WiRITE: General and Practical Wi-Fi based Hand-Writing Recognition

Yanbo Zhang, Weiping Sun and Mo Li, *Fellow, IEEE*

Abstract—Device-free hand-writing systems identify the content that a user writes by hand movement in the air, thus providing an intuitive human computer interface. In this paper, we propose WiRITE, a Wi-Fi hand-writing recognition system built with commodity Wi-Fi APs. Unlike most existing machine learning based hand-writing recognition systems, which are often subject to severe limitations in generality, e.g., high training overhead when adapted across hand-writing alphabets, environments, and users, WiRITE is designed with unique consideration of its generality when applied to practice—being application-transferable, environment-agnostic, and user-independent. With little training overhead, WiRITE behaves inclusively to different users, environments, and applications, stemming from a comprehensive design of signal processing that is built into its core machine learning model. Extensive evaluation is conducted with five users for three applications, i.e., recognizing Digits, English letters, and Chinese characters, in realistic office environment. The experiment results demonstrate that WiRITE provides at least 0.9 accuracy in various combinations of users and applications with 0.93 accuracy in average.

Index Terms—Wi-Fi sensing, gesture recognition, human computer interaction, CSI sanitization, antenna measurement

1 Introduction

RF-based passive sensing has been widely studied for applications including localization and tracking [1]–[3], activity and gesture recognition [4]–[7], human biometric identification [8]–[10], to name a few. Among those is device-free hand-writing, for which passive sensing is supposed to identify the content that a target user writes in the air such as digits and letters. Users are not required to wear extra hardware and natural human-computer interface is provided compared with traditional mechanisms like a keyboard or an e-pencil, particularly when serving emerging VR/AR applications.

Wi-Fi based hand-writing has attracted enormous attention compared with alternative technologies like UWB [11], [12], FMCW [13]–[15], acoustic [16], [17], etc., due mainly to its ubiquitous infrastructure availability that can potentially ease commercialization. Most Wi-Fi based hand-writing systems, so far, are constructed based on channel state information (CSI), which is accessible from commercial Wi-Fi chipsets [18], [19]. CSI essentially carries information about the effects of the wireless medium (multipath reflection, diffraction, scattering, etc.) to a Wi-Fi signal manifested as the attenuation of signal amplitude and the rotation of signal phase on spectrum — which can be used to infer hand-writing gestures.

Direct hand-writing estimation concerns complicated signal processing including signal path separation, denoising, channel parameterization, etc., which is greatly limited by the CSI resolution and precision that commodity Wi-Fi hardware can offer. Many recent works study the adoption of machine learning (ML) based approaches [20]–[22] which take CSI data as input to a classification model for identifying the associated hand-writing gestures. The ML-based approaches largely alleviate the requirement on signal resolution and precision, since it is no longer needed to resolve the multi-path signals but only to ensure recognizable CSI patterns for identifying hand-writing gestures.

ML-based approaches, however, face a major challenge regarding the generality of the trained ML model, which is rooted from the disparity between the distributions of training data and testing data. For example, the hand-writing habits of specific users and the ambient environment during system operation are often different from the conditions where the ML model was trained. That may likely result in different CSI patterns for the same class of hand-writing gestures between training and testing, thus significantly degrading the accuracy. Suffering from the above issue, application of most existing approaches entails extremely high training efforts that include training a general model for all possible combinations of user-environment factors. That incurs huge amount of human efforts in tedious data collection and labeling process for massive amount of samples, often over thousands [4]. Making it worse, whenever the ML model's functions are extended, e.g., enlarging supportable hand-writing alphabets from English letters to include Chinese characters, the excessive human efforts are repetitively incurred, which makes such systems impractical.

In this paper, we propose WiRITE, a Wi-Fi hand-writing recognition system, which addresses the generality issue in ML-based solution. WiRITE adopts a similar concept to domain adaptation, one kind of transfer learning, that transforms Wi-Fi CSI samples, i.e., target domain features, to images containing hand movement trajectories, i.e., source

- *Y. Zhang and M. Li are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798. E-mail: yanbo001@e.ntu.edu.sg, limo@ntu.edu.sg*
- *W. Sun is with 5G/6G Standard Lab, Samsung Research, Seoul, South Korea. E-mail: wp.son@samsung.com*

Manuscript received January 18, 2022; firstly revised September 12, 2022; secondly revised May 15, 2023; accepted April 6, 2023.
(Yanbo Zhang and Weiping Sun are co-first authors; Corresponding author: Weiping Sun)

domain features, so that existing human efforts invested in training image based hand-writing models can be reused in this different domain. That is, WiRITE trains a ML model by exploiting the readily available standard "vanilla" databases of handwritten images. The generality of WiRITE can thus be ensured, considering the fact that these public databases already cover massive amount of training samples of different writing habits and across different target alphabets. For example, 814,255 samples of 52 English letters collected from 3,600 writers are included in EMNIST database [23], and 3,721,874 samples of 7,185 Chinese characters from 1,020 writers are included in CASIA database [24].

To enable the domain adaptation, WiRITE entails a comprehensive design of signal processing and builds that into its core ML model. WiRITE is built with two commodity Wi-Fi APs, one for transmitting and the other for receiving, which operate in their default configurations, i.e., 3x3 MIMO on a 20 MHz channel. The signal processing module in WiRITE targets at eliminating random hardware noise, extracting the path length change of the reflected path, estimating the hand movement trajectory, and translating the trajectory to an image, which collectively project the messy raw CSI samples (target domain) to handwritten images (source domain).

With the above design considerations, WiRITE provides the generality in three-folds: (i) Application-transferrable, i.e., the WiRITE core model can easily be adapted to address new hand-writing alphabets, e.g., Digits, English letters, and Chinese characters with little training overhead to it; (ii) Environment-agnostic, i.e., WiRITE minimizes impacts from hardware and environment noises, and as a result is oblivious to their dynamics when in use; and (iii) User-independent, i.e., WiRITE is robust to the singularities in Wi-Fi signal reflection and scattering as well as the variety in hand-writing habits from different users. The evaluation results conducted with five users for three applications, i.e., Digits, English letters, and Chinese characters, in a typical office environment, demonstrate that WiRITE provides at least 0.9 accuracy in all the combinations of the users and applications with 0.93 accuracy in average.

In summary, we claim the following contributions.

- WiRITE is the first Wi-Fi hand-writing recognition system that advances the state-of-the-art by its generality in users, environments, and applications.
- We propose novel techniques including self-cancellation of CSI noise, direct path nulling, dynamic component extraction and sample augmentation that collaboratively effect to provide the high accuracy and better generality.
- We implement a real-time system with commodity Wi-Fi APs, with which the performance of WiRITE can be corroborated by real world experiments.
- We corroborate WiRITE performance by conducting extensive real world experiments.

The rest of this paper is organized as follows: in Sections 2, 3, and 4, we elaborate on the key design and techniques of WiRITE in supporting generality of applications, environments, and users, respectively. In Section 6, the performance of WiRITE is evaluated with extensive real world experiments. In Section 8, we discuss the related work, and finally, we conclude the paper in Section 9.

2 WiRITE Design

The application-transferability is a key design goal of WiRITE. Different hand-writing applications are well modularized in WiRITE design, whereby transferring to different applications, to WiRITE, is no more than changing a URL linking to corresponding public handwritten image database. In current version, WiRITE can freely switch among Digit, English letter, and Chinese character applications as illustrated in Fig. 1.

The key enabler is the domain adaptation design that projects messy CSI samples (target domain) to handwritten images (source domain), which enables mode switching between different applications by only plugging different online public databases used in training the handwritten image classification model. WiRITE possesses two workflows, i.e., offline and online, which collectively materialize WiRITE. Online workflow mainly takes the task of translation from CSI to handwritten image followed by the classification of the image. Offline workflow concentrates on training general handwritten image classification model to get prepared for online inputs.

2.1 Online

Online workflow involves the process of translating CSI samples collected when a WiRITE user performs hand-writing gesture, to a recognizable handwritten trajectory image, as well as classifying the handwritten image to a final recognition class, i.e., a digit, an English letter, or a Chinese character, in different applications, respectively.

Domain adaptation. As Fig. 1 indicates, the raw CSI samples are processed by three main functional blocks to be translated as a handwritten trajectory image. i) The first block is to extract the *dynamic path information* from the raw CSI samples, which is the information (phase and amplitude) of the reflected path off from the hand of a WiRITE user. A novel *CSI sanitization* function is developed to effectively eliminate the noises caused by side-effects of the hardware of commodity Wi-Fi APs, and a novel path separation method called *dynamic component extraction* is applied to suppress interferences from environment. ii) Given the extracted dynamic component information, the trajectory of the WiRITE user's hand movement is estimated by trajectory estimator. WiRITE targets at correct classification of the hand-writing, and thus is only interested in identifying the movement pattern of the user's hand rather than its precise locations. All possible hand trajectories starting from all the possible locations are derived and input to ML model for a comprehensive final decision. The rationale of the design comes from a well-known feature of image classification, i.e., correct classification is feasible to a scaled and rotated version of an image sharing similar features with the original one. iii) The derived trajectories are transformed to images conforming to the acceptable format of handwritten image classification ML model.

Image classification. The output of domain adaptation, i.e, a set of scaled and rotated handwritten trajectory images, are recognized by a pre-trained image classification ML model. Instead of taking the final classification results, WiRITE makes use of the last softmax layer of the ML model, which

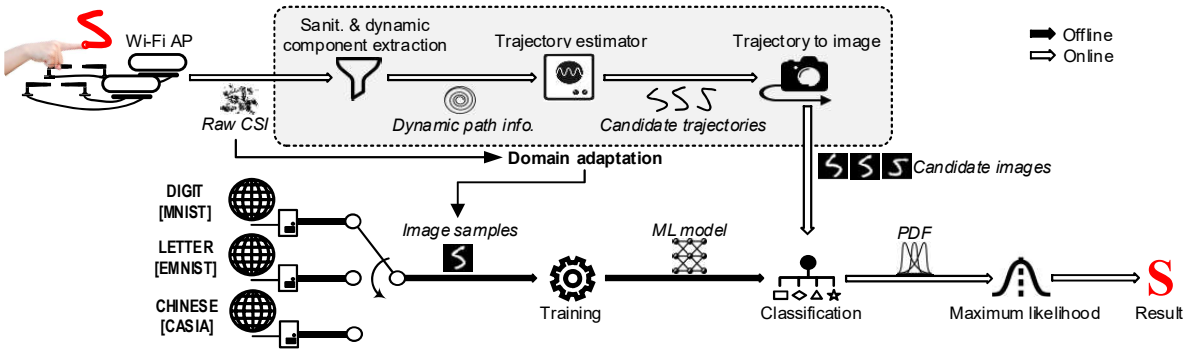


Fig. 1. Overview of WiRITE design.

TABLE 1
Information of three training sets.

	Digit	Letter	Chinese
# training sample	60,000	266,924	40,000
# test sample	10,000	58,796	8,020
# class	10	37	200
# writers	250	500	200
Image size	28x28	28x28	64x64
Database	MNIST [25]	EMNIST [23]	CASIA [24]

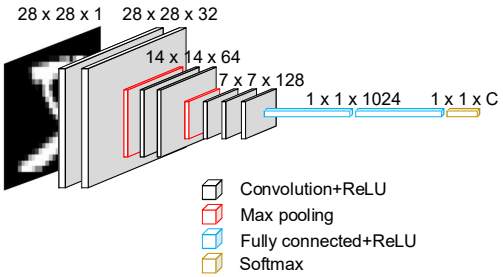


Fig. 2. VGG-10 model adopted in WiRITE.

produces probability density function (PDF) with respect to candidate hand-writing classes. WiRITE exploits this feature in the image classification, by synthesizing the PDFs of all the trajectory images, and makes the final decision based on the synthesized PDF. We call it ML-based maximum likelihood estimation.

2.2 Offline

The task of offline workflow is to provide well-trained image classification ML models to support online classification, where the ML models for different applications are trained separately.

Public image database. There are multiple public databases providing adequate samples of handwritten images for various languages/alphabets. WiRITE benefits from these databases, especially, in i) the generality of the sample images ascribed to the diversified writers, and ii) free training set, which ease application transferability. In this paper, we consider three types of applications, i.e., Digits, English letters, and Chinese characters. The three corresponding public databases are MNIST [25], EMNIST [23], and CASIA [24], respectively. The detailed information regarding the training sets is summarized in Table 1.

ML model architecture. WiRITE trains three different convolutional neural network (CNN) models to support the three applications, respectively. Specifically, the models for Digits and English letters have a common architecture due to the same format of the training samples provided by the public databases, i.e., 28x28 pixel image. This model is a simplified version of VGG-16 [26], a well-known CNN model for image classification, and hence, we call it VGG-10, because it consists of 10 layers (7 convolutional layers and 3 fully-connected layers) as shown in Fig. 2[1], where C indicates number of classes. All the layers use ReLU activation functions as in the original VGG-16. For Chinese characters, we adopt the architecture of M5 model as reported in [27] which is similar in its design nature to the VGG models.

In the following, we will outline the most telling features of WiRITE in light of its design goal in environment agnosticism and user independence.

3 Environment Agnosticism

WiRITE is resilient to environmental variations manifested as CSI hardware noises, and ambient interferences from surroundings. We start with an elaboration on the CSI sanitization technique of WiRITE, followed by an intelligent antenna arrangement scheme capable of mitigating the ambient interferences.

3.1 CSI Sanitization

CSI is complex numbered frequency response of the signal that expresses the signal's amplitude attenuation and phase rotation subject to transmission, reception, and the propagation residing between them. However, besides the channel information, raw CSI is entangled with amplitude and phase noise that are resulted from different sources including the hardware and signal processing modules of both transmitter and receiver [18], [28]–[30]. Specifically, we formulate raw CSI sample H' as follows,

$$H'(k, p, r_n, t_m) = H(k, p, r_n, t_m) \cdot E(k, p, r_n, t_m) \quad (1)$$

1. The simplification is introduced due to the reduction of the hidden layer depth, caused by the input image size, i.e., $224 \times 224 \times 3$ of VGG-16 v.s. $28 \times 28 \times 1$ of WiRITE.

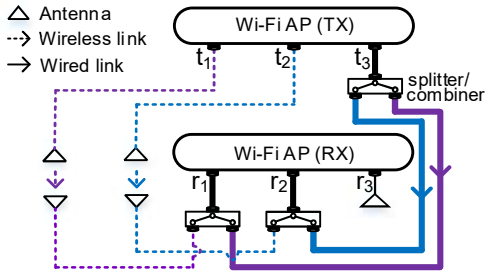
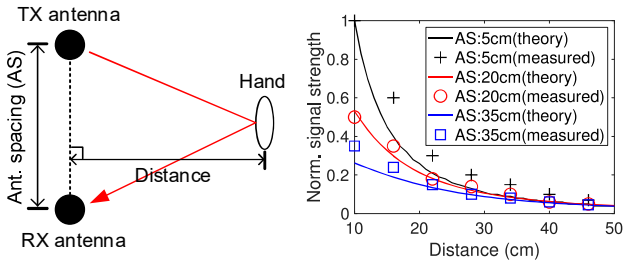


Fig. 3. WiRITE hardware architecture.



(a) Experiment environment. (b) Signal strength of reflected path.

Fig. 4. Impact of antenna spacing on signal strength of reflected paths.

The raw CSI derived from a certain packet (p) represents the channel rotation measured from three dimensions – different subcarrier (k), different Tx antenna (t_m) and different Rx antenna (r_n). The raw CSI is related to the physical multipath channel (H) by the noise factor (E), which can be represented in detail as follows,

$$E(k,p,r_n,t_m) = \varepsilon_1(p,r_n)\, e^{\varepsilon_2(k,p)+\varepsilon_3(p)+\varepsilon_4(r_n,t_m)} \quad (2)$$

where ε_i denotes the ith type of noise classified by its varying dimension. Specifically,

- ε_1 represents the amplitude noise that varies over different packets (p) and Rx antennas (r_n). Such noise is caused by the imperfection of hardware amplifier and automatic gain control (AGC) function.
- ε_2 adds phase noise to H. This type of phase noise is resulted from the mixed effect of packet detection delay (PDD), sampling frequency offset (SFO), and sampling time offset (STO), and varies over different subcarriers (k) and packets (p).
- ε_3 represents another type of phase noise that varies over different packets (p). This type of noise is constant at different subcarriers and is caused by carrier frequency offset (CFO).
- ε_4 represents the third type of phase noise. This noise remains constant over time but could vary with different Tx antennas (t_m) and Rx antennas (r_n) because it is caused by the initial phase offset of transmitter and receiver phase-locked-loops (PLLs).

The noise is difficult to resolve because it comes from various sources and varies with different dimensions. Previous CSI calibration approach [18], [31] removes the phase difference between different measurements but cannot eliminate the hardware noise. The CSI-ratio based de-noising method [32], [33] derives the difference bewteen two wireless links which can hardly be modeled as an deterministic function of channel dynamics. The other filtering-based CSI sanitization method [34], [35] assumes distinctive frequency variation pattern which however is not always the case.

We propose a novel method that achieves self-cancellation of the noise, which is facilitated by a carefully designed hardware architecture as shown in Fig. 3. We introduce three RF splitters and several coaxial cables (RG50) to construct two wired links, i.e., $L(r_1,t_3)$ and $L(r_2,t_3)$, denoted in purple and blue solid lines, respectively. These two *reference links* are used to sanitize links $L(r_1,t_1)$ (dotted purple line) and $L(r_2,t_2)$ (dotted blue line), respectively. That is, we sacrifice two links to make their channel wired, and hence, constant over time, so their CSI variations caused only by the noise, can be used as reference to sanitize the other two wireless links that carry both noises and useful information on wireless channel variations. Mathematically, the ratio between CSI samples of $L(r_1,t_1)$ and $L(r_1,t_3)$ of the same packet at the same subcarrier can be expressed as

$$\frac{H'(r_1,t_1)}{H'(r_1,t_3)} = e^{\varepsilon_4(r_1,t_1)-\varepsilon_4(r_1,t_3)} \frac{H(r_1,t_1)}{H(r_1,t_3)}, \quad (3)$$

since the common ε_1, ε_2 and ε_3 are eliminated by division operation. Since initial PLL offset, ε_4, and the reference link CSI, $H(r_1,t_3)$, are constant in time, $\frac{e^{\varepsilon_4(r_1,t_1)-\varepsilon_4(r_1,t_3)}}{H(r_1,t_3)}$ is also constant in time. Thus, the CSI ratio can be regarded as ground truth CSI of the wireless link, $H(r_1,t_1)$, multiplied by a time-invariant coefficient. To further derive the ground truth CSI, ε_4 and $H(r_1,t_3)$ should be compensated (The splitters and cables used for hardware connection may introduce additional phase offset, but is also time-invariant). The time-invariant coefficient can be measured manually and calibrated accordingly, based on the calibration methods reported previously, e.g., [36], [37]. However, since the time-invariant complex coefficient does not affect the variation pattern of the ground truth CSI — it only introduces constant scaling and rotation to each CSI sample — which WiRITE exploits in hand movement trajectory estimation, the tedious calibration step can be skipped to save manual efforts. Last, we calculate all the CSI ratios according to Eq. 3 with respect to all the subcarriers, and take the average of them as the final sanitized CSI. Note that the averaging operation is applied to eliminate white noise imposed at all subcarriers, which is not explicitly denoted in Eq. 3 for simplicity. In short, for each packet reception, we have two sanitized CSI samples, i.e., H_1^S and H_2^S, obtained by

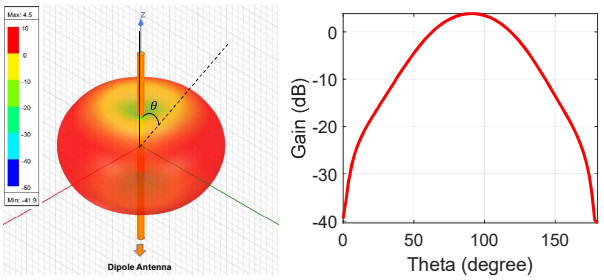
$$H_n^S = \frac{1}{K}\sum_{k=1}^{K} \frac{H'(k,r_n,t_n)}{H'(k,r_n,t_3)}, \quad (4)$$

where n can be 1 or 2. Unless stated otherwise, CSI samples mentioned in the following sections are sanitized CSI samples.

3.2 Antenna Arrangement

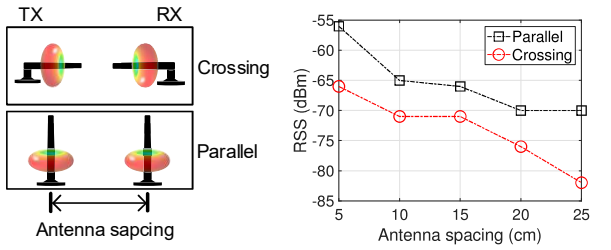
WiRITE is empowered by a carefully designed antenna arrangement, which improves the resilience to ambient interferences, caused by nearby moving objects or human. We focus on two issues: i) The spacing between Tx and Rx antennas, and ii) their orientations.

Antenna spacing. The spacing between Tx and Rx antennas affects the *signal to interference ratio* (SIR); *signal* indicates



(a) The radiation pattern of dipole antenna. (b) The variation of antenna gain at different elevation angle.

Fig. 5. The radiation characteristic of dipole antenna.



(a) Antenna direction. (b) Strength of direct path signal.

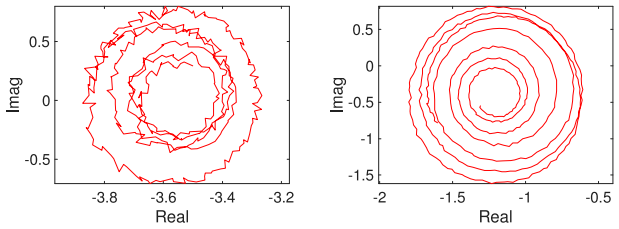
Fig. 6. Impact of antenna orientation on direct path signal strength.

the reflected signal from target user's hand, and *interference* indicates the reflected signal from other moving objects or human. *In WiRITE, we deliberately minimize the spacing between Tx and Rx antennas to maximize the SIR.* That is, consider the free space path loss model, where the received signal strength is manifested as a logarithmic decay (in dB scale) with the increase of the path length. Suppose the user's hand is closer to the antenna pair than the interferer. If we reduce the antenna spacing, the strengths of both signal and interference are shifted to the left along a logarithmic decay curve with a fixed path length offset, resulting in increased gap between them, i.e., SIR.

To verify this, as Fig. 4(a) illustrates, we measure the strength of the reflected path (highlighted in red),[2] when moving a hand away from the antenna pair. We conduct the measurements with different antenna spacing, i.e., 5cm, 20cm, and 35cm, respectively, and compare the measurement results with the free space path loss model [38]. Fig. 4(b) clearly indicates the logarithmic decrease of the normalized signal strength of the reflected path when the hand moves away from the antenna pair. We observe that, with the decrease of the antenna spacing, the curve becomes steeper, which suggests that the small antenna spacing is more favorable for increasing SIR and hence the resilience of the system to the ambient interference.

Antenna orientation. *Flash effect* may arise, if the Tx and Rx antennas are located too close [39]. Typically, the direct path signal is much stronger than the other indirect path signals, and might overwhelm the receiver's analog to digital converter (ADC), preventing it from detecting the minute variations of the indirect signals, which is the major signal

2. The reflected path signal strength is obtained by dynamic component extraction method described in Section 4.1



(a) Parallel arrangement. (b) Crossing arrangement.

Fig. 7. Direct path suppression effect of crossing antenna arrangement.

source of WiRITE. Signal nulling can alleviate flash effect using MIMO [39], [40], while, it entails modifications of the transmitter's precoding matrix, which is not feasible in commodity Wi-Fi devices. WiRITE adopts another approach to suppress the direct path signal. The most widely used class of antenna in Wi-Fi is dipole antenna which WiRITE uses as well. We conducted a simulation to investigate the radiation pattern of dipole antennas. The result is detailed in Fig. 5. As we see from Fig 5(a), the radiation pattern resembles the donut shape where the minimum gain appears at the center of the radiation pattern. Figure 5(b) details the antenna gain at different elevation angle (YoZ plane). We observe that the antenna gain reduces quickly when θ deviates from 90 degrees. For example, when θ equals to 30 degree, the antenna gain is -15dB, and thus the direct path is attenuated by near 20dB when compared to the maximum radiation direction. *We deploy the two dipole antennas crossing each other, adjusting the doughnut-shape radiation pattern, such that the null directions of the two antennas facing to each other, thus suppressing the direct path signal.* We call such deployment 'crossing'. Fig. 6(a) illustrates the 'crossing' antenna arrangement compared with the typical 'parallel' arrangement. We measure the strength of the direct path signal using the two deployments, respectively. Fig. 6(b) shows the measurement results. We observe that irrespective of antenna spacing, 'crossing' arrangement always yields much weaker received signal strength (RSS) than 'parallel'. In particular, for the smallest antenna spacing (5cm), 'crossing' yields about 10 dB signal suppressing, which is comparable to that of MIMO beam nulling reported in [40]. Figs. 7(a) and 7(b) compare the distributions of CSI samples collected with the 5cm antenna spacing when a hand moves away from the antenna pair. It is clear to observe that, with 'crossing', the CSI samples exhibit clear spiral shape conforming to the theoretical expectation, while 'parallel' exhibits a noisy pattern as a result of the flash effect.

Fig. 8 illustrates the final antenna deployment of WiRITE, where the left and right links are composed of 5cm spaced Tx and Rx antennas following the 'crossing' arrangement. The two links are spaced about 50cm, which is determined empirically for best system performance (We provide more detailed discussion on the impact of varied link spacing in Sec. 7). Possible using scenario of such deployment include smart table/desk design where the antennas can be integrated on the surface as PCB patch.

4 User Independence

WiRITE reacts to different users inclusively regardless of

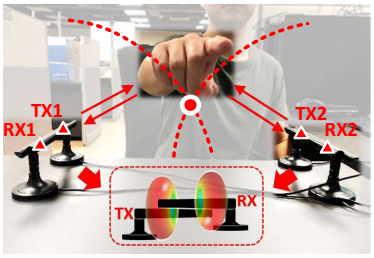
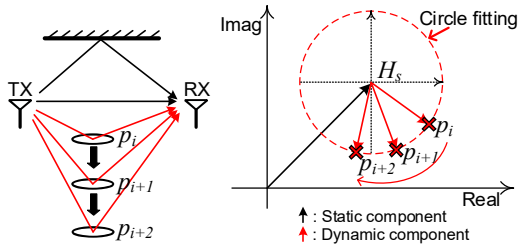


Fig. 8. WiRITE antenna arrangement.



(a) Hand movement. (b) CSI samples in complex plane.

Fig. 9. Illustration of dynamic component rotation caused by hand movement.

their writing habits, hand shapes, and surfaces for Wi-Fi signal reflection. This is enabled by a domain adaptation scheme that projects CSI samples to handwritten trajectory image, followed by generalized ML-based image classification.

4.1 CSIs to Trajectory Images

Each CSI sample may include i) direct path signal, ii) reflected path signal from static objects such as walls, and iii) reflected path signal caused by user's hand. We refer to the first two types of signals as *static* components, and the third type as *dynamic* component.

In the case where there are N paths in total, among which the Nth path is the dynamic component, we can write the CSI as

$$H(p) = \sum_{i=1}^{N-1} a_i e^{\frac{-j2\pi l_i}{\lambda}} + a_N(p)\, e^{\frac{-j2\pi l_N(p)}{\lambda}} \quad (5)$$

where p, λ, a_i and l_i indicate packet index, wavelength, ith path strength, and ith path length, respectively. Because the static components are constant over time, they can be treated as a whole. Accordingly, Eq. 5 can be simplified as the sum of two components, i.e., static and dynamic, expressed as

$$H(p) = a_s e^{\theta_s} + a_d(p)\, e^{\frac{-j2\pi l_d(p)}{\lambda}} = H_s + H_d(p) \quad (6)$$

where we use s and d to indicate static and dynamic components, respectively. Our goal is to extract the $H_d(p)$ for each packet p, since its phase is related to the hand movement.

Consider a case where the dynamic component is reflected by a hand moving away from Tx-Rx antennas as shown in Fig. 9(a). Along with the movement, the length of the reflected path (red arrow) increases, which results in clock-wise phase rotation of the dynamic component. Fig. 9(b) illustrates the CSI samples of packets p_i, p_{i+1}, and p_{i+2}, in complex plane, where the black and red vectors represent H_s and $H_d(p)$, respectively.

Dynamic component extraction. The dynamic component, $H_d(p)$, is extracted based on the above observation: *Within a short period of time, the CSI samples will fall on a circle centered at the static component, H_s, given that the amplitude of the dynamic component is constant.* If we apply *circle fitting* (CF) on several consecutive CSI samples, the center of the fitted circle becomes the estimate of H_s, and the difference between CSI sample, $H(p)$, and the circle center is the estimated dynamic component, $H_d(p)$. That is,

$$\widetilde{H}_d(p) = H(p) - \widetilde{H}_s \quad (7)$$

where $\widetilde{H}_d(p)$ and $\widetilde{H}_s$ are the estimates of $H_d(p)$ and H_s, respectively. $\widetilde{H}_s$ can be obtained by applying CF, i.e., $\mathrm{CF}(\mathbb{S}_H)$, where $\mathbb{S}_H$ indicates the set of CSI samples used in CF.

In practice, however, i) the amplitude of the dynamic component is not constant during a longer time period. It decreases with the increase of the path length, and ii) the static component slowly changes due to the ambient environmental dynamics, e.g., someone else walking around. To deal with the above issue, we adaptively perform CF multiple times to multiple different segments of the CSI samples, by applying *CF windowing*, which continuously identifies the proper segment for CF, referred to as *CF window*, when scanning the entire CSI samples sequentially. Each CF window is selected to satisfy the necessary condition for CF, i.e., the static component and the amplitude of the dynamic component do not change much within a CF window.

Fitted signal to noise ratio. To quantify the quality of a CF window, we define a metric called *fitted signal to noise ratio* ($fSNR$), which is defined as

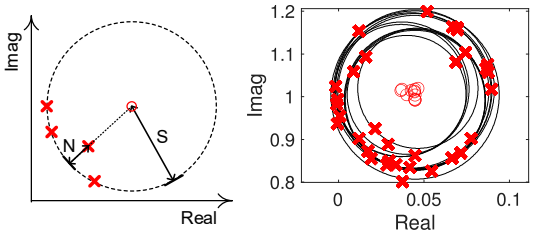
$$fSNR = \frac{rad^2}{MSE}, \quad (8)$$

where rad indicates the radius of the fitted circle, indicating the signal strength of the dynamic component, and MSE indicates the mean squared error, i.e.,

$$MSE = \frac{1}{W}\sum_{i=1}^{W} |rad - |H(i) - \mathrm{CF}(\mathbb{S}_H)||^2, \quad (9)$$

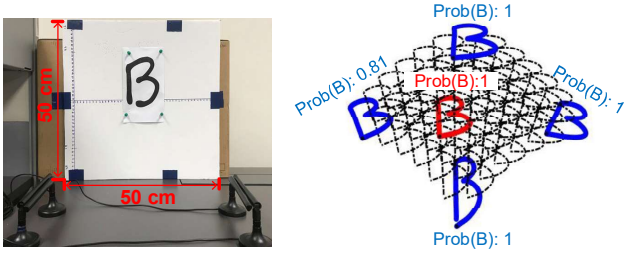
given that a circle is fitted to W consecutive CSI samples. If all the CSI samples perfectly fall on a fitted circle, $fSNR$ will become positive infinity, while if the CSI samples deviate from the circle, $fSNR$ will decrease due to the increase of MSE. Fig. 10(a) illustrates the concept of $fSNR$ with signal (S) and noise (N) components in its definition.

CF windowing. Suppose we have N CSI samples. From the first to $(N-\delta)$th CSI samples, where δ is the smallest size of CF window, we treat each of those samples as the start of a new CF window. We try to identify the end of each CF window, which would be the earliest CSI sample that makes the $fSNR$ greater than a threshold γ. For example, for the first CF window, we apply circle fitting to the first δ CSI samples and see whether the $fSNR{>}\gamma$. If so, the first CF window is determined; otherwise, we progressively increment CF window and check the $fSNR$, until the $fSNR$ exceeds γ, or the CF window exceeds its maximum, ω, or no CSI sample remains, whichever comes first. In the end, all possible CF windows across the whole CSI samples are identified. We perform CF to each identified CF window, and extract the dynamic components of CSI samples. Note



(a) $fSNR$ illustration. (b) CF windowing illustration.

Fig. 10. Illustration of dynamic path extraction by circle fitting.



(a) Detection region. (b) Candidate trajectories.

Fig. 11. Multiple candidate trajectories.

that CF windows are very likely to overlap, that is to say a CSI sample could be the member of multiple CF windows that yield different results of dynamic component. This issue is resolved by letting each CSI sample select the CF window yielding the greatest $fSNR$. CSI samples which do not belong to any CF windows are discarded. Fig. 10(b) illustrates all the identified CF windows denoted as black circles with a set of CSI samples collected when a WiRITE user was writing 'B' in the air. The small red circle indicates the center of each fitted circle.

Trajectory estimation. The basic idea of trajectory estimation is to use the ranging results provided by the two links to intersect the hand reflecting point, as illustrated with Fig. 8. If the initial location of the moving hand is given, we can continuously track the hand movement trajectory based on the path length changes derived from the phase difference of consecutive dynamic components. Specifically, Δ_{l_d}, the path length change from packet p_i to packet p_{i+1}, is obtained by

$$\Delta_{l_d}(p_i) = \frac{\lambda}{2\pi}\left(\angle\widetilde{H}_d(p_i) - \angle\widetilde{H}_d(p_{i+1})\right), \qquad (10)$$

derived from Eq. 6. For Eq. 10 to be effective, we need to ensure that the phase difference should be less than π, as otherwise, the ambiguity will occur among multiple candidates, i.e., $\Delta_\theta \pm \alpha 2\pi$. We ensure this by applying adequately high Wi-Fi transmission rate for sampling, i.e., 333 Hz, which can catch up to 45m/s movement speed.

While the estimation method is able to catch the hand dynamics, without the initial location of the hand, the ground truth trajectory cannot be derived. An inaccurate initial location will result in a scaled and rotated trajectory image, which however may not affect the ML-based image recognition as long as the scaling and rotating factors are bounded. Therefore, WiRITE derives trajectories starting from all possible initial locations in a predefined detection region, which is a 50cm by 50cm rectangle as shown in Fig. 11(a). Fig. 11(b) illustrates the multiple potential trajectory candidates, where the ground truth trajectory is highlighted in red. When all potential variations of the trajectories are input to the ML model, we observe that the ML model performs very well even for the trajectories being scaled and rotated — all the trajectories are classified as the ground truth 'B' with high probabilities as shown in Fig. 11(b). WiRITE ML-based maximum likelihood estimator thus comprehensively makes the final decision considering all the possible trajectories.

Trajectory to image. The hand movement trajectories may have different sizes caused by different writing habits of different users. Trajectory to image function normalizes the trajectories to images with unified size and format. At the same time, trajectories are transformed into images with the same format as that of the training samples as indicated in Table 1. Specifically, each trajectory is translated into black and white (bilevel) image by setting each location as 1, and 0 otherwise. The bilevel image is then size-normalized to fit in a pixel box with size of 20x20 (digit and English letter) or 56x56 (Chinese character) while preserving their width and height aspect ratio. It is further translated and padded with zero pixels to conform to the format of the training samples. The pixels of the image are finally normalized to ensure the maximum and minimum values being 1 and 0, respectively.

4.2 Image Classification

Public database. As formerly stated, three ML models for recognizing Digit, English letter, Chinese character, are readily trained independently. The training sets of the three models are provided by MNIST [25], EMNIST [23], and CASIA [24] public databases, respectively. The detailed information of the training sets is summarized in Table 1. Note that the three vanilla databases already cover vast number of samples and diversified writers, which benefit the generality of the trained ML models.

Sample augmentation. However, with only vanilla training samples, we face two challenges: i) WiRITE derives the final trajectories with multiple scaled and rotated versions of ground truth trajectory image. We need a tuned model to recognize the scaled and rotated versions. ii) Most of the vanilla images are collected from multi-stroke writing on papers, while the hand trajectory images derived from Wi-Fi sensing are from unistroke hand motion in the air. To address above issues, we introduce *sample augmentation* to vanilla training sets. Specifically, we add rotated versions of vanilla images to the training set, where the rotation degrees are $-30°$ and $30°$, respectively.[3]. Also, we manually add some unistroke hand-written images obtained by WiRITE to the training sets for some hand-writing classes selected based on the vanilla ML model recognition results. Fig. 12 illustrates the process of sample augmentation, which consists of sample rotation and manipulation. Section 5 includes more details of the sample augmentation adopted in WiRITE implementation.

3. We did not add scaled versions of the images because we find that even without this procedure, the trained model can effectively recognize the scaled images.

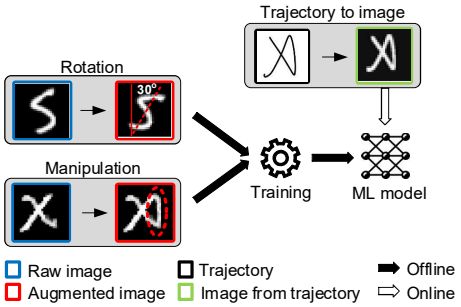


Fig. 12. Sample augmentation used in WiRITE.

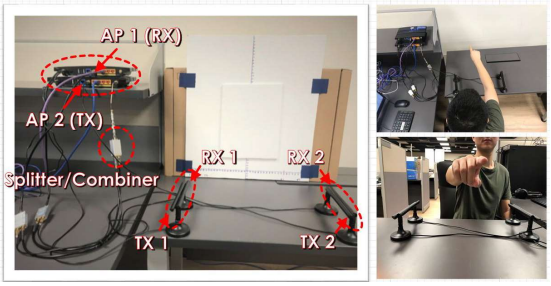


Fig. 13. WiRITE prototype.

Maximum likelihood estimation. The final classification of hand-writing is conducted by ML-based maximum likelihood estimator. Given a set of rotated and scaled versions of trajectory images, the estimator will make final decision by taking all of them into consideration. That is, given the hand-writing class set $\mathbf{C}$, the classification result c is determined by

$$\arg\max_{c\in\mathbf{C}} L(c) = \frac{1}{K}\sum_{i=1}^{K} p_c^i \tag{11}$$

where K, and p_c^i indicate the total number of rotated and scaled trajectory images, and the probability of ith trajectory image belonging to class c, which is provided by the softmax output, respectively.

5 Implementation

Hardware. WiRITE is built with two 802.11n commodity Wi-Fi APs [41] operating in 3x3 MIMO mode on 20 MHz channel at 5 GHz band. As Fig. 13 illustrates, in addition to the Wi-Fi APs, WiRITE's prototype also includes three RF power splitters/combiners used in CSI sanitization. CSI samples are collected by Atheros CSI tool [18] running on the receiving AP and are forwarded to a host PC with 333 Hz sampling rate.

ML model training. Three ML models, i.e., Digit, English, and Chinese, are trained with image samples provided by three vanilla public databases as detailed in Table 1[4] with zero labeling efforts. For Chinese, among the 7,185 characters incorporated in CASIA database, 200 most frequently used characters are selected to reduce the experimenting overhead. The host PC used in training is equipped with Intel Xeon E5-1650 v4 (CPU) and 32 GB memory. Empirically, each ML model reaches its stable state after 20 to 30 epochs, that is, approximately 30 minutes.

4. For English, the number of classes is 37 instead of 52, since 15 lowercase letters, i.e., 'c','i','j','k','l','m','o','p','s','u','v','w','x','y','z', are combined with the respective uppercase letters as a signal class due to the similar shapes between uppercase and lowercase as recommended in [23].

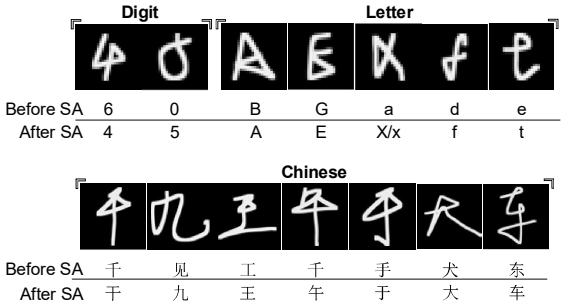


Fig. 14. Sample augmentation effect.

TABLE 2
Sample augmentation details.

	Digit	Letter	Chinese
# augmented classes	2	5	18
# styles per class	5	5	5
# samples per style	10	10	10

Sample augmentation (SA). The three vanilla ML models are retrained with augmented training sets to resolve dissimilarity between unistroke trajectories obtained from WiRITE and their corresponding multi-stroke versions in vanilla databases. Table 2 summarizes the number of augmented classes for the three applications, respectively, which are chosen based on the recognition results of the vanilla ML models. For each class, we manually generate five representative unistroke trajectory images provided by WiRITE's outputs. Each image is added 10 times to the training set to increase the weight on the final ML model. Fig. 14 visualizes examples that compare the recognition results before and after applying SA which suggests its efficacy. We observe that the labels as indicated in 'After SA' are all correct labels, indicating that the augmented samples do help correct the final ML model. The overall extra labeling efforts of SA are only 10 for digits, 25 for English letters, and 90 for Chinese characters, suggesting very high transferability of WiRITE core model when adapted across different applications. With SA, the test accuracies of the three ML models, when applying to the test samples provided by the three vanilla databases, are **0.992** (Digit), **0.962** (English), and **0.94** (Chinese), respectively.

6 Evaluation

Methodology. We evaluate the performance of the proposed system with the following three aspects. *1) User variety: How does the system react inclusively to different users?* We consider the impact from two factors – different hand size/shape and different writing styles. We first evaluate the variation of the end-to-end performance caused by each factor, and then analyse the detailed impact of the user variety on received

signal quality and handwriting trajectory estimation. *2) Environment variety: How does the system perform inclusively to the environmental interference caused by ambient reflections?* Specifically, we consider two kinds of interference – CSI noise and ambient moving dynamics. Raw CSI sample is noisy in both amplitude and phase due to imperfect hardware control. The channel dynamics caused by people moving around may interfere with the CSI time series shaped by the user's handwriting. The impact of the two factors always add up together and mess the handwriting signal. We evaluate the impact of each factor separately. *3) Realistic scenarios: How does the system perform in realistic scenarios?* We evaluate the joint impact of user variety and environment variety on the end-to-end performance in a typical office so that to assess the accuracy and robustness of the proposed system in practice.

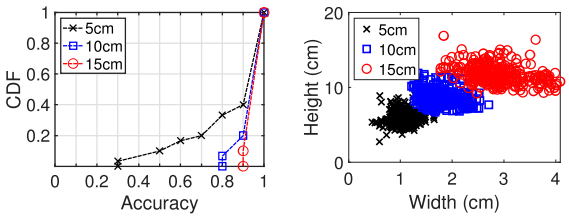
Data collection. Our experiment includes 15 voluntary users who contribute their handwriting data to the system evaluation. We transmit continuous Wi-Fi packets to sample the handwriting activity. For the evaluation of user variety, each user writes English letters following the templates designed with different size and various writing styles. To assess the performance with ambient interference, we collect handwriting samples when the activity is exposed to the environmental dynamics that is intentionally created by a walking interferer. We finally collect handwriting data at a typical office environment without deliberately controlling any possible impact factors (e.g., handwriting size, ambient movement, etc).

Metrics. We apply two metrics to WiRITE evaluation, i.e., *recognition accuracy* that indicates the ratio of the correctly recognized tests to the total number of tests, and *Dynamic Time Warping (DTW) distance* [42] that indicates the level of difference between two hand-writing trajectories — the greater the DTW distance the higher the difference between two trajectory images.

6.1 User Variety

Different users may affect WiRITE's performance due to different writing *sizes* and/or *styles*. Their hand shapes and surfaces are also different, resulting in different *responses* in reflecting or scattering Wi-Fi signals. We run experiment with English letter recognition as a vehicle to examine various user factors to WiRITE performance.

Different sizes. To see the performance limit of WiRITE in terms of trajectory size, we provide reference templates of English letters in squares with three different side lengths, i.e., 5cm, 10cm, and 15cm, to a user, who in the test writes each letter of each template size 10 times. Fig. 15(a) depicts the CDF of the recognition accuracy with respect to the three template sizes. We observe that 5cm side length leads to as low as 0.3 recognition accuracy. When the side length increases, i.e., to 10cm and 15cm, WiRITE achieves the accuracy above 0.9 in most cases. The average accuracies for the three side lengths are 0.87, 0.97, and 0.99, respectively. Fig. 15(b) further presents the statistics of the width and height distributions of the hand-written trajectories observed in the experiment. We see greater heights coming



(a) Average recognition accuracy. (b) Size distribution.

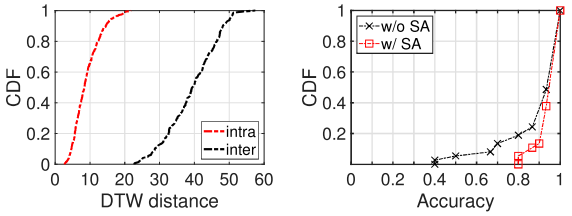
Fig. 15. Different size impact.

from test samples of bigger template sizes[5]. The results suggest that WiRITE is able to provide reasonably high recognition accuracy for hand-writing recognition with over 10cm by 10cm input sizes.

Different styles. WiRITE is inclusive of different writing styles by exploiting the versatility of vanilla databases. We evaluate this by deliberately selecting five templates of different styles for each English letter (in 15cm). The five styles are chosen to make them as different from each other as possible. Each style is written three times, such that we have 15 samples per letter. Fig. 16(a) gives the CDF of DTW distances among those samples. 'intra' indicates the DTW distances obtained by comparing each detected trajectory with its ground truth, while 'inter' means the comparison among the trajectory images of the same letter but across different styles. To obtain the ground truth, we set up a printed template paper for reference aside the experiment testbed. The test user points her finger tip at the template character and moves her hand along the printed trajectory. The hand-moving trajectories are converted to images which serve as ground-truth samples. The ground-truth collection is repeated for different template of each English letter. The statistics show that in the case of 'intra', DTW distance between the detected trajectory and its ground truth is always less than 20, while the differences become greater than 20 for letters of different styles in 'inter', which tells that the differences across styles are indeed much higher than that introduced by Wi-Fi sensing. Fig. 16(b) gives the CDF of WiRITE recognition accuracy. The result shows that with SA, the accuracy is higher than 0.9 in most of the cases, resulting in 0.96 in average (v.s. 0.91 accuracy without SA), which suggests the generality of WiRITE and the efficacy of SA in dealing with the style variation.

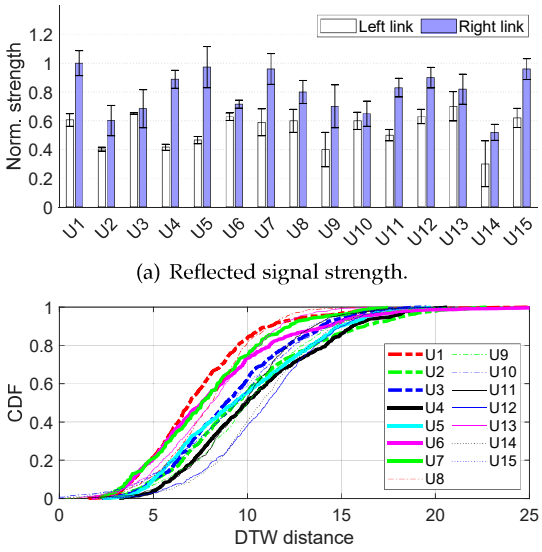
Different users' hands. The performance of WiRITE by its design principle should not be affected much by different users' hands. To verify this, we analyze the WiRITE performance with 15 different users. The users are given the same letter templates (with 15cm side length), and write each template 5 times. Fig. 17(a) summarizes the average reflected signal strength as been received at RX of the left or right pairs, respectively, with respect to the 15 users. We see that, even when the user's hand is located at the same locations, the reflected signal strength varies much. WiRITE exploits the relation between phase and distance, and hence, is robust to the variation of absolute signal

5. The scattered distributions are due to the scaled and rotated versions of the ground truth trajectories.



(a) CDF of DTW distance. (b) CDF of recognition accuracy.

Fig. 16. Recognition accuracy and DTW distance statistics of different styles.



(a) Reflected signal strength.

(b) CDF of DTW distance.

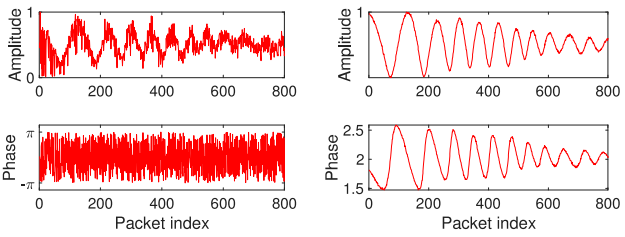
Fig. 17. Recognition accuracy and DTW distance statistics of different users.

strength. Fig. 17(b) gives WiRITE performance by showing the CDF of the DTW distance of the users. Each DTW distance is obtained by comparing detected trajectory to its ground truth. We find that DTW distances of all the users are upper-bounded by 20, which is the upper-bound of those of 'intra' shown in Fig. 16(a), showing the WiRITE's ability in extracting true trajectories regardless of the variation of users' responses to Wi-Fi signals. The average hand-writing recognition accuracies for the users range from 0.94 to 0.98.

6.2 Environment Variety

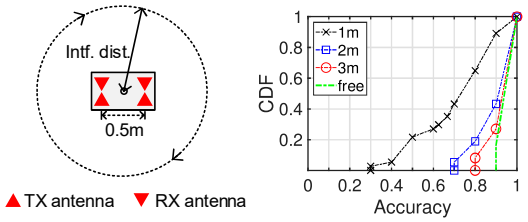
The environmental impacts come from two factors: 1) CSI noises caused by hardware imperfection, and 2) ambient interferences caused by surroundings. We evaluate the resilience of WiRITE to environmental impacts with the two factors.

CSI noises. We collect a sequence of CSI samples when moving a hand away from the Tx-Rx antenna pairs. Figs. 18(a) and 18(b) compare the raw and sanitized CSI samples, respectively, in terms of their normalized amplitudes and phases. We observe that after applying WiRITE's CSI sanitization, which does not leverage any smoothing and filtering process, the CSI samples are corrected and exhibit a clear continuous and repetitive pattern in both



(a) Raw CSI. (b) Sanitized CSI.

Fig. 18. CSI sanitization.



(a) Topology. (b) CDF of recognition accuracy.

Fig. 19. Evaluation of recognition accuracy with ambient interferences.

amplitude and phase domains, which would have been overwhelmed under the noise in raw CSI samples. When the amplitude and phase noise is left unremoved, the hand tracking function would be totally corrupted and cannot recover the continuous trajectory for further recognition.

Ambient interferences. We evaluate WiRITE's resilience to ambient interferences, gained from its crossing antenna arrangement which maximizes the SIR. In the experiment, when the test user moves his hand to write English letters (in 15cm), another person (interferer) walks around the site, away from the center of the antenna arrangement area by 1m, 2m, and 3m, respectively, as depicted in Fig. 19(a). Fig. 19(b) presents how WiRITE recognition accuracy varies when the interferer distance increases from 1m to 3m. When the interferer distance is 1m, the accuracy can be as low as 0.3, with the average accuracy of 0.74, while, with the increase of the interferer distance, the average accuracies are increased to 0.93 and 0.96 in the cases of 2m and 3m, respectively. The result tells that when the interferer distance is more than 1m, the accuracy can be very close to that of non-interference ('free'), i.e., 0.98 in average. Fig. 20 summarizes the recognition accuracy regarding 26 uppercase letters for 1m interference distance. We observe that the most cases still have reasonably high accuracy except five cases, namely, 'A', 'E', 'F', 'Q', and 'W', that achieve lower than 0.7 accuracies. Fig. 21 shows WiRITE's outputs for the five cases. All five cases entail relatively more strikes, and due to that, longer writing time compared to other letters, thus suffering more severe interferences. When the interferer distance increases to 2m, as Fig. 21 shows, the output images are closer to ground truths. Note that in real world, the interferer is mostly located far from the testbed (larger than 1m), so the impact of ambient interferer can be ignored according to our measurement in Fig. 19(b).

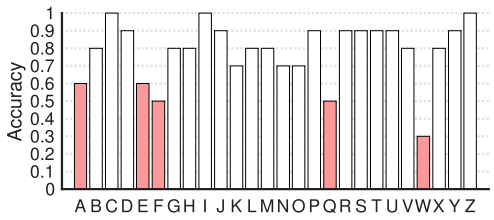


Fig. 20. Recognition accuracy for 1m interference distance.

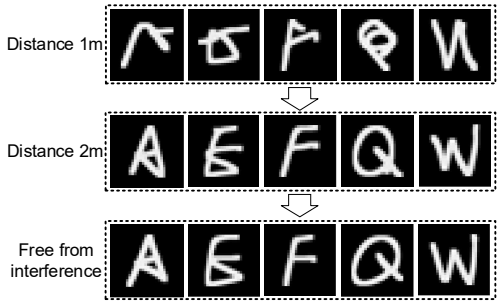


Fig. 21. Illustration of hand-writing images with different levels of ambient interferences.

6.3 Evaluation with Realistic Scenario

We conduct experiments in a typical indoor office environment. A realistic scenario is achieved by ensuring four factors: 1) multiple users, 2) at different locations, 3) with possible ambient interferences, and 4) hand-writing in preferred sizes and styles of individual users.

As Fig. 22 illustrates, we let five WiRITE users at five different locations (marked in red circles) write Digits, English letters, and Chinese characters in their preferred sizes and styles, during normal office hours. The number of normal office members sharing the same office during the experiments was 10 in average, who intermittently generated ambient interferences, when passing by the WiRITE users. Each hand-writing class is written 10 times by each WiRITE user. In particular, for the application of Chinese characters, each user randomly selects 50 Chinese characters among the 200 candidates to reduce evaluation overhead. Note that the experiments was conducted without any re-training of the ML model for adapting to any specific WiRITE user.

Fig. 23(a), 23(b), and 23(c) show the CDF results of WiRITE recognition accuracies for Digits, English letters, and Chinese characters (simple characters that can be completed within a few strokes), respectively, with aggregated results from the five WiRITE users. WiRITE achieves 0.8 or higher accuracies in more than 80% cases for all three applications. The comparative results (with and without SA) also suggest that some hand-writing classes may be subject to low recognition accuracies due mainly to the unistroke and multi-stroke dissimilarity. With SA, WiRITE gains extra 4.4%, 6.8%, and 7% improvements of the recognition accuracy in the three applications, respectively. Table 3 summarizes the average recognition accuracies of the five WiRITE users in the three applications, respectively. The average accuracies are higher than 0.9 in all cases with 0.93 on average.

To show how the writing habits of the five WiRITE users differ, Fig. 24(a) and 24(b) present the statistics of their hand-writing sizes and the DTW distances across different users

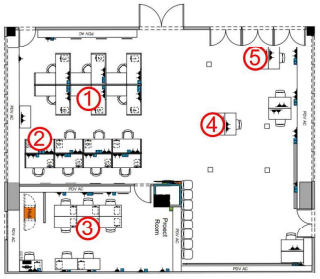


Fig. 22. The realistic experiment environment.

TABLE 3
Average accuracy.

	U1	U2	U3	U4	U5	Average
Digits	0.97	0.92	0.94	0.97	0.92	0.94
English letters	0.95	0.92	0.94	0.94	0.93	0.93
Chinese characters	0.93	0.93	0.94	0.94	0.9	0.92
Average	0.95	0.923	0.94	0.95	0.916	0.93

for a same hand-writing class, respectively. As Fig. 24(a) suggests, in most of the cases, the users write with the height of at least 10cm, which is better supported by WiRITE as discussed in Section 6.1. Fig. 24(b) suggests that the DTW distances range from 5 to 40, showing substantial diversity in their writing styles. With that, the reasonably high WiRITE accuracy across the five users suggest high generality of WiRITE over user variety.

Comparative experiment. We conduct experiment to verify the improvement of WiRITE on dealing with realistic scenarios. The experiment compares our system with a state-of-the-art Wi-Fi based handwriting recognition system called Wi-Reader [32]. We implement the processing flow of Wi-Reader including data partitioning, butterworth de-noising and subcarrier selecting with the CSI obtained from TL-WDR4310 hardware. We measure the recognition accuracy of English handwriting letters of the systems with the impact of 1) varying handwriting styles/strokes and 2) ambient human movement.

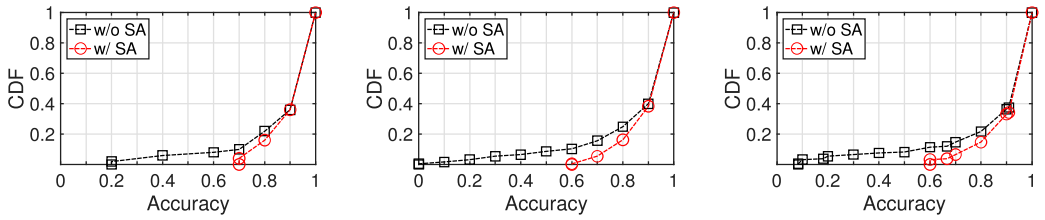
TABLE 4
Recognition accuracy of WiRITE and Wi-Reader at different using scenarios.

	Handwriting Style/Stroke		Ambient Movement	
	same	different	without	with
WiRITE	96%	95%	95%	93%
Wi-Reader [32]	92%	63%	91%	49%

Table 4 presents the comparative results. As we can see, WiRITE provides over 93% averaged accuracy for all four different using scenarios. In comparison, the recognition accuracy of Wi-Reader drops by 29% when the test user performs handwriting in different styles and strokes and by 42% when there exists ambient human movement. The comparison proves the improvement of WiRITE on user generality and environment agnosticism.

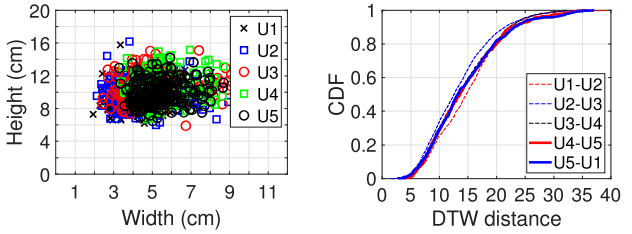
7 Discussion

System latency. We measure the end-to-end system latency to quantify the recognition speed. We wrote 26 English capital letters and recorded the elapsed time during the derivation of the recognition result. According to the result,



(a) Digit (b) English (c) Chinese

Fig. 23. CDF of accuracies of applications for Digit, English letter, and Chinese character recognition.



(a) Size distribution. (b) DTW distance distribution.

Fig. 24. Size and style variety.

the median system latency is 6.84s. For 80% of the tests, the system can recognize the handwriting within 7.8s. We investigate the operation delay of the two sequential blocks – trajectory estimation and image classification. The measurement result shows that trajectory estimation contributes 55% of the total latency. With such observation, we further reduce the system latency by accelerating trajectory estimation by optimizing the channel sampling rate. Without loss of recognition accuracy, we finally reduce the system latency to 5.23s in median. The latency can be further reduced with the support of high-end GPU (currently not adopted).

One model for all applications. Our current design trains separate models for the three different applications. To further improve the application transferability, we may train a single model that applies to the recognition of different applications/domains. A proof-of-concept experiment is conducted to verify the feasibility of applying one model for different applications. Specifically, we train one network model with a macro-dataset that contains 5000 handwriting image samples covering 10 digits, 5 English letters and 5 Chinese characters. The results indicate 91%/92%, 92%/92% and 89%/90% median/averaged accuracy when VGG-16 [26], Inceptionv3 [43] and ResNet50 [44] is adopted, respectively. The accuracy should remain when the model is adopted as the image classifier of WiRITE. We believe the performance can be further improved (for higher recognition accuracy and more identifiable categories) by adopting a more powerful network model.

Simplifying the network model. The current system uses VGG network to highlight the application transferability – the recovered trajectories of different handwriting applications (even complicated ones like Chinese characters) can be directly fed into existing image classification models (where VGG network is a well-known one) for recognition. To implement the system with a simpler network for better efficiency, we can simply replace the VGG model with a CNN model or even a feed-forward network of several layers depending on the complexity of specific classification task. For example, to recognize digit numbers which only contain 10 categories, we can adopt a simple three-layer feed-forward network. To classify Chinese characters which exhibit more complex pattern and contain hundreds of categories, we can try a CNN model with multiple filters for fine-grained feature extraction and identification.

Impact of different link spacing. The recognition accuracy of WiRITE is related to the distance between the two links. We conducted an experiment measuring the recognition accuracy of English letters when the distance varies from 35cm to 65cm at the step of 5cm. The result shows that the accuracy is optimized to be 93% when the distance is set to 50cm, and equals to 85% and 89% when the distances are 45cm and 55cm, respectively. The accuracy is lower at the other distances. To explain the reason, note that with human handwriting activity, there are three types of signal path within the system – the reflected path from human hand (p_1), the direct path within one pair (p_2) and the mutual interference between the two pairs (p_3). The three paths add up at the Rx antenna of each pair. Therefore, both p_2 and p_3 cause interference on p_1 and possibly suppress the hand writing information. When the two links are closely spaced, the mutual interference (i.e., p_3) becomes strong, and could overwhelm the variation of the hand reflected path (i.e., p_1). When the two links are set far from each other, the hand reflected path (i.e., p_1) becomes weak, but the power of the direct path (i.e., p_2) does not change, so the hand moving variation also gets suppressed. Considering the impact of the two interference, the distance between the two links should be wisely selected. For the current system setting, the space of 50cm provides the optimal system performance. We adopt this setup for system implementation.

Impact of hidden interference signal. The possible using scenario of the proposed system could be an indoor room/office where hidden interference signal exists (e.g., Wi-Fi beacon). According to the CSMA/CA channel access control of Wi-Fi, when exposed to nearby hidden interference, the transmission will be pending until no on-air signal is detected. Therefore, for our system, the interference effectively reduces the sampling rate of the continuous hand tracking. We experimentally investigate how the accuracy varies with different sampling rate. The result indicates that the accuracy remains high ($\geq 97\%$) when the sampling rate is no less than 200 Hz, which is testified to be easy to achieve based on our empirical study with real world Wi-Fi usage.

Recognizing consecutive characters. The current design of the system may not be able to handle the recognition of consecutive multiple character inputs. The reason is mainly

because the machine learning model (CNN) adopted in the system is trained with single character dataset (MNIST, EMNIST, CASIA). We may improve the current system towards recognizing consecutive characters in two ways: first, we can design a character segmentation module that can detect the different characters within one word/string and separate them. This module pre-processes the input text and its output (multiple single characters) is fed into the CNN model that handles character recognition. Second, we may replace the current CNN model with a more advanced model (e.g., CRNN[g]) that is able recognize the consecutive sequential characters. But this model incurs larger training overhead to achieve accurate performance. We leave this design as future work.

8 Related Work

Prior works on hand-writing recognition systems fall into four categories from the perspective of data acquisition: camera-based, wearable-based, acoustic signal-based, and RF signal-based.

Camera-based. Camera-based systems have been well studied for over two decades [45], [46]. The rational is to detect the hand from captured images by its appearance and derive the trajectory when it moves. The signal source can be visible light [47], [48] or infrared (depth cameras) [49], [50]. Despite their popularity, these approaches require a dedicated hardware setup and LoS to the user. Also, it may be affected by lighting conditions, the variations of the hand shapes, skin colors, and even textures [50] in separating the hand from cluttered background image. Privacy is also one of the primary issues encountered by this line of approaches.

Wearable-based. Another line of approaches rely on motion sensors embedded in today's smartphones [51], [52] or wearables such as smart watches [53]–[55], wristbands [56], [57], and rings [58], [59]. All such wearable-based systems require an external device to be equipped by the user, which restricts the applicable conditions, and brings inconvenience to users.

Acoustic signal-based. This line of approaches normally benefit from higher ranging/tracking accuracy due to the slower propagation speed of acoustic signals compared to that of RF signals [60]–[64]. EchoWrite [64] proposes an acoustic based system that is able to recognize different finger writing English letters. The authors classify the letters into different groups that are defined with certain combination of basic strokes, which are identified by the Doppler shift profiles extracted from the spectrogram of time domain signal. Although with Wi-Fi signal (which is intrinsically not comparable with acoustic signal in its ranging resolution), WiRITE still advances EchoWrite due to its enlarged recognition diversity (including the three types of input) and explicit handwriting trajectory estimation. Unlike RF signals, acoustic signals inevitably suffer from ambient noise that is inherent to environment (audible and/or inaudible), and interferences coming from other acoustic devices due to the lack of standardized MAC protocol, which intrinsically limits the sensing performance [65].

RF signal-based. High resolution RF sensing entails special hardware to exploit ultra-wideband (GHz) transceivers, multiple antennas, and/or specialized signal modulation (e.g., FMCW) [3], [30], [66], [67]. For example, mmWrite [67] proposes a handwriting recognition system based on accurate hand tracking with using millimeter wave hardware. This system requires the specialized infrastructure that can support 60 GHz front-end, occupies GHz-level large bandwidth, and relies on an antenna array that consist of tens of antenna elements, and thus degrades the ubiquitousness and practicability of the target applications. For handwriting recognition with commercial Wi-Fi hardware, most systems [4], [20]–[22], [32], [33], [68], [69] adopt a similar modality that entails prior training of a ML model to learn the effect of hand motions on wireless signal patterns, and targets only a pre-defined set of motions without any functional extensibility from the application point of view, let alone the burdensome training overhead in tedious data collection and labeling. For example, WiReader [32] designs a real-time handwriting recognition system based on commercial Wi-Fi devices. The system adopts LSTM to classify the feature space extracted from CSI samples obtained after pre-processing. However, WiReader can only support the recognition of 26 English letters with predefined handwriting pattern. In comparison, our proposed system (i.e., WiRITE) is able to recognize hundreds of characters of three different applications. The improvement on application extensibility mainly comes from WiRITE's capability of explicit hand tracking, based on which we can visualize the hand moving trajectory and adopt existing CNN model directly for recognition. WiDraw [70] is the only Wi-Fi system that tracks the trajectory of the hand, and identifies the content with an off-the-shelf text recognition software. However, since it derives the trajectory by detecting the AoA variation of direct path signal, WiDraw requires 30 Wi-Fi transmitters to provide wide range of AoA distribution, which substantially degrades the practicability. In contrast to previous solutions, WiRITE is the first Wi-Fi hand-writing solution that ensures practicability, in that it avoids the manual training sample collections but provides generality and extensibility towards different users, environments, and applications.

9 Conclusion

WiRITE is a general and practical Wi-Fi hand-writing system. The design of WiRITE addresses major challenges in machine learning design and signal processing to achieve the goal of generality, i.e., application transferability, environment agnosticism, and user independence. The performance of WiRITE is corroborated by extensive real world experiments. Future works include applying WiRITE to 802.11ac Wi-Fi APs for finer signal resolution and employing more antennas to enable 3D hand-writing and gesture recognition.

Acknowledgments

This work was supported by Alibaba Group through Alibaba Innovative Research (AIR) Program and Alibaba-NTU Singapore Joint Research Institute (JRI) and Singapore MOE AcRF Tier 2 MOE-T2EP20220-0004. The first two authors contribute equally to this work. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of Alibaba Group and other funding agencies.

REFERENCES

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

[2] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "md-track: Leveraging multi-dimensionality for passive indoor wi-fi tracking," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–16.

[3] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14)*, 2014, pp. 317–329.

[4] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 289–304.

[5] N. Yu, W. Wang, A. X. Liu, and L. Kong, "Qgesture: Quantifying gesture distance and direction with wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–23, 2018.

[6] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, 2013, pp. 27–38.

[7] Y. Liu, Z. Li, Z. Liu, and K. Wu, "Real-time arm skeleton tracking and gesture inference tolerant to missing wearable sensors," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 287–299.

[8] 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*, 2019, pp. 1–13.

[9] C.-Y. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi, and C. Liu, "Extracting gait velocity and stride length from surrounding radio signals," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 2116–2126.

[10] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf wifi," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2015, pp. 267–276.

[11] S. Ahmed and S. H. Cho, "Hand gesture recognition using an ir-uwb radar with an inception module-based classifier," *Sensors*, vol. 20, no. 2, p. 564, 2020.

[12] J. Park and S. H. Cho, "Ir-uwb radar sensor for human gesture recognition by using machine learning," in *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2016, pp. 1246–1249.

[13] S.-J. Ryu, J.-S. Suh, S.-H. Baek, S. Hong, and J.-H. Kim, "Feature-based hand gesture recognition using an fmcw radar and its temporal feature analysis," *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7593–7602, 2018.

[14] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–19, 2016.

[15] X. Cai, J. Ma, W. Liu, H. Han, and L. Ma, "Efficient convolutional neural network for fmcw radar based hand gesture recognition," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 17–20.

[16] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, "Audiogest: enabling fine-grained hand gesture detection by decoding echo signal," in *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, 2016, pp. 474–485.

[17] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 82–94.

[18] 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

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

[20] X. Cao, B. Chen, and Y. Zhao, "Wi-wri: Fine-grained writing recognition using wi-fi signals," in *2016 IEEE Trustcom/BigDataSE/ISPA*. IEEE, 2016, pp. 1366–1373.

[21] J. Xiao, H. Li, and Y. Liu, "Wiwrite: Wi-fi based handwriting recognition like playing lego," in *2019 28th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2019, pp. 1–9.

[22] Z. Fu, J. Xu, Z. Zhu, A. X. Liu, and X. Sun, "Writing in the air with wifi signals for virtual reality devices," *IEEE Transactions on Mobile Computing*, vol. 18, no. 2, pp. 473–484, 2018.

[23] G. Cohen, S. Afshar, J. Tapson, and A. Van Schaik, "Emnist: Extending mnist to handwritten letters," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 2921–2926.

[24] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "Casia online and offline chinese handwriting databases," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 37–41.

[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[27] Y. Zhang, "Deep convolutional network for handwritten chinese character recognition," *Computer Science Department, Stanford University*, 2015.

[28] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, 2015, pp. 269–282.

[29] 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)*, 2016, pp. 165–178.

[30] Y. Xie, Y. Zhang, J. C. Liando, and M. Li, "Swan: Stitched wi-fi antennas," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 51–66.

[31] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the mona lisa: Spot localization using phy layer information," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 2012, pp. 183–196.

[32] Z. Guo, F. Xiao, B. Sheng, H. Fei, and S. Yu, "Wireader: Adaptive air handwriting recognition based on commercial wifi signal," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 483–10 494, 2020.

[33] K. Niu, F. Zhang, Y. Jiang, J. Xiong, Q. Lv, Y. Zeng, and D. Zhang, "Wimorse: A contactless morse code text input system using ambient wifi signals," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9993–10 008, 2019.

[34] R. Zhou, J. Chen, X. Lu, and J. Wu, "Csi fingerprinting with svm regression to achieve device-free passive localization," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2017, pp. 1–9.

[35] 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 *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 243–256.

[36] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Presented as part of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 13)*, 2013, pp. 71–84.

[37] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: Enabling phased array signal processing on commodity wifi access points," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 153–164.

[38] J. L. Volakis, *Antenna engineering handbook*. McGraw-Hill Education, 2007.

[39] F. Adib and D. Katabi, "See through walls with wifi!" in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, 2013, pp. 75–86.

[40] C. Gao, Y. Li, and X. Zhang, "Livetag: Sensing human-object interaction through passive chipless wifi tags," in *15th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 18)*, 2018, pp. 533–546.
[41] O. Project, "Tp-link tl-wdr4310," https://openwrt.org/toh/tp-link/tl-wdr4310, 2018 (accessed Feb. 4, 2020).
[42] K. K. Paliwal, A. Agarwal, and S. S. Sinha, "A modification over sakoe and chiba's dynamic time warping algorithm for isolated word recognition," *Signal Processing*, vol. 4, no. 4, pp. 329–333, 1982.
[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *arXiv e-prints*, p. arXiv:1512.00567, Dec. 2015.
[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[45] J. M. Rehg and T. Kanade, "Visual tracking of high dof articulated structures: an application to human hand tracking," in *European conference on computer vision*. Springer, 1994, pp. 35–46.
[46] T. Starner and A. Pentland, "Real-time american sign language recognition from video using hidden markov models," in *Motion-based recognition*. Springer, 1997, pp. 227–243.
[47] L. Jin, D. Yang, L.-X. Zhen, and J.-C. Huang, "A novel vision-based finger-writing character recognition system," *Journal of Circuits, Systems, and Computers*, vol. 16, no. 03, pp. 421–436, 2007.
[48] A. Schick, D. Morlock, C. Amma, T. Schultz, and R. Stiefelhagen, "Vision-based handwriting recognition for unrestricted text input in mid-air," in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 2012, pp. 217–220.
[49] Microsoft, "Kinect for windows," https://developer.microsoft.com/en-us/windows/kinect/, 2020.
[50] C. Wang, Z. Liu, and S.-C. Chan, "Superpixel-based hand gesture recognition with kinect depth camera," *IEEE transactions on multimedia*, vol. 17, no. 1, pp. 29–39, 2014.
[51] S. Agrawal, I. Constandache, S. Gaonkar, R. Roy Choudhury, K. Caves, and F. DeRuyter, "Using mobile phones to write in air," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*, 2011, pp. 15–28.
[52] T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, and J. Song, "E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, 2011, pp. 260–273.
[53] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, 2015, pp. 9–14.
[54] S. Shen, H. Wang, and R. Roy Choudhury, "I am a smartwatch and i can track my user's arm," in *Proceedings of the 14th annual international conference on Mobile systems, applications, and services*, 2016, pp. 85–96.
[55] J. Gong, Z. Xu, Q. Guo, T. Seyed, X. Chen, X. Bi, and X.-D. Yang, "Wristext: One-handed text entry on smartwatch using wrist gestures," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–14.
[56] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis, "Risq: Recognizing smoking gestures with inertial sensors on a wristband," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, 2014, pp. 149–161.
[57] H. Truong, S. Zhang, U. Muncuk, P. Nguyen, N. Bui, A. Nguyen, Q. Lv, K. Chowdhury, T. Dinh, and T. Vu, "Capband: Battery-free successive capacitance sensing wristband for hand gesture recognition," in *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, 2018, pp. 54–67.
[58] J. Gummeson, B. Priyantha, and J. Liu, "An energy harvesting wearable ring platform for gesture input on surfaces," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, 2014, pp. 162–175.
[59] S. Nirjon, J. Gummeson, D. Gelb, and K.-H. Kim, "Typingring: A wearable ring platform for text input," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 227–239.
[60] A. Wang and S. Gollakota, "Millisonic: Pushing the limits of acoustic motion tracking," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–11.
[61] H. Chen, F. Li, and Y. Wang, "Echotrack: Acoustic device-free hand tracking on smart phones," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.
[62] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using active sonar for fine-grained finger tracking," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 1515–1525.
[63] Y. Wang, J. Shen, and Y. Zheng, "Push the limit of acoustic gesture recognition," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 566–575.
[64] K. Wu, Q. Yang, B. Yuan, Y. Zou, R. Ruby, and M. Li, "Echowrite: An acoustic-based finger input system without training," *IEEE Transactions on Mobile Computing*, vol. 20, no. 5, pp. 1789–1803, 2020.
[65] C. Cai, R. Zheng, and M. Hu, "A survey on acoustic sensing," *arXiv preprint arXiv:1901.03450*, 2019.
[66] T. Wei and X. Zhang, "mtrack: High-precision passive tracking using millimeter wave radios," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015, pp. 117–129.
[67] S. D. Regani, C. Wu, B. Wang, M. Wu, and K. R. Liu, "mmwrite: passive handwriting tracking using a single millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 291–13 305, 2021.
[68] N. Xiao, P. Yang, Y. Yan, H. Zhou, and X.-Y. Li, "Motion-fi: Recognizing and counting repetitive motions with passive wireless backscattering," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2024–2032.
[69] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "Wifinger: Talk to your smart devices with finger-grained gesture," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 250–261.
[70] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015, pp. 77–89.

Yanbo Zhang receives the B.E. degree in communication from Harbin Institute of Technology, Harbin, China in 2017. He is currently working toward the PhD degree in computer science and engineering at Nanyang Technological University. His research interests include wireless communication and sensing systems, internet of things and mobile computing.

Weiping Sun received his B.S. degree in network engineering from Dalian University of Technology, Dalian, China, in 2006, and his Ph.D. degree in electrical engineering from Seoul National University, Seoul, South Korea, in 2017. He had worked as a postdoc research fellow in Nanyang Technological University, Singapore, from 2018 to 2020. He joined Samsung Research, Seoul, South Korea, in 2020, where he is a staff engineer, working on the next generation cellular system. His research interests include RF-based sensing, PHY/MAC designs of cellular networks, WLAN, WPAN, and coexistence among heterogeneous networks.

Mo Li (M'06) received the B.S. degree in computer science and technology from Tsinghua University, Beijing, China, in 2004, and the Ph.D. degree in computer science and engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2009. He is currently an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include networked and distributed sensing, wireless and mobile, cyber-physical systems, smart city, and urban computing.