This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

MetaSense: Boosting RF Sensing Accuracy using Dynamic Metasurface Antenna

Guohao Lan, *Member, IEEE,* Mohammadreza F. Imani, *Member, IEEE,* Zida Liu, José Manjarrés, Wenjun Hu, *Member, IEEE,* Andrew S. Lan, *Member, IEEE,* David R. Smith, *Senior Member, IEEE,* and Maria Gorlatova *Member, IEEE*

***Abstract*—Conventional radio frequency (RF) sensing systems rely on either frequency diversity or spatial diversity to ensure high sensing accuracy. Such reliance introduces several practical limitations that hinder the pervasive deployment of existing solutions. To circumvent this prevalent reliance, we present MetaSense, a system that leverages *antenna pattern diversity* for fine-grained RF sensing. MetaSense incorporates the *Dynamic Metasurface Antenna* (DMA) and the *auxiliary-assisted ensemble multi-mask learning framework* (AEMML) in its design. The DMA is a novel type of antenna that can provide a diverse set of uncorrelated radiation patterns in a low-cost and low-complexity manner. The AEMML is a quality-aware learning framework that can dynamically assess and aggregate the heterogeneous channel measurements from different antenna patterns to ensure high sensing accuracy. It also incorporates a transfer learning model that allows it to generalize to new sensing conditions with few training instances required. We prototype MetaSense and demonstrate its effectiveness on a writing motion recognition task using a custom-designed two-dimensional DMA. The results show that MetaSense achieves 92% to 98% accuracy in classifying ten miniature writing motions, outperforming a non-tunable antenna by 20% in all scenarios. Moreover, when deployed in new sensing positions where limited training instances are available, MetaSense requires as few as five training instances per class to achieve over 90% accuracy.**

***Index Terms*—Metamaterials, Metasurface, Reconfigurable Intelligent Surfaces, Wireless Sensing, Ensemble learning.**

I. Introduction

Leveraging signal fluctuations to detect environmental dynamics is a well-studied area in physics known as diffusing wave spectroscopy [1]. This concept has been applied in the RF sensing domain, where the variations in the wireless signal are used to capture the environmental changes caused by the motion of interest. To achieve high accuracy in RF sensing, the fundamental issue is to obtain *a high dimension of uncorrelated* input that provides sufficient information about the monitoring target. In wireless systems, this can be achieved by having enough *spatial* or *frequency diversity*. First, leveraging spatial diversity, we can add more transceiver pairs at independent locations, where the local wireless signals are affected by the monitored motion differently. Examples are the use of Multiple-Input Multiple-Output (MIMO) [2] and antenna arrays [3] for sensing. Second, we can rely on frequency diversity to transmit the sensing signal in a wideband. Examples are the WiFi-based solutions [4] that transmit signal at a 20/40MHz bandwidth, as well as the radar-based [5] and ultra-wideband-based [6] solutions that require a GHz bandwidth for sensing.

In practice, however, reliance on either frequency diversity or spatial diversity is difficult and costly. Custom-built devices such as Doppler-radar [5], [7] and antenna array [3] are costly in both hardware and signal processing. The wide frequency band requirement makes radio frequency components, e.g., amplifiers and oscillators, more complex and expensive than those of a narrow-band device. Moreover, increasing the number of antennas not only makes the device cumbersome, but also increases the complexity in digital signal processing [3]. Commodity devices, such as WiFi infrastructures, are more pervasive and widely deployed. Unfortunately, WiFi-based solutions are known to degrade in performance due to the multipath issues [8], [9].

In this work, instead of adding more transceivers or extending the signal bandwidth, we propose the use of *antenna pattern diversity* to boost RF sensing performance. However, achieving configurable antenna patterns with high diversity is non-trivial. Conventional array antennas require power-hungry phased shifters, amplifiers, and other RF components to generate different antenna patterns [10], and thus are costly and complex when a large scale of antenna elements is needed. To move beyond these limitations, we exploit the *Dynamic Metasurface Antenna* (DMA) to ensure antenna pattern diversity for RF sensing. The DMA is a novel class of antennas that can effectively and rapidly change their radiation patterns from a simplified hardware platform [11], [12]. Instead of using conventional antenna elements, the key enablers of DMA are the *metamaterial elements* which are artificial materials engineered to allow the manipulation of electromagnetic waves in a deliberate and controlled manner [13]. The DMA is embedded with a set of sub-wavelength-sized metamaterial elements on its top layer. Each of the embedded metamaterial elements passively radiates portion

G. Lan, D. R. Smith, and M. Gorlatova are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 (email: guohao.lan@duke.edu; drsmith@duke.edu; maria.gorlatova@duke.edu)

M. F. Imani is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 (email: mohamad.imani@gmail.com).

Z. Liu is with the School of Electrical Engineering and Computer Science, Pennsylvania State University, University Park, PA 16802 (email: zjl5310@psu.edu).

J. Manjarrés is with the Department of Engineering, Olivet Nazarene University, Bourbonnais, IL 60914 (email: jemanjarres@olivet.edu).

W. Hu is with the Department of Electrical Engineering, Yale University, New Haven, CT 06520 (email: wenjun.hu@yale.edu).

A. S. Lan is with the College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA 01003 (email: andrewlan@cs.umass.edu).

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

of energy from the antenna's waveguide into the wireless channel, and thus, by tuning the resonance frequency of each individual element, the overall radiation pattern of DMA can be effectively controlled [14].

Without the need for complex, costly, and energy hungry RF components in its design, DMA is an emerging technology for realizing large scale adaptive antenna arrays in a smaller form-factor with simple and low-cost design [15]. Moreover, metasurface and metamaterial-based antenna designs have shown great performance in eliminating the mutual coupling effects among antenna elements [16], [17] to ensure higher antenna diversity [18]. In light of the great promise, a recent report forecasts the commercial market of DMA-based sensing and communication devices to exceed ten billion dollars by 2030 [19]. Indeed, DMA has attracted a lot of attention from both academia [20], [21] and industry [22]–[24]. Industry leaders, such as LG and NETGEAR, have included metamaterial-based antennas in smartphones and routers [25], Huawei and PARC have started to deploy metasurface antennas for 5G communications [26], [27].

Despite the appealing benefits of DMA, its applications and challenges for fine-grained RF sensing systems have not yet been studied. In this paper, we present MetaSense, the first end-to-end DMA-based RF sensing system. Leveraging the antenna pattern diversity of the DMA, MetaSense embraces a high dimension of uncorrelated channel measurement to develop more accurate and robust RF sensing solutions. Bringing this high-level concept into a practical system requires overcoming several challenges:

- As there is no off-the-shelf DMA hardware available, the first challenge is in prototyping an effective DMA to provide a large set of distinct antenna patterns. In this paper, we design and implement a single-port, two-dimensional DMA with 98 metamaterial elements operating at microwave frequencies (i.e., 17.5-22GHz). We design a simple Arduino-based controller to configure the DMA's antenna pattern. Our prototype provides hundreds of uncorrelated antenna patterns to boost the measurement dimension.

- The antenna pattern diversity of DMA offers a fruitful measurement for sensing. Intrinsically coupled to this capability is the challenge in designing a learning mechanism that can *properly assess and compare the sensing quality of different antenna patterns*, and can *dynamically aggregate them based on the estimated quality in runtime*. This is essential as different antenna patterns are unequal in signal resolvability and sensing performance. Indeed, as will be shown in Section VIII-D, the accuracy varies from 64% to 82% given different DMA pattern configurations. Moreover, the performance changes dynamically with the sensing conditions, and is hard to pre-estimate without actual channel measurement. In this paper, borrowing the concept of certainties from information theory, we propose the normalized entropy as the auxiliary feature to assess the sensing quality of different DMA antenna patterns. In addition, we design the quality-aware Auxiliary-assisted Ensemble Multi-Mask Learning (AEMML) that can dynamically aggregate the heterogeneous DMA measurements to boost the sensing accuracy.

- Lastly, existing learning-based RF sensing systems [28], [29] leverage the deep neural networks (DNNs) to boost the recognition accuracy. However, a large labeled dataset is required to achieve good sensing performance. Moreover, when the pre-trained DNN model is deployed in a new location or environment, its sensing accuracy will degrade significantly due to the domain shift problem [30], as the radio signals used for sensing are subject to environment and location changes [28], [29]. Thus, existing DNN-based RF sensing systems try to generalize the classification model by collecting datasets across a large number of environments [28], [29], which is expensive and inefficient. To move beyond this limitation, we employ transfer learning to generalize the proposed AEMML framework. Our solution significantly reduces the number of training instances required to extend MetaSense to new sensing locations and environments, allowing it to achieve over 90% accuracy with only five training instances per class.

The main contributions of this paper are:

- We investigate the use of the antenna pattern diversity of DMA to boost RF sensing accuracy. It achieves fine-grained RF sensing with '*a single transceiver device working at a single frequency*'. Our solution paves the way for future low-cost and low-complexity RF sensing systems, where limited transceivers and bandwidth are available or accessible.

- We present the first end-to-end system design for DMA-based RF sensing which includes a signal processing pipeline to handle noise and misalignment issues in the DMA signal, a robust segmentation algorithm for motion detection, as well as the quality-aware learning framework that can assess the sensing quality of different DMA patterns and dynamically aggregate them to boost the sensing accuracy at runtime.

- Using the two-dimensional DMA we designed and implemented, we evaluate the performance of MetaSense. We consider the miniature writing motion recognition as a case study. Specifically, we leverage a programmable drawing robot to generate miniature writing movements with high randomness. This robot-based setup allows us to conduct comprehensive and repeatable experiments when human interactions are restricted due to the COVID-19. We also take the MNIST handwritten digits dataset [31] as the reference to design different drawing patterns. Extensive experiments show that MetaSense achieves 92% to 98% accuracy in different settings, outperforming the non-tunable antenna by 20% in all scenarios. Moreover, by dynamically aggregating the inputs from diverse DMA masks, our quality-aware multi-mask learning framework achieves up to 12.5% accuracy improvement compared to the best conventional classifier.

- Our transfer learning-based framework enables efficient system adaptation to new sensing conditions with limited training instances required. Our evaluation shows that, together with the help of the DMA antenna pattern diversity, MetaSense requires as few as five training instances per class to achieve over 90% accuracy when deployed in new sensing locations, which outperforms the conventional method by 18%.

The rest of this paper is organized as follows. Related work is reviewed in Section 2. Section 3 presents a primer on DMA

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

and the use of DMA for sensing. Section 4 provides the system overview of MetaSense. Section 5 details the design of our DMA hardware. Section 6 introduces the signal processing pipeline. Section 7 presents the quality-aware multi-mask sensing framework. We present the evaluation in Section 8 and discuss the limitations and future directions in Section 9. We conclude the work in Section 10.

II. Related Work

A. Wireless Sensing

Our work is related to existing efforts that leverage RF signal, visible light, and sound for sensing.

WiFi-based works rely on frequency diversity to obtain detailed phase and amplitude information from each of the subcarriers [32]. To further improve the sensing performance, more advanced methods are stitching multiple antennas in a single-device [3], [33] or leveraging multiple transceivers [2]. These solutions are cumbersome and costly in both hardware and signal processing. More recently, attempts have been made to achieve a single transceiver solution [34]. However, they still rely on a wide bandwidth to ensure good sensing performance.

Radio Frequency Identification (RFID) is also promising for sensing. However, existing methods need to attach the RFID tags to the sensing target [35], [36] and require multiple dedicated readers and antennas (e.g., RF-IDraw [35] needs eight antennas and two readers, Tagoram [36] needs four antennas).

Radar-based systems, including the ultra-wideband (UWB) [37], Doppler-radar [5], and the frequency modulated carrier wave (FMCW) radar [7], [38], rely on frequency diversity to achieve good performance, as their sensing resolution is proportional to the bandwidth that the signal sweeps. For instance, FMCW radars need to sweep the sensing signal in a total bandwidth of 1.69GHz [38], while UWB-based systems require 1GHz bandwidth [37]. Recent works in mmWave [39] based sensing make use of a higher frequency band to avoid the multipath issue. They are promising given their high sensing resolution, but are more complex and expensive than the DMA in achieving a high dimension of antenna patterns.

Instead of having more transceivers or extending the signal bandwidth, MetaSense applies *antenna pattern diversity* to obtain a high dimension of uncorrelated channel measurement for sensing. It achieves high sensing accuracy with only a single transceiver pair operating at a single frequency.

B. DMA-based Imaging

Recent works have proposed the use of DMA for computational imaging [11], [12]. However, they focus on *static object* imaging using *one-dimensional metasurface* with limited antenna pattern diversity. To improve the measurement dimension, they require a wide frequency band to achieve frequency-diverse antenna patterns (e.g., 8GHz [11]) and multiple transceivers (e.g., four transceivers [12]). By contrast, MetaSense enables *fine-grained dynamic motion sensing* using a single transceiver pair with a single carrier frequency.

C. DMA-based Sensing

Leveraging the antenna pattern diversity of DMA for sensing has also been proposed recently [40], [41]. However, the authors only focus on the niche case of binary motion detection as a proof-of-concept (e.g., detecting the presence of motion). By contrast, for the first time, MetaSense demonstrates the use of DMA for fine-grained miniature motion sensing. We build on the literature but advance it by addressing a set of specific challenges that lacked adequate attention in the past, namely, *a complete processing pipeline* to handle noise and misalignment in the DMA signal, *a robust segmentation mechanism* to extract the motion signal, as well as *a quality-aware multi-mask learning framework* that can properly assess and aggregate the high-dimensional measurements of the DMA to improve sensing accuracy at runtime. Using our two-dimensional DMA, we demonstrate the effectiveness of MetaSense on a writing motion recognition task.

D. Domain Adaptation in RF Sensing

As the radio signals are subject to environment and location changes [28], [29], it results in the domain shift problem [30] when using the pre-trained classifier for recognition. Existing works require collecting a large labelled dataset across different sensing scenarios to generalize the recognition model [28], [29], [42]. By contrast, MetaSense incorporates the transfer learning [43] to address the domain adaptation problem in RF sensing. Our evaluation indicates that MetaSense requires only five training instances per class to achieve over 90% accuracy when deployed in new sensing locations.

III. DMA-based RF Sensing

A. Background of the DMA

The Dynamic Metasurface Antenna (DMA) is a novel antenna that offers controllable radiation pattern diversity from a simplified hardware platform [11], [12]. The key enablers of DMA are the metamaterials. Metamaterials were initially proposed as artificial media that were engineered to allow the manipulation of electromagnetic waves in a deliberate and controlled manner [13]. This notion was later adapted to planar counterparts, thus metasurfaces [15]. More recently, metasurfaces excited by a guided mode (instead of a plane wave) have been considered, giving rise to metasurface antennas. A DMA, which is a subclass of metasurface antennas, is usually a single-port waveguide exciting a set of sub-wavelength-sized metamaterial radiators integrated into its top layer. Each of the embedded metamaterial elements radiates a portion of the energy from the waveguide into free space, and therefore, the overall radiation pattern of the DMA is the superposition of the contributions from all excited elements. The electromagnetic response of each metamaterial element can be altered to control the amplitude and the phase of the radiated signal (hence, *dynamic* metasurface antenna). The operation of each element is programmable using simple external electronic controls. Thus, by varying the electromagnetic response of the metamaterial elements and switching different sets of elements to radiate, the DMA provides dynamic radiation

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

pattern diversity without the need for power-hungry phase shifters, amplifiers, and other RF components that are required in conventional phased array antennas. More importantly, metamaterial-based antenna designs can dramatically reduce the inter-element coupling effect that occurs in conventional large-scale dense antenna arrays [17], and thus ensure higher antenna efficiency and diversity [18].

B. RF Sensing Primer

Below, we revisit the wireless channel theory to get some intuition in using wireless signal for motion sensing. Considering the case where a stationary transmitter radiates a sinusoidal signal at frequency f to a receiver, while the sensing object is in motion within the transmission range. For instance, as shown in Figure 1(a), a stationary transmitter emits a signal to a receiver while a subject is in motion. The wireless channel between the transmitter and the receiver can be modeled as $Y(f,t) = X(f,t) \times H(f,t)$, where $X(f,t)$ is transmitted signal, $Y(f,t)$ is the received signal, and $H(f,t)$ is the wireless channel response [44]. For a given $X(f,t)$, $Y(f,t)$ is directly affected by the $H(f,t)$, which is sensitive to environment variations. Any small perturbations in the environment will notably change the structure of the wireless channel and cause variations in $H(f,t)$. Due to the multipath effect, $H(f,t)$ is modeled as the weighted sum of all paths' channel responses. If there are N paths, $H(f,t)$ can be expressed as [32], [44]:

$$H(f,t) = \sum_{k=1}^{N} w_k A_k(f,t) e^{-j2\pi \frac{d_k}{\lambda}}, \quad (1)$$

where w_k is the corresponding weight of the kth path, $A_k(f,t)$ is the complex-valued representation of the amplitude and the initial phase offset, d_k is the path length, λ is the wavelength, and $e^{-j2\pi \frac{d_k}{\lambda}}$ is the phase offset due to the propagation delay. Moreover, we can divide the N paths into two groups, the static and the dynamic paths [9], [45]. As shown in Figure 1(a), $H_S(f,t)$ represents the overall channel response of the static paths, which include the line-of-sight path and the paths reflected by the static objects in the environment. When the subject moves by a small distance, $H_S(f,t)$ does not change. $H_D(f,t)$ is the channel response of the dynamic paths that are reflected by the moving subject. As shown in Figure 1(a), the moving person creates a path length change of Δd in the dynamic path, which then leads to a phase change of $e^{-j2\pi \frac{\Delta d}{\lambda}}$ in $H_D(f,t)$. The overall channel response $H(f,t)$ is the sum of $H_S(f,t)$ and $H_D(f,t)$. As shown in Figure 1(b), the subject's movement results in a variation of ΔH in the overall channel response. By capturing the wireless channel variation, we can detect the motion of interest.

C. Boosting Sensing Accuracy Using DMA

1) Motivation in DMA-based sensing: based on Equation 1, we can further define the received signal $Y(f,t)$ as:

$$Y(f,t,\mathbf{u}) = \sum_{k=1}^{N} w_k \cdot g(f,\mathbf{u}) \cdot \alpha(f,\psi_k), \quad (2)$$

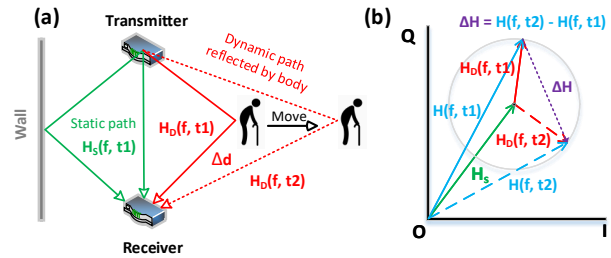


Fig. 1: (a) Channel variations due to human movements; (b) The phasor representation of the variations.

where N is the number of paths, $\mathbf{u}$ is the location of the receiver, and $g(f,\mathbf{u})$ is a term determined by frequency f and location $\mathbf{u}$. For the kth path, $\alpha(f,\psi_k)$ is the product of the transmitter antenna pattern, $\alpha_t(f,\psi_k)$, and the receiver antenna pattern, $\alpha_r(f,\psi_k)$, in direction ψ_k, and w_k is the corresponding weight. Note that the basic principle of boosting sensing accuracy is to have a high dimension of uncorrelated measurement of $Y(f,t,\mathbf{u})$ that can provide sufficient information about the sensing target. Without the reliance on multiple transceivers and a wide frequency band, the remaining variables we can tune in Equation 2 are the *carrier frequency* f and the *antenna pattern* $\alpha(f,\psi_k)$:

- **Frequency hopping.** The first potential solution is switching the carrier frequency. For instance, assuming a 20MHz bandwidth, we can change f among the center frequencies of the 3 and 24 non-overlapping channels of the 2.4GHz and 5GHz WiFi, respectively, to boost measurement dimension (non-overlapping channels are required to ensure low signal correlation). However, achieving fast channel switching is non-trivial. For WiFi-based systems, the default channel switching mechanism in 802.11 protocol induces several seconds of delay [46], which is far beyond the sensing requirement.
- **Antenna pattern diversity.** Alternatively, if we can program the antenna to rapidly change its pattern, we can ensure measurement diversity by having different $\alpha(f,\psi_k)$. This motivates us to use the DMA as either the transmitter or the receiver, or even both, for RF sensing. For instance, with the DMA as the transmitter, we can generate a variety of radiation patterns on the transmitter, $\alpha_t(f,\psi_k)$, to probe the N multiple paths with different weights. As will be shown in Section VIII-B, using a single DMA as the transmitter, we can easily obtain a 200-dimension of uncorrelated channel measurements to boost the sensing accuracy.

IV. System Overview

System design. Figure 2 shows the overview of MetaSense which contains: (1) the *DMA transmitter* and (2) the *Sensing unit*. The DMA transmitter is incorporated with our custom designed DMA and a mask controller which dynamically configures the DMA to send wireless signals with different antenna patterns (Section V). The reflected wireless signal captured by the dipole antenna is used as the input for sensing. The sensing unit includes two major components: the signal processing pipeline (Section VI) and the quality-aware multi-mask sensing framework (Section VII). The former

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

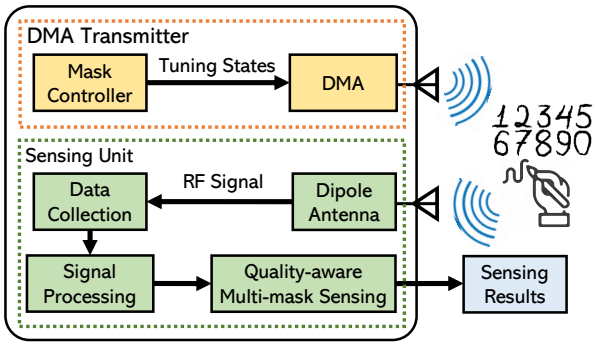


Fig. 2: The overview of MetaSense which contains a *DMA transmitter* and a *sensing unit*.

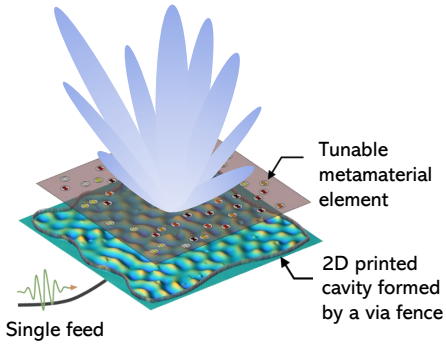


Fig. 3: Schematic of the DMA with an example of the radiation pattern generated by the metamaterial elements.

is a complete processing pipeline which handles denoising, segmentation, and motion alignment for the DMA signal; the latter is a recognition framework that can dynamically assess and aggregate the diverse measurements from DMA to boost sensing accuracy.

Fine-grained writing movement sensing. As a case study, in this paper MetaSense is designed to recognize the ten Arabic numerals written by a tiny drawing robot (details will be given in Section VIII). The motions of the robot will cause variations in the signal captured by the receiver. As the motions in writing different digits affect the wireless signal differently, we use those differences to recognize the digit that has been written. Recognition of the writing motions can facilitate many contactless human-computer interaction applications. For instance, a user can draw digits in the air to interact with smart-home appliances (e.g., smart TV) in a contact-free and non-intrusive manner. The use of the robot allows us to generate *miniature movement* with *high randomness*. As will be shown in Section VIII, the robot ensures *2,500 possible ways* in drawing the same digit in a 2cm×2cm drawing area which is *300 times smaller* than the 35cm×35cm gesture moving area considered in related work [47], [48]. The random and minute robot movements make our task more challenging than recognizing human drawing. Moreover, this robot-based measurement methodology allows for the exact replication of all our experiments by other research groups, and permits data collection when human interactions are restricted, such as throughout COVID-19 shelter-in-place orders. Note that MetaSense is not limited to the application of robot writing recognition as presented in this paper. With minor tuning efforts, the same design can be easily adapted to other contexts, such as activity recognition and gesture recognition.

V. DMA Transmitter

A. Hardware Design

Figure 3 shows the schematic design of our DMA. Overall, the DMA is a single-fed, electrically-large cavity with controllable metamaterial elements patterned into the front radiating surface. The device is excited by a single coaxial probe which feeds the radio waves into a planar cavity formed by an irregularly shaped via cage. The radio waves bounce around inside the cavity before leaking out through the metamaterial elements. These radiating elements thus project the wave formed inside the cavity into the wireless channel. The superposition of the waves from all the radiating metamaterial elements forms the overall radiation pattern. To realize a dynamic response, each metamaterial element is loaded with a PIN diode, giving rise to a binary response (radiating or not radiating). The radiating status for each of the metamaterial elements are addressed independently by applying simple DC voltage signals. Thus, by selecting different sets of elements to radiate, we can create distinct radiation patterns in a simple programmable fashion.

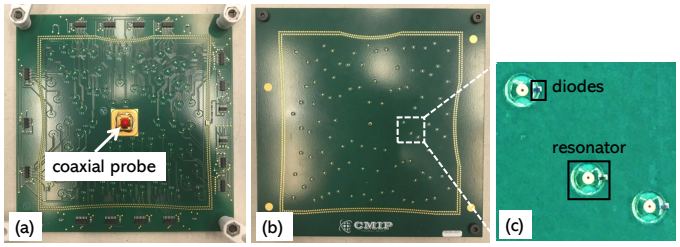


Fig. 4: DMA hardware design and implementation. (a) Back-end of the device. A coaxial probe is used to feed the radio wave into the device. (b) Front-end of the DMA. There are 96 metamaterial elements patterned randomly on the surface. (c) The details of a metamaterial element.

B. Implementation

Figure 4 shows the implementation of the DMA. The device has a form-factor of approximately 15cm×15cm×3mm. The front-end is embedded with 96 metamaterial elements. Each of the elements is an electrically-small, complementary electric-LC resonator. This metamaterial element design has been proven to exhibit high radiation efficiency while maintaining low Ohmic losses and low cross-polarized radiation [49]. In addition, a PIN diode is added to the resonator to control its radiating state. In our implementation, each of the 96 metamaterial elements is controlled externally by the DC voltage provided by an Arduino microcontroller. The tuning states of all the elements determine the overall radiation pattern that will be generated by the front-end metasurface. Thus, by binary tuning the DC voltage applied to the elements (i.e., 0V for not radiating and 5V for radiating), the DMA allows $2^{96} = 7.9 \times 10^{28}$ radiation patterns with a single RF chain. In this paper we call different tuning states of the 96

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

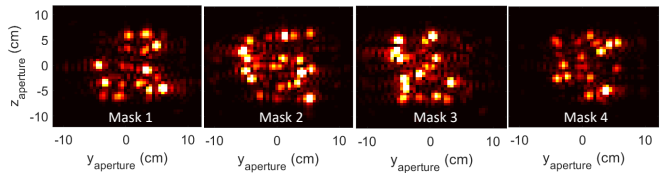


Fig. 5: Example of four different DMA masks. Each of the bright spots is a radiating metamaterial element.

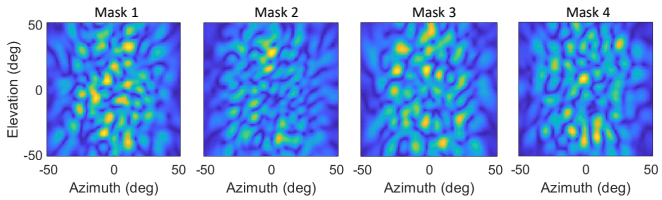


Fig. 6: The four resulting radiation patterns for different DMA masks.

elements as different DMA ***masks***. Each ***mask*** configuration corresponds to a different DMA radiation pattern. As an example, Figure 5 shows four DMA masks. Each of the bright spots in these plots corresponds to a radiating metamaterial element. Figure 6 shows the resulting radiation patterns of the four DMA masks. Clearly, the radiation patterns change with the mask tuning state. In addition to pattern diversity, the device is capable of changing the mask configuration at a MHz rate. When different DMA masks are used for sensing, the changes in the motion being monitored are almost negligible within the time duration of a few hundreds mask switches.

VI. Signal Processing

In this section, we present the design of the signal processing pipeline which is used to prepare the raw DMA signal for the recognition.

A. Denoising

The raw wireless signal captured by the dipole antenna is noisy. The noise results from the ambient human movements as well as minor imperfections of the antenna hardware.

We apply the Discrete Wavelet Transform (DWT) [50] to filter time-domain and frequency-domain noise contained in the raw signal. DWT is able to resolve the signal at different frequency ranges and provides good resolution in both time and frequency domains. It performs a hierarchical transformation that transforms the raw signal into multiple frequency levels called wavelet levels. For each wavelet level, DWT calculates the *detail coefficients* and the *approximation coefficients* which correspond to the high and low frequency components in the signal, respectively. The key insight in DWT-based noise filtering is to modify the coefficients of the signal based on the estimated cut-off thresholds in different wavelet levels. Below, we describe the denoising procedure.

First, we apply the Daubechies D4 wavelet on the raw signal to compute the level 5 coefficients. The selection of level 5 is based on the sampling frequency we used and the frequency of the targeted motion. Since we sample the wireless signal from each of the DMA masks at 500Hz (details are given

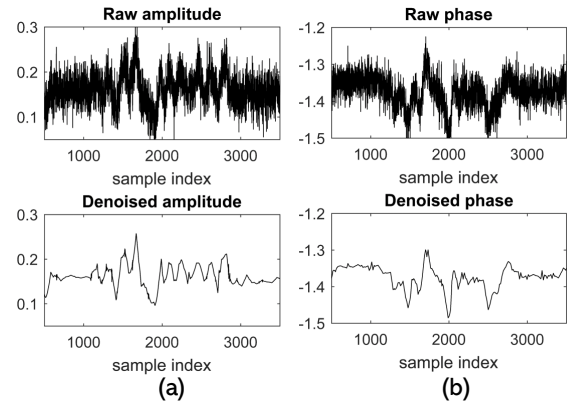


Fig. 7: Example of denoising: (a) and (b) compare the amplitude and the phase of the raw and the denoised DMA signal.

in Section VIII-A), the highest frequency component in the measured signal is 250Hz. Moreover, based on the fast Fourier transform (FFT), we notice that the frequency of the writing motions is bounded by 7Hz. During DWT decomposition, the frequency span halves every DWT level [50], and thus, the level 5 coefficients represent the frequency range of [0, $250/2^5$] Hz, i.e., [0, 7.8] Hz, which accommodates the targeted frequency range of [0, 7] Hz. Second, we apply the soft-thresholding method [51] to calculate the cut-off threshold based on the Stein's unbiased risk estimate. Then, we compare the decomposed level 5 coefficients with the estimated threshold and set all detail coefficients with values below the threshold to 0. Finally, we apply inverse DWT on the modified coefficients to reconstruct the denoised signal. As an example, Figures 7 (a) and (b) compare the signal before and after the DWT denoising. The signal is smooth after denoising.

B. Motion Detection and Segmentation

After denoising, we identify and extract the motion signal from the entire time-series data. Following the widely used assumption [47], [52], [53], we assume that there is a short pause before and after each motion. As an example, Figure 8 shows the amplitude and the phase of the denoised wireless signal which contains two writing motion segments. The statistical properties of the signal (both amplitude and phase) are stable within the pause duration, but change abruptly within the motion segments. To extract the motion segments from the time-series data, we first apply changepoint analysis [54] on the denoised signal to identify the changepoints. Formally, consider a data sequence, $y_{1:n} \triangleq (y_1, ..., y_n)$. A changepoint is said to occur at sample index τ, such that the statistical properties of the split data sequences $\{y_1, ..., y_\tau\}$ and $\{y_{\tau+1}, ..., y_n\}$ are different. In practice, there could be d changepoints in the target signal. They split the original time-series into $d+1$ segments, with the ith segment containing the data sequence of $y_{(\tau_{i-1}+1):\tau_i}$, where τ_i is the sample index for the ith changepoint. The changepoints are identified by finding d data points that minimize the target function $\sum_{i=1}^{d+1} \left[\mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right)\right] + \beta d$, where $\mathcal{C}$ is the cost function and βd is the penalty term. In our implementation, the input data sequence $y_{1:n}$ is a two-dimensional time-series which

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

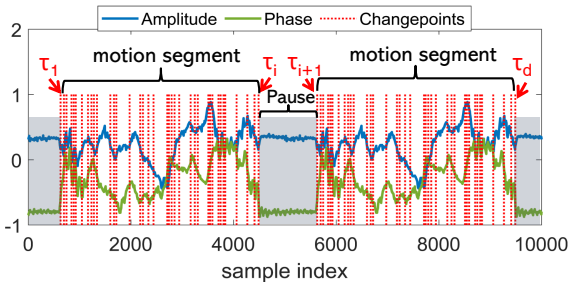


Fig. 8: Example of motion detection and segmentation. Two motions separated by a pause. The red dotted lines indicate the sample indices of the identified changepoints.

contains the amplitude and the phase of the denoised DMA signal. We use the root mean square (RMS) as the statistic for the cost function, and apply the pruned exact linear time algorithm [54] to detect the changepoints. As the number of changepoints d is unknown, we add the penalty term βd to the target function to avoid over-fitting [55]. As shown in Figure 8, the changepoint detection algorithm estimates a set of sample indices, $\tau_{1:d}$, at which the RMS of the signal has changed abruptly. The start and the end index of the motion segment are estimated by finding any two adjacent indexes, τ_i and τ_{i+1} in $\tau_{1:d}$, that satisfy $(\tau_{i+1} - \tau_i) > K$. The value of K equals to the number of samples in the shortest possible pause. In our implementation, given the sampling frequency of 500Hz and a shortest pause duration of two seconds, K is set to 1000. Following this rule, as shown in Figure 8, we can easily identify indices τ_i and τ_{i+1} as the start and the end points of the pause. The index of the first changepoint τ_1 and the estimated index τ_i will be the start and the end index of the first motion segment, respectively. The segmentation performance of the proposed algorithm is evaluated in Section VIII-C.

C. Motion Alignment

The variations in the writing motions and experimental setups lead to three types of misalignment in the segmented signals: variations in the transmission power of the DMA, variations in writing speed, and variations in the writing size. As shown in Figure 9, the variations result in either *temporal misalignment* or *amplitude misalignment* in the received signal. Specifically, different writing speeds and writing sizes lead to varying signal duration which makes the same writing signal mismatched in the time dimension (i.e., temporal misalignment). Similarly, different power levels affect the resolution of the received signal and cause amplitude shifts in the same writing pattern (i.e., amplitude misalignment). We apply signal transformation techniques to resolve the signal misalignment. The alignment process is illustrated in Figure 10. First, we apply the Z-score transformation [56] on the original signal to minimize the amplitude variation. The Z-score transformation returns the z-score for each data sample in the original signal, such that the transformed signal follows the standard normal distribution. Figures 10(a) and (b) compare the signal before and after the Z-score transformation. The amplitude shifts are eliminated and the two signals are converted to the same

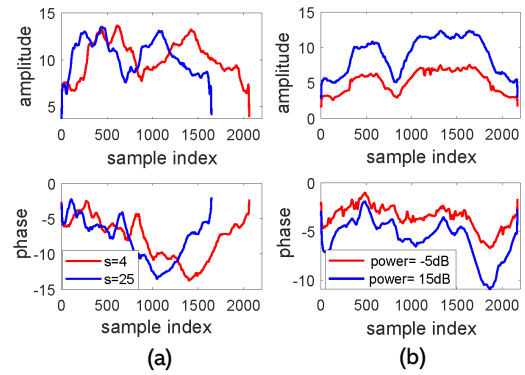


Fig. 9: Example of the signal misalignment: (a) the *temporal misalignment* due to different writing speeds s, (b) the *amplitude misalignment* due to different power levels.

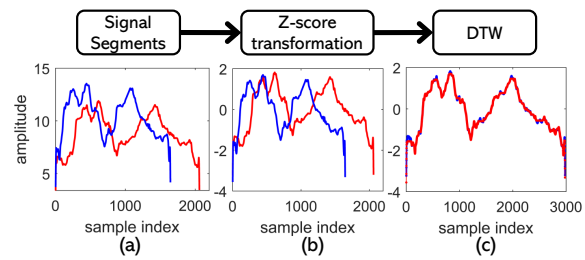


Fig. 10: Alignment process. (a) Original signal. (b) Signal after Z-score. (c) Signal after Z-score and DTW.

scale. Then, we apply the Dynamic Time Warping (DTW) [57] to cope with the temporal mismatch. The final outputs are shown in Figure 10(c), in which the misalignment in the two signals is minimized. The signal alignment process allows us to minimize the negative impacts from different practical parameters on the recognition performance.

VII. Quality-aware Multi-mask Sensing Framework

The antenna pattern diversity of the DMA provides a fruitful measurement for sensing. To ensure robust and high recognition accuracy, we propose the Auxiliary-assisted Ensemble Multi-Mask Learning (AEMML) framework which can properly assess and compare the sensing quality of different DMA masks, and dynamically aggregate them based on the estimated quality in runtime.

A. Auxiliary-assisted Ensemble Multi-mask Learning

Figure 11 shows the overview of the AEMML which incorporates three major components in its design: (1) Convolutional Neural Network (CNN) based *mask learner* for feature extraction and first-level recognition, (2) *auxiliary feature* for runtime DMA mask sensing quality assessment, and (3) *stacking-based multi-mask learning* which dynamically aggregates the heterogeneous recognition results from the mask learners to boost the sensing accuracy.

DMA mask learner. The AEMML contains m independent *mask learners*, where m equals to the total number of DMA masks used for sensing. The mask learners are homogeneous

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

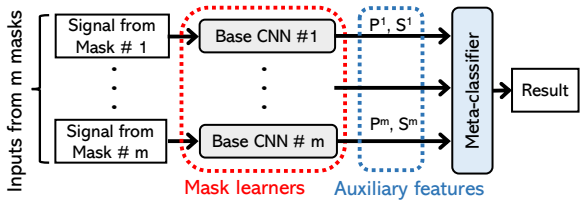


Fig. 11: The architecture of the Auxiliary-assisted Ensemble Multi-Mask Learning (AEMML).

TABLE I: The network architecture of the mask learner.

Layer	Size In	Size Out	Filter
conv1	64 × 32 × 2	64 × 32 × 64	3 × 3, 1
conv2	64 × 32 × 64	62 × 30 × 64	3 × 3, 1
pool	62 × 30 × 64	31 × 15 × 64	2 × 2, 2
Flatten	31 × 15 × 64	29760	
fc1	29760	512	
fc2	512	10	

in architecture but trained independently using the signal from different DMA masks, such that a particular learner is designed to learn the features for a specific DMA configuration. The mask learners adopt CNN-based architecture given its ability in learning a proper data representation from high-dimensional input [58]. The network architecture of the mask learner is shown in Table I, which consists of two convolutional layers (*conv1* and *conv2*), one pooling layer (*pool*), one flatten layer, and two fully connected layers (*fc1* and *fc2*). We use this shallow design to avoid over-fitting. We apply the spline interpolation on the pre-processed signal to make the input data have the same length of 2048. Then, we reshape it to the size of 64×32. The size of the final input to the first convolutional layer is 64×32×2, where the two channels correspond to the amplitude and the phase signals, respectively.

Runtime mask assessment. Given the antenna pattern diversity of the m DMA masks, the m mask learners are unequal in their recognition performance. One way to assess and compare the performance of different mask learners is to leverage the confidence values reported by the classifiers [59]. In our design, borrowing the concept of certainties from information theory, we propose the *normalized entropy* as the metric to assess the sensing quality of different mask learners. For the kth mask learner, the normalized entropy $\mathcal{S}_i^k$ measures the learner's confidence on the classification of the ith instance, and is defined as $\mathcal{S}_i^k(\mathcal{P}_i^k) = -\sum_{i=1}^{|\mathcal{J}|} \frac{p_i \log p_i}{\log |\mathcal{J}|}$, where $\mathcal{P}_i^k = \{p_1, ..., p_{|\mathcal{J}|}\}$ is the probability vector of the kth mask learner on the ith instance, and $|\mathcal{J}|$=10 is the set of ten possible writing digits. The normalized entropy values are between 0 and 1. A value close to 0 indicates that the mask learner is confident about its classification on the instance, whereas a value close to 1 means that it is not confident. The normalized entropy is used as the auxiliary feature to assess the quality and confidence of the heterogeneous mask learners.

Stacking-based multi-mask learning. Although the normalized entropy captures the sensing quality of different DMA masks, aggregating the m sensing outputs to boost the final prediction accuracy is not straightforward. A naive solution is to use the normalized entropy as the weight and apply either weighted average or weighted majority voting to combine the m outputs. However, these methods are sensitive to the biases

Algorithm 1 AEMML training and classification

Training input: (1) Dataset $\mathbb{D} = \{\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_m\}$, in which $\mathcal{D}_k = \{(x_1^k, y_1), ..., (x_n^k, y_n)\}$ is the pre-processed RF signal for the kth mask; (2) m mask learners: $\mathsf{CNN}_1, \mathsf{CNN}_2, ..., \mathsf{CNN}_m$; and (3) the meta-classifier META.

1: # **Training of the m mask learners**
2: **for** $k = 1, ..., m$ **do**
3: $\quad h_k = \mathsf{CNN}_k(\mathcal{D}_k)$; ▷ train CNN_k using data $\mathcal{D}_k$
4: # **Training of the meta-classifier**
5: $\mathfrak{D} = \emptyset$; ▷ initiate a new dataset for META
6: **for** $i = 1, ..., n$ **do** ▷ iterate over the n training instances
7: $\quad$ **for** $k = 1, ..., m$ **do** ▷ iterate over the m masks
8: $\quad\quad \mathcal{P}_i^k = h_k(x_i^k)$; ▷ probability vector of the kth CNN
9: $\quad\quad \mathcal{S}_i^k = -\sum_{j=1}^{|\mathcal{J}|} \frac{\mathcal{P}_i^k(j) \log \mathcal{P}_i^k(j)}{\log |\mathcal{J}|}$; ▷ normalized entropy
10: $\quad \mathfrak{D} = \mathfrak{D} \cup ((\mathcal{P}_i^1, ..., \mathcal{P}_i^m), (\mathcal{S}_i^1, ..., \mathcal{S}_i^m), y_i)$;
11: $h' = \mathsf{META}(\mathfrak{D})$; ▷ train META using $\mathfrak{D}$
12: # **Multi-mask stacking for recognition**
Classification output: $y' = h'(h_1(x^1), ..., h_m(x^m))$;

of the mask learners with respect to their heterogeneous input signals [60], and result in higher prediction error. To reduce the biases of the mask learners and boost the final prediction accuracy, we borrow the concept of stacking [61] and use a meta-classifier for multi-mask aggregation. In our design, the meta-classifier is a shallow neural network with three fully connected layers. As shown in Figure 11, the meta-classifier can be considered as a second-level recognizer that is trained to combine the predictions of the first-level mask learners. It takes the probability vectors and normalized entropy of the m mask learners as the input, and outputs the aggregated result.

Putting all together. The details of the AEMML are shown in Algorithm 1. The training contains two major steps. First, we use the signal of the kth DMA mask, $\mathcal{D}_k = \{(x_1^k, y_1), ..., (x_n^k, y_n)\}$, to train the kth mask learner. The dataset $\mathcal{D}_k$ contains n training instances, in which x_i^k is the pre-processed signal of the ith writing instance, and y_i is the corresponding label. In the second step, we use the probability vectors, i.e., $\{\mathcal{P}_i^1, ..., \mathcal{P}_i^m\}$, that output from the m mask learners (i.e., base CNNs) to train the meta-classifier META. In addition, we calculate the normalized entropies, $\{\mathcal{S}_i^1, ..., \mathcal{S}_i^m\}$, from the probability vectors as the auxiliary information to quantify the classification performance of the mask learners. This runtime mask assessment allows the meta-classifier to dynamically adjust the stacking weights during the classification. As the mask learners are unequal in accuracy, AEMML is more robust against this variance when compared to the standard stacking methods which assign equal weights to the base CNNs during the aggregation.

B. Transfer Learning for New Sensing Domains

The proposed AEMML framework takes advantages of the superior feature learning and classification capabilities of CNNs to ensure good sensing performance. Such capabilities, however, rely on the availability of abundant labeled training instances that cover diverse sensing conditions. Moreover, as shown in Equation 2, the radio signals captured by the receiver are not only affected by the motion of the sensing object, but also the environment and physical location where the receiving wireless signals are measured. Consequently, when the sensing

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

location or environment changes, the features of the receiving signal also change. This is known as the domain shift problem [30], which significantly degrades the sensing accuracy of pre-trained DNN models. Thus, existing DNN-based RF sensing systems try to generalize the classification model by collecting datasets across a large number of environments [29], which is expensive and inefficient.

To efficiently generalize AEMML to different locations and environments in a data-efficient manner, we employ transfer learning [43] to transfer knowledge from a pre-trained source domain (e.g., a known environment or location) to a new target domain (e.g., a new deployment location). Specifically, we divide the AEMML architecture into *general layers* and *domain-specific layers*. The general layers include the first two convolutional layers of the mask learner, i.e., *conv1* and *conv2*. The domain-specific layers include the two fully connected layers of the mask learner, i.e., *fc1* and *fc2*, and all layers in the meta-classifier. The underlying principle of our design is that the low-layers of the DNNs are known to learn features that are not specific to the training dataset or task, but are general and applicable to datasets or tasks in different domains. On the other hand, features computed by the last layers of the DNNs depend greatly on the training dataset and task [43]. When deployed in the target domain (e.g., a new sensing environment or location), the AEMML model inherits the general layers directly from a pre-trained model (e.g., AEMML model that is well-trained with sufficient training instances collected at a specific location) and only fine-tunes the domain-specific layers with a small number of new instances collected from the target location (e.g., five or ten instances per class). This transfer learning-based AEMML design significantly reduces the number of model parameters that need to be trained for different sensing domains, and dramatically reduces the number of required training instances to adapt the framework to new domains without sacrificing recognition accuracy.

VIII. Evaluation

A. Experimental Setup

Hardware setup. We use the DMA as the transmitter and a dipole antenna as the receiver. As shown in Figure 12(a), to form a single-device design, the DMA and the dipole antenna are placed closely together. The signal of the DMA is transmitted at a single frequency of $f = 19.4$GHz with 10dBm transmission power. The power is *20 and 100 times lower* than the default 23dBm and 30dBm power used in WiFi (5GHz channel) and mmWave-based sensing systems [39], [62], respectively. We configure the DMA with 40 randomly selected masks to enable 40 distinct radiation patterns (which we find is sufficient to ensure high accuracy in this case study). The wireless signal is measured by the dipole antenna at a sampling rate of 20KHz (with each DMA mask sampled at 500Hz). The measurement is a 40-dimensional complex-valued time-series, where each dimension corresponds to the receiving signal (amplitude and phase) of a particular mask.

Data collection. As shown in Figure 12(b), we use the drawing robot from Line-us [63] to perform digit writing. The robot is controlled wirelessly by a G-code [64] program.

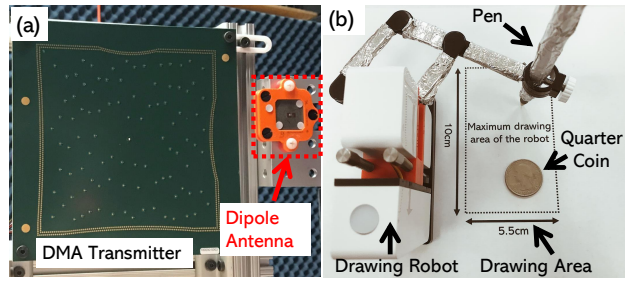


Fig. 12: (a) The setup of the DMA transmitter and dipole antenna, and (b) the drawing robot used in the experiment.

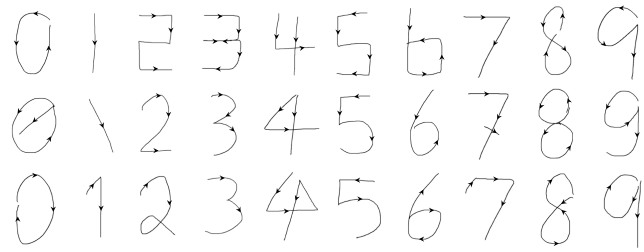


Fig. 13: The ten digits that the robot is programmed to draw. We consider three variants for each of the digits. The arrows on the digits indicate the motion of the pen.

As the writing habits vary among users, we take the MNIST handwritten digits dataset [31] as the reference to design different drawing patterns. Figure 13 shows the ten digits that the robot is programmed to draw. We consider *three variants* for each of the digits. The arrows on the digits indicate the movement of the pen during drawing. As shown in Figure 14, we consider *five deployments* of the sensing device and the robot: three line-of-sight distances (i.e., P1, P2, and P3) and three angles (i.e., P1, P4, and P5). For each of the deployments, we program the robot to draw 20 times for each variant of the ten digits. In total, we collect $10\times3\times20 = 600$ drawing instances for each of the five deployments, resulting in 3,000 instances in total. We use a central controller to coordinate the DMA mask switching, the robot drawing, and the signal measurement of the dipole antenna.

Robot control. The use of the robot allows us to generate *miniature movement with high randomness*. The robot is programmed to draw digits at random *size*, *speed*, and *starting position* (i.e., the position where the pen starts to draw). For each of the digit variants, we consider ten distinct speed levels, ten distinct drawing sizes, and 25 distinct starting positions, which results in *10×10×25=2,500 possible ways* to draw the same digit. Moreover, as shown in Figure 12(b), the largest and smallest drawing sizes of our robot are 5.5cm×10cm (i.e., 5.5cm horizontal and 10cm vertical) and 2cm×2cm, respectively, which are *20 to 300 times smaller* than the 35cm×35cm gesture moving area considered in related works [47], [48], and make our task more challenging.

B. Sensing Signal Correlation

Below, we examine the sensing signal correlation of the conventional WiFi CSI-based system, and compare it with that of our DMA-based solution. Note that we are not comparing the

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

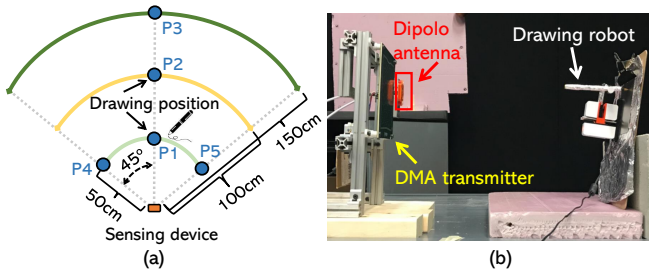


Fig. 14: (a) The five different deployments of the drawing robot. (b) The setup at P1. The robot is placed at a line-of-sight distance of 50cm from the sensing device.

recognition accuracy of WiFi-based and DMA-based sensing systems directly, as they are using radio signal with different frequencies. Instead, as an approximation, we compare their sensing signals in both dimension and correlation. This is because a high dimension of *uncorrelated signal* is critical in boosting the final recognition accuracy [34]. For a given dimension of signals, the lower their correlation, the more complementary information can be captured, and thus the higher the final recognition accuracy.

Strong correlation in WiFi CSI. First, we set up a pair of WiFi transceivers equipped with the TL-WND3800 wireless adapters. Both transceivers have two antennas, and thus form four transmission links. The transceivers are separated with one meter line-of-sight distance to ensure good signal quality. We use the Atheros CSI tool [65] to collect the WiFi CSI at 300Hz rate. Both transceivers are configured at the 5GHz channel with 20MHz bandwidth for communication, and thus, result in a *56-dimensional* time-series measurement (each dimension corresponds to one of the 56 subcarriers) for each of the four links. We apply the DWT-based denoising method to filter out the noise in the raw CSI measurement (as noise is always uncorrelated), and then calculate the correlation matrix. Figure 15 shows the correlation matrices of the CSI amplitude for the four transmission links. We can see that the subcarriers in the same link have high positive correlation (i.e., with correlation coefficient above 0.8), and the correlations are higher between successive subcarriers.

Weak correlation in DMA signal. In comparison, we configure the DMA with 200 randomly selected masks as the transmitter, and use a dipole antenna as the receiver to measure the wireless channel. The signal from each of the masks is measured at 300Hz sampling rate. This provides us a *200-dimensional* time-series measurement. We filter out the noise in the measurement and calculate the correlation matrix. Figure 16 shows the correlation matrices among the 200 dimension of DMA measurements in amplitude and phase, respectively. Distinct from the WiFi CSI, the results indicate weak correlation (i.e., correlation coefficient below 0.2) among the 200 DMA masks.

Advantage of DMA in RF sensing. The high correlation among the WiFi subcarriers indicates a high redundancy in the CSI measurement, which severely limits the signal diversity and recognition performance of WiFi-based system, especially in the 'sensing dead zone' [8], [9]. To obtain a high

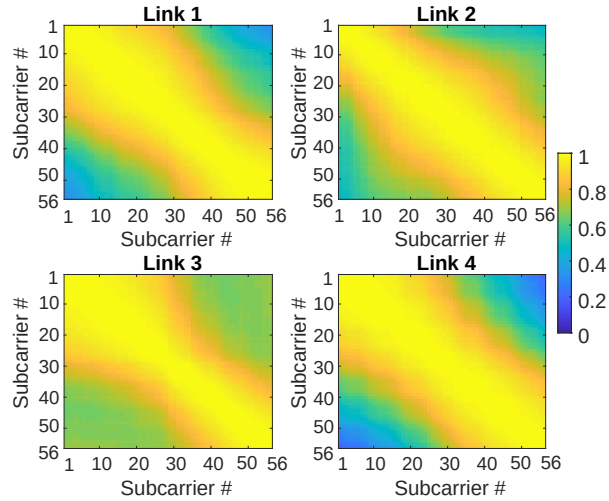


Fig. 15: Correlation matrices of 56 CSI subcarriers for the four WiFi links.

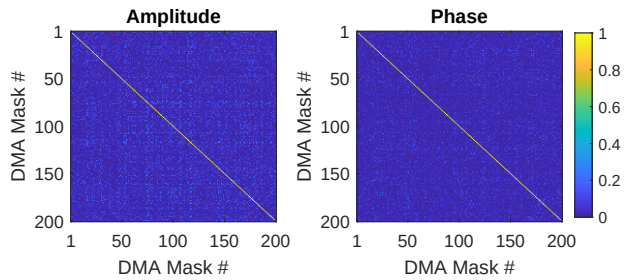


Fig. 16: Correlation matrices of the amplitude and phase of the 200 randomly selected DMA masks, respectively.

dimension of uncorrelated sensing inputs, existing solutions are stitching multiple antennas in a single-device [3], [33] or leveraging multiple transceivers [2]. The state-of-the-art solution SWAN [3] is stitching 12 antennas on a single device, and thus, can provide 12×12=144 streams of uncorrelated channel measurement using a pair of devices. By contrast, using a single DMA as the transmitter, we can easily obtain a 200-dimensional uncorrelated input for sensing (and thus, a pair of DMAs can easily boost the dimension to 40,000). As will be shown in the following evaluation, the weak correlation ensures *largely disjoint failure conditions* among different DMA masks, and thus, boosts the final sensing accuracy.

C. Performance of Motion Segmentation

Below, we evaluate the segmentation algorithm presented in Section VI-B. We are interested in the *detection ratio*, defined as the total number of correctly detected and segmented writing motions divided by the total number of actual motions the robot has performed. If the algorithm fails to detect the appearance of a writing motion, we consider it as an error. Moreover, if it detects the motion but mistakenly segments it, we also count it as an error. The segmentation performance is quantified by the variation in both amplitude and phase of the captured wireless signal, as the algorithm relies on the RMS of the signal to detect the changepoints. The heatmap in Figure 17(a) shows the detection ratio of the ten digits across 40 masks. A darker area in the heatmap indicates a

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

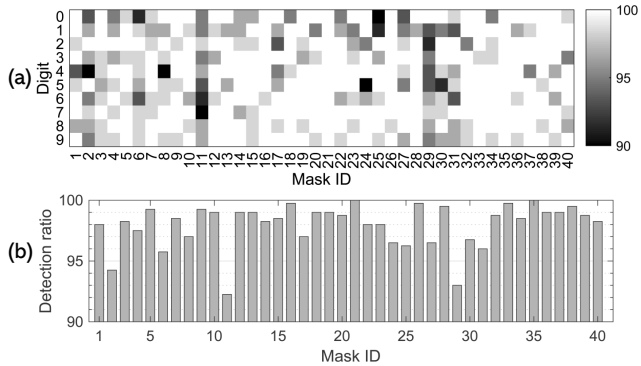


Fig. 17: (a) Heatmap of the detection ratio given different DMA masks and digits. (b) The average detection ratio per mask over all 10 digits.

higher error rate. The error rates vary for different DMA masks depending on their radiation pattern and signal quality. For instance, masks #11 and #29 have an 8% detection error, whereas masks #21 and #35 achieve 100% detection rate over all 10 digits. The overall detection ratio of our algorithm is 98% averaged over all masks and digits. The errors are mainly due to the mistakes in signal segmentation, especially when the writing motion is minute (i.e., the drawing size of the digit is 2cm×2cm).

D. Recognition Performance of AEMML

1) Methodology: We compare the performance of the proposed AEMML with the following baseline methods:

- Conventional machine learning algorithms, i.e., Decision Tree (DT) and Support Vector Machines (SVM), that have been widely used in RF-based recognition [4]. The SVM is configured with the Gaussian kernel function to ensure good performance in dealing with nonlinear features [66]. Unlike AEMML which uses the pre-processed signal as the input for training and classification, we extract two sets of features from the phase and amplitude as the input for DT and SVM: we consider four statistical features, peak factor, wave factor, coefficient dispersion, and autocorrelation coefficient, which are introduced in [47], to capture the *time-domain* characteristics; we adopt the widely used Daubechies D4 wavelet to decompose the raw signal and extract the wavelet detail coefficients in the 5th level as the features.
- To investigate the advantages of AEMML in dynamic multi-mask stacking, we implement the Ensemble Multi-Mask Classifier (EMML) as the variant for comparison. For EMML, the auxiliary feature is not used by the meta-classifier and the mask learners are considered equally in the stacking process.

The final classification results are obtained using 3-folds cross-validation where each fold contains 200 instances (i.e., 20 instances for each of the ten digits). For both AEMML and EMML, the three folds are used for training the base CNNs, training the meta-classifier, and testing, respectively. For DT and SVM, only one fold is needed for training and one fold is used for testing. Below, we evaluate our system in the following four aspects: first, we take the deployment at position P1 as an example to study the impact of DMA antenna diversity on the recognition performance. Second, we compare the AEMML with the other three classifiers to prove its advantages in dynamic multi-masks stacking. Lastly, we study how the device positioning, i.e., distance and orientation, affects the recognition accuracy.

2) Advantages of DMA antenna diversity: First, to study how the recognition accuracy can be improved by the antenna diversity, we randomly select m masks ($m \leq 30$) from the original 40-mask measurements for training and testing. As the selection of the m masks will affect the recognition result, we repeat the experiment 40 times (each time with randomly selected m masks) and report the averaged result as the final accuracy. We examine the case where only the amplitude or the phase is used as the input for classification, as well as the case where both amplitude and phase are used. The results are shown in Figure 18. First, *relying on a fixed antenna pattern for sensing gives the worst accuracy*. As shown in Figure 18(a), in the single mask cases, the average recognition accuracy for SVM and DT can be as low as 50%, and the accuracy for both AEMML and EMML is only 70%. However, for all scenarios, the sensing accuracy increases with the number of DMA masks used – i.e., *the accuracy increases with the increase in antenna pattern diversity*. As discussed, the recognition accuracy of a single DMA mask is quantified by the channel variation. With diverse antenna patterns, we are more likely to find several masks that can provide *complementary and disjoint features* to ensure good sensing performance. As an illustration, Figure 19 shows the recognition accuracy of the 40 mask learners. The accuracy varies from 64% to 82% given different DMA configurations. This confirms the heterogeneity of the sensing signals from different DMA masks. The accuracy variance also indicates *largely disjoint failure conditions* among the 40 mask learners in the recognition. However, despite the heterogeneity of recognition accuracy, there are eight masks that achieve more than 80% accuracy. Thus, by leveraging DMA's antenna diversity, we can dramatically improve the accuracy.

3) Advantages of AEMML: Figure 18 also compares the recognition performance among the four classifiers. In all scenarios, AEMML and EMML outperform the conventional classifiers (i.e., SVM and DT). This is expected, as the former two methods adopt our CNN-based mask learner for feature extraction and recognition, whereas the latter two rely on manually crafted features (i.e., statistical features and DWT coefficients) for classification. For AEMML and EMML, their gain in performance comes from the ability of CNN in automatic feature learning and data representation [28], [67].

Moreover, we can see that AEMML achieves over 6% improvement compared to the standard ensemble method (i.e., EMML), and up to 12.5% improvement compared to the best conventional classifier (i.e., SVM). The advantage of AEMML is most distinct when the number of masks used is small (i.e., $m = 5$), and diminishes with more masks used. To explain, Figure 19 shows the uneven recognition accuracy for the 40 base CNNs. As the conventional methods weight the m masks equally, with a small number of masks used, the performance of the classifier is more likely to be affected by the 'bad' mask learners (e.g., mask learner #11, #26, or #36).

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

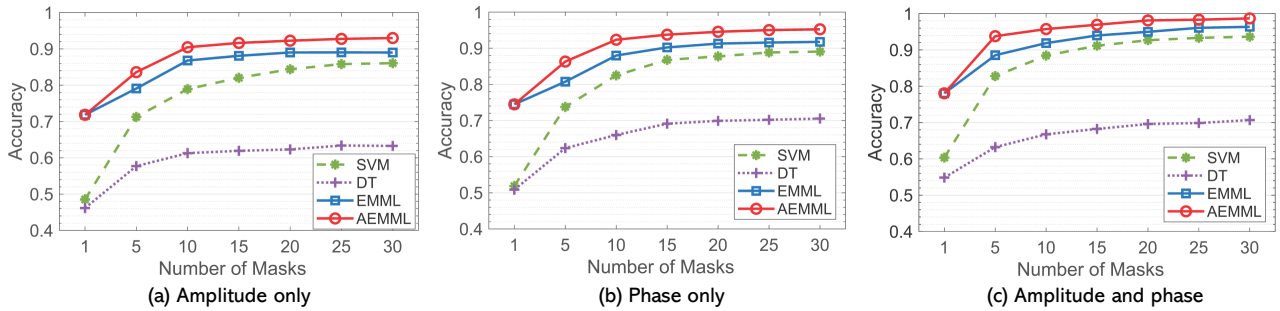


Fig. 18: Recognition accuracy at position P1 for different classifiers and different number of DMA masks (m).

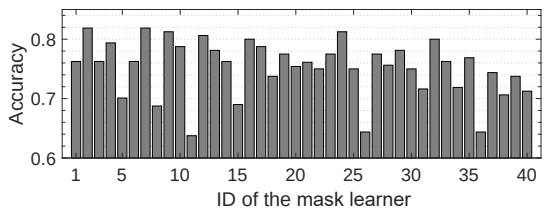
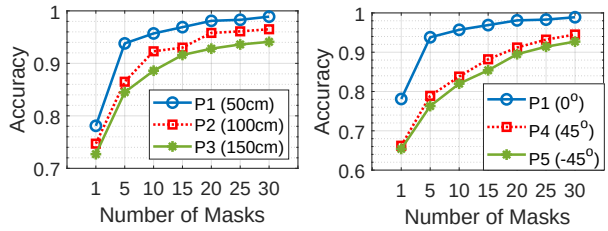


Fig. 19: Recognition accuracy for the 40 mask learners. The accuracy varies among different DMA masks.



(a) Accuracy at different distances (b) Accuracy at different angles

Fig. 20: The accuracy of AEMML given different locations of the sensing target: (a) distance and (b) angle.

By contrast, the auxiliary feature enables the AEMML to learn and discriminate the quality of different masks on the specific recognition task, and allows it to dynamically integrate the multi-mask outputs in a better way. Thus, AEMML is more robust as it learns to assign less weight to the 'bad' masks and more weight to the 'good' masks. The negative impact of 'bad' masks diminishes when more masks are combined. The result demonstrates the superiority of AEMML over the conventional methods.

4) Effects of target location: Below, we evaluate the performance of AEMML given different locations. We fix the position of the sensing device and place the robot at five different positions shown in Figure 14. The distance and angle determine the *sensing coverage* of the system. The results are shown in Figure 20.

Distance coverage. As the receiving signal strength attenuates with the propagation distance, the recognition accuracy also decreases with the distance. In our experiment, given 10dBm transmission power, when the distance increases from 50cm to 150cm, MetaSense experiences a minor 4.8% accuracy decrease with $m = 30$ masks used for sensing. We can still achieve 94% accuracy in recognizing the miniature robot motion at 150cm. Note that the 10dBm transmission power used in this experiment is *20 and 100 times lower* than the default 23dBm and 30dBm power used in WiFi (5GHz channel) and mmWave-based sensing system [39], respectively. Therefore, by increasing the DMA transmission power, we can expect a longer sensing distance.

Angle coverage. In the single mask case we notice a 12% and a 13% decrease in the accuracy when the drawing robot has a 45° or -45° angle difference with the sensing device, respectively. This is because the DMA is not configured to generate directional antenna patterns, and most of the energy is radiated towards the direct front of the antenna (i.e., 0°). Thus, the signal that is reflected by the drawing robot becomes weaker when there is a large angle difference between the robot and the sensing device (i.e., 45° or -45°). However, by leveraging DMA's radiation diversity, the imperfection can be resolved with more mask used. Overall, we achieve over 93% accuracy with $m = 30$ at all locations.

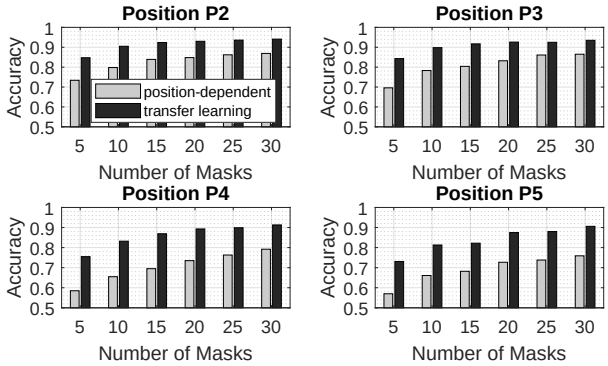
E. Performance in New Sensing Locations with Limited Training Samples

Below, we evaluate the performance of AEMML in scenarios where limited number of training samples are available. Specifically, we consider position P1 as the source domain with sufficient training instances available (20 instances per class), and the other four positions P2-P5 as target domains with only a few-shot training instances (i.e., 5-shot and 10-shot cases where only five and ten instances are available per class, respectively). The experiment simulates the practical scenario where the system is extended to sense motion performed at new positions or deployed in new environments with limited training samples provided by the user.

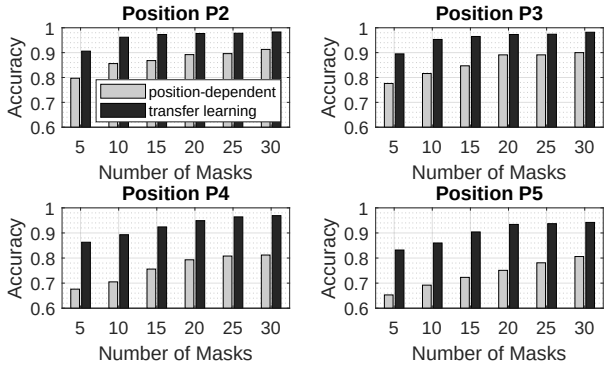
We compare the sensing accuracy of two training strategies: (1) we only use the few-shot samples, i.e., 5-shot and 10-shot, from each of the four target positions to train the AEMML and test it using the remaining data collected from the target position. This represents the *position-dependent* training strategy; (2) we first pre-train the AEMML model using the source domain dataset collected from position P1. Then, we incorporate AEMML with the transfer learning model introduced in Section VII-B. Specifically, we transfer the general layers of the pre-trained model and only fine-tune the domain-specific layers using the few-shot instances from the target domain (e.g., new sensing position). We also use the remaining data collected from the target position as the testing dataset. This represents the *transfer learning* training strategy. Figure 21 compares the accuracy of AEMML at the four target positions with different number of masks used as the

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

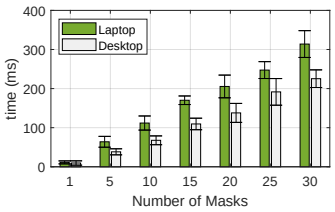


(a) Accuracy of AEMML in 5-shot scenario.

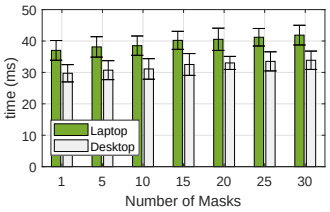


(b) Accuracy of AEMML in 10-shot scenario.

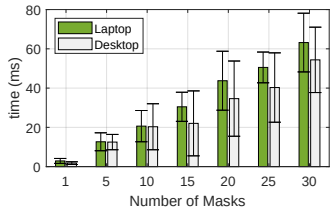
Fig. 21: The accuracy of AEMML at different sensing locations with a different number of training instance available: (a) 5-shot scenario, and (b) 10-shot scenario.



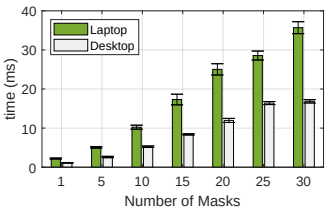
(a) Denoising



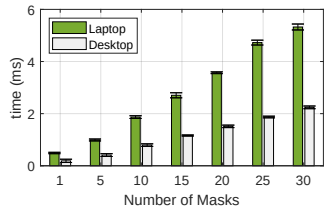
(b) Motion detection and segmentation



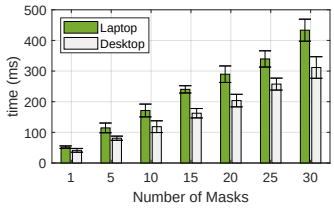
(c) Motion alignment



(d) AEMML on CPU



(e) AEMML on GPU



(f) End-to-end latency

Fig. 22: The average computation latency of different system components on two platforms: (a) denoising, (b) motion detection and segmentation, (c) motion alignment, (d) AEMML running on CPU, (e) AEMML running on GPU, and (f) end-to-end latency with AEMML running on GPU. The error bars indicate the standard deviation of the measured latency for the 500 trials. Overall, with signals from 30 masks used as the input, the end-to-end latency of MetaSense is 433ms and 311ms when running on the laptop and the desktop, respectively.

sensing input. Overall, the proposed transfer learning-based strategy outperforms the position-dependent strategy by 7-18% and 8-19% in the 5-shot and 10-shot scenarios, respectively. As shown in Figure 21, together with the help of the DMA antenna pattern diversity, MetaSense can still achieve over 90% accuracy with as few as five training instances available per class when deployed in new sensing positions. The results demonstrate the capability of MetaSense to generalize to new sensing environments and locations in a data-efficient manner.

F. System Profiling

Below, we provide a comprehensive profiling of the system in terms of computation latency and runtime memory usage. Specifically, we use a desktop equipped with an Intel i7-8700k CPU and an Nvidia GTX 1080 GPU to simulate an edge server, and leverage a laptop embedded with an Intel i7-7700HQ CPU and an Nvidia GTX 1050 GPU to simulate the next-generation home appliances. Moreover, we only enable a single CPU core among the four cores of the laptop to approximate the computational power of a smart TV. In practice, Samsung SMART TVs are equipped with 1.3GHz Quad core processor and the AI Quantum-series processor [68], while Sony has incorporated the X1-series processor in their smart TV [69]. Both of them are more powerful than the single core laptop CPU we considered in this measurement and should achieve a lower latency. We have torn down the system pipeline into four computing stages: (1) denoising, (2) motion detection and segmentation, (3) motion alignment, and (4) classification using the AEMML framework. We realize the first three system components in Matlab and deploy them on the CPU. The Matlab-based implementation ensures good computational efficiency in signal processing where large arrays and matrices are involved, and can be easily deployed on both low-end IoT devices and on an edge server [70]. The AEMML framework is implemented using Keras 2.3 on top of the TensorFlow 2.0 framework and is tested on both CPU and GPU.

1) Computation latency: To examine the computation latency, we run 500 trials of the end-to-end system pipeline and report the average computation latency for each of these computing stages.

Figure 22 shows the average latency on the two platforms when a different number of DMA masks is used for sensing. The error bars in the plots indicate the standard deviation of the measured latency for the 500 trials. We have three major observations. First, the latency for all the computing stages

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

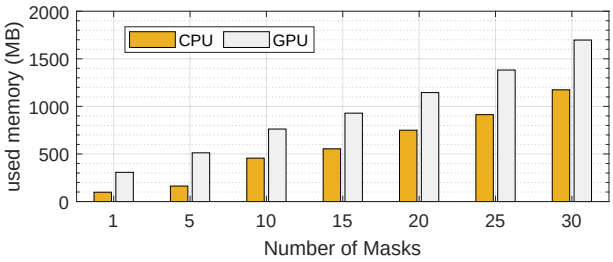


Fig. 23: Runtime memory usage (in MB) of the AEMML when running on CPU and GPU, respectively.

increases with the number of masks used. This is because, when a higher dimension of DMA signal is used as the input, it requires more computations during the signal processing and more inferences in the AEMML which make the computation more intensive. Second, as shown in Figures 22(d) and (e), the latency of the AEMML is significantly lower when running on the GPU than when running on the CPU. For both platforms we examine, GPU achieves at least five times speedup over CPU regardless of the number of masks used. Lastly, the desktop achieves the lowest latency in all the computing stages. Overall, with the signal from 30 masks used as the input, the end-to-end latency of MetaSense is 433ms and 311ms when running on the laptop and desktop, respectively.

2) Memory usage: We also evaluate the runtime memory usage of the AEMML framework. We subtract the memory usage before the framework is loaded and only report the memory that is allocated to the framework and the inference data (the preprocessed signal for recognition). Note that the memory usage of DNNs is determined by the input data size and the network model (e.g., the weight parameters and activations). For a given number of masks used as the input, the memory usage of AEMML is deterministic. Figure 23 compares the memory usage of AEMML when running on CPU and GPU. Even with 30 masks used as the inputs, the AEMML framework requires a modest memory usage of 1174MB and 1697MB when running on the CPU and GPU, respectively.

IX. Discussion and Future Directions

Evaluation with real subjects in different sensing applications: in our current evaluation, we leverage the programmable robot to generate repeatable and miniature digit writing movements. This robot-based setup also allows us to conduct comprehensive and reproducible experiments when human interactions are restricted. However, despite the high randomness of the robot, the movements of the robotic arm constitute only a finite set of real-world writing motions. In practice, the writing movements of real subjects are more diverse and have higher degrees of freedom. Thus, one of the future directions is to evaluate the proposed system with real subjects. Moreover, in addition to the writing recognition considered in the current work, we believe that MetaSense can be easily adapted to many other sensing applications with minor tuning efforts. In future work, we will investigate the use of MetaSense in applications such as daily activity recognition [32], respiration detection [8], and user authentication [2].

Passive and energy-neutral operation: thanks to its small form-factor and hardware simplicity, the DMA can be easily embedded into walls and daily small objects. Instead of using the DMA as an active RF sensor, we believe that another promising use of the DMA in RF sensing is to leverage it as a passive signal reflector to improve the sensing performance of existing deployments. For instance, as a reflector with antenna reconfigurability, the metasurface antenna can assist high-frequency bands solutions, e.g., millimeter-wave-based systems [71], to extend the non-line-of-sight sensing coverage by reflecting the original signal into the desired direction. Moreover, for a single pair of Wi-Fi devices with limited antenna pattern diversity [34], [72], the RF signal backscattered from the DMA can help the Wi-Fi receiver obtain a higher dimensional channel measurement (i.e., the DMA backscatters the same incoming RF signal with different antenna patterns, and thus ensures signal diversity in the receiving signal). Indeed, a similar concept has been envisioned recently to boost the wireless communication performance [20], [21], [71], [73], [74], where a metasurface is used to configure the electromagnetic behavior of a wireless environment. In addition, as metamaterial elements can be made out of passive elements that do not require any active power sources for transmission [75], the DMA can be potentially powered by energy harvesting solutions [76], [77] to achieve energy-neutral operation.

X. Conclusion

This paper presents MetaSense, the first system that achieves fine-grained RF sensing with a single transceiver pair and a single frequency. It exploits the antenna pattern diversity of the Dynamic Metasurface Antenna (DMA) to ensure high-dimensional sensing measurements. We implement MetaSense and evaluate its performance on a fine-grained writing recognition task. Our experiments show that MetaSense can achieve over 93% accuracy in different settings, outperforming the non-tunable antenna by 20% in all scenarios. Moreover, when deployed in new sensing positions where limited training data are available, MetaSense requires as few as five training instances per class to achieve over 90% accuracy.

Acknowledgment

The work of M. Gorlatova is partially supported by NSF grant CSR-1903136 and the Lord Foundation of North Carolina. The work of D. R. Smith is partially supported by the Advanced Research Projects Agency-Energy (ARPA-E), US Department of Energy, under Grant DE-AR0000937. The work of W. Hu is partially supported by NSF grant CNS-1763212. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] D. Pine, D. Weitz, P. Chaikin, and E. Herbolzheimer, "Diffusing wave spectroscopy," *Physical Review Letters*, vol. 60, no. 12, pp. 1134–1137, 1988.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

[2] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT," in *Proceedings of ACM MobiHoc*, 2017.

[3] Y. Xie, Y. Zhang, J. C. Liando, and M. Li, "SWAN: Stitched WiFi antennas," in *Proceedings of ACM MobiCom*, 2018.

[4] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys*, vol. 52, no. 3, pp. 46:1–46:36, 2019.

[5] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.

[6] M. Gowda, A. Dhekne, S. Shen, R. R. Choudhury, L. Yang, S. Golwalkar, and A. Essanian, "Bringing IoT to sports analytics," in *Proceedings of NSDI*, 2017.

[7] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proceedings of NSDI*, 2014.

[8] 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 ACM UbiComp*, 2016.

[9] K. Niu, F. Zhang, J. Xiong, X. Li, E. Yi, and D. Zhang, "Boosting fine-grained activity sensing by embracing wireless multipath effects," in *Proceedings of ACM CoNext*, 2018.

[10] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 949–973, 2015.

[11] J. Hunt, T. Driscoll, A. Mrozack, G. Lipworth, M. Reynolds, D. Brady, and D. R. Smith, "Metamaterial apertures for computational imaging," *Science*, vol. 339, no. 6117, pp. 310–313, 2013.

[12] T. Sleasman, M. F. Imani, J. N. Gollub, and D. R. Smith, "Dynamic metamaterial aperture for microwave imaging," *Applied Physics Letters*, vol. 107, no. 20, p. 204104, 2015.

[13] J. B. Pendry, D. Schurig, and D. R. Smith, "Controlling electromagnetic fields," *Science*, vol. 312, no. 5781, pp. 1780–1782, 2006.

[14] D. R. Smith, O. Yurduseven, L. P. Mancera, P. Bowen, and N. B. Kundtz, "Analysis of a waveguide-fed metasurface antenna," *Physical Review Applied*, vol. 8, no. 5, p. 054048, 2017.

[15] C. L. Holloway, E. F. Kuester, J. A. Gordon, J. O'Hara, J. Booth, and D. R. Smith, "An overview of the theory and applications of metasurfaces: The two-dimensional equivalents of metamaterials," *IEEE Antennas and Propagation Magazine*, vol. 54, no. 2, pp. 10–35, 2012.

[16] E. Saenz, I. Ederra, R. Gonzalo, S. Pivnenko, O. Breinbjerg, and P. de Maagt, "Coupling reduction between dipole antenna elements by using a planar meta-surface," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 2, pp. 383–394, 2009.

[17] X. M. Yang, X. G. Liu, X. Y. Zhou, and T. J. Cui, "Reduction of mutual coupling between closely packed patch antennas using waveguided metamaterials," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 389–391, 2012.

[18] S. Soltani and R. D. Murch, "A compact planar printed MIMO antenna design," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 3, pp. 1140–1149, 2015.

[19] "Metamaterial devices market 'to exceed 10BN dollars by 2030'," https://optics.org/news/10/8/32, Aug. 2019.

[20] C. Liaskos, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "Using any surface to realize a new paradigm for wireless communications," *Communications of the ACM*, vol. 61, no. 11, pp. 30–33, 2018.

[21] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 129, 2019.

[22] "Kymeta," https://www.kymetacorp.com/.

[23] "Echodyne," https://echodyne.com/.

[24] "Metawave," https://www.metawave.co/.

[25] S. R. Das, "Metamaterials arrive in cellphones," *IEEE Spectrum*, Oct. 2009, https://spectrum.ieee.org/telecom/wireless/metamaterials-arrive-in-cellphones.

[26] "Huawei and NTT Docomo achieves a new breakthrough in 5G mmWave long-distance mobility trial over 39 GHz band," https://www.huawei.com/en/press-events/news/2017/12/NTT-DOCOMO-5G-mmWave-Trial, Dec. 2017.

[27] "A faster, more resilient 5G network," https://www.parc.com/blog/a-faster-more-resilient-5g-network%E2%80%A8/.

[28] X. Li, Y. Zhang, I. Marsic, A. Sarcevic, and R. S. Burd, "Deep learning for RFID-based activity recognition," in *Proceedings of ACM SenSys*, 2016.

[29] 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 ACM MobiCom*, 2018.

[30] M. Sugiyama and A. J. Storkey, "Mixture regression for covariate shift," in *Advances in Neural Information Processing Systems*, 2007, pp. 1337–1344.

[31] "The MNIST database of handwritten digits," http://yann.lecun.com/exdb/mnist/.

[32] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proceedings of ACM MobiCom*, 2015.

[33] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proceedings of NDSI*, 2013.

[34] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "MD-track: Leveraging multi-dimensionality in passive indoor WiFi tracking," in *Proceedings of ACM MobiCom*, 2019.

[35] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: virtual touch screen in the air using RF signals," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 235–246, 2015.

[36] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proceedings of ACM MobiCom*, 2014.

[37] A. Dhekne, M. Gowda, Y. Zhao, H. Hassanieh, and R. R. Choudhury, "Liquid: A wireless liquid identifier," in *Proceedings of ACM MobiSys*, 2018.

[38] Y. Tian, G.-H. Lee, H. He, C.-Y. Hsu, and D. Katabi, "RF-based fall monitoring using convolutional neural networks," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 137:1–137:24, 2018.

[39] C. Xu, Z. Li, H. Zhang, A. S. Rathore, H. Li, C. Song, K. Wang, and W. Xu, "WaveEar: Exploring a mmWave-based noise-resistant speech sensing for voice-user interface," in *Proceedings of ACM MobiSys*, 2019.

[40] P. del Hougne, M. F. Imani, T. Sleasman, J. N. Gollub, M. Fink, G. Lerosey, and D. R. Smith, "Dynamic metasurface aperture as smart around-the-corner motion detector," *Scientific Reports*, vol. 8, p. 6536, 2018.

[41] G. Lan, M. F. Imani, P. Del Hougne, W. Hu, D. R. Smith, and M. Gorlatova, "Wireless sensing using dynamic metasurface antennas: Challenges and opportunities," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 66–71, 2020.

[42] X. Wang, X. Wang, and S. Mao, "RF sensing in the internet of things: A general deep learning framework," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 62–67, 2018.

[43] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proceedings of NeurIPS*, 2014.

[44] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[45] N. Patwari, J. Wilson, S. Ananthanarayanan, S. K. Kasera, and D. R. Westenskow, "Monitoring breathing via signal strength in wireless networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1774–1786, 2014.

[46] S. Shi, Y. Xie, M. Li, A. X. Liu, and J. Zhao, "Synthesizing wider WiFi bandwidth for respiration rate monitoring in dynamic environments," in *Proceedings of IEEE INFOCOM*, 2019.

[47] 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.

[48] T. Yu, H. Jin, and K. Nahrstedt, "WritingHacker: Audio based eavesdropping of handwriting via mobile devices," in *Proceedings of ACM UbiComp*, 2016.

[49] I. Yoo, M. F. Imani, T. Sleasman, and D. R. Smith, "Efficient complementary metamaterial element for waveguide-fed metasurface antennas," *Optics Express*, vol. 24, no. 25, pp. 28 686–28 692, 2016.

[50] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.

[51] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[52] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proceedings of IEEE INFOCOM*, 2015.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

[53] D. Ma, G. Lan, M. Hassan, W. Hu, M. Upama, A. Uddin, and M. Youssef, "SolarGest: Ubiquitous and battery-free gesture recognition using solar cells," in *Proceedings of ACM MobiCom*, 2019.

[54] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

[55] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal Processing*, vol. 85, no. 8, pp. 1501–1510, 2005.

[56] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[57] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of ACM KDD*, 1994.

[58] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[59] B. Zadrozny and C. Elkan, "Transforming classifier scores into accurate multiclass probability estimates," in *Proceedings of ACM KDD*, 2002.

[60] K. M. Ting and I. H. Witten, "Stacked generalization: When does it work?" in *Proceedings of IJCAI*, 1997.

[61] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[62] Z. Li, B. Chen, Z. Yang, H. Li, C. Xu, X. Chen, K. Wang, and W. Xu, "FerroTag: a paper-based mmWave-scannable tagging infrastructure," in *Proceedings of ACM SenSys*, 2019.

[63] "Line-us drawing robot," https://www.line-us.com/.

[64] "G-code," https://reprap.org/wiki/G-code.

[65] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity WiFi," in *Proceedings of ACM MobiCom*, 2015.

[66] S.-i. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks*, vol. 12, no. 6, pp. 783–789, 1999.

[67] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of IEEE CVPR*, 2016.

[68] "Samsung AI Quantum processor," https://www.samsung.com/sg/tvs/qled-tv/technology/.

[69] "Sony X1 processor," https://www.sony.com/electronics/picture-quality#SonyProcessors.

[70] "MATLAB for IoT applications," https://www.mathworks.com/solutions/internet-of-things.html.

[71] K. W. Cho, M. H. Mazaheri, J. Gummeson, O. Abari, and K. Jamieson, "A reconfigurable metamaterial surface for mmWave networks," in *Proceedings of ACM HotMobile*, 2021.

[72] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single WiFi link," in *Proceedings of ACM MobiSys*, 2018.

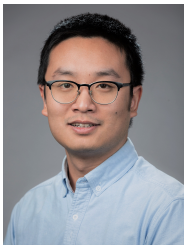
[73] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, 2018.

[74] L. Chen, W. Hu, K. Jamieson, X. Chen, D. Fang, and J. Gummeson, "Pushing the physical limits of IoT devices with programmable metasurfaces," in *Proceedings of the USENIX NSDI*, 2021.

[75] D.-H. Kwon and S. A. Tretyakov, "Arbitrary beam control using passive lossless metasurfaces enabled by orthogonally polarized custom surface waves," *Physical Review B*, vol. 97, no. 3, p. 035439, 2018.

[76] R. J. Vullers, R. Van Schaijk, H. J. Visser, J. Penders, and C. Van Hoof, "Energy harvesting for autonomous wireless sensor networks," *IEEE Solid-State Circuits Magazine*, vol. 2, no. 2, pp. 29–38, 2010.

[77] M. Gorlatova, A. Wallwater, and G. Zussman, "Networking low-power energy harvesting devices: Measurements and algorithms," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1853–1865, 2012.

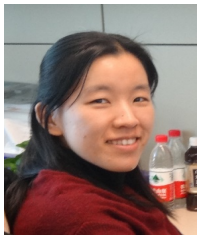


Guohao Lan received the B.E. degree in Software Engineering from Harbin Institute of Technology, China, in 2012, and the M.S. degree in Computer Science from KAIST, South Korea, in 2015, and the Ph.D. degree in Computer Science and Engineering from the University of New South Wales, Australia, in 2018. He is currently a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. His research interests include wireless sensing and mobile computing.

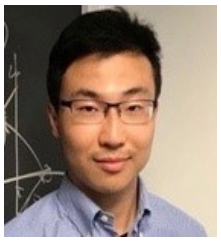
Mohammadreza F. Imani received the B.S.E. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2007 and the M.S.E. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2010 and 2013, respectively. From 2014 to 2018, he has served as a Postdoctoral Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. Currently, he is an assistant professor in the School of Electrical, Computer, and Energy Engineering at Arizona State University, Tempe, Az. His research interests include analytical and applied electromagnetics, metamaterials and metasurfaces, microwave imaging and sensing, and MIMO communication systems. He has authored and co-authored over 70 journal and conference papers and holds 1 granted and 5 pending patent applications.

Zida Liu is currently a Ph.D. student in Computer Science and Engineering department at Pennsylvania State University. He obtained his M.S. degree in the Electrical and Computer Engineering department from Duke University in May 2020. And he got his B.Eng. degree in Telecommunications from Huazhong University of Science and Technology, China, in 2018. His research interests lie in mobile computing, wireless sensing, and Internet of Things.

José Manjarrés received his B.Sc. degrees in Electronics Engineering and Mathematics from Universidad del Norte, Colombia, in 2015, and his Ph.D. (Cum Laude) degree in Electrical and Electronics Engineering from Universidad del Norte, Colombia, in 2020. He is an Assistant Professor of Electrical and Computer Engineering at Olivet Nazarene University. His research interests include wearable systems, applied machine learning, and biomedical sensing.



Wenjun Hu is an assistant professor of Electrical Engineering and Computer Science at Yale University, having previously been a researcher at Microsoft Research Asia and a postdoctoral research associate at the University of Washington. She received her BA and PhD in Computer Science from the University of Cambridge. She is a recipient of a Google Faculty Research Award, the 2018 SIGCOMM Test-of-Time Award, and the 2009 IEEE Bennett Prize.



Andrew Lan is an assistant professor in the College of Information and Computer Sciences at University of Massachusetts Amherst. He received his Ph.D. degree in Electrical and Computer Engineering from Rice University and has worked as a postdoctoral research associate at Princeton University. His research interests lie at the intersection of artificial intelligence, data science, and education, highlighted by a best student paper award at IEEE Big Data 2020. He has also organized a series of workshops on machine learning for education; see ml4ed.cc for details.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3070225, IEEE Internet of Things Journal

David R. Smith received the B.S. and Ph.D. degrees in physics from the University of California at San Diego, San Diego, CA, USA, in 1988 and 1994, respectively. He is currently the James B. Duke Professor of Electrical and Computer Engineering at Duke University and Director of the Center for Metamaterials and Integrated Plasmonics. His research interests include the theory, simulation, and characterization of unique electromagnetic structures, including photonic crystals and metamaterials, as well as applications of such materials. Dr. Smith has provided key experimental demonstrations in the metamaterials field, including the first demonstration of a negative index metamaterial in 2000, and the first demonstration of a metamaterial "invisibility cloak" in 2006. Dr. Smith is a co-recipient of the Descartes Scientific Research Prize (2005) awarded by the European Union and the James C. McGroddy Prize for New Materials, awarded by the American Physical Society (2013). Dr. Smith was elected as a Fellow of the National Academy of Inventors in 2016. Since 2009, Dr. Smith has been listed as a "Highly Cited Researcher" by Clarivate Analytics, in the field of Physics.

Maria Gorlatova is a Nortel Networks Assistant Professor at Duke University Department of Electrical and Computer Engineering in Durham, North Carolina. Dr. Gorlatova's research focuses on enabling next-generation Internet of Things and augmented reality deployments. She earned her Ph.D. in Electrical Engineering from Columbia University. Prior to joining Duke, she was an Associate Research Scholar and an Associate Director of the EDGE Lab at Princeton University Department of Electrical Engineering. In addition, Dr. Gorlatova has extensive industry experience in a variety of R&D and business-oriented roles, and has previously been affiliated with Telcordia Technologies, IBM, and D. E. Shaw Research. Dr. Gorlatova is a recipient of multiple awards including the NSF CAREER Award, the ACM SenSys Best Student Demonstration Award, the IEEE Communications Society Young Author Best Paper Award, the ACM/IEEE IPSN Best Research Artifact Award, and the IEEE Communications Society Award for Advances in Communications.

2327-4662 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Massachusetts Amherst. Downloaded on April 23,2021 at 18:35:30 UTC from IEEE Xplore. Restrictions apply.