fig. 9(a) shows the calibrated CSI phase.

Many researchers use CSI phase difference from two antennas to realize WiFi sensing. Assume $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ are two CSI phase fingerprints received by two antennas, respectively. The CSI phase difference is defined as

$$\begin{aligned} \boldsymbol{d} &= \boldsymbol{a_1} - \boldsymbol{a_2} \\ &= (|\angle h_{11} - \angle h_{21}|, |\angle h_{12} - \angle h_{22}|, \cdots, |\angle h_{1N} - \angle h_{2N}|) \end{aligned} \tag{17}$$

Fig. 9(b) shows the phase difference from two antennas.

Some researchers transformed CSI phase to image to realize WiFi sensing, so that image processing technology can be applied. Fig. 9(c) is an example of CSI phase image, on which the x-axis represents the subcarriers, the y-axis represents the packet/time, and the color represents the calibrated phase value.

Localization. Rao et al. [146] used CSI phase fingerprint to achieve localization over time. Wan et al. [164] and Tong et al. [165] realized calibration-free CSI phase fingerprinting localization by generating the theoretical CSI phase fingerprint database and achieved domain independence. MoLoc [86] realized localization in the mobile ship environment based on CSI amplitude and phase fingerprint. MFDL [170] and Fei et al. [31] used the Fresnel zone model to achieve localization, which determined the elliptic region that the target located in according to the CSI phase.

Activity recognition. CARM [101] used CSI amplitude, CSI phase and spectrogram for human activity recognition robust to environments and users. WiFit [103] achieved bodyweight exercise monitoring robust to environments and human using features extracted from CSI amplitude, CSI phase difference and Doppler velocity spectrum. TL-HAR [127] transformed CSI amplitude and phase to images for multiple human activity recognition. Shi et al. [107] leveraged CSI amplitude and phase to achieve activity recognition across different environments.

Gesture recognition. JADA [109] and WiADG [110] leveraged CSI phase difference to recognize gestures across environments. Yang et al. [112] leveraged CSI phase difference to recognize gestures across environments and human. MobileDA [116] trained a simple student model that adapted to the new environment by cross domain distillation and evaluated on gesture recognition using CSI phase difference. DANGR [114] used both CSI amplitude and phase for cross environment gesture recognition. AirDraw [168] used CSI phase to achieve learning-free in-air handwriting by passive gesture tracking and achieved domain independence. FingerDraw [169] tracked centimeter-level finger drawings by leveraging the CSI-quotient model and the Fresnel model using CSI amplitude and phase.

Miscellaneous. AR-Alarm [93] utilized CSI phase difference as the salient signal to achieve intrusion detection in different environments and environmental dynamics without calibration efforts. Regani et al. [94] utilized CSI phase to achieve driver authentication over time and environmental dynamics.

F. CSI propagation

LiFS [23] estimated the target locations without offline training by modeling the CSI measurements of multiple wireless links as a set of power fading based equations using only the subcarriers not affected by multipath. S-Phaser [24] utilized calibrated CSI phase to compute the direct path length between an AP and terminals and used a geometric positioning algorithm to determine the target location. They are propagation model based localization systems. They do not require explicit pre-deployment effort, hence they are domain independent.

G. CSI AoA or ToF

Some works use AoA and/or ToF to achieve cross domain WiFi sensing, since AoA and ToF are domain agnostic. AoA and ToF can be calculated from calibrated CSI phase. The fundamental principle of AoA estimation is that incident signals from different angles incur different phase differences across the antennas in an array, as shown in Fig 10. Therefore, the phase shifts across antennas and subcarriers can be used to obtain the AoA and the equivalent TOF estimations.

SpotFi [161] calculated ToF and AoA from CSI and estimated the target location by using the direct path AoA estimates and RSS measurements from all the APs.

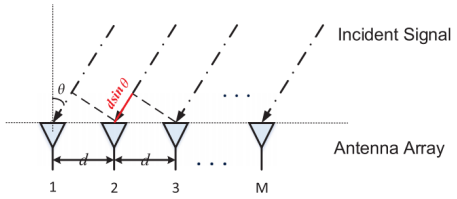
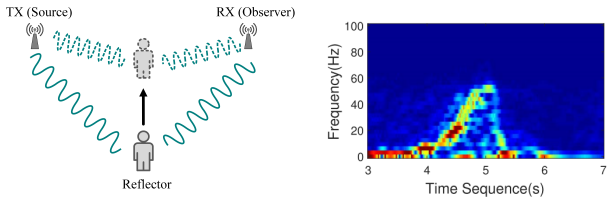


Fig. 10. AoA estimation.



(a) Doppler shift (b) DFS

Fig. 11. Doppler effect and DFS.

Chronos [162] calculated ToF from CSI to realize location tracking. SiFi [163] calculated ToA from CSI to realize location tracking. MaTrack [32] used AoA calculated from CSI to achieve human tracking. IndoTrack [36] combined Doppler shift with the AoA spectrum to estimate human velocity and location, thus realize human tracking. DeFi [166] used AoA and ToF from CSI to realize location tracking. Widar2.0 [35] built a unified model accounting for AoA, ToF and Doppler shift together to realize human tracking with a single WiFi link. Karanam et al. [37] estimated AoA using the received signal magnitude measurements and realized localization and tracking. Li et al. [151] leveraged Angle Difference of Arriva (ADOA) calculated from CSI to achieve cross location activity recognition. WiDeo [167] calculated each reflection's amplitude, ToF and AoA to realize motion tracing.

H. CSI DFS

Human movements can cause changes in the length of the reflection path, resulting in Doppler frequency shifts (DFS). By measuring the signal frequency, the direction, speed, and distance involved with the human movement can be deduced. In the configuration of device-free WiFi sensing, the transmitter (as the source) and the receiver (as the observer) are statically deployed. A person (as the reflector) performing actions alters the wireless transmission frequencies, inducing Doppler effect. As OFDM divides the bandwidth into multiple subcarriers and modulates data in each subcarrier, there is the possibility to identify small Doppler shifts from WiFi CSI [171]. Fig. 11 shows an example of spectrogram of frequency shifts caused by a human passing a WiFi transmission link. Researchers made use of DFS to achieve various WiFi sensing tasks.

Localization & tracking. Widar [33] estimated human moving velocity and location by modeling the relationships between CSI dynamics and locations and velocities exploiting Doppler shifts. Widar2.0 [35] built a unified model accounting for AoA, ToF and Doppler shifts together to realize domain independent human tracking with a single WiFi link. IndoTrack [36] combined DFS with AoA spectrum to estimate human velocity and location, thus realize domain independent human tracking.

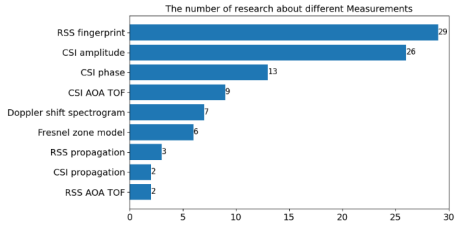


Fig. 12. Number of papers by the measurements.

Activity recognition. WiFit [103] achieved bodyweight exercise monitoring robust to environments and human, using features extracted from CSI amplitude, CSI phase difference and Doppler velocity spectrum. WiDrive [105] used DFS calculated from CSI to realize in-car driver activity recognition adaptive to drivers and vehicles.

Gesture recognition. Widar3.0 [111] achieved cross domain gesture recognition by deriving a domain independent feature BVP from DFS, adaptive to environments, locations, orientations and persons. Kang et al. [117] realized gesture recognition based on DFS through multi-source unsupervised domain adaptation, considering environment, gesture performer, location and orientation. Niu et al. [118] proposed movement fragments and relative motion direction changes as two position independent features that could be extracted from DFS and developed a robust gesture recognition system.

People counting. Di Domenico et al. [121] focused on crowd counting without retraining in new environments by analyzing the shape of the Doppler spectrum of the received WiFi signal which was correlated to the number of people.

I. Remarks

The majority of works leverage RSS fingerprint and CSI amplitude to realize WiFi sensing, which have xx works and xx works, respectively. The following are CSI phase, CSI AoA or ToF, and DFS, in which xx works use CSI phase, xx works use CSI AoA or ToF, and xx works use DFS.

V. By the Methods

To deal with the cross domain issue, many methods have been proposed for various applications. We category them in Table IV. The methods of WiFi sensing can be classified to learning-based and model-based. Learning-based methods try to learn the relationship between the features and the labels, using training samples of CSI measurements and the corresponding ground truth labels, while model-based methods are based on the physical theory of signal processing. To solve the cross domain issue, learning-based methods try to extract domain independent features or transfer the knowledge from the source domains to the target domains, while model-based methods are often domain independent but quite challenging.

TABLE IV
THE METHODS AND THE RELATED WORKS.

Category	Method	Application	Related work
Learning based	Radio map updating	Localization	[99], [130], [131], [136], [139], [143], [158], [159]
	Feature extraction	Localization	[84], [124], [140], [147], [157]
		Tracking	[148]
		Activity recognition	[100], [101], [103]–[105], [107], [108], [149]
		Gesture recognition	[111]–[113], [115], [118], [152]
		Fall detection	[91]
		Intrusion detection	[93]
		User authentication	[94], [95]
		People counting	[120], [121]
	Feature mapping	RSS localization	[85], [96], [97], [122], [123], [132]–[135], [138], [141], [142], [154], [156]
		CSI localization	[86], [87], [144]–[146]
		Gesture recognition	[114], [116]
	Pseudo labeling	Gesture recognition	[129]
	Data synthesization	Activity recognition	[128]
		Gesture recognition	[89], [90], [114], [124]
		Gait recognition	[90]
	Fine tuning	Tracking	[125]
		Activity recognition	[106], [127], [150]
		User authentication	[153]
		Fall detection	[119]
	Adversarial network	Activity recognition	[88], [102], [126]
		Gesture recognition	[109], [110], [117]
Model based	Propagation model	Localization	[23], [24], [83], [98], [137], [164], [165]
		Gesture recognition	[168]
		Respiration monitoring	[92]
	AoA or ToF	Localization	[160]–[163]
		Tracking	[32], [35]–[37], [166]
		Activity recognition	[151]
		Gesture recognition	[167]
	DFS	Tracking	[33], [35], [36]
	Fresnel zone model	Localization	[31], [170]
		Activity recognition	[42]
		Gesture recognition	[169]
		Gait recognition	[155]

A. Radio map updating

WiFi localization often leverages the radio map, i.e. fingerprint database, to localize the target, by comparing the real-time signal values with the ones stored in the radio map. Accurate localization requires a up-to-date radio map, however, the radio map varies with time and various environmental changes. To solve this problem, the radio map needs to be updated to reflect the current status, as shown in Fig. 13.

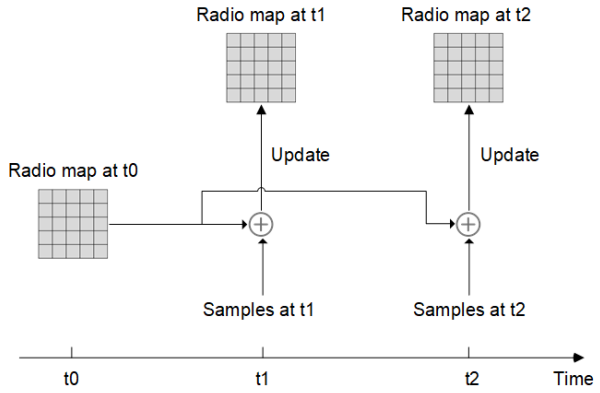


Fig. 13. Radio map updating.

Yin et al. [130], [131] adapted the radio map for indoor localization by applying a regression analysis to learn the temporal predictive relationship between the RSS values received by sparsely located reference points and that received by the mobile device. TrHMM [136] introduced a semi-supervised HMM to transfer the learned model from one time to another by applying regression analysis at time 0, rebuilding the radio map at time t and using EM on unlabeled traces at time t. AcMu [139] pinpointed mobile devices with trajectory matching and employed them as mobile reference points to collect real-time RSS values. With these real-time reference data, AcMu adapted the radio map by learning an underlying relationship of RSS dependency between different locations. AP-Sequence [143] divided the area into small regions and assigned an AP sequence to each region as a fingerprint and achieved low overhead fingerprint map maintenance over time

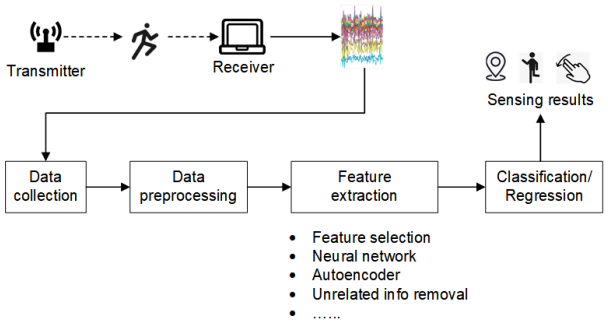


Fig. 14. Feature extraction.

and heterogeneous devices. LAAFU [158] updated the radio map with Gaussian process and implicit crowdsourced signals in the presence of altered APs. Using subset sampling, LAAFU identified any altered APs and filtered them out before a location decision. FitLoc [99] collected RSS changes of a few locations in the new area and integrated it with the radio map of the original area and transferred the radio map into a subspace where the distribution distances were minimized, by utilizing Fisher Linearly Discriminant Analysis (FLDA) and exploiting the Bregman Divergence as a regularization to measure the RSS distribution distances. CRISLoc [159] tackled the challenge of altered APs with a joint clustering and outlier detection method to find them, and used a transfer learning approach to reconstruct the CSI radio map on the basis of the outdated fingerprints and a few fresh measurements.

B. Feature extraction

Feature extraction is one of the most common methods to achieve cross domain WiFi sensing, which tries to extract the features robust to different domains, as shown in Fig. 14, via different ways, including feature selection, neural network, autoencoder, unrelated information removal, etc. The extracted features are then classified by a classifier or predicted by a regressor to infer the sensing results.

Localization & tracking. HED [140] combated AP sequence disorders caused by AP changes and RSS variance in localization, using the relative values of RSS and the order number of a given AP in a sorted fingerprint. AutoFi [84] automatically calibrated the localization profiles in an unsupervised manner, by online removing profile contaminants introduced by environment changes. It used an autoencoder to preserve critical features of fingerprints and reproduced them later in new localization profiles. WiDeep [157] combined stacked denoising autoencoders and a probabilistic framework to handle the noise in RSS due to heterogeneous devices and captured the complex relationship between the RSS heard by the mobile phone and its location. FiDo [124] was able to localize different users with labeled CSI data from only one or two users, leveraging a data augmenter that introduced data diversity using a Variational Autoencoder (VAE) and a domain-adaptive classifier that adjusted itself to newly collected unlabeled data using a joint classification-reconstruction structure. HAIL [147] leveraged both AP rank and RSS values to achieve high adaptability of indoor localization and devised a neural network to measure the similarities of RSS fingerprints. Kim et al. [148] used the maximum RSS combined with PDR to achieve smartphone tracking and overcame the RSS variance problem due to device type, device placement, user direction, and environmental changes over time.

Activity recognition. Chang et al. [100] transformed CSI into images to achieve activity recognition. They proposed a location-dependency removal method based on SVD to eliminate the background CSI and extract the channel information of signals reflected by human bodies. CARM [101] recognized activities based on a CSI-speed model that quantified the relation between CSI dynamics and human movement speeds and a CSI-activity model that quantified the relation between human movement speeds and human activities. CARM performed data fusion on multiple WiFi links to provide better robustness under different environments. FALAR [149] leveraged CSI to recognize activities regardless of locations and slight environmental changes. FALAR used a Kernel Density Estimation (KDE) based motion extraction method to eliminate slight environmental changes and reconstructed CSI data by CSVD to discard most location information. WiFit [103] was a bodyweight exercise monitoring system robust to environments and human. It counted the exercises based on the Doppler displacement calculated from the Doppler velocity spectrum. The exercise was classified with SVM using features extracted from CSI amplitude, CSI phase difference and Doppler velocity spectrum. WiSDAR [104] enabled spatial diversity aware activity recognition by extending the multiple antennas of WiFi devices to construct multiple separated antenna pairs and employed a deep learning framework that integrated the hidden features from both temporal and spatial dimensions. WiDrive [105] was a real-time in-car driver activity recognition system based on CSI DFS and Hidden Markov Models with Gaussian Mixture emissions Model (HMM-GMM). An online adaptation algorithm was used to update the parameters of the classifier to adapt for different drivers and vehicles. MatNet-eCSI [107] adopted Matching Network with enhanced CSI to perform one-shot learning to recognize human activities in a new environment and required only one training sample from the new environment. HARMN-EF [108] accomplished environment independent activity recognition by leveraging MatNet and the proposed CSI-CE method, which enhanced the activity dependent features whilst mitigated the behavior unrelated information.

Gesture recognition. Yang et al. [112] recognized gestures via a Siamese recurrent convolutional architecture based on CNN and RNN using CSI phase. The Siamese framework used transferable pairwise loss to remove structured noise such as individual heterogeneity and various measurement conditions. WiHand [152] leveraged the low rank and sparse decomposition and separated gesture signals from background information, thus making it resilient to location variation. Widar3.0 [111] achieved cross domain gesture recognition by extracting the domain independent feature BVP and developing a one-fits-all model that required only one-time training but could adapt to different data domains. Wang et al. [113] alleviated the efforts of retraining in a new scenario by learning

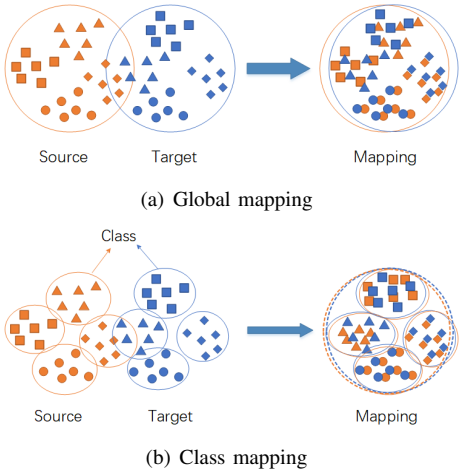
a universal similarity evaluation ability utilizing a deep similarity evaluation network. Ma et al. [115] presented a gesture recognition system based on similarity evaluation, which could recognize new type of gestures, or gestures performed in a new scenario or by a new user, using very few number of or even one sample of the new gesture. Niu et al. [118] developed a model to quantify the relationship between signal frequency and target position, motion direction and speed. Based on this model, they proposed movement fragments and relative motion direction changes as two position independent features and developed the gesture recognition system.

Miscellaneous. FallDeFi [91] detected falls based on CSI amplitude spectrogram and used a sequential forward selection algorithm to single out features that were resilient to environment changes. AR-Alarm [93] detected human intrusion by extracting a robust feature using the ratio between the dynamic and the static CSI profiles of the environment and achieved intrusion detection in different environments without calibration efforts. Regani et al. [94] authenticated the in-car driver by identifying the unique radio biometric information recorded in CSI. To deal with the in-car changing environments, they built a long-term driver radio biometric database and employed a machine learning algorithm. Shi et al. [95] presented user authentication by recognizing human activities, using CSI amplitude, relative amplitude and STFT holograms. They proposed a deep-learning based approach with an unsupervised domain discriminator to mitigate the impact of varying location and environmental changes. Di Domenico et al. [120] focused on crowd counting without retraining in new environments, by identifying a set of differential CSI feature candidates and selecting the most effective ones via minimization of the summation of the Davies-Bouldin indexes. They [121] presented another crowd counting system by analyzing the shape of the Doppler spectrum of the received WiFi signal which was correlated to the number of people.

C. Feature mapping

Feature mapping is the most common method to bridge the source and the target domains. Many methods have been proposed in the category of feature mapping to achieve domain alignment and then domain adaptation. The basic process of feature mapping is illustrated in Fig. 15. The features in the source domains and the target domains are mapped to a shared space, in which the divergence between them is minimized, or the features in the target domains are transformed to the source domains. After feature mapping, the source domain data and the target domain data follow the similar probability distribution. The sensing model built with the source data can then be adapted to the target domain. There are different ways of feature mapping. In Fig. 15(a), the features from the source and the target domains are globally mapped and aligned, while in Fig. 15(b), the features are mapped and aligned according to their labels. Class mapping achieves better adaptation effect than global mapping, in the cost of some labeled data, while global mapping does not require labeled data from the target domain.

RSS localization. The idea of feature mapping started with the works of RSS localization. LeManCoR [132] used Manifold co-Regularization to build a mapping function between the signal space and the location space and adapted the mapping function over different time periods. TransMapping [96] transferred the learning model trained for one area of a building to another, by learning a mapping function between the signal space and the location space based on manifold learning. A low-dimensional manifold was shared between the data collected in different areas in an environment as a bridge to propagate the knowledge across the whole environment. Pan et al. [133] found a latent space by Maximum Mean Discrepancy Embedding (MMDE) and projected the data in related domains onto this latent space. The distance between distributions of the data in different domains in this latent space was minimized. TCA [134] learned the transfer components across domains in a RKHS using MMD. With the new representations in this pace, machine learning methods could train classifiers or regression models in the source domain for use in the target domain. Zheng et al. [156] employed an alternating optimization approach to iteratively learn feature mappings, to map different devices' data to a well-defined low-dimensional feature space, and multi-task regression models for localization on multiple devices. LuMA [135] learned a mapping between the source domain and the target domain in a low-dimensional space, so that the knowledge could be transferred from the source domain to the target domain, to handle the problem caused by device and time. Wang et al. [97] built a localization model in a multi-floor building, collecting labeled data on one floor and unlabeled data on other floors. By co-embeding the different floors' data in a common low-dimensional manifold and aligning the unlabeled data with the labeled data, the labels could be propagated to the unlabeled data. KARMA [154] utilized a one-time fingerprint of the space and applied a set of causality calibration functions in real-time to compensate for the change in the factors at testing time. EIL [138] revised the RSS values affected by changed environments through a linear transformation and achieved time-independent localization. TLCS [122], [123] employed a transferring function to transfer the distorted RSS changes across different categories of targets into a latent



(a) Global mapping

(b) Class mapping

Fig. 15. Feature mapping.

feature space, where the distributions of the distorted RSS changes from different categories of targets were unified. TKL-WinSMS [142] designed a localization model using transfer kernel learning, which learned a domain invariant kernel by matching the source and target distributions in RKHS. MTLF [141] encoded metric learning in transfer learning, in which instance weights were learned and exploited to bridge the distributions of different domains, while Mahalanobis distance was learned to maximize the intra-class distances and minimize the inter-class distances. Liu et al. [85] reshaped data distributions of RSS fingerprints in the target domain based on the transformation matrix from the source domain to achieve low overhead localization robust to environmental dynamics and different environment.

CSI localization. MSDFL [144] converted the fingerprints of the changed environments to the original environment by constructing a conversion model based on Multi-layer Perceptron (MLP). Rao et al. [145] utilized transfer deep learning to learn new feature representations from the CSI amplitude samples as fingerprints, which could simultaneously minimize the intra-class differences, maximize inter-class differences, and minimize the distribution differences between fingerprint database and testing samples. Rao et al. [146] also used CSI phase for device-free localization. In view of the unpredictable nature of calibrated CSI phase over time, they adopted a transfer deep supervised neural network (TDSNN) method combining deep neural networks and transfer learning to solve it. MoLoc [86] designed a CSI fingerprint roaming model to automatically learn the predictive fingerprint variation pattern and transfer the online fingerprint measurement to adapt to dynamic ship motions for localization in the mobile ship environment. AdapLoc [87] adapted the original localization model to the changed environment by mapping the source domain and target domain into a shared space and minimizing the distances of the fingerprints from the same location and maximizing the distance of the fingerprints from different locations.

Gesture recognition. MobileDA [116] allowed a teacher network trained in the server to distill the knowledge for a student network running on the edge device by cross domain distillation, thus to effectively train a simple student model that adapted to the new environment by diminishing domain disparity. DANGR [114] recognized gestures based on a deep adaptation network using CSI. In order to shrink the domain discrepancies in different environments, DANGR adopted domain adaptation based on multi-kernel MMD scheme, which matched the mean-embeddings of abstract representations across domains in RKHS.

D. Pseudo labeling

Pseudo labeling is a common method in semi-supervised learning, which annotates the unlabeled samples with pseudo labels according to the labeled samples, as illustrated in Fig. 16. WiSign [129] recognized sign language with sparsely labeled training data with semi-supervised learning framework, which labeled the unlabeled samples using the knowledge from labeled samples.

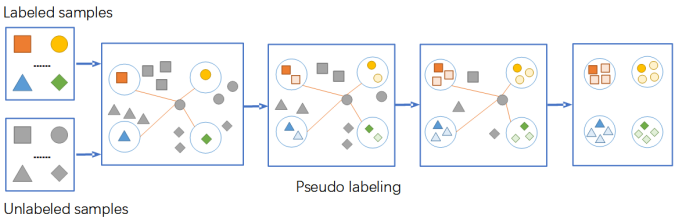


Fig. 16. Pseudo labeling.

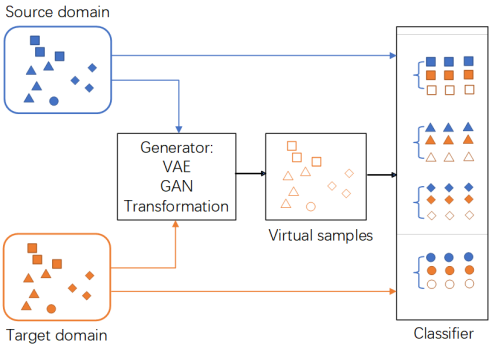


Fig. 17. Data synthesization.

E. Data synthesization

Learning-based methods need a large number of data to train the models. To facilitate this need, some methods proposed to synthesize virtual samples for the target domains. The process of data synthesization can be illustrated in Fig. 17.

FiDo [124] was able to localize different users with labeled data from only one or two users, leveraging a data augmenter that introduced data diversity using a Variational Autoencoder (VAE) and a domain adaptive classifier that adjusted itself to newly collected unlabeled data using a joint classification-reconstruction structure. Zhang et al. [128] proposed an activity recognition system that synthesized variant activity data through CSI transformation methods to mitigate the impact of activity inconsistency and subject-specific issues. WiAG [89] recognized the gestures irrespective of user position and orientation. It required the training samples of all the gestures in one domain and automatically generated virtual samples for all the gestures in other domains through a translation function. CrossSense [90] enabled training measurements to be collected once and used across sites, by employing an Artificial Neural Network (ANN) to train a roaming model that generated synthetic training samples of gestures or gaits for each target environment from the original. DANGR [114] recognized gestures based on a deep adaptation network using CSI. DANGR first exploited GAN to augment the dataset, then adopted domain adaptation based on multi-kernel MMD scheme to shrink the domain discrepancies in different environments, which matched the mean-embeddings of abstract representations across domains in a reproducing kernel Hilbert space.

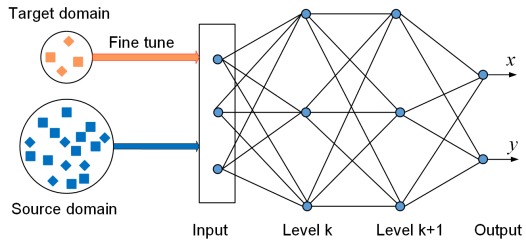


Fig. 18. Fine tuning.

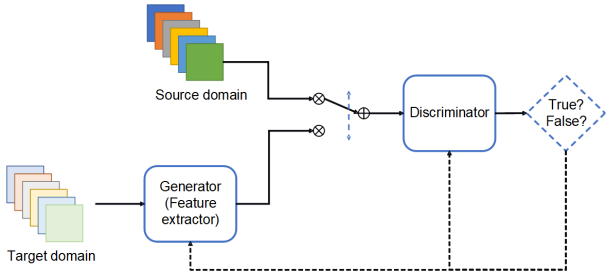


Fig. 19. Adversarial network.

F. Fine tuning

Fine tuning is a simple yet effective method, which retrains the pre-trained model with a small number of samples from the target domain, as illustrated in Fig. 18.

FreeTrack [125] realized device-free tracking based on CSI through DNN and particle filtering. To combat with human diversity, fine tuning on DNN was employed to enhance the adaptivity of the method. TL-HAR [127] recognized multiple human activities by employing packet-level classification and image transformation together with transfer learning. They reduced training complexity through transfer learning that inferred knowledge from a pre-trained model. Sheng et al. [106] achieved cross-scene action recognition by integrating spatial features learned from the CNN into the BLSTM model. The off-the-shelf model was used as the pre-trained model and fine-tuned in the new scenario. WiLISensing [150] leveraged CNN and fine tuning to recognize activities in a position without training or with very few training samples. Jung et al. [153] verified identity based on the gesture signals captured by CSI. A CNN-based feature extractor was first pre-trained using the CSI from one position, and transferred to recognize signals from another position via a rapid retraining process, utilizing the kernel and the range (KAR) space projection learning. TL-Fall [119] was fall detection system, which was trained with labeled data to derive knowledge in the source environment. With the derived knowledge, the fall detection model working in the target environment could be trained with only a few labeled data.

G. Adversarial network

In recent years there is an increasing trend of using adversarial networks to achieve across domain, as shown in Fig. 19. The key of an adversarial network consists of a generator and a discriminator. The generator generates fake samples which are as similar as the real samples, the discriminator tries to distinguish the real and the fake samples. The adversarial network reaches balance between the generation and the discrimination. For cross domain WiFi sensing, we can use the target domain data as the fake samples. After adversarial training, the cross domain features can be extracted.

EI [102] achieved environment independent activity recognition by exploiting adversarial networks. EI was composed of a CNN feature extractor, an activity recognizer, a domain discriminator, and several constraints. By exploiting adversarial networks, EI could remove environment and subject specific information and learn transferable features of activities. CsiGAN [126] enabled activity recognition adaptive to user changes based on semi-supervised Generative Adversarial Networks (GAN) to meet the scenarios that unlabeled data from left out users were very limited. CsiGAN introduced a complement generator, which used limited unlabeled data to produce diverse fake samples to train a robust discriminator. WiCAR [88] was an in-car activity recognition framework that was able to remove domain specific information while retaining the activity related information to the maximum extent. WiCAR used a deep learning architecture integrated with multi-adversarial domain adaptation to adapt to the new domains, including driving conditions, car models and human subjects. WiADG [110] and JADA [109] exploited unsupervised joint adversarial domain adaptation to realize gesture recognition without collecting and labeling training data in new environments. They trained a target encoder and fine-tuned the source encoder through adversarial learning to map both unlabeled target data and labeled source data to a domain-invariant feature space such that a domain discriminator could not distinguish the domain labels of the data. Kang et al. [117] proposed gesture recognition based on DFS through multi-source unsupervised domain adaptation, considering environment, gesture performer, location and orientation. They applied an adversarial learning scheme together with feature disentanglement and an attention scheme to achieve the goal.

H. Propagation model

According to wireless communication principles, the power fading [23] between the two transceivers is mainly related to the propagation fading, diffraction fading and target absorption fading, as illustrated in Fig. 20. Many works used WiFi propagation models to realize sensing, to avoid the labor-intensive training and make sensing robust in the face of domain heterogeneity.

WiGEM [137] modeled RSS as a Gaussian Mixture Model (GMM) and used Expectation Maximization (EM) to learn the maximum likelihood estimates of the model parameters, thus localized a device based on the maximum a posteriori estimate. Ohara et al. [98] transferred a RSS model obtained in the source environment to the end user environment. With the transferred models, they could construct a positioning model for the end user environment inexpensively. LiFS [23] estimated the target locations by modeling the CSI measurements of multiple wireless links as a set of power fading based

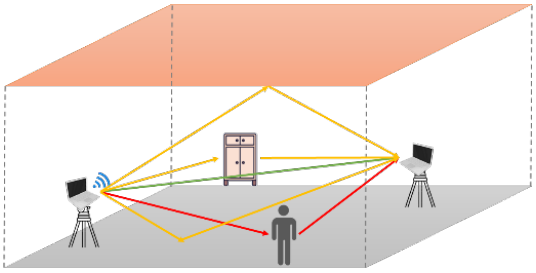


Fig. 20. Propagation model.

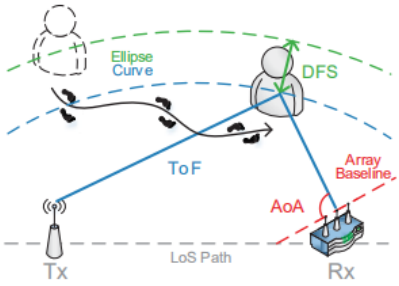


Fig. 21. AoA, ToF and DFS.

equations using only the subcarriers affected by multipath. CRIL [83] coupled RSS and Inertial Navigation System (INS) for localization. It employed a Kalman filter to fuse the localization results from the RSS and the INS system and updated the parameters of the RSS-based channel model utilizing the fused results. S-Phaser [24] utilized calibrated CSI phase to compute the direct path length between an AP and terminals and used a geometric positioning algorithm to determine the target location. Wan et al. [164] and Tong et al. [165] presented calibration-free CSI phase fingerprinting localization via hybrid inter-antenna distance and context-aware fingerprinting recognition. The fingerprinting database was constructed theoretically without site survey based on the signal propagation model and geometric methods.

AirDraw [168] used CSI phase to realize learning-free in-air handwriting by passive gesture tracking. They denoised CSI measurements by the ratio between two close-by antennas, and separated the reflected signal from noise by performing PCA, and proposed a signal calibration algorithm for tracking correction by eliminating the static components unrelated to hand motion. RespiRadio [92] detected a person's respiration rate in dynamic ambient environments by synthesizing a wider-bandwidth WiFi radio. With the synthesized WiFi radio, RespiRadio could identify the path reflected by the breathing person and analyzed the periodicity of the signal power measurements only from this path to infer the respiration rate.

I. AoA ToF DFS

Many works made use of AoA and/or ToF as well as DFS to realize human localization and tracking.

ArrayTrack [160] calculated the target location by computing the AoA information from the frames transmitted by the target and overheard by multiple APs. SpotFi [161] combined CSI values across subcarriers to jointly estimate AoA and ToF of each path, identified AoA of direct path between the target and AP, and used the direct path AoA estimate as well as RSS information to calculate the target location. Chronos [162] computed the sub-nanosecond ToF and then computed the distance between each antenna of the AP and the target, hence localize the target. SiFi [163] used a super-resolution algorithm to estimate the ToA spread, applied a clustering algorithm to identify the direct path and removed the symbol time offset at the same time, and estimated the location by using a weighted iterative least square model to combat sampling frequency offset. MaTrack [32] identified the target's angle and tracked the walking human by detecting the subtle reflection signals from the human body and further differentiating them from those reflected signals from static objects. IndoTrack [36] proposed Doppler-MUSIC to extract accurate Doppler velocity information from CSI and Doppler-AoA to determine the absolute trajectory of the walking human by jointly estimating velocity and location via probabilistic co-modeling of spatial-temporal Doppler and AoA information. Widar [33] simultaneously estimated human moving velocities and locations by modeling the relationships between CSI dynamics and locations and velocities exploiting Doppler shifts. Widar2.0 [35] enabled single-link human tracking by leveraging AoA, ToF, Doppler shifts and attenuation together. Karanam et al. [37] estimated the AoA of signal paths arriving at a receiver array using only the received signal magnitude, to track a moving target. DeFi [166] estimated the target location by refining the AoA of the target reflection path based on CSI. They eliminated the static paths using a background elimination algorithm, and separated out the target reflection path from the motion paths according to the AoA and equivalent ToF measurements.

Li et al. [151] proposed a cross location activity recognition system leveraging Angle Difference of Arriva (ADOA) calculated from CSI using BLSTM. WiDeo [167] mined the backscatter reflections from the environment that WiFi transmissions naturally produced, calculated each reflection's amplitude, ToF and AoA, to trace where reflecting objects were located and how they were moving.

J. Fresnel zone model

In indoor environment, the Fresnel zones exist for WiFi radio propagation. Fresnel zones refer to the concentric ellipses with foci in a pair of transceivers, as shown in Fig. 22. Assume P_1 and P_2 are two transceivers, the Fresnel zones containing n ellipses can be constructed by [58]

$$|P_1Q_n| + |Q_nP_2| - |P_1P_2| = n\lambda/2 \tag{18}$$

where Q_n is a point in the n-th ellipse. The innermost ellipse is defined as the 1st Fresnel zone and the n-th Fresnel zone corresponds to the elliptical annuli between the $(n-1)$-th and n-th ellipses.

MFDL [170] and Fei et al. [31] leveraged the Fresnel zone model to achieve device-free localization that could determine which elliptic region the target located in according to the CSI phase. Zhang et al. [42] leveraged the Fresnel

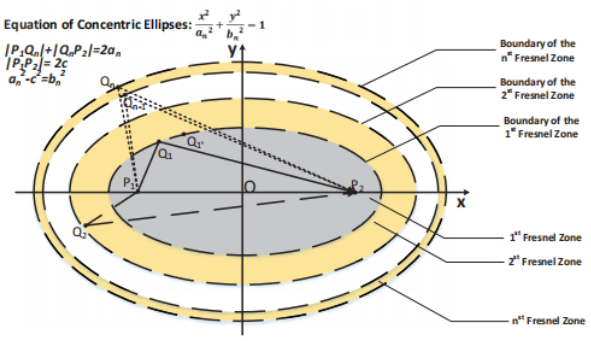


Fig. 22. Fresnel zone model.

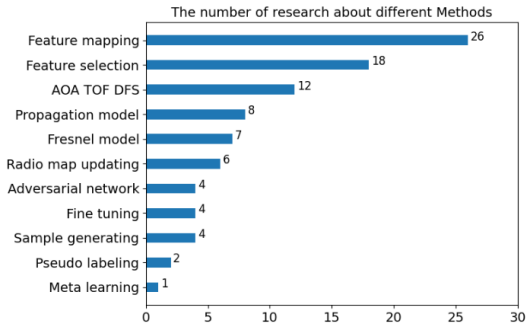


Fig. 23. Number of papers by the methods.

diffraction model and proposed a diffraction-based sensing model to quantitatively determine the signal propagation wrt. a target's motions in the first Fresnel zone. The model linked the received signal variation patterns with motions, hence could be used to detect human activities. FingerDraw [169] tracked finger drawings by leveraging the CSI-quotient model and the Fresnel model, which used the channel quotient between two antennas of the receiver to cancel out the noise in CSI amplitude and the random offsets in CSI phase, and quantified the correlation between CSI value dynamics and object displacement. WiDIGR [155] achieved direction-independent gait recognition based on the Fresnel model, using a series of signal processing techniques to eliminate the differences among induced signals caused by walking in different directions and generated a high-quality direction-independent signal spectrogram.

K. Remarks

The majority of the cross domain sensing methods are learning-based. In our survey xx works are based on learning and xx works are based on model. In the category of learning-based methods, feature mapping is the most commonly used method, having xx works, next is feature extraction, having xx works. In the category of model-based works, AoA-ToF-DFS xx works and propagation has xx works, which are the most commly used.

Fig. 24 shows the evolution of cross domain WiFi sensing over the years. Along the time, different methods are proposed to cope with the cross domain issue.

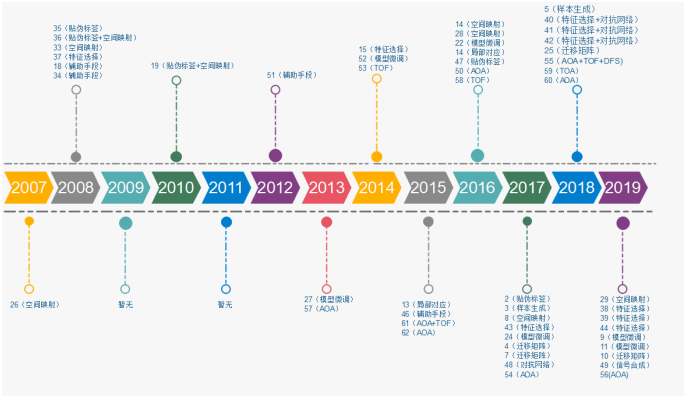


Fig. 24. Evolution of cross domain WiFi sensing.

VI. By the Applications

We also categorize the cross domain WiFi sensing works by their applications.

A. Localization

Localization is probably the mostly researched topic in WiFi sensing. Researchers have been trying to solve the problem of localization across different domains.

RSS. Most of the early works used RSS fingerprints for localization, including Yin et al. [130], [131], LeMan-CoR [132], TransMapping [96], Pan et al. [133], [134], TrHMM [136], Zheng et al. [156], LuMA [135], Wang et al. [97], KARMA [154], EIL [138], Wang et al. [122], [123], AcMu [139], HED [140], LAAFU [158], FitLoc [99], TKL-WinSMS [142], Xu et al. [141], Liu et al. [85], WiDeep [157]. AP-Sequence [143] used AP sequence derived from relative RSS as the fingerprint for localization. HAIL [147] leveraged both AP rank and RSS values to achieve localization. Some works used RSS propagation model for localization. WiGEM [137] used RSS propagation model for indoor localization. Ohara et al. [98] used RSS variance model for indoor localization. CRIL [83] used RSS channel model and INS for indoor localization. ArrayTrack [160] utilized RSS AoA information calculated by multiple APs using the frames sent out by the mobile device to infer target locations.

CSI. In recent years, the works using CSI for localization increased dramatically, mostly using CSI amplitude. AutoFi [84], MSDFL [144], Rao et al. [145], FiDo [124], Zhou et al. [87] used CSI amplitude fingerprint to realize localization. CRISLoc [159] used CSI amplitude fingerprint for smartphone-based localization. MoLoc [86] designed a CSI fingerprint roaming model for localization in the mobile ship environment. Some works used CSI phase to realize localization. Rao et al. [146] used CSI phase fingerprint for localization. Wan et al. [164] and Tong et al. [165] presented CSI phase fingerprinting localization, in which the fingerprint database was constructed theoretically without site survey based on the signal propagation model and the geometric method. MoLoc [86] designed a CSI fingerprint roaming model for localization in the mobile ship environment. MFDL [170] and Fei et al. [31] used the Fresnel

TABLE V
THE APPLICATIONS AND THE RELATED WORKS.

Application	Measurement	Related work
Localization	RSS fingerprint	[85], [96], [97], [99], [122], [123], [130]–[136], [138]–[143], [147], [154], [156]–[158]
	RSS propagation	[83], [98], [137]
	RSS AoA	[160]
	CSI amplitude	[84], [86], [87], [124], [144], [145], [159]
	CSI phase	[31], [86], [146], [164], [165], [170]
	CSI propagation	[23], [24]
	CSI AoA or ToF	[161]–[163]
Tracking	RSS fingerprint	[148]
	CSI amplitude	[125]
	CSI AoA or ToF	[32], [35]–[37], [166]
	CSI DFS	[33], [36]
Activity recognition	CSI amplitude	[42], [88], [100]–[102], [104], [106]–[108], [126]–[128], [149], [150]
	CSI phase	[101], [107], [127]
	CSI AoA or ToF	[151]
	CSI DFS	[103], [105]
Gesture recognition	CSI amplitude	[89], [90], [113]–[115], [129], [152], [169]
	CSI phase	[109], [110], [112], [114], [116], [168], [169]
	CSI AoA or ToF	[167]
	CSI DFS	[111], [117], [118]
Fall detection	CSI amplitude	[91], [119]
	CSI DFS	[91]
Respiration monitoring	CSI amplitude	[92]
Intrusion detection	CSI phase	[93]
Gait recognition	CSI amplitude	[90]
	CSI DFS	[155]
User authentication	CSI amplitude	[95], [153]
	CSI phase	[94]
People counting	CSI amplitude	[120]
	DFS	[121]

zone model to achieve localization, which determined the elliptic region that the target located in according to the CSI phase. LiFS [23] estimated the target locations by modeling the CSI measurements of multiple wireless links as a set of power fading based equations using only the subcarriers affected by multipath. S-Phaser [24] utilized calibrated CSI phase to compute the direct path length between an AP and terminals and used a geometric algorithm to determine the target location. SpotFi [161] calculated ToF and AoA from CSI and estimated the target location by using the direct path AoA estimates and RSS measurements from all the APs. Chronos [162] used ToF to realize localization. SiFi [163] achieved localization based on ToA from CSI.

1) RSS fingerprinting: Yin et al. [130], [131] adapted the radio map for localization along the time dimension by offsetting the environmental dynamics using regression analysis. They applied model trees to learn the predictive relationship between the RSS received by the reference points (RP) and that received by the mobile device. This prediction model was then used for localization based on the online RSS from the mobile device and the RPs. The limitation was that they needed to deploy several RPs additionally. The methods achieved the average accuracy over different times of about 40% 0.5m, 76% 1.5m, 90% 3m with 9 APs and 8 RPs.

Pan et al. proposed different methods for cross domain RSS localization. LeManCoR [132] adapted the mapping function between the RSS and the location from one time period to another based on Manifold Co-Regularization. The limitation was that they needed to deploy several RPs additionally and required a small number of new data. The method achieved the average accuracy over different times of about 40% 1.5m, 75% 3m with 10 RPs. TransMapping [96] transferred the learning model trained for one area of a building to another. The method learned a mapping function between the signal space and the location space by solving an optimization problem based on manifold learning. A low-dimensional manifold was shared between the data collected in different areas as a bridge to propagate the knowledge across the whole environment. The limitation was that they required a small number of labeled data in the target area and the areas are in the same WiFi environment. The average accuracy in the target area was 3.45m, about 50% 1.5m and 70% 3m. Maximum Mean Discrepancy Embedding (MMDE) [133] aimed to find a low-dimensional latent space, in which the distance between data distributions in different domains was minimized. They projected the data in related domains onto this latent space, where they could apply standard learning algorithms to train classification or regression models. The experiments on RSS localization in two different days achieved approximately 55% 1.5m and 85% 3m in the target domain. Transfer Component Analysis (TCA) [134] learned the transfer components across domains in a reproducing kernel Hilbert space using Maximum Mean Discrepancy (MMD). With the new representations in this subspace, standard machine learning algorithms could train classifiers or regression models in the source domain for use in the target domain. RSS localization in two different times achieved an accuracy of around 2.5m in the target domain.

Zheng et al. proposed a transferred Hidden Markov Model (TrHMM) [136] to transfer the learned HMM model from one time period to another. They first applied regression analysis at time 0 to learn the temporal predictive correlations between the RSS; then rebuilding the radio map at time t; then using EM on the unlabeled traces at time t to improve the HMM model. Experiments using one time data training and two times data testing achieved the correct rate of more than 80% in the target domains. To solve the multi-device localization problem, Zheng et al. proposed a latent multi-task learning algorithm [156]. They employed an alternating optimization approach to iteratively learn feature mappings, to map different devices' data to a well-defined low-dimensional feature space, and multi-task regression models for the devices. Experiments using one device training and another device testing achieved the accuracy of about 5.3m for the target device with a small number of labeled data from the target device.

LuMA [135] was to handle the transfer learning problem

TABLE VI
CROSS DOMAIN LOCALIZATION.

Work	Year	Type	Method	Measurement	Domain	Evaluation	Performance
Yin et al. [130], [131]	2005, 2008	Device based	Radio map updating: model tree	RSS fingerprint	Time	9 AP, 8 RP, 6 times	CDF: ≈40% 0.5m, ≈76% 1.5m, ≈90% 3m
LeManCoR [132]	2007	Device based	Feature mapping: manifold co-regularization	RSS fingerprint	Time	10 RP, 6 times	CDF: ≈40% 1.5m, ≈75% 3m
TransMapping [96]	2008	Device based	Feature mapping: manifold learning	RSS fingerprint	Cross environment	2 areas in same building	Accuracy: 3.45m, CDF: ≈50% 1.5m, ≈70% 3m
MMDE [133]	2008	Device based	Feature mapping: dimensionality reduction	RSS fingerprint	Time	2 days	CDF: ≈55% 1.5m, ≈85% 3m
TrHMM [136]	2008	Device based	Radio map updating: HMM	RSS fingerprint	Time	3 times	Correct rate>80%
Zheng et al. [156]	2008	Device based	Feature mapping: alternating optimization	RSS fingerprint	Target device	2 devices	Accuracy: ≈5.3m
LuMA [135]	2008	Device based	Feature mapping: manifold alignment	RSS fingerprint	Time, target device	2 devices, 2 times	CDF for time: ≈65% 3m, CDF for device: ≈60% 3m
Wang et al. [97]	2010	Device based	Feature mapping: manifold alignment	RSS fingerprint	Cross environment	2 out of 3 floors	CDF: ≈40% 2m, ≈55% 3m, ≈60% 4m, ≈70% 5m
TCA [134]	2011	Device based	Feature mapping: dimensionality reduction	RSS fingerprint	Time	2 times	Accuracy: ≈2.5m
WiGEM [137]	2011	Device based	Propagation model: GMM	RSS propagation	Time, target device	2 testbeds, 4 devices	Accuracy: 7-10m/4-5m, CDF: ≈20%/≈35% 3m, ≈25%/≈70% 5m
ArrayTrack [160]	2013	Device based	AoA	RSS AoA	Agnostic	1 testbed	Static: 3AP 317cm, 6AP 38cm; Moving: 3AP 57cm, 6AP 31cm
KARMA [154]	2014	Device based	Feature mapping	RSS fingerprint	Orientation, target device	2 phones, 2 user orientations	Accuracy: 2x improvement
EIL [138]	2014	Device free	Feature mapping: linear transformation	RSS fingerprint	Time	3 times	CDF: 90% 0.5 0.6m
TLCS [122], [123]	2014, 2015	Device free	Feature mapping: linear transformation	RSS fingerprint	Human	4 target categories	CDF: 50% 0.5m, 80% 0.7m
Ohara et al. [98]	2015	Device free	Propagation model	RSS propagation	Cross environment	3 source environments, 1 target environment	Accuracy: 1.84m
AcMu [139]	2015	Device based	Radio map updating	RSS fingerprint	Time	20 days over 6 months	Accuracy: 2x improvement
SpotFi [161]	2015	Device based	AoA	CSI AoA, CSI ToF, RSS	Agnostic	A floor, 6 AP	Accuracy: 0.4m, CDF: 80% 1.8m
Chronos [162]	2016	Device based	ToF	CSI ToF	Agnostic	1 AP	LOS 65cm, NLOS 98cm
HED [140]	2016	Device based	Feature extraction	RSS fingerprint	Time, AP	Over 6 months	Accuracy: 1.88m/1.91m over days, 2.03m/2.19m over months
LiFS [23]	2016	Device free	Propagation model	CSI propagation	Agnostic	3 environments, 4TX-7RX	Accuracy: 0.7m, CDF: 80% 1.2m
LAAFU [158]	2017	Device based	Radio map updating: GPR	RSS fingerprint	AP	AP power adjustment	Accuracy: 20% improvement, CDF 35% 3m, 70% 6m
FitLoc [99]	2017	Device free	Feature mapping: subspace learning	RSS fingerprint	Cross environment	2 users, outdoors, similar furnitured indoors	Accuracy: 0.89m outdoors, 1.4m indoors
AutoFi [84]	2017	Device free	Feature extraction: contaminant removal, autoencoder	CSI amplitude fingerprint	Environmental dynamics	Open windows, open doors	Correct rate: windows 84.9%, doors 90.2%
TKL-WinSMS [142]	2017	Device based	Feature mapping: transfer kernel learning	RSS fingerprint	Time	6 months	Accuracy: 1.86m, CDF: 30% 1m, 60% 2m, 80% 3m
MTLF [141]	2017	Device based	Feature mapping: metric transfer learning	RSS fingerprint	Time	2 times	Accuracy: ≈5m
CRIL [83]	2017	Device based	Propagation model: with INS	RSS propagation	Environmental dynamics	Changes of channel states	Simulation accuracy ≤1m; Experiment accuracy ≤3m
MFDL [170]	2017	Device free	Fresnel zone model	Fresnel zone model	Agnostic	1TX-2RX	Outdoor accuracy: 45cm, indoor accuracy: 50-75cm
Liu et al. [85]	2018	Device based	Feature mapping: matrix transformation	RSS fingerprint	Environmental dynamics, cross environment	3 sources, 1 target	CDF: 80% 1m, 90% 2m, 95% 3m
AP-Sequence [143]	2018	Device based	Radio map updating	RSS fingerprint: AP sequence	Time, target device	2 sites, 1 month, 3 smartphones	Accuracy: 4 7.6m
SiFi [163]	2018	Device based	ToA	CSI ToA	Agnostic	Different environments, 1 AP	Accuracy: 0.93m in offices, 1.29m in corridors, 1.83m in NOLS
S-Phaser [24]	2019	Device based	Propagation model	CSI propagation	Agnostic	1 AP	Accuracy: 1.5m
WiDeep [157]	2019	Device based	Feature extraction	RSS fingerprint	Target device	2 testbeds, 5 users with different phones	Accuracy: 2.64m, 1.21m
Wan et al. [164], Tong et al. [165]	2019, 2020	Device based	Propagation model	CSI phase fingerprint	Agnostic	2 testbeds, 4/5 APs	CDF: 80% 0.3m
MSDFL [144]	2019	Device free	Feature mapping: MLP transformation	CSI amplitude fingerprint	Time	5 days	Accuracy: 1.96m, CDF: 60% 2m
Rao et al. [145]	2020	Device free	Feature mapping: deep transfer learning	CSI amplitude fingerprint	Time	5 days	Accuracy: 1.4 1.5m, CDF: 60% 1.5m, 100% 2.5m
Rao et al. [146]	2020	Device free	Feature mapping: deep transfer learning	CSI phase fingerprint	Time	6 days	Accuracy: ≈1.3m, CDF: 50% 1.5m, 100% 2.5m
FiDo [124]	2020	Device free	Sample generating, feature extraction	CSI amplitude fingerprint	Human	9 users, 2 with labels	Recall 84.6%, precision 85%
CRISLoc [159]	2020	Device based	Radio map updating: subspace learning	CSI amplitude fingerprint	AP	1 or 2 AP alteration	Accuracy: ≈0.29m, 1 altered AP: +5.4cm, 2 altered APs: +8.6cm
HAIL [147]	2020	Device based	Feature extraction	RSS fingerprint	Target device, time	3 devices, 10 days	Accuracy: 1.5m/2.0m
MoLoc [86]	2020	Device free	Feature mapping: roaming	CSI amplitude phase fingerprint	Environmental dynamics	4 rooms in 2 ships, 4 users, 5 times	Accuracy: 0.68m, Correct rate: 92.8%
Fei et al. [31]	2020	Device free	Fresnel zone model	Fresnel zone model	Agnostic	2 rooms, 5 users	Accuracy: 0.336m, 0.312m
AdapLoc [87]	2021	Device free	Feature mapping: semi domain adaptation	CSI amplitude fingerprint	Environmental dynamics, AP	4 rooms	Smaller room: ≈1.5m, larger room: ≈ 2m

in localization caused by target device and time. LuMA was a dimensionality reduction method based on manifold alignment, which learned a mapping between a source data set and a target data set in a low-dimensional space. With 10% labeled data from the target domain, LuMa achieved about 65% within 3m for a different time and about 60% within 3m for a different device.

Wang et al. [97] established a localization model in a multi-floor building, collecting labeled data on one floor and unlabeled data on other floors. They co-embeded these different floors' data in a common low-dimensional manifold and aligned the unlabeled data with the labeled data so that they could propagate the labels to the unlabeled data. Localization experiments of one source and one target out of 3 floors achieved the accuracy of about 40% 2m, 55% 3m, 60% 4m, 70% 5m.

KARMA [154] was an online dynamic calibration methodology, which utilized a one-time fingerprint of the space and applied a set of causality calibration functions in real-time to compensate for the changes at testing time. Experimental studies demonstrated that KARMA could improve the localization quality by a factor of 2x to fingerprinting without considering changes.

EIL [138] revised the RSS values affected by changed environments through a linear transformation by eliminating the interference of environment on RSS over time and used only object caused RSS for localization. Localization after training for a short time, 10 days and one month kept a range of 0.5m to 0.6m localization errors for 90% locations.

Wang et al. [122], [123] proposed a transferring compressive sensing based device free localization approach TLCS, which employed linear transformation to transfer the distorted RSS changes across different categories of targets into a latent feature space, where the distributions of the distorted RSS changes from different categories of targets were unified. Experiments with 4 categories of targets achieved the CDF of 50% 0.5m and 80% 0.7m.

AcMu [139] provided automatic radio map self-updating for localization over time. By pinpointing mobile devices with trajectory matching, they employed them as mobile reference points to collect real-time RSS when they were static. With these fresh reference data, they adapted the radio map by learning an underlying relationship of RSS dependency between different locations. The method leveraged inertial sensors in the mobile devices. Experiments for 20 days across 6 months showed 2× improvement on localization accuracy by maintaining an up-to-date radio map.

HED [140] aimed to combat AP sequence disorders caused by coming-and-going of APs and RSS variance with time. HED used the relative values of RSS and the order number of a given AP in a sorted fingerprint to capture the inherent geographical relationship among different APs. Experiments achieved 1.88m and 1.91m over a few days, 2.03m and 2.19m over a few months.

LAAFU [158] achieved localization and automatic radio map update in the presence of altered APs without extra site survey, by employing implicit crowdsourced signals. Using subset sampling, LAAFU identified altered APs and filtered them out before a location decision was made. The radio map was updated with Gaussian Process Regression (GP) and crowdsourcing. In the expriments, LAAFU achieved 20% accuracy improvement compared with the traditional schemes.

FitLoc [99] localized multiple targets over various areas based on RSS, especially in outdoor environments and similar furnitured indoor environments. By collecting RSS of only a few locations in the new area and integrating it with the radio map of the original area, FitLoc transferred the radio map into a subspace, where the distribution distances were minimized, utilizing Fisher Linearly Discriminant Analysis (FLDA) and Bregman Divergence. Experiments achieved the average accuracy of 0.89 outdoors and 1.4 indoors for two targets.

TKL-WinSMS [142] realized adaptive localization in dynamic environments over time. It used APs as online reference points and extracted real-time RSS among them. With these online RSS and the RSS from target mobile devices as unlabeled target data, TKL-WinSMS used transfer kernel learning to learn a domain invariant kernel by matching the source and target distributions in the reproducing kernel Hilbert space. The resultant kernel could be used as input for the SVR training. The experiments over 6 months achieved the accuracy of 1.86m, and 30% 1m, 60% 2m, 80% 3m.

MTLF [141] was a metric transfer learning framework to encode metric learning into transfer learning. In MTLF, instance weights were learned and exploited to bridge the distributions of different domains, while Mahalanobis distance was learned simultaneously to maximize the intra-class distances and minimize the inter-class distances for the target domain. Experiments over different times achieved the accuracy of about 5m.

To achieve low overhead localization robust to environmental dynamics and different environments, Liu et al. [85] proposed to reshape the distributions of RSS fingerprints in the target domain based on the transformation matrix from the source domain, so that the data belonging to the same cluster would be logically closer to each other, whereas others would be further apart from each other. Experiments with 3 source domains achieved CDF of 80% 1m, 90% 2m and 93% 3m for the target domain.

AP-Sequence [143] divided the area into small regions and assigned an AP sequence to each region as a fingerprint. To achieve low overhead in fingerprint map maintenance over time and heterogenous device, AP-Sequence leveraged dynamic region partitioning and generated fingerprints based on relative RSS differences among various APs. Experiments in 2 sites over one month with 3 different smartphones achieved the average accuracy of 4-7.6m.

WiDeep [157] was a device-based RSS fingerprinting localization system that achieved a robust accuracy in the presence of noise and device heterogenity. WiDeep combined a stacked denoising autoencoders deep learning model and a probabilistic framework to handle the noise and captured the relationship between RSS heard by the mobile phone and its location by building a deep neural network for each reference point. The results of 5 participants using different phones over different days achieved a mean accuracy of 2.64m and 1.21m

for the larger and the smaller testbeds.

HAIL [147] was an adaptive indoor localization approach, which leveraged both AP rank and RSS values to achieve robustness to device heterogeneity and RSS variance over time. A back-propagation neural network (BPNN) was devised to measure the fingerprints similarities based on RSS values. Experiments with 3 different devices over 10 days achieved the accuracy of about 1.5m in a blockage environment and 2.0m in a blockage free environment.

2) CSI fingerprinting: AutoFi [84] automatically calibrated the localization profiles in an unsupervised manner, to address the issue of environment changes. AutoFi online removed profile contaminants introduced by environment changes by analyzing the variations in CSI profiles at a reference location. It applied an autoencoder to preserve critical features of fingerprints robust to the varying environments. The localization accuracy was 84.9% for opening the windows and 90.2% for opening the doors.

Rao et al. proposed MSDFL [144], which was was a CSI fingerprinting system based on matrix similarity. It converted the CSI amplitude fingerprints of the changed environments over time to the original environment by constructing a conversion model based on Multi-layer Perceptron (MLP). The conversion model was trained using the no-target fingerprints. Experiments achieved the accuracy of 1.96m and CDF of 60% within 2.0m. To cope with localization accuracy reduction caused by time, Rao et al. [145] further utilized transfer deep learning to learn new feature representations from the CSI amplitude fingerprints, which simultaneously minimized intra-class differences, maximized inter-class differences, and minimized distribution differences. The experiments in 5 days achieved 1.4–1.5m accuracy, CDF of 60% 1.5m and approximately 100% 2.5m. Rao et al. also proposed DFPhaseFL [146], which purely used CSI phase for device free fingerprinting localization. In view of the unpredictable nature of calibrated CSI phase over time, they combined deep neural networks and transfer learning to solve it. The experiments in 6 days achieved the mean localization error of about 1.3m, about 50% having an error within 1.5m, approximately 100% within 2.5m.

FiDo [124] was a WiFi-based domain-adaptive system able to localize different users with labelled data from only one or two example users. FiDo contained a data augmenter that introduced data diversity using a variational autoencoder and a domain-adaptive classifier that adjusted itself to newly collected unlabelled data using a joint classification-reconstruction structure. The experiments with 9 volunteers, 2 out of which had labeled data, achieved the recall of 84.6% and precision of 85%.

CRISLoc [159] was a CSI amplitude fingerprinting localization system based on smartphones. CRISLoc detected altered APs with a joint clustering and outlier detection method. A transfer learning approach was levaraged to reconstruct the CSI fingerprint database on the basis of the outdated fingerprints and a few fresh measurements. An enhanced KNN approach was used to pinpoint the location of a smartphone. Experiments showed that CRISLoc achieved a mean error of around 0.29m, and the mean error increased by 5.4cm and 8.6cm upon the movement of one or two APs.

To adapt to dynamic ship motions for localization in the mobile ship environment, MoLoc [86] designed a CSI fingerprint roaming model to automatically learn the predictive fingerprint variation pattern and transfer the online fingerprint measurement. The fingerprint roaming model was trained by an unsupervised learning strategy using unlabeled onboard collected data combined with ship sensors. Evaluations on real-world cruise ships demonstrated the localization accuracy of 92.8% when the distinction iwas 0.4 m and the mean localization error of 0.68m.

AdapLoc [87] adapted the source localization model to the target environment via semi-supervised domain adaptation, with a small number of CSI amplitude fingerprints from the target domain. The adaptation was achieved by mapping the CSI fingerprints from the source and the target domains into a shared space and minimizing the distribution divergence between them, meanwhile minimizing the distance of the fingerprints at the same locations and maximizing the distance of the fingerprints at different locations. Evaluations in four testbeds with multiple environmental changes demonstrated the accuracy of about 1.5m in smaller rooms and 2.0m in larger rooms.

3) Propagation model: WiGEM [137] modeled RSS as a Gaussian Mixture Model (GMM) and used Expectation Maximization (EM) to learn the maximum likelihood estimates of the model parameters, thus localized a device based on the maximum a posteriori estimate. It avoided the labor-intensive training and made the location estimates robust in the face of device heterogeneity and time varying phenomena. Experiments in two testbeds with four mobile devices showed that accuracy of 7-10m, about 20% 3m, 25% 5m and 4-5m, about 35% 3m, 70% 5m in the two testbeds.

Ohara et al. [98] realized device-free localization by detecting the person passing between a transmitter and a receiver. By transferring a signal strength variance model obtained in the source environment to the target environment, they could construct a localization model for the target environment. The method needed the floor plans of the source and target environments and unlabeled data in the target environment. Experiments with 3 source environments achieved the accuracy of 1.84m in the target environment.

LiFS [23] estimated the target locations by modeling the CSI measurements of multiple wireless links as a set of power fading based equations using only the subcarriers not affected by multipath. LiFS calculated a target's location without requiring offline training, but it required multiple transceivers and the locations of transmitters needed to be known. Experiments in three environments achieved the median accuracy of 0.7m and 80% 1.2m, the median accuracy of 0.5m under LOS and 1.1m under NLOS.

CRIL [83] coupled RSS channel modeling localization system and Inertial Navigation System (INS) for adaptive localization. It employed a Kalman filter to fuse the localization results from the RSS and the INS system and updated the parameters of the RSS channel model utilizing the fused results. Simulations achieved the accuracy of less then 1m and

experiments achieved the accuracy of less than 3m in dynamic environments.

S-Phaser [24] utilized CSI to compute the direct path length between the AP and the terminal. Using the feature of multiple channels, they designed the interpolation elimination method to calibrate signal phases from CSI, and then used the Chinese residue theorem or broadband angle ranging to compute the real distance of the direct path from the calibrated phases. S-Phaser then used a geometric positioning algorithm to determine the target location. The experiments achieved the accuracy of 1.5m with a single AP, 60% within 1.5m under LOS and 40% within 1.5m under NLOS.

Wan et al. [164] and Tong et al. [165] presented a calibration-free CSI phase fingerprinting localization system via hybrid inter-antenna distance and context-aware fingerprinting recognition. The fingerprinting database was constructed theoretically without site survey by associating the AoA with the CSI phase difference based on the signal propagation model and geometric methods. The localization accuracy was improved by parsing the user's trajectory. Experiments achieves 80% localization errors within 0.3m.

4) AoA ToF: ArrayTrack [160] utilized the AoA information calculated from RSS by multiple APs using the frames sent out by the mobile device to locate the mobile device. When the mobile device transmitted a frame on the air, multiple ArrayTrack APs overheard the transmission and each computed AoA information from the mobile device's incoming frame. ArrayTrack aggregated the AoA data to estimate the mobile device location. ArrayTrack was implemented on the Rice WARP FPGA platform with 16 antennas for each AP. For static devices, the accuracy was 317cm for 3 APs and 38cm for 6 APs. For moving devices, the accuracy was 57cm for 3 APs and 31cm for 6 APs (more specific, 90%, 95% and 98% within 80cm, 90cm and 102cm). The limitation of ArrayTrack was that it needed a large number of antennas and need LOS to work accurately.

SpotFi [161] combined CSI values across subcarriers and antennas to jointly estimate AoA and ToF of each path, identified AoA of direct path between the localization target and AP, and used the direct path AoA estimate as well as RSS information to calculate the location of the target. The experiments with 6 APs showed the median accuracy of 40cm and 80% within 1.8m.

Chronos [162] enabled a single AP to localize clients to within tens of centimeters. Chronos computed sub-nanosecond ToF by stitching together information across multiple WiFi frequency bands. By multiplying ToF with the speed of light, a MIMO AP computed the distance between each of its antennas and the client, hence localizing it. Chronos achieved a median error of 65cm in LOS and 98cm in NLOS.

SiFi [163] was a ToA-based indoor localization system. It first collected CSI from the three antennas of the AP, then it used a super-resolution algorithm to estimate the ToA spread, afterwards it applied a clustering algorithm to identify the direct path and removed the symbol time offset; finally it estimated the location by using a weighted iterative least square model to combat sampling frequency offset. Experiments achieved the accuracy of 0.93m in offices, 1.29m in corridors and 1.83m in NOLS.

5) Fresnel zone model: MFDL [170] achieved device free localization of a moving person based on the Fresnel zone model. It estimated the Fresnel phase difference across multiple subcarriers, computed the layer numbers in the corresponding Fresnel zones, and calculated the intersection of two Fresnel zones to obtain the target location. Experiments achieved 45cm median error outdoors and 50-75cm median error in three indoor environments.

Fei et al. [31] proposed a device free localization system based on Fresnel zone model of radio communication that could determine which elliptic region the target located in according to the CSI phase. Results showed that the average error was 33.6cm in the meeting room and 31.2cm in the office room.

B. Tracking

Most cross domain tracking works used domain independent features such as AoA and ToF, some of them used CSI fingerprint or RSS fingerprint. Kim et al. [148] used maximum RSS combined with PDR to realize smartphone-based pedestrian tracking. FreeTrack [125] used CSI amplitude fingerprint for tracking. MaTrack [32] used AoA calculated from CSI for device-free tracking. IndoTrack [36] combined Doppler frequency shift with the AoA spectrum to estimate the human velocity, location, and thus realize human tracking. DeFi [166] achieved localization based on AoA and ToF from CSI. Widar2.0 [35] built a unified model accounting for AoA, ToF and Doppler shifts together to realize human tracking with a single WiFi link. Karanam et al. [37] estimated AoA using only the received signal magnitude measurements and realized localization and tracking. Widar [33] estimated human moving velocities and locations by modeling the relationships between CSI dynamics and locations and velocities exploiting Doppler shifts.

1) Fingerprinting: Kim et al. [148] analyzed the RSS variance problem when using smartphones for WiFi fingerprinting, including device type, device placement, user direction, and environmental changes over time. To overcome the RSS variance problem, they proposed a smartphone-based tracking system, using the location where the maximum RSS of a AP was observed, combined with PDR by a particle filter. Experiments of walking with different configurations achieved the accuracy of less than 2.4m, and 80% 3.0m.

FreeTrack [125] proposed a device-free tracking method based on CSI amplitude fingerprint through deep neural networks (DNN) and particle filtering (PF). To combat with environmental variants such as human diversity, fine tuning on DNN was proposed to enhance the adaptivity of the method. Experiments of training with one person and testing on two different persons achieved the accuracy of 0.63m and 0.72m.

2) AoA ToF DFS: Widar [33] simultaneously estimated human moving velocity and location by modeling the relationship between CSI dynamics and user location and velocity, capturing the geometrical constraints between the signal propagation path length change rates (PLCRs) and user location

TABLE VII
CROSS DOMAIN TRACKING.

Work	Year	Type	Method	Measurement	Domain	Evaluation	Performance
Kim et al. [148]	2012	Device based	Feature extraction: maximal RSS, PDR	RSS fingerprint, accelerometer, compass	Time, orientation, target device	Different configurations	Accuracy<2.4m, CDF 80% $\leq$3.0m
FreeTrack [125]	2020	Device free	Fine tuning	CSI amplitude fingerprint	Human	1TX-1RX, 3 users	Accuracy: 0.63m/0.72m
Widar [33]	2016	Device free	DFS	CSI PLCR	Agnostic	1TX-2RX, 5 users	Accuracy: 0.24m/0.36m
Widar2.0 [35]	2018	Device free	AoA, ToF, DFS	CSI AoA, ToF, DFS	Agnostic	1TX-1RX, 3 rooms, 6 users	Accuracy 0.75m
MaTrack [32]	2016	Device free	AoA	CSI AoA	Agnostic	2TX-2RX/4RX (locations known), 3 rooms	Accuracy$\approx$0.6m
IndoTrack [36]	2017	Device free	AoA, DFS	CSI AoA, DFS	Agnostic	1TX-2RX, 2 rooms, 5 users, 3 weeks	Accuracy 0.35m
Karanam et al. [37]	2018	Device based Device free	AoA	CSI AoA	Agnostic	1TX-3RX, open/closed areas	Accuracy: device-based 0.2m, device-free 0.27m
DeFi [166]	2018	Device free	AoA ToF	CSI AoA, ToF	Agnostic	1TX-4RX, 2 rooms	Accuracy 0.6m

and velocity. The experiments showed that Widar achieved a median location error of 24cm given initial positions and 36cm without them and a mean relative velocity error of 11%. Widar2.0 [35] enabled device free human tracking by leveraging AoA, ToF, Doppler shifts and attenuation together based on CSI. It tracked moving targets and required LOS to work properly. Using a single WiFi link, Widar2.0 achieved a median tracking accuracy of about 0.75m in different environments with different humans.

MaTrack [32] proposed Dynamic-MUSIC to identify the moving target's angle and track the walking human by detecting the subtle reflection signals from the human body and further differentiating them from those reflected signals from static objects. With two receivers, MaTrack achieved a median accuracy about 0.6m when the human was walking.

IndoTrack [36] proposed Doppler-MUSIC to extract accurate Doppler velocity information from CSI and Doppler-AoA to determine the absolute trajectory of the walking human by jointly estimating velocity and location via probabilistic co-modeling of spatial-temporal Doppler and AoA information. Experiments achieved a 35cm median error in human trajectory estimation.

Based on the spatial correlation of the received signal magnitude, Karanam et al. [37] estimated the AoA of signal paths arriving at a receiver array, for localization of fixed passive/active objects and tracking of moving passive/active targets. The method needed LOS to work properly. The experiments using 3 receivers achieved the accuracy of 20cm for active moving targets and 26.75cm for passive moving targets.

DeFi [166] estimated the target location by refining the AoA of the target reflection path based on CSI. DeFi eliminated the static paths using a background elimination algorithm, and separated out the target reflection path from the motion paths according to the AoA and equivalent ToF. AoA measurements were incorporate into a particle filter to realize target location estimation. Experimental results achieved a median error of less than 0.6m.

C. Activity recognition

To deploy human activity recognition in real-world applications, for example health care at home, a pair of transmitter and receiver should cover at least a whole room. Collecting the training samples of all the activities at all the locations in a whole room is labor-intensive thus impractical. It is desired to construct an activity recognition model working across all the locations with only partially collected activity samples. This is a major challenge for CSI-based human activity recognition in a large scale. Environments have significant impact on activity recognition. This issue is attracting more and more attention, as it is the prerequisite to large scale applications of activity recognition. However, the research on environment independent or cross domain activity recognition is still in its infancy.

The works on across domain activity recognition used CSI, including amplitude, phase, AoA, ToF and DFS. FALAR [149], EI [102], CsiGAN [126] and Sheng et al. [106] used CSI amplitude to realize activity recognition. CARM [101] used CSI amplitude, CSI phase and spectrogram for activity recognition. WiFit [103] was a bodyweight exercises monitoring system robust to environments and human using features extracted from CSI amplitude, CSI phase difference and Doppler velocity spectrum. WiSDAR [104] used CSI amplitude and spectrogram for activity recognition. Chang et al. [100] transformed CSI into image for activity recognition. TL-HAR [127] used both CSI amplitude and phase and transformed them to images for multiple human activity recognition. Zhang et al. [42] leveraged the Fresnel diffraction model to determine the signal propagation wrt. a target's motions in the first Fresnel zone, hence to detect human activities using CSI amplitude. WiCAR [88] was an in-car activity recognition framework based on deep learning integrated with multi-adversarial domain adaptation to adapt to new driving conditions, car models and human subjects. WiDrive [105] was a real-time in-car driver activity recognition system based on CSI DFS and Hidden Markov Models with Gaussian Mixture emissions Model (HMM-GMM). An online adaptation algorithm was used to update the parameters

TABLE VIII
CROSS DOMAIN ACTIVITY AND GESTURE RECOGNITION.

Application	Work	Year	Method	Measurement	Domain	Evaluation	Performance
Activity recognition	Chang et al. [100]	2016	Feature extraction: SVD	CSI image	Cross environment	6 activities, 5 rooms (leave-one-out)	Accuracy: 76.67%, 91.67%, 97.50%, 98.33%, 96.67%
	CARM [101]	2017	Feature extraction: DWT, HMM, Multi-link	CSI amplitude, phase, spectrogram	Cross environment, human	8 activities, 5 environments (1 training), 25 users	Accuracy: 1 link$\geq$72%; 3 links$\geq$85%
	FALAR [149]	2018	Feature extraction: KDE, CSVD, NMF	CSI amplitude	Location	5 activities, 8 locations (5 training)	Accuracy 90.6%
	EI [102]	2018	Adversarial network	CSI amplitude	Cross environment, human	6 activities, 6 rooms, 11 users	Accuracy $\approx$75%
	WiFit [103]	2018	Feature extraction	CSI amplitude, phase difference, Doppler velocity spectrum	Cross environment, human	3 activities, 2 environments, 20 users	Counting accuracy$>$98%, classification accuracy$>$90%
	WiSDAR [104]	2019	Feature extraction	CSI amplitude, spectrogram	Cross environment, human	8 activities, 6 users, 4 rooms (1 training)	Accuracy: weak$\approx$96%, rich 80 90%, new user$>$85%
	TL-HAR [127]	2019	Fine tuning	CSI image	Human	10 users	Accuracy: single link 96.7%, multi-link 99.1%
	CsiGAN [126]	2019	Adversarial network	CSI amplitude	Human	Dataset of SignFi and FallDeFi	Accuracy: 84.17%, 86.27%
	Zhang et al. [42]	2019	Fresnel zone model	CSI amplitude	Agnostic	3 activities, 11 users; 9 activities, 20 users	Accuracy: $>$95%, 92.1%
	WiCAR [88]	2019	Adversarial network	CSI amplitude spectrogram	Cross environment, environmental dynamics, human	8 activities, 6 car models, 6 driving conditions, 8 users	Accuracy 95%
	WiDrive [105]	2019	Feature extraction: HMM-GMM	CSI DFS	Cross environment, human	1 new car, 6 new users	Accuray: new car$\approx$80%, new users$>$83%
	Sheng et al. [106]	2020	Fine tuning	CSI amplitude	Cross environment	4 activities, 3 rooms (1 training), 6 users	Accuracy$>$90%
	MatNet-eCSI [107]	2020	Feature extraction: CCFE, MatNet	CSI amplitude, phase	Cross environment	7 activities, 5 users, 3 environments (1 training, 1 testing)	Accuracy: 46 89%
	HAR-MN-EF [108]	2020	Feature extraction: CSI-CE, MatNet	CSI amplitude	Cross environment	6 activities, 5 users, 3 environments (2 training, 1 testing)	Accuracy: 60.8 75.1%
	WiLISensing [150]	2020	Fine tuning	CSI amplitude	Location	6 activities, 36 locations, 1 room	Accuracy$>$90%
	Li et al. [151]	202	ADOA	ADOA	Location	6 activities, 24 locations, 1 room	Accuracy 97.3%
	Zhang et al. [128]	2021	Sample generating	CSI amplitude image	Human	10 activities, 5 users	Accuracy$\approx$90%

of the classifier to adapt for different drivers and vehicles. MatNet-eCSI [107] used both amplitude and phase for activity recognition. HAR-MN-EF [108] used CSI amplitude to accomplish environment-independent activity recognition by leveraging MatNet and the proposed CSI-CE method. WiLISensing [150] leveraged CNN and fine tuning to recognize activities in a position without training or with very few training samples. Li et al. [151] proposed a cross location activity recognition system leveraging Angle Difference of Arriva (ADOA) calculated from CSI using BLSTM. Zhang et al. [128] used CSI amplitude image to achieve activity recognition.

Chang et al. [100] transformed CSI into images to achieve action recognition. They proposed a location-dependency removal method based on Singular Value Decomposition (SVD) to eliminate the background CSI and effectively extract the channel information of signals reflected by human bodies. Experimental results of leave-one-out testing in identifying 6 actions across 5 different rooms achieved the accuracy of 76.67%, 91.67%, 97.50%, 98.33% and 96.67%.

CARM [101] was a CSI-based human activity recognition and monitoring system robust to different environments and human. CARM was based on two models: a CSI-speed model that quantified the relation between CSI dynamics and human movement speeds and a CSI-activity model that quantified the relation between human movement speeds and human activities. They performed data fusion on multiple WiFi links to provide better robustness under different environments. Experiments with 8 activities from 25 volunteers in 5 environments achieved an accuracy of more than 72% for one link and more than 85% for 3 links on a new environment and a new person.

FALAR [149] leveraged CSI to recognize activities regardless of locations and slight environmental changes. They used a Kernel Density Estimation (KDE) based motion extraction method to eliminate slight environmental changes and Class Estimated Basis Space Singular Value Decomposition (CSVD) to discard most location information. The extraction and classification process was realized by Non-negative Matrix Factorization (NMF). Experiments with 5 activities in 8 locations (in which 5 training and 8 testing) achieved the accuracy of 88.6%.

EI [102] was an environment independent activity recognition system composed of a CNN feature extractor, an activity recognizer, a domain discriminator, and several constraints. By exploiting adversarial networks, EI could remove environment and subject specific information and learn transferable features of activities. Experiments on WiFi with labeled data from 22 source domains (11 volunteers in 3 rooms) and unlabeled data from 18 target domains (10 volunteers in 3 rooms) achieved the accuracy of approximately 75%.

WiFit [103] was a bodyweight exercises monitoring system. It segmented and counted the exercises based on the Doppler displacement calculated from the Doppler velocity spectrum. The exercise was classified with SVM using normalized standard deviation of CSI amplitude, normalized standard deviation of CSI phase difference, entropy of CSI histogram, change direction of Doppler Displacement, Doppler velocity intensity, and normalized valid Doppler velocity range. Experiments of 3 exercise in 2 environments with 20 users showed that WiFit

achieved more than 98% counting accuracy and more than 90% of classification accuracy.

WiSDAR [104] enabled spatial diversity aware activity recognition by extending the multiple antennas of modern WiFi devices to construct multiple separated antenna pairs. They generated spectrograms for each antenna pair and proposed a CNN-LSTM model to integrate the hidden features from both temporal and spatial dimensions. The experiments of 8 activities under different antenna topologies demonstrated an accuracy of around 96% in the environment with weak multipath effects, 80 90% in the environment with rich multipath effects, over 85% for strange people.

TL-HAR [127] recognized multiple human activities by employing classification and image transformation together with transfer learning. They transformed CSI amplitude and phase to images to capture correlation among subcarriers, and used CNN to extract representative features for classification. They reduced training complexity through transfer learning that inferred knowledge from a pre-trained Inception V3 model. The experiments achieved the recognition accuracy of 96.7% and 99.1% for single and multiple links.

CsiGAN [126] enabled activity recognition adaptive to user changes based on semi-supervised Generative Adversarial Networks (GAN) to meet the scenarios that unlabeled data from left out users were very limited. CsiGAN introduced a complement generator, which used limited unlabeled data to produce diverse fake samples to train a robust discriminator. Experiments with a gesture dataset SignFi and an activity dataset FallDeFi achieved the accuracy of 84.17% and 86.27% on a left-out user.

Zhang et al. [42] leveraged the Fresnel diffraction model and proposed a diffraction-based sensing model to quantitatively determine the signal propagation with respect to a target's motions in the first Fresnel zone. The model linked the CSI amplitude variation patterns with motions, hence could be used to detect human activities. Experiments of 11 users with 3 activities achieved the accuracy of 95% above, 20 users with 9 activities achieved the accuracy of 92.1%.

WiCAR [88] was an in-car activity recognition framework that was able to remove domain-specific information while retaining the activity related information to the maximum extent. WiCAR used a deep learning architecture integrated with multi-adversarial domain adaptation to adapt to the new domains, including driving conditions, car models and human subjects. Evaluations on 8 activities 6 driving conditions 6 car models 8 volunteers achieved the recognition accuracy of around 95% in untrained domains.

WiDrive [105] was a real-time in-car driver activity recognition system based on CSI. WiDrive consisted of a feature extractor to construct DFS, a activity classifier based on Hidden Markov Models with Gaussian Mixture emissions Model (HMM-GMM), and an online adaptation algorithm to update the parameters of the classifer to adapt for different drivers and vehicles. The experiments of a new car showed the accuracy of about 80% and 6 new users showed the accuracy of more than 83%.

Sheng et al. [106] achieved cross-scene action recognition by integrating spatial features learned from the CNN into the BLSTM model. The off-the-shelf model was used as the pre-trained model and fine-tuned in the new scenario. Experiments of 6 users with 4 activities training in 1 room and testing in 2 other rooms achieved the accuracy of more than 90%.

Shi et al. [107] proposed an activity recognition scheme using Matching Network with enhanced CSI (MatNet-eCSI) to successfully perform one-shot learning to recognize human activities in a new environment. They proposed a CSI correlation feature enhancement (CCFE) method to enhance the activity-dependent information and eliminate the activity-unrelated information. Experiments with 5 users in 3 environments achieved the accuracy of 46-89% for one training and one testing scenario.

Shi et al. [108] proposed the HAR-MN-EF scheme to accomplish environment-independent activity recognition by leveraging MatNet and the proposed CSI-CE method, which enhanced the activity-dependent features whilst mitigated the behavior-unrelated information. Experiments with 6 activities of 5 users in 3 environments showed an accuracy of 60.8-75.1% for training with 2 environments and testing in a new environment without retraining.

WiLISensing [150] leveraged CNN and fine tuning to recognize activities in a position without training or with very few training samples. Experiments at 36 location in one room showed WiLISensing achieved the accuracy above 90% in recognizing six activities.

Li et al. [151] proposed a cross location activity recognition system leveraging Angle Difference of Arriva (ADOA) calculated from CSI using BLSTM. Experiments with 6 activities at 24 locations in one room achieved 97.3% recognition accuracy in a supervised learning framework.

Zhang et al. [128] proposed an activity recognition system that synthesized variant activity data through CSI transformation methods to mitigate the impact of activity inconsistency and subject-specific issues. Experiments of 10 activities from 5 users with 1TX-2RX showed synthetic data improved performance by up to 34.6% and achieved accuracy of around 90%.

D. Gesture recognition

Gesture recognition can be based on CSI amplitude, phase, AoA, ToF and DFS. WiSign [129] used CSI amplitude to realize sign language recognition. WiAG [89] and WiHand [152] used CSI amplitude to recognize gestures. Wang et al. [113] used CSI image for gesture recognition. CrossSense [90] used both CSI fingerprint and RSS fingerprint for gesture recognition and localization. Han et al. [114] used both CSI amplitude and phase for gesture recognition. Ma et al. [115] used CSI amplitude to construct radio images for gesture recognition. Kang et al. [117] proposed gesture recognition based on DFS through multi-source unsupervised domain adaptation. Niu et al. [118] proposed movement fragments and relative motion direction changes extracted from DFS as position-independent features and developed a robust gesture recognition system accordingly. WiADG [109], [110], Yang et al. [112], MobileDA [116] used CSI phase difference for gesture recognition. Han et al. [114] used both CSI amplitude

TABLE IX
CROSS DOMAIN ACTIVITY AND GESTURE RECOGNITION.

Application	Work	Year	Method	Measurement	Domain	Evaluation	Performance
Gesture recognition	WiDeo [167]	2015	AoA ToF	AoA, ToF, amplitude from backscatter	Agnostic	WARP software radios	Location accuracy 0.8m, motion tracing accuracy 0.07m
	WiSign [129]	2017	Pseudo labeling	CSI amplitude	Human	7 users	Accuracy≈87%
	WiAG [89]	2017	Sample generating: translation function	CSI amplitude	Location, orientation, environmental dynamics	6 gestures, 10 users, 5 positions, 8 days	Accuracy 91.4%
	WiADG [110], JADA [109]	2018	Adversarial network	CSI phase difference	Cross environment	6 gestures, 2 rooms	WiADG: 83.3%, 66.6%; JADA: 87.8%, 90.3%
	CrossSense [90]	2018	Sample generating: ANN	CSI amplitude	Cross environment, environmental dynamics	3 rooms, 100 users, 40 gestures	Accuracy>90%
	Yang et al. [112]	2019	Feature extraction: Siamese	CSI phase difference	Cross environment, human	6 gestures, 10 users, 2 rooms	Environment: 65.5%, 75.3%; human: 85.7%, 83.5%
	WiHand [152]	2019	Feature extraction: LRSD	CSI amplitude	Location	5 gestures, 26 users, 9 locations	Accuracy 93%
	Widar3.0 [111]	2019	Feature extraction: BVP	DFS	Cross environment, human, location, orientation	16 users, 5 locations, 5 orientations, 3 environments	Accuracy: 82.6 92.4%
	Wang et al. [113]	2020	Feature extraction: similarity evaluation	CSI amplitude image	Cross environment, human	22 users, environments	Accuracy 89%
	Ma et al. [115]	2020	Feature extraction: similarity evaluation	CSI amplitude image	Cross environment, human	4 users, 2 environments	Accuracy>90%
	DANGR [114]	2020	Sample generating: GAN; Feature mapping: MK-MMD	CSI amplitude, phase	Cross environment	4 gestures, 4 environments	Accuracy 94.5%
	AirDraw [168]	2020	Propagation model	CSI phase	Agnostic	5 users	Mean error<2.2cm
	FingerDraw [169]	2020	Fresnel zone model	CSI amplitude, phase	Agnostic	20 users, 3 environments	Tracking accuracy 1.27cm, recognition accuracy>93%.
	MobileDA [116]	2020	Distillation, feature mapping	CSI phase difference	Cross environment	6 gestures, 2 rooms	Accuracy≈90.2%
	Kang et al. [117]	2021	Adversarial network	DFS	Cross environment, human, location, orientation	16 users, 5 locations, 5 orientations, 3 environments	Accuracy 85.7%
	Niu et al. [118]	2021	Feature extraction: fragments, orientation change	DFS	Cross environment, human, location, orientation	8 gestures, 10 users, 3 environments, 5 locations, 5 orientations	Accuracy 98.6%

and phase for gesture recognition. AirDraw [168] used CSI phase to realize a learning-free in-air handwriting system by passive gesture tracking. FingerDraw [169] tracked centimeter-level finger drawings by leveraging the CSI-quotient model and the Fresnel model based on CSI amplitude and phase. WiDeo [167] calculated each reflection's amplitude, ToF and AoA to realize motion tracing. Widar3.0 [111] achieved cross domain gesture recognition by deriving a domain independent feature BVP obtained from Doppler frequency shift.

WiDeo [167] mined the backscatter reflections from the environment that WiFi transmissions naturally produced, calculated each reflection's amplitude, ToF and AoA, to trace where reflecting objects were located and how they were moving. The experiments using the WARP software radios achieved a median localization accuracy of 0.8m and motion tracing accuracy of 7cm.

WiSign [129] used CSI amplitude of selected subcarriers to recognize sign language with sparsely labeled training dataset. Two solutions based on transfer learning and semi-supervised learning were proposed. The semi-supervised learning framework labeled the unlabeled instances using the knowledge from labeled instances. The transfer learning-based solution calculated the similarities between two labeled instances and choosing auxiliary data which were similar to the new user's training data. Evaluations showed a mean prediction accuracy of 87.01% and 87.38% for the transfer learning-based approach and semi-supervised learning-based approach.

WiAG [89] recognized the gestures irrespective of user position, orientation and environmental dynamics. It required the CSI training samples of all the gestures in one domain and automatically generated virtual samples for all the gestures in other domains through a translation function. The evaluation results showed the average gesture recognition accuracy of 91.4%.

WiADG [110] and JADA [109] identified human gestures across environments via adversarial domain adaptation. They trained a target encoder and fine-tuned the source encoder through adversarial learning to map both unlabeled target data and labeled source data to a domain-invariant feature space such that a domain discriminator could not distinguish the domain labels of the data. The results in 2 indoor environments validated that WiADG achieved gesture recognition accuracy of 83.3% and 66.6% and JADA achieved 87.8% and 90.3% under spatial dynamics.

CrossSense [90] enabled WiFi training measurements to be collected once and used across sites, by employing an Artificial Neural Network (ANN) to train a roaming model that generated synthetic training samples for each target environment from the original. The roaming model was trained offline using the samples collected from the original and each target environment. Experiments involving 3 rooms and 100 user achieved the accuracy of more than 90% for CSI and more than 80% for RSS in different environments and environmental changes.

Yang et al. [112] achieved cross domain gesture recognition via a Siamese recurrent convolutional architecture using CSI phase difference. They adopted CNN to extract spatial features and BLSTM to capture temporal dynamics of gestures. The

Siamese framework used transferable pairwise loss to remove structured noise such as individual heterogeneity and various measurement conditions. The experiments demonstrated the accuracy of 65.3% and 75.3% for training in one room and testing in the other, 85.7% and 83.5% for traing with 5 users and testing with the other 5 users.

WiHand [152] leveraged the low rank and sparse decomposition (LRSD) algorithm and separated gesture signal from background information, thus making gesture recognition resilient to location variation. Extensive evaluations of CSI amplitude based gesture recognition achieved the average accuracy of 93% across different locations.

Widar3.0 [111] achieved cross domain gesture recognition by extracting the domain independent feature body-coordinate velocity profile BVP. On this basis, they developed a one-fits-all model that required only one-time training but could adapt to different data domains. In experiments with 1TX-6RX across locations, orientations, environments and user diversity, Widar3.0 achieved 82.6-92.4% cross domain recognition accuracy. Widar3.0 required one transmitter and at least three receivers to work.

Wang et al. [113] alleviated the efforts of retraining in a new scenario by learning a universal similarity evaluation ability utilizing a deep similarity evaluation network based on CSI image. In the experiment, 2 users wrote 26 capital letters from A to Z in the air with hand gestures in two scenarios, thus collected 4 datasets. Training in one dataset and testing in the other three achieved the recognition accuracy of 89%.

Ma et al. [115] presented a device-free gesture recognition system based on CSI amplitude image, which could recognize new type of gestures, or gestures performed in a new scenario or by a new user. They designed a deep network which could not only learn discriminative deep features, but also learn a transferable similarity evaluation ability from the training set and apply the learned knowledge to the new testing conditions. Extensive experiments on sign language recognition and letter recognition in the air conducted by four users in two scenarios demonstrate an accuracy of more than 90%, using very few number of new samples.

DANGR [114] was a CSI gesture recognition scheme based on deep adaptation networks. DANGR first exploited GAN to augment the dataset, then adopted domain adaptation based on multi-kernel Maximum Mean Discrepancy scheme to shrink the domain discrepancies in environments, which matched the mean-embeddings of abstract representations across domains in a reproducing kernel Hilbert space. Experiments in four environments yielded mean 94% accuracy of gesture recognition with one training and the other three testing.

AirDraw [168] was a learning-free in-air handwriting system by passive gesture tracking using CSI phase. They denoised CSI measurements by the ratio between two close-by antennas, separated the reflected signal from noise by performing PCA, and proposed a signal calibration algorithm for tracking correction by eliminating the static components unrelated to hand motion. Experiments of AirDraw with 5 users could track user's hand trace with a median error lower than 2.2cm within a tracking range by 1.5m.

FingerDraw [169] tracked centimeter-level finger drawings by leveraging the CSI-quotient model and the Fresnel zone, which used the channel quotient between two antennas of the receiver to cancel out the noise in CSI amplitude and the random offsets in CSI phase, and quantified the correlation between CSI value dynamics and object displacement. Experiments with one transmitter and two receivers in three environments with 20 users showed the overall median tracking accuracy of 1.27cm, and the recognition accuracy of drawing ten digits in the air of over 93%.

MobileDA [116] allowed a teacher network trained in the server to distill the knowledge for a student network running in the edge device by cross-domain distillation, thus to effectively train a simple sutdent model that adapted to the new environment by diminishing domain disparity. Leveraging unlabeled data in the new environment, the student model amended the feature learning to be domain invariant by statistical metrics. Evaluations on CSI gesture recognition achieved the accuracy of about 90.2% in a new environment.

Kang et al. [117] proposed gesture recognition based on DFS through multi-source unsupervised domain adaptation, considering environment, gesture performer, location and orientation. They applied an adversarial learning scheme together with feature disentanglement modules to remove gesture irrelevant factors and introduced an attention scheme to promote positive transfer between source and target domain. Experiments were conducted with Widar3.0 dataset using only two links' data. The accuracy was 87.8%, 91.8%, 92.5%, 87.1% and 85.7% across environments, subjects, locations, orientations and four influential factors respectively.

Niu et al. [118] analyzed the Doppler frequency shift in WiFi sensing and developed a model to quantify the relationship between signal frequency and target position, motion direction and speed. Based on this model, they proposed movement fragments and relative motion direction changes as two position-independent features, which could be extracted from the time frequency spectrogram. On this basis, they designed a suite of position-independent gestures and developed the gesture recognition system accordingly. Evaluations showed more than 96% recognition accuracy without any training, robust to different locations, orientations, environments and persons.

E. Miscellaneous cross domain sensing applications

TL-Fall [119] used CSI amplitude to detect falls. FallDeFi [91] used CSI amplitude to generate spectrogram to detect falls. RespiRadio [92] used CSI amplitude for respiration rate monitoring. AR-Alarm [93] utilized CSI phase difference to detect intrusion. CrossSense [90] made use of both CSI fingerprints and RSS fingerprints for gesture recognition and gait recognition. WiDIGR [155] achieved direction-independent gait recognition based on the Fresnel model and DFS and a series of signal processing techniques. Jung et al. [153] used CSI for identity verification based on the gesture signals of handwritten signature. Shi et al. [95] presented device-free user authentication by recognizing human activities, using CSI amplitude, relative amplitude and STFT holograms. Regani et al. [94] used CSI phase to authenticate

TABLE X
MISCELLANEOUS CROSS DOMAIN WIRELESS SENSING APPLICATIONS.

Application	Work	Year	Method	Measurement	Domain	Evaluation	Performance
Fall detection	FallDeFi [91]	2018	Feature extraction	CSI amplitude, spectrogram	Cross environment, environmental dynamics, human, time	5 rooms, 3 users, different days	Source accuracy>93%, target accuracy near 80%
	TL-Fall [119]	2019	Fine tuning	CSI amplitude	Cross environment	4 rooms (1 training), 10 users	Sensitivity 86.83%, specificity 84.71%
Respiration monitoring	RespiRadio [92]	2019	Propagation model: synthesized wideband	CSI amplitude	Environmental dynamics	1 room, 3 users	Estimation error 0.152bpm
Intrusion detection	AR-Alarm [93]	2017	Feature extraction	CSI phase difference	Cross environment, environmental dynamics	2 rooms, 4 users, 2 months	TDR>93%
Gait recognition	CrossSense [90]	2018	Sample generating: ANN	CSI amplitude	Cross environment, environmental dynamics	3 rooms, 100 users	Accuracy>80%
	WiDIGR [155]	2020	Fresnel zone model	CSI spectrogram	Orientation	3 environments, 3 6 users, 0° 90°	Accuracy 78.28 92.83%
User authentication	Regani et al. [94]	2020	Feature extraction: frequency hopping	CSI phase	Time, environmental dynamics	7 users, 2 months	Accuracy: 2 driver 99.13%, 3 driver 84.33%, 7 driver 53.85%
	Jung et al. [153]	2019	Fine tuning	CSI amplitude	Location	2 locations, 4 orientations, 50 identities	Accuracy 94.8%
	Shi et al. [95]	2020	Feature extraction	CSI amplitude, relative amplitude, hologram	Environmental dynamics, location	2 rooms, 10 activities, 10/5 users, 4 locations, 3 furniture changes	Environmental dynamics: 83.6 87.3%, location: 81.2 91.3%
People counting	Di Domenico et al. [120]	2016	Feature extraction	CSI amplitude	Cross environment	1 room training, 2 rooms testing	Error≤2: 91%, 81%
	Di Domenico et al. [121]	2016	Feature extraction	CSI DFS, RSS	Cross environment	1 room training, 2 rooms testing	Error≤2: 99% and 92%

the driver. Di Domenico et al. [120] focused on crowd counting without retraining in new environments, by identifying a set of differential CSI feature candidates and selecting the most effective ones via minimization of the summation of the Davies-Bouldin indexes. They [121] presented another crowd counting system by analyzing the shape of the Doppler spectrum of the received WiFi signal which was correlated to the number of people.

1) Fall detection: FallDeFi [91] detected fall based on WiFi CSI amplitude, using STFT to extract time-frequency features and a sequential forward selection algorithm to single out features that were resilient to environment changes. It achieved a 93% accuracy. When the environment changes, it achieved an accuracy close to 80%.

TL-Fall [119] was a transfer-learning based device-free fall detection system. The fall detection model was trained with labeled data to derive knowledge in the old environment. With the derived knowledge, the fall detection model working in the new environment could be trained with only a few labeled data. The experimental results demonstrated 86.83% fall detection sensitivity and 84.71% fall detection specificity on average.

2) Respiration monitoring: Respiration rate monitoring is beneficial for the diagnosis of a variety of diseases, such as heart failure and sleep disorders. RespiRadio [92] could detect a person's respiration rate in dynamic ambient environments by synthesizing a wider-bandwidth WiFi radio. With the synthesized WiFi radio, RespiRadio could identify the path reflected by the breathing person and then analyzed the periodicity of the signal power measurements only from this path to infer the respiration rate. The results in a dynamics environment demonstrated the overall estimation error of 0.152 breaths per minute (bpm).

3) Intrusion detection: AR-Alarm [93] was an adaptive human intrusion detection system using WiFi CSI phase difference of two antennas. By utilizing a robust feature of the ratio between the dynamic and static CSI profiles of the environment and self-adaptive learning mechanism, AR-Alarm achieved real-time intrusion detection in different environments without calibration efforts. In the experiments in 2 rooms, AR-Alarm achieved the true detection rate of more than 93% for different environments and environmental changes.

4) Gait recognition: CrossSense [90] enable WiFi training measurements to be collected once and used across sites, by employing an Artificial Neural Network (ANN) to train a roaming model that generated synthetic training samples for each target environment. The roaming model was trained offline using the samples collected from the original and each target environment. Experiments involving 3 rooms and 100 user achieved the accuracy of more then 80% for different environments and environmental changes.

WiDIGR [155] was a direction-independent gait recognition system. Based on the Fresnel model, WiDIGR used a series of signal processing techniques to eliminate the differences among induced signals caused by walking in different directions and generated a high-quality direction-independent signal spectrogram. The experimental results in different environments and orientations achieved the accuracy ranging from 78.28% for a group of six subjects to 92.83% for a group of three.

5) User authentication: Regani et al. [94] proposed in-car driver authentication by identifying the unique radio biometric information recorded in CSI. To deal with the in-car changing environments, they built a long-term driver radio biometric database and created machine learning models that made the system adaptive to new in-car environments. The accuracy achieved 99.13% for two drivers, 84.33% for three drivers and 53.85% for seven drivers.

Jung et al. [153] proposed identity verification based on the gesture signals of handwritten signature captured by CSI at different positions. A CNN-based feature extractor was first pre-trained using the CSI from one position, and transferred to recognize signals collected from another position via a rapid retraining process, utilizing the kernel and the range (KAR) space projection learning for computational efficiency. Experiments with 2 positions, 4 orientations and 50 identities achieved an accuracy of 94.8%.

Shi et al. [95] presented device-free user authentication by recognizing human activities, using CSI amplitude, relative amplitude and STFT holograms. They proposed a deep-learning based approach with an unsupervised domain discriminator to mitigate the impact of varying location and environmental changes. Experiments in two scenarios achieved the accuracy of 83.6-87.3% for environmental dynamics, 81.2-91.3% for location variations.

6) People counting: Di Domenico et al. [120] focused on crowd counting without retraining in new environments, by identifying a set of differential CSI feature candidates and selecting the most effective ones via minimization of the summation of the Davies-Bouldin indexes. Experiments by training in a room and testing in two different rooms showed that more than 91% and 81% estimation errors were less or equal than 2 persons respectively. They [121] presented another crowd counting system by analyzing the shape of the Doppler spectrum of the received WiFi signal which was correlated to the number of people. The results in two rooms showed that more than 84% and 71% estimation errors were less or equal than 1 person, and the errors were less or equal than 2 persons for 99% and 92% of the times.

F. Remarks

The majority of the cross domain sensing works are for localization, then for gesture recognition, tracking, and activity recognition. There are xx works for localization, xx works for gesture recognition, xx works for tracking, and xx works for activity recognition.

VII. Discussions and Challenges

A. Discussions

1) Domain agnostic method:

2) Domain independent features:

3) Domain adaptation:

B. Challenges

1) Unsupervised domain adaptation:

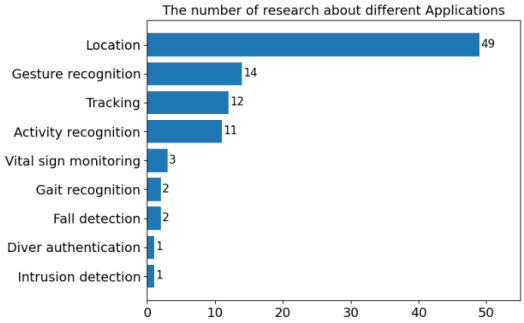


Fig. 25. Number of papers by the applications.

2) Domain generalization: The current sensing models are based on the multipath analysis to find out the human impact on the received signals. They add limitations to the deployment environments. There should not be other persons or moving objects around, otherwise other people or moving object's actions are also captured by the received signals. When being deployed in real scenarios, theses limitations should be eliminated.

3) Multiple user sensing: "When multiple persons are sharing the same physical space, the received signals will contain all the impacts by all the persons' motions. Existing solutions have difficulty in multi-taget sensing. Although FMCW and antenna array-based solutions [172] can track the hand gestures of multiple people simultaneously, leveraging a directional antenna array. For other wireless signals, a promising way to isolate concurrent activities of different people is to separate sensing space with a complex web of wireless links in the area. Nevertheless, it is still challenging to address the effect of multiple users and recognize different actions conducted by different users with a limited number of wireless links."

"Regardless the recent advancement of CSI-based sensing technology, monitoring two or more objects simultaneously is still a crucial challenge. In [54], the authors presented the first study to address gesture tracking for multiple users. They tested simultaneously performed gestures and studied their impact on wireless signals. Tracking signals for random motions and unfixed orientations is a challenging task and needs to be investigated in literature."

4) Complex deep learning: "Some wireless sensing systems exploit deep learning approaches to automatically extract features. The deep learning approaches, require not only an extensive training set but also a comparable computation and storage capacity to perform training. It is a big burden on the users to collect a large number of training samples and maybe not computable on resources limited devices such as wearable and edge devices."

5) Lack of standard datasets: "Currently, most wireless sensing studies evaluate their performance using their dataset. Researchers have to recruit some volunteers to collect wireless signal streams. Moreover, the experimental environments are often chosen according to the particular targets of the applications. Consequently, the system performance often depends

on the deployment and the collection process, which makes it difficult for comparison among different studies."

C. Future work

1) Domain generalization:

2) Few shot learning:

3) Continuous learning in complicated environment:

VIII. Conclusions

In this article, a survey of the state-of-the-art cross domain WiFi sensing has been conducted. We category the works according to the domains, the measurements, the applications and the method. There are still several challenges to address in the future.

References

[1] S. Sigg, S. Shi, and Y. Ji, "Rf-based device-free recognition of simultaneously conducted activities," in *2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, ser. UbiComp'13 Adjunct. ACM, 2013, pp. 531–540.

[2] S. Orphomma and N. Swangmuang, "Exploiting the wireless rf fading for human activity recognition," in *2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2013, pp. 1–5.

[3] Fugui Qi, Zhao Li, Fulai Liang, Hao Lv, Qiang An, and Jianqi Wang, "A novel time-frequency analysis method based on HHT for finer-grained human activity using SFCW radar," in *2016 Progress in Electromagnetic Research Symposium (PIERS)*, 2016, pp. 2536–2539.

[4] P. Markopoulos, S. Zlotnikov, and F. Ahmad, "Adaptive radar-based human activity recognition with l1-norm linear discriminant analysis," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. PP, pp. 1–1, January 2019.

[5] B. Islam, M. T. Islam, J. Kaur, and S. Nirjon, "Lorain: Making a case for lora in indoor localization," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2019, pp. 423–426.

[6] F. Zhang, Z. Chang, K. Niu, J. Xiong, B. Jin, Q. Lv, and D. Zhang, "Exploring lora for long-range through-wall sensing," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 2, June 2020.

[7] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Computing Surveys*, vol. 46, no. 2, December 2013.

[8] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[9] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, Jun. 2019. [Online]. Available: https://doi.org/10.1145/3310194

[10] R. C. Shit, S. Sharma, D. Puthal, P. James, B. Pradhan, A. v. Moorsel, A. Y. Zomaya, and R. Ranjan, "Ubiquitous localization (ubiloc): A survey and taxonomy on device free localization for smart world," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3532–3564, 2019.

[11] W. Liu, Q. Cheng, Z. Deng, H. Chen, X. Fu, X. Zheng, S. Zheng, C. Chen, and S. Wang, "Survey on csi-based indoor positioning systems and recent advances," in *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2019, pp. 1–8.

[12] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 3, pp. 1629–1645, 2020.

[13] H. F. T. Ahmed, H. Ahmad, and A. C.V., "Device free human gesture recognition using wi-fi csi: A survey," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103281, 2020.

[14] Y. He, Y. Chen, Y. Hu, and B. Zeng, "Wifi vision: Sensing, recognition, and detection with commodity mimo-ofdm wifi," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8296–8317, 2020.

[15] K. Ngamakeur, S. Yongchareon, J. Yu, and S. U. Rehman, "A survey on device-free indoor localization and tracking in the multi-resident environment," *ACM Computing Surveys*, vol. 53, no. 4, July 2020.

[16] X. Zhu, W. Qu, T. Qiu, L. Zhao, M. Atiquzzaman, and D. O. Wu, "Indoor intelligent fingerprint-based localization: Principles, approaches and challenges," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 4, pp. 2634–2657, 2020.

[17] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *2012 IEEE INFOCOM*, 2012, pp. 2210–2218.

[18] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," in *2010 ACM SIGCOMM*. ACM, 2010, pp. 159–170.

[19] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wifi," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: ACM, 2015, p. 53–64. [Online]. Available: http://doi.acm.org/10.1145/2789168.2790124

[20] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based indoor localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1300–1309, 2013.

[21] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Youssef, "MonoPHY: Mono-stream-based device-free WLAN localization via physical layer information," in *2013 IEEE WCNC*, 2013, pp. 4546–4551.

[22] I. Sabek and M. Youssef, "MonoStream: A minimal-hardware high accuracy device-free WLAN localization system," *Computer Science*, 2013.

[23] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "LiFS: Low human-effort, device-free localization with fine-grained subcarrier information," in *MobiCom'2016*. ACM, 2016, pp. 243–256.

[24] S. Han, Y. Li, W. Meng, C. Li, T. Liu, and Y. Zhang, "Indoor localization with a single Wi-Fi access point based on OFDM-MIMO," *IEEE Systems Journal*, vol. 13, no. 1, pp. 964–972, 2019.

[25] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2017.

[26] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, 2016.

[27] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.

[28] X. Wang, X. Wang, and S. Mao, "CiFi: Deep convolutional neural networks for indoor localization with 5GHz Wi-Fi," in *2017 IEEE ICC*, 2017, pp. 1–6.

[29] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 346–10 356, 2017.

[30] R. Zhou, M. Hao, X. Lu, M. Tang, and Y. Fu, "Device-free localization based on CSI fingerprints and deep neural networks," in *15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2018, pp. 1–9.

[31] H. Fei, F. Xiao, H. Huang, and L. Sun, "Indoor static localization based on fresnel zones model using cots wi-fi," *Journal of Network and Computer Applications*, vol. 167, p. 102709, 2020.

[32] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei, "Dynamic-MUSIC: Accurate device-free indoor localization," in *ACM UbiComp'2016*. ACM, 2016, pp. 196–207.

[33] K. Qian, C. Wu, Z. Yang, C. Yang, and Y. Liu, "Decimeter level passive tracking with wifi," in *3rd Workshop on Hot Topics in Wireless*, ser. HotWireless'16. ACM, 2016, pp. 44–48.

[34] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi," in *18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. Mobihoc'17. ACM, 2017, pp. 1–10.

[35] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single wi-fi link," in *The 16th Annual International Conference on Mobile Systems, Applications, and Services (Mobisys)*, 2018, p. 350–361.

[36] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, "IndoTrack: Device-free indoor human tracking with commodity Wi-Fi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, September 2017.

[37] C. Karanam, B. Korany, and Y. Mostofi, "Magnitude-based angle-of-arrival estimation, localization, and target tracking," in *The 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN'2018)*, 04 2018, pp. 254–265.

[38] S. Shi, S. Sigg, L. Chen, and Y. Ji, "Accurate location tracking from CSI-based passive device-free probabilistic fingerprinting," *IEEE*

Transactions on Vehicular Technology, vol. 67, no. 6, pp. 5217–5230, June 2018.

[39] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *MobiCom*, 2014, pp. 617–628.

[40] D. Zhu, N. Pang, G. Li, and S. Liu, "WiseFi: Activity localization and recognition on commodity off-the-shelf WiFi devices," in *IEEE HPCC/SmartCity/DSS*, 2016.

[41] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li, "Wi-chase: A wifi based human activity recognition system for sensorless environments," in *IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2017, pp. 1–6.

[42] F. Zhang, K. Niu, J. Xiong, B. Jin, T. Gu, Y. Jiang, and D. Zhang, "Towards a diffraction-based sensing approach on human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 3, no. 1, pp. 1–25, March 2019.

[43] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2019.

[44] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "WiAct: A passive WiFi-based human activity recognition system," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 296–305, 2020.

[45] C. Han, K. Wu, Y. Wang, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," in *IEEE INFOCOM*, 2014, pp. 271–279.

[46] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2017.

[47] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "Rt-fall: A real-time and contactless fall detection system with commodity wifi devices," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 511–526, February 2017.

[48] C. Feng, S. Arshad, and Y. Liu, "Mais: Multiple activity identification system using channel state information of wifi signals," *International Conference on Wireless Algorithms, Systems, and Applications*, pp. 419–432, 2017.

[49] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *19th Annual International Conference on Mobile Computing*, ser. MobiCom'13. ACM, 2013, pp. 27–38.

[50] W. He, K. Wu, Y. Zou, and Z. Ming, "WiG: Wifi-based gesture recognition system," in *24th International Conference on Computer Communication and Networks (ICCCN)*, 2015, pp. 1–7.

[51] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger: Talk to your smart devices with finger-grained gesture," in *UbiComp'2016*. ACM, 2016, pp. 250–261.

[52] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. Ni, "We can hear you with wi-fi!" *IEEE Transactions on Mobile Computing*, vol. 15, pp. 2907–2920, November 2016.

[53] Q. Zhou, J. Xing, J. Li, and Q. Yang, "A device-free number gesture recognition approach based on deep learning," in *2016 12th International Conference on Computational Intelligence and Security (CIS)*, 2016, pp. 57–63.

[54] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using wifi," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys'18, 2018, p. 401–413.

[55] H. Abdelnasser, K. A. Harras, and M. Youssef, "Ubibreathe: A ubiquitous non-invasive wifi-based breathing estimator," in *16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc'15. ACM, 2015, pp. 277—-286.

[56] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf wifi devices," *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2466–2479, October 2016.

[57] J. Shang and J. Wu, "Fine-grained vital signs estimation using commercial wi-fi devices," in *8th Wireless of the Students, by the Students, and for the Students Workshop*. ACM, 2016, pp. 30—-32.

[58] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: Do user location and body orientation matter?" in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'2016)*. ACM, 2016, p. 25–36.

[59] X. Wang, C. Yang, and S. Mao, "Phasebeat: Exploiting csi phase data for vital sign monitoring with commodity wifi devices," in *IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 1230–1239.

[60] X. Wang, C. Yang, and S. Mao, "Tensorbeat: Tensor decomposition for monitoring multi-person breathing beats with commodity wifi," *ACM Transactions on Intelligent Systems and Technology*, vol. 9, February 2017.

[61] M. I. Khan, M. A. Jan, Y. Muhammad, D.-T. Do, A. ur Rehman, C. X. Mavromoustakis, and E. Pallis, "Tracking vital signs of a patient using channel state information and machine learning for a smart healthcare system," *Neural Computing and Applications*, January 2021.

[62] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using WiFi," in *2014 IEEE INFOCOM*, 2014, pp. 361–369.

[63] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only wifi power measurements," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 7, pp. 1381–1393, July 2015.

[64] E. Cianca, M. D. Sanctis, and S. D. Domenico, "Radios as sensors," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 363–373, April 2017.

[65] R. Zhou, X. Lu, Y. Fu, and M. Tang, "Device-free crowd counting with wifi channel state information and deep neural networks," *Wireless Networks*, vol. 26, 07 2020.

[66] S. Doong, "Counting human flow with deep neural network," in *Hawaii International Conference on System Sciences*, 2018.

[67] Y. Yang, J. Cao, X. Liu, and X. Liu, "Wi-Count: Passing people counting with COTS WiFi devices," in *ICCCN 2018*, August 2018, pp. 1–9.

[68] F. Xiao, Z. Guo, Y. Ni, X. Xie, S. Maharjan, and Y. Zhang, "Artificial intelligence empowered mobile sensing for human flow detection," *IEEE Network*, vol. 33, no. 1, pp. 78–83, January 2019.

[69] R. Zhou, Z. Gong, X. Lu, and Y. Fu, "Wiflowcount: Device-free people flow counting by exploiting doppler effect in commodity wifi," *IEEE Systems Journal*, vol. 14, no. 4, pp. 4919–4930, 2020.

[70] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Omnidirectional coverage for device-free passive human detection," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1819–1829, 2014.

[71] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with WiFi," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, 2015.

[72] J. Lv, W. Yang, L. Gong, D. Man, and X. Du, "Robust WLAN-based indoor fine-grained intrusion detection," in *2016 IEEE GLOBECOM*, December 2016, pp. 1–6.

[73] H. Zhu, F. Xiao, L. Sun, R. Wang, and P. Yang, "R-TTWD: Robust device-free through-the-wall detection of moving human with WiFi," *IEEE Journal on Selected Areas in Communications*, vol. PP, no. 99, pp. 1–1, 2017.

[74] F. Hong, X. Wang, Y. Yang, Y. Zong, Y. Zhang, and Z. Guo, "Wfid: Passive device-free human identification using wifi signal," in *13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous'2016)*. ACM, 2016, pp. 47–56.

[75] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: Wifi-based person identification in smart spaces," in *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*, ser. IPSN '16. IEEE Press, 2016.

[76] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *UbiComp'2016*. ACM, 2016, pp. 363–373.

[77] Y. Chen, W. Dong, Y. Gao, X. Liu, and T. Gu, "Rapid: A multimodal and device-free approach using noise estimation for robust person identification," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 1, no. 3, September 2017.

[78] K. Wu, "Wi-Metal: Detecting metal by using wireless networks," in *IEEE ICC*. IEEE, 2016, pp. 1–6.

[79] Y. Zou, Y. Wang, S. Ye, K. Wu, and L. M. Ni, "TagFree: Passive object differentiation via physical layer radiometric signatures," in *IEEE PerCom*. IEEE, 2017, pp. 237–246.

[80] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang, "Towards in-baggage suspicious object detection using commodity WiFi," in *IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2018, pp. 1–9.

[81] A. Hanif, M. S. Chughtai, A. A. Qureshi, A. Aleem, F. Munir, M. Tahir, and M. Uppal, "Non-obtrusive detection of concealed metallic objects using commodity WiFi radios," in *IEEE GLOBECOM*. IEEE, 2018, pp. 1–6.

[82] B. Zhou, Z. Chen, Z. Gong, and R. Zhou, "Detection of suspicious objects concealed by walking pedestrians using wifi," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[83] S. Cai, W. Liao, C. Luo, M. Li, X. Huang, and P. Li, "CRIL: An efficient online adaptive indoor localization system," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4148–4160, May 2017.

[84] X. Chen, C. Ma, M. Allegue, and X. Liu, "Taming the inconsistency of wi-fi fingerprints for device-free passive indoor localization," in *IEEE INFOCOM - IEEE Conference on Computer Communications*, 2017, pp. 1–9.

[85] K. Liu, H. Zhang, J. K. Ng, Y. Xia, L. Feng, V. C. S. Lee, and S. H. Son, "Toward low-overhead fingerprint-based indoor localization via transfer learning: Design, implementation, and evaluation," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 898–908, 2018.

[86] M. Chen, K. Liu, J. Ma, X. Zeng, Z. Dong, G. Tong, and C. Liu, "MoLoc: Unsupervised fingerprint roaming for device-free indoor localization in a mobile ship environment," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11 851–11 862, 2020.

[87] R. Zhou, H. Hou, Z. Gong, Z. Chen, K. Tang, and B. Zhou, "Adaptive device-free localization in dynamic environments through adaptive neural networks," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 548–559, 2021.

[88] F. Wang, J. Liu, and W. Gong, "WiCAR: WiFi-based in-car activity recognition with multi-adversarial domain adaptation," in *2019 IEEE/ACM 27th International Symposium on Quality of Service (IWQoS)*, 2019, pp. 1–10.

[89] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using wifi," in *The 15th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys'17. ACM, 2017, pp. 252—-264.

[90] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "CrossSense: Towards cross-site and large-scale WiFi sensing," in *The 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 305–320.

[91] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "Falldefi: Ubiquitous fall detection using commodity wi-fi devices," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, January 2018.

[92] S. Shi, Y. Xie, M. Li, A. X. Liu, and J. Zhao, "Synthesizing wider wifi bandwidth for respiration rate monitoring in dynamic environments," in *IEEE INFOCOM - IEEE Conference on Computer Communications*, 2019, pp. 181–189.

[93] S. Li, X. Li, K. Niu, H. Wang, Y. Zhang, and D. Zhang, "Ar-alarm: An adaptive and robust intrusion detection system leveraging csi from commodity wi-fi," in *Enhanced Quality of Life and Smart Living*, M. Mokhtari, B. Abdulrazak, and H. Aloulou, Eds. Cham: Springer International Publishing, 2017, pp. 211–223.

[94] S. D. Regani, Q. Xu, B. Wang, M. Wu, and K. J. R. Liu, "Driver authentication for smart car using wireless sensing," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2235–2246, 2020.

[95] C. Shi, J. Liu, N. Borodinov, B. Leao, and Y. Chen, "Towards environment-independent behavior-based user authentication using wifi," in *2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2020, pp. 666–674.

[96] S. J. Pan, S. Dou, Y. Qiang, and J. T. Kwok, "Transferring localization models across space," in *The 23rd AAAI Conference on Artificial Intelligence*, July 2008.

[97] H.-Y. Wang, V. W. Zheng, J. Zhao, and Q. Yang, "Indoor localization in multi-floor environments with reduced effort," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2010, pp. 244–252.

[98] K. Ohara, T. Maekawa, Y. Kishino, Y. Shirai, and F. Naya, "Transferring positioning model for device-free passive indoor localization," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp)*. ACM, 2015, pp. 885–896.

[99] L. Chang, X. Chen, Y. Wang, D. Fang, J. Wang, T. Xing, and Z. Tang, "Fitloc: Fine-grained and low-cost device-free localization for multiple targets over various areas," *IEEE/ACM Transactions on Networking*, vol. 25, no. 4, pp. 1994–2007, 2017.

[100] J.-Y. Chang, K.-Y. Lee, Y.-L. Wei, K. C.-J. Lin, and W. Hsu, "Location-independent wifi action recognition via vision-based methods," in *The 24th ACM International Conference on Multimedia*, ser. MM'16. Association for Computing Machinery, 2016, p. 162–166.

[101] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[102] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *The 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom'18. ACM, 2018, pp. 289—-304.

[103] S. Li, X. Li, Q. Lv, G. Tian, and D. Zhang, "WiFit: Ubiquitous bodyweight exercise monitoring with commodity wi-fi devices," in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2018, pp. 530–537.

[104] F. Wang, W. Gong, and J. Liu, "On spatial diversity in wifi-based human activity recognition: A deep learning-based approach," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2035–2047, 2019.

[105] Y. Bai, Z. Wang, K. Zheng, X. Wang, and J. Wang, "WiDrive: Adaptive WiFi-based recognition of driver activity for real-time and safe takeover," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 2019, pp. 901–911.

[106] B. Sheng, F. Xiao, L. Sha, and L. Sun, "Deep spatial-temporal model based cross-scene action recognition using commodity wifi," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[107] Z. Shi, J. A. Zhang, Y. D. R. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2020.

[108] Z. Shi, J. A. Zhang, R. Xu, Q. Cheng, and A. Pearce, "Towards environment-independent human activity recognition using deep learning and enhanced csi," in *2020 IEEE Global Communications Conference (GLOBECOM 2020)*, 2020, pp. 1–6.

[109] H. Zou, J. Yang, Y. Zhou, and C. J. Spanos, "Joint adversarial domain adaptation for resilient wifi-enabled device-free gesture recognition," in *The 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 202–207.

[110] H. Zou, J. Yang, Y. Zhou, L. Xie, and C. J. Spanos, "Robust wifi-enabled device-free gesture recognition via unsupervised adversarial domain adaptation," in *The 27th International Conference on Computer Communication and Networks (ICCCN)*, 2018, pp. 1–8.

[111] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *The 17th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys'19. ACM, 2019, pp. 313—-325.

[112] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from wifi: A siamese recurrent convolutional architecture," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 763–10 772, 2019.

[113] J. Wang, Q. Gao, X. Ma, Y. Zhao, and Y. Fang, "Learning to sense: Deep learning for wireless sensing with less training efforts," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 156–162, 2020.

[114] Z. Han, L. Guo, Z. Lu, X. Wen, and W. Zheng, "Deep adaptation networks based gesture recognition using commodity wifi," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–7.

[115] X. Ma, Y. Zhao, L. Zhang, Q. Gao, M. Pan, and J. Wang, "Practical device-free gesture recognition using wifi signals based on metalearning," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 228–237, 2020.

[116] J. Yang, H. Zou, S. Cao, Z. Chen, and L. Xie, "MobileDA: Toward edge-domain adaptation," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6909–6918, 2020.

[117] H. Kang, Q. Zhang, and Q. Huang, "Context-aware wireless based cross domain gesture recognition," *IEEE Internet of Things Journal*, pp. 1–1, 2021.

[118] K. Niu, F. Zhang, X. Wang, Q. Lv, H. Luo, and D. Zhang, "Understanding wifi signal frequency features for position-independent gesture sensing," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021.

[119] L. Zhang, Z. Wang, and L. Yang, "Commercial wi-fi based fall detection with environment influence mitigation," in *16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2019, pp. 1–9.

[120] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi, "A trained-once crowd counting method using differential wifi channel state information," in *2016 WPA*. ACM, 2016, pp. 37–42.

[121] S. D. Domenico, G. Pecoraro, E. Cianca, and M. D. Sanctis, "Trained-once device-free crowd counting and occupancy estimation using WiFi: A doppler spectrum based approach," in *2016 WiMob*, 2016, pp. 1–8.

[122] J. Wang, X. Chen, D. Fang, C. Q. Wu, T. Xing, and W. Nie, "Poster abstract: Implications of target diversity for organic device-

free localization," in *The 13th International Symposium on Information Processing in Sensor Networks*, ser. IPSN'14, 2014, p. 279–280.

[123] J. Wang, X. Chen, D. Fang, C. Q. Wu, Z. Yang, and T. Xing, "Transferring compressive-sensing-based device-free localization across target diversity," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2397–2409, 2015.

[124] X. Chen, H. Li, C. Zhou, X. Liu, D. Wu, and G. Dudek, "FiDo: Ubiquitous fine-grained WiFi-based localization for unlabelled users via domain adaptation," in *Proceedings of The Web Conference 2020*, ser. WWW'20, 2020, p. 23–33.

[125] R. Zhou, M. Tang, Z. Gong, and M. Hao, "Freetrack: Device-free human tracking with deep neural networks and particle filtering," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2990–3000, 2020.

[126] C. Xiao, D. Han, Y. Ma, and Z. Qin, "Csigan: Robust channel state information-based activity recognition with gans," *IEEE Internet of Things Journal*, vol. PP, pp. 1–1, August 2019.

[127] S. Arshad, C. Feng, R. Yu, and Y. Liu, "Leveraging transfer learning in multiple human activity recognition using wifi signal," in *IEEE 20th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2019, pp. 1–10.

[128] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-lstm for human activity recognition using wifi signal," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4628–4641, 2021.

[129] J. Shang and J. Wu, "A robust sign language recognition system with sparsely labeled instances using wi-fi signals," in *IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, October 2017, pp. 99–107.

[130] Jie Yin, Qiang Yang, and Lionel Ni, "Adaptive temporal radio maps for indoor location estimation," in *3rd IEEE International Conference on Pervasive Computing and Communications*, 2005, pp. 85–94.

[131] J. Yin, Q. Yang, and L. M. Ni, "Learning adaptive temporal radio maps for signal-strength-based location estimation," *IEEE Transactions on Mobile Computing*, vol. 7, no. 7, pp. 869–883, 2008.

[132] S. J. Pan, J. T. Kwok, Y. Qiang, and J. J. Pan, "Adaptive localization in a dynamic wifi environment through multi-view learning," in *The 22nd AAAI Conference on Artificial Intelligence*, July 2007.

[133] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *The 23rd AAAI Conference on Artificial Intelligence*, July 2008, pp. 677–682.

[134] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[135] Z. Sun, Y. Chen, J. Qi, and J. Liu, "Adaptive localization through transfer learning in indoor wi-fi environment," in *The 7th International Conference on Machine Learning and Applications*, 2008, pp. 331–336.

[136] V. W. Zheng, E. W. Xiang, Q. Yang, and D. Shen, "Transferring localization models over time," in *The 23rd AAAI Conference on Artificial Intelligence*, 2008, pp. 1421–1426.

[137] A. Goswami, L. E. Ortiz, and S. R. Das, "Wigem: A learning-based approach for indoor localization," in *The 7th Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT'11. Association for Computing Machinery, 2011.

[138] L. Chang, D. Fang, Z. Yang, X. Chen, J. Wang, W. Nie, and T. Xing, "EIL: An environment-independent device-free passive localization approach," in *The 13th International Symposium on Information Processing in Sensor Networks (IPSN)*. IEEE, 2014, pp. 291–292.

[139] C. Wu, Z. Yang, C. Xiao, C. Yang, Y. Liu, and M. Liu, "Static power of mobile devices: Self-updating radio maps for wireless indoor localization," in *IEEE Conference on Computer Communications (INFOCOM)*, 2015, pp. 2497–2505.

[140] Y. Gu, M. Chen, F. Ren, and J. Li, "HED: Handling environmental dynamics in indoor WiFi fingerprint localization," in *IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–6.

[141] Y. Xu, S. J. Pan, H. Xiong, Q. Wu, R. Luo, H. Min, and H. Song, "A unified framework for metric transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 6, pp. 1158–1171, 2017.

[142] H. Zou, Y. Zhou, H. Jiang, B. Huang, L. Xie, and C. Spanos, "Adaptive localization in dynamic indoor environments by transfer kernel learning," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[143] J. Jun, L. He, Y. Gu, W. Jiang, G. Kushwaha, V. A, L. Cheng, C. Liu, and T. Zhu, "Low-overhead wifi fingerprinting," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 590–603, March 2018.

[144] X. Rao and Z. Li, "MSDFL: a robust minimal hardware low-cost device-free WLAN localization system," *Neural Computing and Applications*, pp. 1–18, 2019.

[145] X. Rao, Z. Li, and Y. Yang, "Device-free passive wireless localization system with transfer deep learning method," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, 10 2020.

[146] X. Rao, Z. Li, Y. Yang, and S. Wang, "DFPhaseFL: a robust device-free passive fingerprinting wireless localization system using csi phase information," *Neural Computing and Applications*, vol. 32, 09 2020.

[147] J. Xue, J. Liu, M. Sheng, Y. Shi, and J. Li, "A wifi fingerprint based high-adaptability indoor localization via machine learning," *China Communications*, vol. 17, no. 7, pp. 247–259, 2020.

[148] Y. Kim, H. Shin, and H. Cha, "Smartphone-based wi-fi pedestrian-tracking system tolerating the rss variance problem," in *IEEE International Conference on Pervasive Computing and Communications*, 2012, pp. 11–19.

[149] J. Yang, H. Zou, H. Jiang, and L. Xie, "Fine-grained adaptive location-independent activity recognition using commodity wifi," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1–6.

[150] X. Ding, T. Jiang, Y. Li, W. Xue, and Y. Zhong, "Device-free location-independent human activity recognition using transfer learning based on cnn," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[151] Y. Li, T. Jiang, X. Ding, and Y. Wang, "Location-free csi based activity recognition with angle difference of arrival," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[152] Y. Lu, S. Lv, and X. Wang, "Towards location independent gesture recognition with commodity wifi devices," *Electronics*, vol. 8, no. 10, p. 1069, 2019.

[153] J. Jung, J. Kim, and K. Toh, "Transfer learning of wi-fi handwritten signature signals for identity verification based on the kernel and the range space projection," in *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 2586–2590.

[154] P. Sharma, D. Chakraborty, N. Banerjee, D. Banerjee, S. K. Agarwal, and S. Mittal, "KARMA: Improving wifi-based indoor localization with dynamic causality calibration," in *The 11th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2014, pp. 90–98.

[155] L. Zhang, C. Wang, M. Ma, and D. Zhang, "Widigr: Direction-independent gait recognition system using commercial wi-fi devices," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1178–1191, 2020.

[156] Q. Y. Vincent Wenchen Zheng, Sinno Jialin Pan and J. J. Pan, "Transferring multi-device localization models using latent multi-task learning," in *The 23rd AAAI Conference on Artificial Intelligence*, 2008, pp. 1427–1432.

[157] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, "WiDeep: WiFi-based accurate and robust indoor localization system using deep learning," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2019, pp. 1–10.

[158] S. He, W. Lin, and S.-H. G. Chan, "Indoor localization and automatic fingerprint update with altered ap signals," *IEEE Transactions on Mobile Computing*, vol. 16, no. 7, pp. 1897–1910, 2017.

[159] Z. Gao, Y. Gao, S. Wang, D. Li, and Y. Xu, "Crisloc: Reconstructable csi fingerprinting for indoor smartphone localization," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[160] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI'13)*. Lombard, IL: USENIX, 2013, pp. 71–84.

[161] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 269–282, aug 2015.

[162] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI'16)*. Santa Clara, CA: USENIX Association, March 2016, pp. 165–178.

[163] W. Gong and J. Liu, "Sifi: Pushing the limit of time-based wifi localization using a single commodity access point," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, March 2018.

[164] Y. Wan, X. Tong, Y. Peng, and X. Tian, "Calibration-free localization based on fingerprinting of channel state information," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, 2019, pp. 394–398.

[165] X. Tong, Y. Wan, Q. Li, X. Tian, and X. Wang, "Csi fingerprinting localization with low human efforts," *IEEE/ACM Transactions on Networking*, pp. 1–14, 2020.

[166] L. Zhang, Q. Gao, X. Ma, J. Wang, T. Yang, and H. Wang, "DeFi: Robust training-free device-free wireless localization with wifi," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8822–8831, Sep. 2018.

[167] K. Joshi, D. Bharadia, M. Kotaru, and S. Katti, "Wideo: Fine-grained device-free motion tracing using rf backscatter," in *The 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI'15)*, 2015.

[168] Z. Han, Z. Lu, X. Wen, J. Zhao, L. Guo, and Y. Liu, "In-air handwriting by passive gesture tracking using commodity wifi," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2652–2656, 2020.

[169] D. Wu, R. Gao, Y. Zeng, J. Liu, L. Wang, T. Gu, and D. Zhang, "Fingerdraw: Sub-wavelength level finger motion tracking with wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 4, no. 1, pp. 1–27, March 2020.

[170] R. Gao, H. Wang, D. Wu, K. Niu, E. Yi, and D. Zhang, "A model based decimeter-scale device-free localization system using COTS Wi-Fi devices," in *2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2017 ACM International Symposium on Wearable Computers*. ACM, 2017, pp. 241–244.

[171] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity Wi-Fi for interactive exergames," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI'17. ACM, 2017, pp. 1961–1972.

[172] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, "Enabling identification and behavioral sensing in homes using radio reflections," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13.