MetaSketch: Wireless Semantic Segmentation by Reconfigurable Intelligent Surfaces

Jingzhi Hu, *Graduate Student Member, IEEE*, Hongliang Zhang, *Member, IEEE*, Kaigui Bian, *Senior Member, IEEE*, Zhu Han, *Fellow, IEEE*, H. Vincent Poor, *Life Fellow, IEEE*, and Lingyang Song, *Fellow, IEEE*

***Abstract*— Semantic segmentation is a process of partitioning an image into segments for recognizing regions of humans and objects, which can be widely applied in scenarios such as healthcare and safety monitoring. To avoid privacy violation, using radio frequency (RF) signals instead of photos for semantic segmentation has gained increasing attention. However, traditional human and object recognition by using RF signals is a passive signal collection and analysis process without changing the radio environment. The recognition accuracy is restricted significantly by unwanted multi-path fading, and/or the limited number of independent channels between RF transceivers. This paper introduces MetaSketch, a novel RF-sensing system that performs semantic recognition and segmentation for humans and objects by making the radio environment reconfigurable. A metamaterial-based reconfigurable intelligent surface is incorporated to diversify the information carried by RF signals. Using compressive sensing techniques, MetaSketch reconstructs a point cloud consisting of the reflection coefficients of humans and objects at different spatial points, and recognizes the semantic meaning of the points by using symmetric multilayer perceptron groups. Our evaluation results show that MetaSketch is capable of generating favorable radio environments, extracting exact point clouds, and labeling the semantic meaning of the points with an average error rate of less than 1% in an indoor space.**

***Index Terms*— RF sensing, reconfigurable intelligent surface, semantic segmentation, compressive sensing.**

I. Introduction

In COMPUTER vision, semantic segmentation seeks to partition the pixel set of an image into subsets, with each subset having the same semantic meaning. Owing to its wide applications in public safety and healthcare monitoring scenarios, semantic segmentation has garnered significant interest recently as a powerful tool for simultaneous recognition and localization of humans and objects. Generally, semantic segmentation is conducted over images captured by video cameras and is used to obtain meaningful representations for the images to simplify and facilitate further potential analyses [1].

However, using video cameras to collect images for semantic segmentation inevitably introduces privacy concerns. As a potential solution, recently, using radio frequency (RF) signals for profiling humans and objects has attracted increasing attention. Many RF-sensing systems based on WiFi or millimeter-wave signals have been proposed for recognizing humans and objects [2]–[6], or generating images that can be further used as materials for semantic segmentation [7]–[9]. Nevertheless, the systems in these works only passively adapt to the radio environment. Due to the complicated and uncontrollable nature of radio environments, the accuracy and flexibility of the systems can be affected significantly [10], [11].

Recently, reconfigurable intelligent surfaces (RISs) have been developed as a promising solution to actively customize the undesirable propagation channels into favorable ones [12]. RISs can be used as passive multiple-input multiple-output (MIMO) transmission systems to improve data rates and spatial resolution [13], [14], and also have the potential to improve the accuracy of RF sensing [15]. An RIS is composed of an intelligent controlling circuit and a 2D metamaterial surface which contains a massive number of sub-wavelength electrically controllable elements [16]. Applied with different controlling voltages, an element is able to impose different phase shifts to the signals reflected by it. Thus, by programming its elements, an RIS can reconfigure radio propagation channels, which can focus RF signals for better receiver SNRs [17] or enable RF signals to carry diverse information about humans and objects [15]. Equipped with an RIS, an RF sensing system can potentially achieve more accurate semantic recognition and segmentation results than those of traditional RF-sensing systems.

In this paper, we present MetaSketch, an RIS-based RF-sensing system that can extract a spatial point cloud of reflection coefficients from the received RF signals and perform semantic segmentation on the point cloud to recognize humans and objects. Specifically, via programming the configurations of the RIS, MetaSketch creates independent propagation channels to facilitate the point cloud extraction.

Manuscript received 2 August 2021; revised 17 November 2021 and 10 January 2022; accepted 13 January 2022. Date of publication 26 January 2022; date of current version 12 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61941101 and Grant 61829101, in part by the National Science Foundation (NSF) under Grant CNS-2107216 and Grant EARS-1839818, and in part by Toyota. The associate editor coordinating the review of this article and approving it for publication was R. He. *(Corresponding author: Lingyang Song.)*

Jingzhi Hu and Lingyang Song are with the School of Electronics, Peking University, Beijing 100871, China (e-mail: jingzhi.hu@pku.edu.cn; lingyang.song@pku.edu.cn).

Hongliang Zhang and H. Vincent Poor are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08540 USA (e-mail: hongliang.zhang92@gmail.com; poor@princeton.edu).

Kaigui Bian is with the School of Computer Science, Peking University, Beijing 100871, China (e-mail: bkg@pku.edu.cn).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: zhan2@uh.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3144340.

Digital Object Identifier 10.1109/TWC.2022.3144340

After deployment, MetaSketch requires no video camera to obtain images for segmentation, and thus it is privacy-protecting and has various potential applications in healthcare and security scenarios.

The main challenge of building an RF sensing system for semantic segmentation without cameras is the absence of a method to directly capture humans and objects and to match them to a certain set of received signals. To handle this issue, MetaSketch extracts point clouds directly from processing the RF signals by compressive sensing and performs semantic segmentation based on the point cloud. The design of MetaSketch is structured around three components that together provide an architecture for using compressive sensing and semantic segmentation for RIS-based RF-sensing systems: (1) a *radio environment reconfiguration module* that creates independent propagation channels by using an RIS to facilitate compressive sensing; (2) a *point cloud extraction module* that extracts reflection coefficients of different spatial points; and (3) a *semantic segmentation module* that labels the point clouds with their semantic meanings to recognize humans and objects.

We implement a prototype system and evaluate its semantic segmentation capability over daily scenarios which involve a human and a set of practical objects. Experimental results show that MetaSketch can extract point clouds in space from RF signals and perform semantic segmentation accurately with an average error rate of less than 1%, given the setup of a human and multiple objects in a 1.6 m^3 indoor space represented by 400 evenly distributed points. In summary, the contributions of this work can be listed as follows.

- We propose a novel RF-sensing system named MetaSketch, which has radio environment reconfiguration capability and can perform semantic segmentation for point clouds extracted from received RF signals.
- We design efficient algorithms for MetaSketch to optimize the reconfiguration of the radio environment, to extract reflection coefficient point clouds, and to recognize the point clouds with their semantic meanings.
- We implement a prototype system, and the evaluation results show that MetaSketch is capable of labeling the semantic meaning of the spatial points with an average error rate of less than 1% in an indoor setting.

The rest of the paper is organized as follows. Section II reviews related work on RF-sensing systems and video-image-based semantic segmentation. Section III provides preliminaries for understanding the design of MetaSketch. In Section IV, we describe the system model, including the system components and a protocol to coordinate them. In Section V, we describe the component modules of MetaSketch. In Section VI, we elaborate on the implementation of MetaSketch, and Section VII provides the evaluation results. Finally, we draw conclusions in Section VIII.

II. Related Work

In this section, we summarize related work, including the existing literature on RF-sensing systems and the image-based semantic segmentation technique.

A. RF-Sensing Systems

Recent years have witnessed much interest in RF-sensing systems for human and object recognition. Most existing systems passively adapt to the radio environment and obtain the sensing results by analyzing the influence of human bodies and objects on the RF signals. Different systems have been designed for people localization [18], [19] and particular posture and gesture identification [20]–[24]. Further, RF-sensing has also proven to be feasible for imaging humans and objects with the help of MIMO techniques [7], [25].

In comparison, RIS-based RF sensing systems are capable of reconfiguring multiple RF propagation channels into mutually independent ones [11]. As RF signals traveling over independent channels generally carry more information than those on coherent ones, the proposed system can potentially achieve higher accuracy compared with existing systems [23], [26]. Specifically, in [15], the authors have proposed an RIS-based RF sensing system to recognize human postures with high accuracy. Furthermore, in [27], the authors proposed a millimeter-wave RF sensing system aided by an RIS for imaging an object, which verifies the performance advantages provided by an RIS in RF imaging. Moreover, in [28], the authors proposed RIS-aided RF systems for the localization of user devices.

Nevertheless, it remains a challenge to design an effective RF semantic segmentation system, which can directly recognize the semantic meanings of different spatial points based on received RF signals. Comparing the MetaSketch and the existing works on RIS-aided RF sensing systems, we highlight that the main differences are:

- To the best of the author's knowledge, MetaSketch is the first RF-sensing system model proposed to realize complicated computational vision approaches, such as semantic segmentation, by using pure RF signals.
- Technically, the novelty of MetaSketch lies in that it jointly utilizes the RIS-aided radio environment reconfiguration, compressive sensing, and semantic segmentation to recover a semantically labeled point cloud in space.

B. Image-Based Semantic Segmentation

In image-based semantic segmentation, each segment is a set of pixels of the image which collectively represent one semantically meaningful object. Most semantic segmentation approaches process images captured by 2D video cameras. These approaches are increasingly mature due to the availability of deep convolutional neural networks (CNNs), such as fully CNNs [29], region-based CNNs [30], and deep convolutional encoder-decoders [31].

Different from the increasingly mature semantic segmentation in 2D, the semantic segmentation in 3D is immature and has received much research attention. The authors in [32] proposed to construct human skeletons in 3D from 2D images by using CNNs combined with kinematic skeleton fitting. Further, the authors in [33] explored deep learning architectures capable of reasoning about geometric data other than 2D images such as 3D point clouds and meshes. This work lays the foundation for semantic recognition and

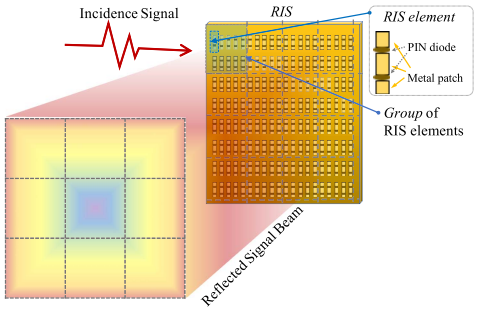


Fig. 1. RIS elements and signal reflection on the RIS.

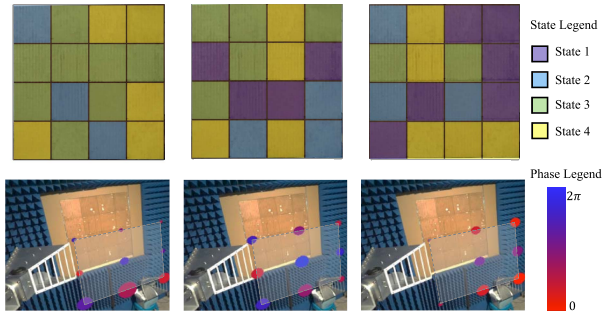


Fig. 2. Configurations of the RIS and the corresponding reflected signals at different positions.

segmentation in RIS-based RF-sensing systems, where the reflection coefficients of spatial points are captured as a point cloud.

III. Preliminaries

In this section, we provide three preliminaries for understanding the feasibility of MetaSketch's design.

A. Feasibility of Reconfiguring Radio Environments by RIS

An RIS is an artificial thin film of reconfigurable electromagnetic materials, which is composed of a massive number of uniformly distributed *RIS elements*. As shown in Fig. 1, the RIS elements are arranged in a two-dimensional array, and each RIS element can adjust its response to the incident RF signals by leveraging positive-intrinsic-negative (PIN) diodes [34]. We refer to the different responses as the *states* of the RIS element. Thus, each state of an RIS element has a unique reflection coefficient for the incident signals, which can be represented by a complex number. The amplitude and phase of the reflection coefficient indicate the amplitude ratio and the phase shift between the reflected and incident signals, respectively [35]. Moreover, to reduce the controlling complexity, *group*s of adjacent RIS elements are controlled together, which means that the states of the elements in a group are set the same. Therefore, the states of all the RIS elements can be indicated by the states of the groups, which is referred to as the *configuration* of the RIS.

Through changing its configuration, the RIS is able to modify the waveforms of the reflected signals and form directional beams [36]. With this beamforming capability, the RIS can reconfigure the radio environment and generate diverse reflection beam patterns in the plane in front of it, as shown in Fig. 2. To show the RIS's capability of reconfiguring the radio environment, we conduct a pilot testing by changing the configuration of the RIS and measuring the reflected signals at 9 different positions on a plane around 1.2 m in front of the RIS. In Fig. 2, the configurations of the RIS are depicted in the upper part, where the four colors indicate four different states of groups. The corresponding reflected signal values are visualized as colored solid circles in the lower part. Specifically, the size of each circle is proportional to the signal amplitude, and the color represents the signal phase. It can be observed that by changing the configuration of the RIS, the amplitude and phase of the reflected signals in the environment can be effectively modified.

B. Feasibility of Extracting Point Cloud via Compressive Sensing

In the following, we demonstrate the feasibility of extracting a point cloud of reflection coefficients from RF signals by using compressive sensing. Specifically, we aim to restore the reflection coefficients at multiple spatial points from the received signals with a limited number of RIS's configurations, which can be expressed mathematically as follows.

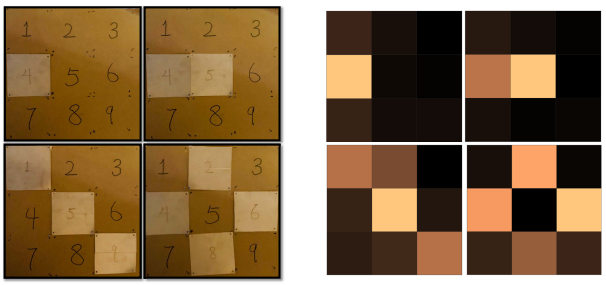
Denote the reflection coefficients at the M spatial points as an M-dim vector $\boldsymbol{\eta}$. Assume that the RIS takes K configurations, and the K corresponding received signals form a received signal vector $\boldsymbol{y}$. Then, denote the mapping between the $\boldsymbol{\eta}$ and $\boldsymbol{y}$ by $\boldsymbol{H}$, which is a $K \times M$ matrix determined by the K configurations of the RIS. Thus, $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\eta} + \boldsymbol{e}$, with $\boldsymbol{e}$ being the noises at the receiver with element-wise variance ϵ, and the point cloud extraction problem can be expressed as to reconstruct $\boldsymbol{\eta}$ based on $\boldsymbol{y}$ and $\boldsymbol{H}$. The point cloud extraction problem is an *underdetermined linear system problem* as $M \gg K$. This is because, in general, the number of spatial points in a point cloud, i.e., M, needs to be large to obtain high spatial resolution. Nevertheless, the number of RF measurements, i.e., K, needs to be limited so that the measurement duration is short enough for real-time measuring.

The underdetermined linear system problems have infinite solutions, which means additional constraints on $\boldsymbol{\eta}$ are needed to reconstruct $\boldsymbol{\eta}$ effectively. In our considered scenarios, the additional constraint we used to help reconstruct $\boldsymbol{\eta}$ is that $\boldsymbol{\eta}$ is a sparse vector. This is because *firstly*, a majority of spatial points do not contain reflectors. *Secondly*, based on [7], only a small number of spatial points have reflectors whose surfaces can reflect signals towards the wireless receiver. Therefore, only a small number of spatial points have non-zero reflection coefficients, which indicates $\boldsymbol{\eta}$ to be sparse.

Given the constraint of $\boldsymbol{\eta}$ being sparse, we can adopt the compressive sensing technique to extract $\boldsymbol{\eta}$ from received signal vector $\boldsymbol{y}$ effectively. The compressive sensing technique seeks to solve the sparsest $\boldsymbol{\eta}$ which satisfies $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\eta} + \boldsymbol{e}$ [37]. Specifically, $\boldsymbol{\eta}$ can be obtained by solving the following l_1-norm minimization problem [38],

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \|\boldsymbol{\eta}\|_1 \quad \text{s.t. } \|\boldsymbol{H}\boldsymbol{\eta} - \boldsymbol{y}\|_2 \leq \epsilon. \tag{1}$$

To verify the feasibility of (1) in practice, we use the RIS-based RF-sensing to reconstruct the reflection



(a) (b)

Fig. 3. (a) The signal plane with 10 cm × 10 cm metal patches at part of the 9 positions; and (b) illustrations of the reconstructed absolute values of the reflection coefficients at different positions by solving (1).

coefficients of 9 rectangle spatial grids. Specifically, we place a 10 cm × 10 cm metal patch at each of the 9 positions on the plane shown in Fig. 2 in turn and obtain 9 received signal vectors when the RIS takes 5 configurations. Based on the received signal vectors, $\boldsymbol{H}$ can be obtained. Then, we place multiple metal patches on the 9 grids, obtain received signal vector $\boldsymbol{y}$, and solve (1) to obtain $\hat{\boldsymbol{\eta}}$, which indicates the average reflection coefficients of the 9 grids with respect to the metal patch.

Fig. 3 (a) shows the photos where the light (yellow) regions are the metal patches, and the dark (brown) regions are the cardboards which have a negligible impact on RF signals. Fig. 3 (b) shows the reconstructed amplitudes of the reflection coefficients at the 9 grids, where the brightness of the color is proportional to the amplitude values. Comparing Figs. 3 (a) and (b), we can observe that solving (1) successfully reconstructs the reflection coefficients at the 9 grids, which verifies the feasibility of extracting point clouds via compressive sensing with the help of an RIS.

C. Feasibility of Semantic Segmentation for Point Cloud

In this paper, we aim to label each point in the point cloud with its semantic meaning. However, as the point clouds are essentially *sets* and should be invariant to changing order, they are different from traditional data structures for semantic segmentation such as pixel images [33]. Thus, traditional semantic segmentation methods based on CNN [29], [39] are not suitable, which rely on spatially ordered input for regional feature extraction and are not insensitive to changing order.

In order to handle the unordered properties of point cloud data, the input data need to be treated symmetrically. As shown in [33], this can be done by using multi-layer perceptrons (MLPs) with shared parameters to treat the feature vector of each point in the point cloud. As shown in Fig. 4, an MLP contains multiple layers of *neurons*. Each neuron takes inputs from the connected neurons in the former layer, handles them by weighted summation with bias and an activation function (e.g., *sigmoid*), and outputs the result value to the next layer. By this means, an MLP is able to obtain a representative feature vector of a point from its position and the extracted reflection coefficient vector. Moreover, since we use the point cloud data for semantic segmentation, it will be more privacy-protecting than existing RIS-aided sensing systems, such as [40], which use images of human as training data.

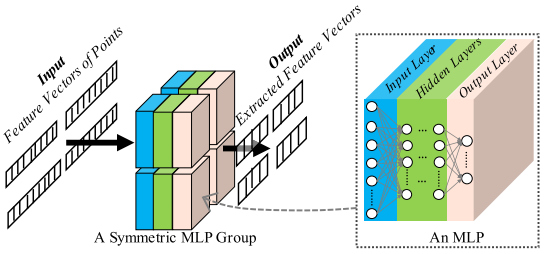


Fig. 4. Illustration on a symmetric MLP group to process the feature vectors in a point cloud.

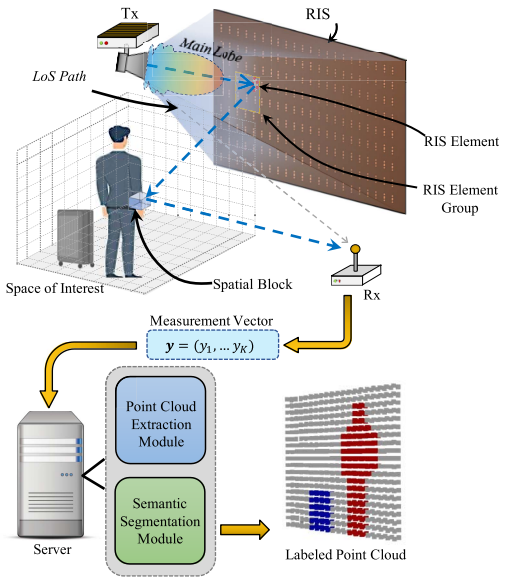


Fig. 5. Component modules of MetaSketch: radio environment reconfiguration is shown in the top, and point cloud extraction and semantic segmentation modules are shown in the bottom.

IV. System Model

Based on the preliminaries in Section III, we introduce the system model of MetaSketch in this section. The MetaSketch is a sensing system that can perform semantic recognition and segmentation for humans and objects based on RF signals. To achieve this goal, MetaSketch uses an RIS to make the radio environment reconfigurable and obtain spatial point clouds of humans and objects by using compressive sensing techniques. In the following, we first describe the components of the MetaSketch and then provide a protocol that coordinates the components.

A. System Components

As illustrated in Fig. 5, MetaSketch contains the following three component modules:

- **Radio environment reconfiguration module**: This module contains a pair of RF transceivers and an RIS, where the transmitter (Tx) and receiver (Rx) have single antennas. The Tx antenna is directional and has its main lobe pointed

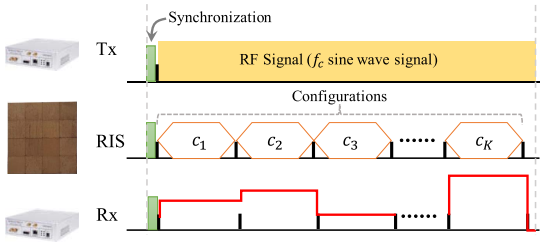


Fig. 6. Illustration on the data collection phase.

toward the RIS. Thus, the majority of the transmitted signals arrive at the RIS, while few of them directly reach the Rx antenna through the line-of-sight (LoS) path. The transmitted signals are reflected by the RIS and then reach the humans and objects in different spatial blocks, carrying the information of them to the Rx. Besides, the Rx antenna is omnidirectional to receive the reflected signals from different positions in the space of interest.

- **Point cloud extraction module**: This module is implemented in the server connected to the Rx and uses the compressive sensing technique to extract the reflection coefficient point cloud from the baseband signals from the Rx.
- **Semantic segmentation module**: This module is also implemented in the server and takes the point cloud obtained in the previous module as the input. This module adopts symmetric MLP groups to label each point in the point cloud with its semantic meaning for human and object recognition.

B. Coordination Protocol

In the following, we propose a protocol to coordinate the component modules of MetaSketch to perform RF-sensing, point cloud extraction, and semantic segmentation. In the protocol, the timeline is slotted and divided into *cycles*, and MetaSketch operates in a synchronized and periodic manner. Each cycle is constituted of two phases: *data collection* and *signal processing* phases. In the following parts, we describe the data collection and signal processing phases in detail.

1) Data Collection Phase: As illustrated in Fig. 6, in the data collection phase, the RIS changes configuration sequentially. The receiver measures the received signals during each configuration and stores them as a vector, referred to as the *measurement vector*. Specifically, at the beginning of a data collection phase, the Tx first transmits a starting signal to the RIS and the Rx for synchronization. Then, the Tx starts to transmit a sine wave signal with frequency f_c, and the RIS changes from the first to the K-th configuration sequentially, which are denoted by $\boldsymbol{c}_1$ to $\boldsymbol{c}_K$. Here, K is the total number of configurations in a data collection phase as shown in Fig. 6, and $\boldsymbol{c}_k$ ($k \in [1, K]$) is an L-dim vector with L being the number of groups of the RIS. Moreover, the K configurations of the RIS constitute a *measurement matrix* $\boldsymbol{C}$, i.e., $\boldsymbol{C} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_K)$. While the RIS can adopt a random $\boldsymbol{C}$, we propose a method to obtain an optimized configuration matrix in Section V-A. At the end of a data collection phase, the Rx generates $\boldsymbol{y}$ by taking the averages of the received signals within each duration of the K configurations. Then, the Rx sends $\boldsymbol{y}$ to the server for point cloud extraction and semantic segmentation.

2) Signal Processing Phase: The signal processing phase follows the data collection phase. Specifically, after receiving the measurement vector generated by the Rx, the server first invokes the point cloud extraction module to extract the point cloud of the humans and objects from the measurement vector. Then, the generated point cloud is processed by the semantic segmentation module, which provides each point with the label representing its semantic meaning. The algorithms used in the point cloud extraction and semantic segmentation modules will be explained in Sections V-B and V-C, respectively.

V. Problem Formulation and Algorithm Design for MetaSketch

In this section, we describe the problem formulation and algorithm design for MetaSketch's three component modules.

A. Radio Environment Reconfiguration

In this section, we describe how the radio environment reconfiguration module of MetaSketch derives its configuration matrix. While random configuration matrices are available, to facilitate the point cloud extraction and semantic segmentation, it requires the configuration matrix to be optimized. Specifically, the information about humans and objects is contained in the reflection coefficients at different spatial blocks. Denote the reflection coefficients by an M-dim vector $\boldsymbol{\eta}$, where M is the cardinality of a set of pre-assigned spatial blocks in the space of interest whose reflection coefficients we aim to restore. The j-th element of $\boldsymbol{\eta}$ indicates the average reflection coefficient in the j-th spatial block. Consequently, we need to optimize $\boldsymbol{C}$ so that $\boldsymbol{\eta}$ can be restored from $\boldsymbol{y}$ with the highest accuracy.

In MetaSketch, though M can be large, most spatial blocks are empty, and thus tend to have zero reflection coefficients. Besides, for the spatial blocks that contain parts of humans and objects, only those with specific angles can reflect the signals towards the Rx antenna and have non-zero reflection coefficients. Therefore, $\boldsymbol{\eta}$ is a sparse vector and can be solved by using the compressive sensing technique. Based on [41], to minimize the loss between the reconstructed $\boldsymbol{\eta}$ and the actual one, we can minimize the average mutual coherence (AMC) of $\boldsymbol{H}$, which is defined as

$$\mu(\boldsymbol{H}) = \frac{1}{M(M-1)} \sum_{m,m' \in [1,M], m \neq m'} \frac{|\boldsymbol{h}_m^T \boldsymbol{h}_{m'}|}{\|\boldsymbol{h}_m\|_2 \cdot \|\boldsymbol{h}_{m'}\|_2}. \quad (2)$$

Here, $\boldsymbol{h}_m \in \mathbb{C}^K$ and $\boldsymbol{h}_m$ is the m-th column of $\boldsymbol{H}$, where $\mathbb{C}$ indicates the set of complex numbers. The i-th element of $\boldsymbol{h}_m$ ($i \in [1, K]$) indicates the influence of an object with normalized reflection coefficient 1 in the m-th spatial block on the received signals, under the k-th configuration of the RIS. In (2), measurement matrix $\boldsymbol{H}$ is determined by $\boldsymbol{C}$, and we can obtain the value of $\boldsymbol{H} = \boldsymbol{g}(\boldsymbol{C})$ according to the Appendix.

Based on (2), we can formulate the optimization problem for the radio environment reconfiguration as the following

Algorithm 1 Configuration Matrix Optimization

Input : Initial random configuration matrix $\boldsymbol{C}^{(0)}$; Initial population size in the genetic algorithm (GA) N_P; Number of generations in the GA N_G; Small value σ to ensure convergence.

Output: Optimal AMC μ^* and configuration matrix $\boldsymbol{C}^*$.

1 Set $\boldsymbol{C}^* = \boldsymbol{C}^{(0)}$, and compute initial $\mu^* = \mu(\boldsymbol{H}^*)$ based on (2) with $\boldsymbol{H}^* = \boldsymbol{g}(\boldsymbol{C}^*)$ as in Appendix;

2 Set the number of consecutive iterations with no improvements as $N_{\text{non}} = 0$ and current frame index $k = 1$;

3 **while** *True* **do**

4 Transform $\boldsymbol{C}^*$ to zero-one matrix $\tilde{\boldsymbol{D}}$ by (13) in Appendix, and denote the k-th row of $\tilde{\boldsymbol{D}}$ as $\tilde{\boldsymbol{d}}_k$ and the other rows as $\tilde{\boldsymbol{D}}_{-k}$;

5 Invoke pattern search algorithm [42] to solve $\tilde{\boldsymbol{d}}_k^* = \arg\min_{\tilde{\boldsymbol{d}}_k \in [0,1]^{L \cdot N_s}} \mu([\tilde{\boldsymbol{d}}_k, \tilde{\boldsymbol{D}}_{-k}]\boldsymbol{A})$, where $\boldsymbol{A}$ is defined in Appendix;

6 Round up $\tilde{\boldsymbol{d}}_k^*$ to discrete configuration vector $\boldsymbol{c}_k'$ by $(\boldsymbol{c}_k')_l = \arg\max_{j \in [1, N_s]}((\tilde{\boldsymbol{d}}_k^*)_{(L-1)N_s + j})$;

7 Invoke genetic algorithm [43] to solve $\boldsymbol{c}_k^* = \arg\max_{\boldsymbol{c}_k \in [1, N_s]^L} \mu(\boldsymbol{g}(\boldsymbol{c}_k, \boldsymbol{C}_{-k}^*))$ with an initial population containing $\boldsymbol{c}_k'$, and denote the resulting AMC as $\mu^{*\prime}$;

8 If $\mu^{*\prime} < \mu^* - \sigma$, update $\mu^* = \mu^{*\prime}$ and the k-th row of $\boldsymbol{C}^*$ to be $\boldsymbol{c}_k^*$; otherwise, set $N_{\text{non}} = N_{\text{non}} + 1$;

9 If $N_{\text{non}} < K$, set $k = \text{mod}(k+1, K) + 1$; otherwise, return μ^* and $\boldsymbol{C}^*$;

10 **end**

AMC minimization problem:

$$\text{(P1)} \quad \min_{\boldsymbol{C}} \ \mu(\boldsymbol{H}), \tag{3}$$

$$s.t. \ \boldsymbol{H} = \boldsymbol{g}(\boldsymbol{C}), \tag{4}$$

$$c_{k,l} \in \{1, \ldots, N_s\}, \quad \forall k \in [1, K], \ l \in [1, L], \tag{5}$$

where N_s denotes the number of states of each RIS element and $c_{k,l}$ denotes the state of the l-th group in the k-th configuration. To solve (P1), we propose a configuration matrix optimization algorithm, which is described in Algorithm 1. The computational complexity and convergence of the proposed algorithm can be analyzed as follows.

1) Computational Complexity Analysis: To evaluate the scalability of Algorithm 1, we analyze its computational complexity with respect to L, N_{s}, M, and K. We consider the worst case where the maximum number of iterations is reached. It can be observed in Algorithm 1 that the computational complexity is dominated by Step 5 and Step 7, where the pattern search and genetic algorithms are invoked.

For the pattern search algorithm, in the worst-case scenario, the computational complexity is determined by that of an iteration. In each iteration, the algorithm polls each element of the optimization variable vector and applies a deviation on the element to find a vector improving the objective function value. As optimization variable vector $\tilde{\boldsymbol{d}}_k$ has $L \times N_{\text{S}}$ elements, the objective function is evaluated for $\mathcal{O}(L \times N_{\text{s}})$ times. Besides, the computational complexity to evaluate the objective function $\mu([\tilde{\boldsymbol{d}}_k, \tilde{\boldsymbol{D}}_{-k}]\boldsymbol{A}))$ is determined by $\boldsymbol{H} = \tilde{\boldsymbol{D}}\boldsymbol{A}$ and (2), which is $\mathcal{O}(KMN_{\text{s}}L + KM^2)$. Therefore, the computational complexity of the pattern search in Step 5 is $\mathcal{O}(KMN_{\text{s}}L + KM^2)$.

Moreover, for the genetic algorithm, as the number of iteration, i.e., N_G, is fixed, the computational complexity is determined by that of an iteration. Specifically, in each iteration, the algorithm evolves the optimization variable vectors in the current population by cross-over and mutation. Then, it evaluates the results in terms of the objective function values to decide whether to replace the current vectors with the evolved ones. Since the cross-over and mutation are simple element-wise operations, they have a low computational complexity of $\mathcal{O}(L)$. Therefore, the computational complexity of the genetic algorithm is dominated by the evaluation of the objective function, which is also $\mathcal{O}(KMN_{\text{s}}L + KM^2)$. In summary, the computational complexity of Algorithm 1 is $\mathcal{O}(KMN_{\text{s}}L + KM^2)$.

2) Convergence Analysis: In each iteration of Algorithm 1, (P1) is firstly relaxed to continuous optimization problem $\tilde{\boldsymbol{d}}_k^* = \arg\min_{\tilde{\boldsymbol{d}}_k \in [0,1]^{L \cdot N_s}} \mu([\tilde{\boldsymbol{d}}_k, \tilde{\boldsymbol{D}}_{-k}]\boldsymbol{A})$, which is solved by using the pattern search algorithm in Step 5. Based on [44], the pattern search is guaranteed to converge to a global optimal point satisfying the first-order necessary conditions under linear constraints. Then, in Step 6, the solution obtained by the pattern search algorithm, i.e., $\tilde{\boldsymbol{d}}_k^*$ is then rounded up to be $\boldsymbol{c}_k'$, which is close to the optimal k-th configuration and used as the initial point for the following genetic algorithm.

Then, in Step 7, based on [45], given a sufficiently large number of generations, the genetic algorithm converges to a global optimum. Therefore, $\boldsymbol{c}_k^*$ is a globally optimal k-th configuration for the RIS given $\boldsymbol{C}_{-k}^*$. Moreover, in Step 8, the configuration matrix $\boldsymbol{C}^*$ is updated when $\boldsymbol{c}_k^*$ results in a lower AMC value than the current best, which ensures that after each iteration, μ^* is monotonically decreasing.

Based on the definition of μ in (2), μ^* has a lower bound of 0. Therefore, Algorithm 1 is bound to converge since μ^* cannot decrease to lower than zero, which indicates the number of iterations to be finite. Given that N_G is sufficiently large, Algorithm 1 converges to a locally optimal configuration matrix, whose AMC value cannot be further reduced by changing its row vectors individually.

B. Point Cloud Extraction

The point cloud extraction consists of two steps, which are the *measurement matrix construction* and the *reconstruction of reflection coefficients*.

1) Measurement Matrix Construction: To perform point cloud extraction, we need first to estimate $\boldsymbol{H}^*$ as accurately as possible given $\boldsymbol{C}^*$ obtained in Algorithm 1, so that the accuracy of point cloud extraction is maximized. Though this estimation can be done by theoretical analysis in Appendix, i.e., $\boldsymbol{H}^* = \boldsymbol{g}(\boldsymbol{C}^*)$ as in (P1), the precise $\boldsymbol{H}^*$ may be slightly different from the theoretical calculation. This is due to that

the influence of environmental scattering and the LoS path of RF signals are not handled.

To obtain a precise $\boldsymbol{H}^*$, we first assign a set of M spatial blocks, which is denoted by $\mathcal{M} = \{(x_m, y_m, z_m)|m \in [1, M]\}$. Then, we position a metal patch at the center of each spatial block, and collect the measurement vectors using the protocol described in Section IV-B with the RIS adopting $\boldsymbol{C}^*$. When the metal patch is at the m-th ($m \in [1, M]$) spatial block, the collected measurement vector is denoted as $\hat{\boldsymbol{y}}_{\boldsymbol{C}^*,m}$. Following that, we remove the metal patch and obtain the measurement vector for the environmental scattering and LoS path, which is denoted as $\boldsymbol{y}^B$. Based on the superposition property of the RF signals, the channel gain for the propagation channel reflected at the m-th spatial block can be obtained by

$$\boldsymbol{h}_m^* = (\hat{\boldsymbol{y}}_{\boldsymbol{C}^*,m} - \boldsymbol{y}^B)/\eta_m^M, \tag{6}$$

where η_m^M is the reflection coefficient of the metal patch at the m-th spatial block towards the Rx antenna. Moreover, for normalization, we set $\eta_m^M = 1$. Furthermore, in (6), by subtracting $\boldsymbol{y}^B$ from $\hat{\boldsymbol{y}}_{\boldsymbol{C}^*,m}$, the influence of the LoS path and environmental scattering is removed. It is also worth noticing that though the construction of the measurement matrix by (6) is accurate, it can be kind of time-consuming when K and M are large. To improve the time efficiency of constructing the measurement matrix, the channel estimation method in [46] can be adopted. Based on (6), we obtain measurement matrix $\boldsymbol{H}^* \in \mathbb{C}^{K\times M}$ corresponding to $\boldsymbol{C}^*$ by $\boldsymbol{H}^* = [\boldsymbol{h}_1^*, \dots \boldsymbol{h}_M^*]$.

2) Reconstruction of Reflection Coefficients: Measurement matrix $\boldsymbol{H}^*$ can be used to extract the point cloud of reflection coefficients as shown in (1), which is equivalent to reconstructing $\boldsymbol{\eta}$ from a measurement vector $\boldsymbol{y}$. Specifically, denote the collected measurement vector in a cycle as $\boldsymbol{y}$, and the following equation holds:

$$\boldsymbol{y} - \boldsymbol{y}^B = \boldsymbol{H}^*\boldsymbol{\eta} + \boldsymbol{e}, \tag{7}$$

where $\boldsymbol{\eta} \in \mathbb{C}^M$ is the relative reflection coefficient vector of the pre-assigned spatial blocks with respect to metal patches, and $\boldsymbol{e}$ indicates a noise vector with element-wise variance ϵ. Then, based on the compressive sensing technique described in Section III-B, with known $\boldsymbol{y}$, $\boldsymbol{y}^B$, and $\boldsymbol{H}^*$, we can then extract the point cloud, i.e., $\boldsymbol{\eta}$, by solving the following l_1-norm minimization problem:

$$\text{(P2)} \quad \min_{\boldsymbol{\eta}\in\mathbb{C}^M} \ \|\boldsymbol{\eta}\|_1, \tag{8}$$

$$\|\boldsymbol{H}^*\boldsymbol{\eta} - \boldsymbol{y} + \boldsymbol{y}^B\|_2 \le \epsilon. \tag{9}$$

Problem (P2) can be recast as a *second-order cone program* problem and solved by convex optimization tools in [47].

Based on $\boldsymbol{\eta}^*$ obtained by solving (P2), the generation of the point cloud, which is a set of feature vectors with spatial positions, is described as follows. Denote the point cloud as $\mathcal{P}$ which contains M elements. Each element in $\mathcal{P}$ is a 5-dim real-valued feature vector, which is composed of its spatial position and the reconstructed reflection coefficient of it, i.e.,

$$\boldsymbol{p}_m = (x_m, y_m, z_m, \mathrm{Re}(\boldsymbol{\eta}_m^*), \mathrm{Im}(\boldsymbol{\eta}_m^*)), \quad \forall m \in [1, M]. \tag{10}$$

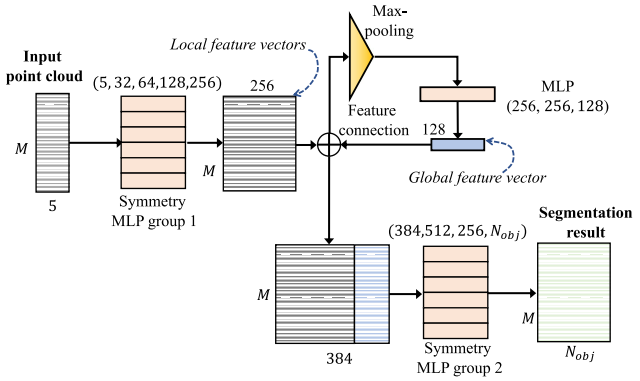


Fig. 7. Diagram of the semantic segmentation algorithm.

Here, the first three dimensions indicate the coordinates of the center of the m-th spatial block, and $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary parts of a complex value, respectively.

C. Semantic Segmentation

The semantic segmentation module takes the extracted point cloud $\mathcal{P}$ as input and outputs a set $\tilde{\mathcal{P}}$, which contains the elements with spatial positions and semantic labels, i.e.,

$$\tilde{\mathcal{P}} = \{(x_m, y_m, z_m, \boldsymbol{b}_m)|m \in [1, M]\} \tag{11}$$

where N_{obj} denotes the total number of possible semantic labels, and $\boldsymbol{b}_m \in [0, 1]^{N_{\mathrm{obj}}}$ denotes the estimated probabilities for the point to have the semantic labels with $\sum_{i=1}^{N_{\mathrm{obj}}} b_{m,i} = 1$. Without loss of generality, we denote the mapping performed by the semantic segmentation module by $\boldsymbol{f}_S^{\boldsymbol{\theta}} : \mathcal{P} \to \tilde{\mathcal{P}}$ with $\boldsymbol{\theta}$ being the parameter vector.

The aim of the semantic segmentation algorithm is to solve the optimal $\boldsymbol{\theta}$ minimizing the differences between the predicted semantic label, i.e., b_m, and the ground truth label, i.e., $\hat{b}_m$. It can be observed that compared with the works on imaging [48] and localization [49], MetaSketch is featured by its capability of recognizing and labeling different objects. To handle this problem, we adopt the supervised learning technique to train $\boldsymbol{\theta}$ given a training data set [50].

As described in Section III-C, $\boldsymbol{f}_S^{\boldsymbol{\theta}}$ needs to be symmetric. Moreover, the process of labeling the semantic meaning of each point needs to consider both local and global information. This is because knowing the semantic meaning of the point cloud as a whole facilitates figuring out the semantic meaning of each point [33]. To satisfy the above requirements and solve the semantic segmentation problem, we design the semantic segmentation algorithm based on [33]. Specifically, $\boldsymbol{f}_S^{\boldsymbol{\theta}}$ is modeled as a specially designed neural network depicted in Fig. 7, which contains *symmetric MLP groups* and *feature-gathering connections*. In this case, $\boldsymbol{\theta}$ indicates the connection weights and biases in the neural network.

1) Symmetric MLP Groups: We process M points in $\mathcal{P}$ by 2 symmetric MLP groups, each of which contains M symmetric MLPs, as shown in Fig. 7 where the numbers in the brackets indicate the numbers of neurons in different layers. As the MLPs within a symmetric MLP group have the same

structure and parameters, the results of the symmetric MLP group are invariant to the permutation of the input points.

2) Feature-Gathering Connections: Feature-gathering connections refer to the *max-pooling layer* and the concatenation of the local feature vectors and the global feature vector, which are depicted in Fig. 7. Specifically, in the max-pooling layer, the maximum value in each dimension of the M input feature vectors is picked to form the output vector, which thus has 256 dimensions. By this means, the max-pooling layer reduces the number of parameters and aggregates the information, which also alleviates overfitting. The output of it is then processed by a three-layer MLP to generate the global feature vector. Then, the global feature vector is concatenated to the local feature vector of each point. By this means, the feature vector of each point now contains both local and global information.

We adopt the average *cross-entropy (CE) loss* [50] as the optimization objective in the algorithm, which is defined by $\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{|\mathcal{D}_t|M} \sum_{(\mathcal{P},\hat{\boldsymbol{b}}_1,\ldots,\hat{\boldsymbol{b}}_M)\in\mathcal{D}_t} \sum_{m=1,\ldots,M} \mathrm{CE}(\boldsymbol{b}_m; \hat{\boldsymbol{b}}_m)$, where $\{(x_m, y_m, z_m, \boldsymbol{b}_m)\}_{m\in\{1,\ldots,M\}} = \boldsymbol{f}_S^{\boldsymbol{\theta}}(\mathcal{P})$. Here, $\mathcal{D}_t$ denotes the training data set. Then, the *Adam* algorithm [50] is invoked to solve $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$.

VI. Implementation

In this section, we present the implementation of MetaSketch, including the RIS, the RF transceivers, and the server.

A. Building the RIS

The RIS has a size of $69 \times 69 \times 0.52$ cm^3 and is composed of 16 independently controllable groups which are tightly paved in squares. The RIS is specially designed for the incident signals of 3.198 GHz, which is referred to as the *working frequency*. Each group contains $12 \times 12 = 144$ RIS elements arranged in a two-dimensional array, and thus the total number of RIS elements is 2304. The side length of an RIS element is $\delta = 1.5$ cm which is around 0.16 times the wavelength of 3.198 GHz signals. To be specific, each RIS element has the size of $1.5 \times 1.5 \times 0.52$ cm^3 and is composed of 4 rectangle copper patches printed on a dielectric substrate (Rogers 3010) with a dielectric constant of 10.2 and 3 PIN diodes (BAR 65-02L). Any two adjacent copper patches are connected by a PIN diode, and each PIN diode has two operation states, i.e., ON and OFF, which are controlled by applying bias voltages through the via holes. When the applied bias voltage is 1.2 V (or 0 V), the PIN diode is at the ON (or OFF) state. The detailed information of the RIS can be found in [15].

As there are 3 PIN diodes in an RIS element, the total number of possible states of an RIS element is 8. We simulate the S_{21} parameters, i.e., the *forward transmission gain*, of the RIS element in different states for normal-direction incident RF signals in the CST software, Microwave Studio, Transient Simulation Package [51]. At its designed working frequency of 3.198 GHz, four states of an RIS element have phase shift values with an interval equaling to $\pi/2$. We pick these four states to be the *available state set* $\mathcal{S}_a$, i.e., $\mathcal{S}_a = \{\hat{s}_1, \hat{s}_2, \hat{s}_3, \hat{s}_4\}$. Specifically, the four selected states have the phase values equaling to $\pi/4$, $3\pi/4$, $5\pi/4$ and $7\pi/4$, respectively.

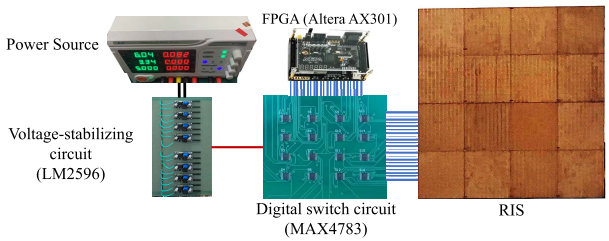


Fig. 8. Diagram of RIS control circuit.

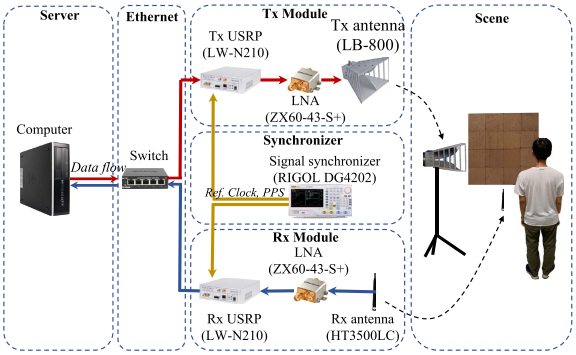


Fig. 9. Components of the RF transceiver and the server of MetaSketch.

As described in Section III-A, the RIS elements within the same group are in the same state. The states of the 16 groups are controlled by the RIS control circuit. As shown in Fig. 8, it contains a direct current (DC) power source, 16 voltage-stabilizing circuits (LM2596), 16 digital switch circuits (MAX4783), and a field-programmable gate array (FPGA) (ALTERA AX301). The DC power source is connected to the voltage-stabilizing circuits, and the input voltage to the voltage-stabilizing circuits is about 6 V. The voltage-stabilizing circuits stabilize the input voltage and reduce it to a 1.2 V output. Then, the digital switch circuits are single-pole double-throw and control the PIN diodes to work under 0 V or the stabilized 1.2 V biases.

B. Building the RF Transceiver and Server

The rest of the MetaSketch, i.e., the RF transceiver and the server, are built and connected as shown in Fig. 9. The details of each component are provided below.

1) USRP Devices: We implement the Tx and Rx based on two USRPs (LW-N210), which are capable of converting baseband signals to RF signals, or vise versa. The USRP is composed of the hardwares including the RF modulation/demodulation circuits and baseband processing units and can be controlled by using the GNU packet in Python [52].

2) Low-Noise Amplifiers (LNAs): Since the RF signals are reflected twice (on RIS and on objects) before reaching the Rx antenna, they suffer from large attenuation in signal strength, which results in low SNR and degrades the measurement precision. To handle this issue, two LNAs (ZX60-43-S+) connect the Tx and Rx USRPs and the antennas, which can amplify the transmitted and received RF signals by 15 dB.

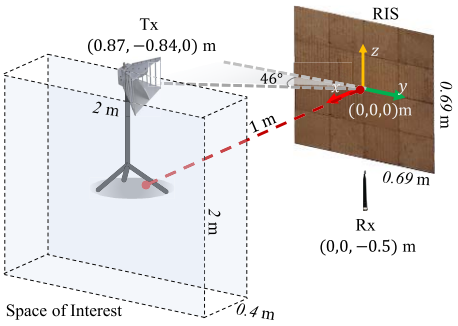


Fig. 10. Environment layout in the experiments.

3) Tx and Rx Antennas: The Tx antenna is a directional double-ridged horn antenna (LB-800), and the Rx antenna is an omnidirectional vertical antenna (HT3500LC). The polarizations of both the Tx and Rx antennas are linear and aligned to be vertical to the ground.

4) Signal Synchronizer: For the Rx USRP to obtain the relative phases and amplitudes of the received signals with respect to the transmitted signals of the Tx USRP, we employ a signal source (DG4202) to synchronize the frequency and phase of the Tx and Rx USRPs. The signal source provides the reference clock signal and the pulses-per-second (PPS) signal to the USRPs, which ensure the modulation and demodulation of the USRPs to be coherent.

5) Ethernet Switch: The Ethernet switch connects the USRPs and a server to form a local Ethernet, where the controlling signals and received baseband signals are exchanged.

6) Server: The server controls the two USRPs by using the GNU packet in Python, extracts the measurement vectors from the received baseband signals, and performs point cloud extraction and semantic segmentation.

VII. Simulation and Experimental Evaluation

In this section, we demonstrate the experimental setup for MetaSketch and evaluate the performance of its component modules by using simulation results and experimental results.

A. Experimental Setup

We describe the experimental setup in four aspects: the environmental layout to test MetaSketch, the setting of the RF transceivers, the collected data, and the evaluation metrics.

1) Environmental Layout: The environmental layout of MetaSketch is shown in Fig. 10. To be specific, the origin of coordinate is at the center of the RIS, and thus the RIS is in the y-z plane. Besides, the z-axis is vertical to the ground and pointing upwards, and the x- and y-axes are parallel to the ground. The Tx and Rx antennas are located at $(0.87, -0.84, 0)$ m and $(0, 0, -0.5)$ m, respectively.

The humans and objects are in the space of interest, which is a $0.4 \times 2 \times 2$ m^3 cuboid region located at 1 m away from the RIS. Besides, the space of interest is regularly divided into $M = 400$ spatial blocks each with size $0.4 \times 0.1 \times 0.1$ m^3. Moreover, since the space of interest is behind the Tx antenna, and the Tx antenna is a directional horn antenna, no LoS path from the Tx antenna to the space of interest exists.

2) Transceiver Setting: To obtain a high SNR and reduce the error of point cloud extraction in (P2), the Tx power of the USRP device is set to be the maximum of 20 dBm. Thus, given that the gain of LNA is 15 dB, the actual transmitting power of the Tx Module is 35 dBm.

Besides, the operating frequency of the USRP devices in MetaSketch is set to be the same as the working frequency of the RIS, which is $f_c = 3.198$ GHz, in order to obtain the designed four phase shifts with a $\pi/2$ interval. This is because if the USRP devices operate at other frequencies, the RIS elements will not be able to effectively impose the phase shifts on the reflected signals. Therefore, it results in a high AMC value of the measurement matrix and the degradation of the MetaSketch's performance.[1]

3) Collected Data: The optimized configuration matrix of RIS, i.e., $\boldsymbol{C}^*$, is obtained by solving (P1) in the server and is uploaded to the FPGA. In the data collection phase, the RIS changes its configuration every 0.1 seconds. To obtain the corresponding measurement matrix, i.e., $\boldsymbol{H}^*$, we set a 0.1×0.1 m^2 metal patch at the center of each spatial block sequentially given RIS adopting $\boldsymbol{C}^*$, as described in Section V-B1.

We first generate a set of 64 point clouds with semantic labels as the ground truth set. Specifically, we arrange a human and up to 4 objects in the space of interest according to each of the point clouds. The 4 objects include a bottle, a laptop, and a suitcase. We measure the received signals following the protocol in Section IV-B. Using measurement matrix $\boldsymbol{H}^*$ corresponding to $\boldsymbol{C}^*$, the point cloud extraction module processes the received signals and extracts point clouds by solving (P2). The ground truth set and the corresponding extracted point clouds constitute training data $\mathcal{D}_t$ for the semantic segmentation algorithm.

In the collected training data, each point is represented by a 5-dim vector and a label. The first three dimensions indicate the coordinates of the point; the next 2 dimensions indicate the real and imaginary values of the regenerated reflection coefficients of the point. The label is a 5-dim one-hot vector indicating the semantic meaning of the point.

4) Evaluation Metrics: We adopt the following three evaluation metrics.

a) AMC: As defined in (2), the AMC evaluates the average coherence between every two columns in the measurement matrix. A lower AMC indicates the propagation channels via different spatial blocks are more independent of each other. The AMC is inversely proportional to the reconstruction performance of the compressive sensing method [41].

b) Average cross-entropy loss: We adopt average cross-entropy loss as the objective metric to train the semantic segmentation module as described in Section V-C. The average cross-entropy loss evaluates the divergence between the estimated semantic labels and the ground truth.

[1]Since the design of MetaSketch is not specific to a certain RIS, the proposed MetaSketch can adapt to an RIS with a different working frequency.

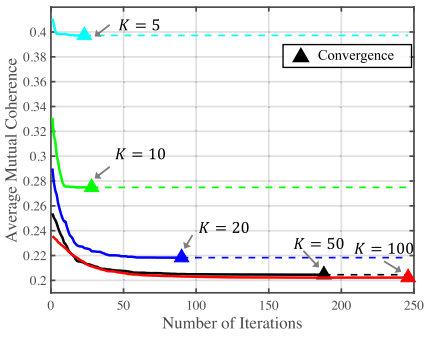


Fig. 11. AMC of measurement matrix versus the number of iterations in Algorithm 1.

c) Average error rate: For each semantic label, the error rate is defined as the ratio between the points which belong to this label in truth but are labeled incorrectly and the total number of points belonging to this label in the ground truth. We adopt the average error rate of the N_{obj} labels as an intuitive metric to evaluate the performance of MetaSketch.

B. AMC Minimization by Radio Environment Reconfiguration

Fig. 11 shows the resulting AMC of $\boldsymbol{H}^*$ obtained by solving (P1) versus the number of iterations in Algorithm 1, under different numbers of configurations, K. It can be observed that the AMC decreases with the number of iterations, which verifies the effectiveness of the proposed configuration matrix optimization algorithm. Besides, it can also be seen that the converged optimal AMC of $\boldsymbol{H}^*$ decreases with K, which can be explained as follows. To reduce the AMC of $\boldsymbol{H}^*$, it requires the columns of $\boldsymbol{H}^*$ to distribute their large elements into different dimensions. As K determines the number of dimensions of $\boldsymbol{h}_m^*$, large K increases the probability to have the large elements at different dimensions and thus potentially results in a lower AMC. Therefore, as K increases, the AMC value of $\boldsymbol{H}^*$ decreases, which can lead to higher accuracy for the compressive sensing technique to extract point clouds.

Specifically, we then compare the mutual coherence of the measurement matrices corresponding to the random and optimized configuration matrices in the $K = 10$ case. The configuration matrix in Fig. 12 (a) is $\boldsymbol{C}^*$ obtained by Algorithm 1, and Fig. 12 (c) shows the mutual coherence of the corresponding $\boldsymbol{H}^*$. Besides, the configuration matrix in Fig. 12 (b) is a random configuration matrix where the elements are generated following a uniform distribution on $[1, 4]$. Fig. 12 (d) shows the coherence of column vectors of $\boldsymbol{H}$ for the random $\boldsymbol{C}$. Moreover, the overall AMC in the two cases are showed by the black planes, and their values are provided by the tags in Figs. 12 (c) and (d). Comparing Figs. 12 (c) and (d), we observe that Algorithm 1 effectively optimizes the configuration matrix and reduces the mutual coherence of the measurement matrix, resulting in a lower AMC than that of a random configuration matrix. Based on the discussion in Section V-A, the configuration matrix in Fig. 12 (a) achieves higher accuracy of point cloud extraction than that in Fig. 12 (b).

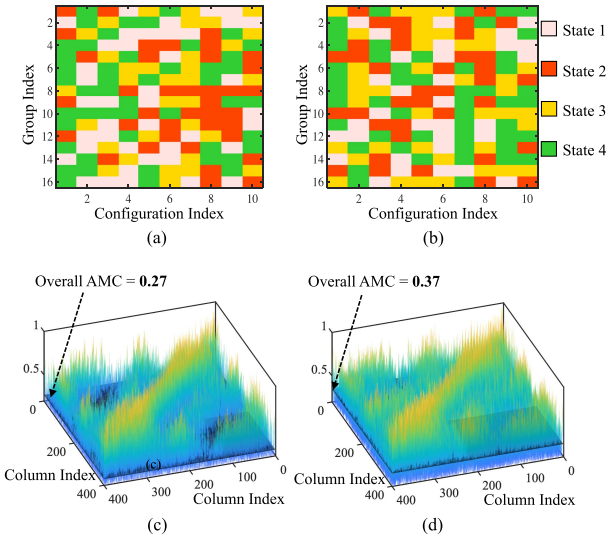


Fig. 12. (a) and (b) illustrate the optimized and random RIS configuration matrices in $K = 10$ case. (c) and (d) show the mutual coherence values of the measurement matrices corresponding to the configuration matrices in (a) and (b), respectively. The black planes in (c) and (d) indicate the overall AMC, and their quantitative values are provided in the tags.

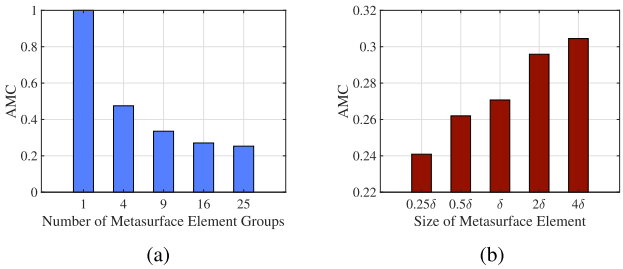


Fig. 13. AMC of the optimized measurement matrix $\boldsymbol{H}^*$ given (a) different numbers of RIS element groups and (b) different sizes of an RIS element. In (b), $\delta = 1.5$ cm denotes the size of the RIS element used in the implementation.

Figs. 13 (a) and (b) show the influence of the number of RIS element groups and the size of an RIS element on the AMC value of $\boldsymbol{H}^*$. It can be observed in Fig. 13 (a) that as the number of RIS element groups increases, the AMC value decreases. This is because more groups of controllable RIS elements indicate that the RIS has a stronger beamforming capability, which enables different spatial blocks to have less coherent influence on the received signals. Therefore, based on the discussion in Section V-A, the performance of the MetaSketch in terms of point cloud extraction and semantic segmentation improves with the number of RIS elements.

Besides, it can be observed in Fig. 13 (b) that the AMC value increases with the size of the RIS element, which can be explained intuitively as follows. Given a fixed number of RIS elements, increasing the size of RIS elements enlarges the RIS. Relatively speaking, it can be considered as the space of interest is shrunken while the size of the RIS remains the same. In this case, the spatial blocks become more close to each other, which makes the measurement vectors of different

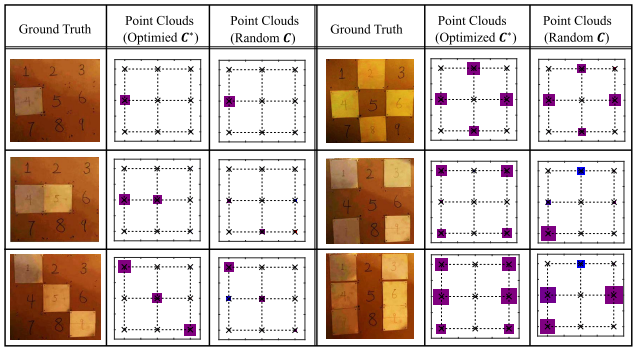


Fig. 14. Extracted point clouds when the RIS adopts the optimized and random configuration matrices, i.e., $\boldsymbol{C}^*$ and $\boldsymbol{C}$, respectively. The size of a square indicates the amplitude of the extracted reflection coefficient in the spatial block, and the color of a square indicates its phase.

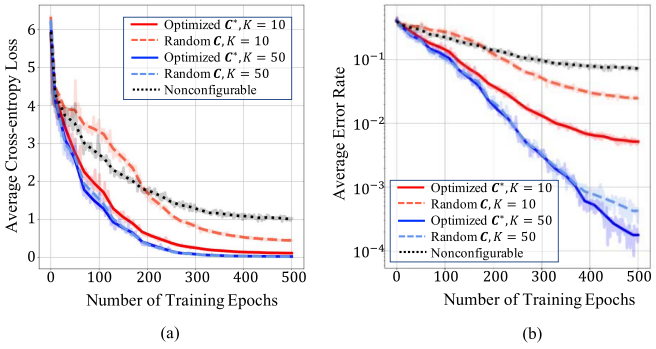


Fig. 15. (a) Average cross-entropy loss and (b) average error rate versus the number of training epochs given optimized and random configuration matrices with different K. The shaded areas illustrate the values of each epoch, and the thick lines indicate the average values for the last 20 epochs.

spatial blocks more coherent. Therefore, the AMC value becomes higher as the size of RIS element increases.

C. Evaluation on Point Cloud Extraction

Fig. 14 show the photos and the corresponding extracted point clouds given the optimized and the random configuration matrices, i.e., $\boldsymbol{C}^*$ and $\boldsymbol{C}$. In the photos, the light (yellow) regions are the metal patches, and the dark (brown) regions are the cardboards which have a negligible impact on RF signals. It can be observed that when the number of metal patches is small, the point cloud extraction module can successfully reconstruct the point clouds which reflect the ground truth well given both $\boldsymbol{C}^*$ and $\boldsymbol{C}$. When the number of metal patches is larger than 4, the point clouds may not be in accordance with ground truth in the random configuration matrix case. Nevertheless, the point clouds obtained when using the optimized RIS configuration matrix reflect the ground truth more accurately than those obtained when using a random RIS configuration matrix, which verifies the effectiveness of optimizing the radio environment reconfiguration.

D. Evaluation on Semantic Segmentation

Figs. 15 (a) and (b) show the average cross-entropy loss and average error rate versus the number of training epochs of the semantic segmentation algorithm. The red and blue lines are obtained when $K = 10$ and $K = 50$, respectively, where the solid lines are the results given $\boldsymbol{C}^*$ and the dash lines are the results given $\boldsymbol{C}$. It can be observed that as the number of training epochs increases, both the average cross-entropy loss and the average error rate decrease. Besides, when K is small ($K = 10$), it can be observed that using $\boldsymbol{C}^*$ can help the semantic segmentation module to achieve a much lower average cross-entropy loss and average error rate. Specifically, when $K = 10$, after about 350 epochs of training, MetaSketch can perform semantic segmentation with an average error rate of less than 1%. If the number of configurations is sufficiently large, e.g., $K = 50$, the average error rate can be further reduced to less than 0.1% after 400 epochs of training. Nevertheless, when K becomes larger, the duration of the data collection phase increases, which may make the assumption of humans and objects being static during the data collection phase impractical.

Moreover, in Figs. 15 (a) and (b), the black dot lines are obtained in a nonconfigurable radio environment, where the configuration of the RIS is fixed to all states being $\hat{s}_1$, i.e., $K = 1$ and $\boldsymbol{C} = \mathbf{1}$. Compared with the system with a nonconfigurable radio environment, the average cross-entropy loss and the average error rate of MetaSketch are significantly lower, which verifies the effectiveness of the radio environment reconfiguration as well as the proposed system.

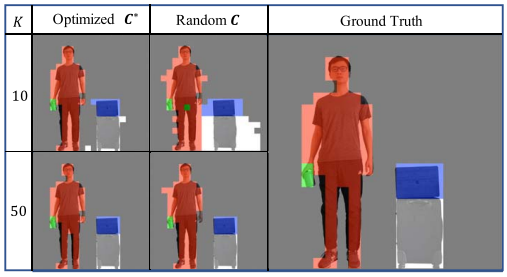


Fig. 16. Semantic segmentation results for human and objects given that the RIS adopts the optimized and the random configuration matrices with $K = 10$ and $K = 50$. The number of training epochs is 500. The human, suitcase, laptop, and bottle are labeled by red, white, blue, and green, respectively.

Fig. 16 shows the semantic segmentation results after 500 training epochs overlaid on the ground truth,[2] given that the RIS adopts the optimized and random configuration matrices with different K. Besides, Fig. 17 provides detailed information about the training process. Specifically, Figs. 17 (a) and (b) are the ground truth photo and the ground truth labeled point cloud, respectively, and Fig. 17 (c) shows the semantic segmentation results in different training epochs, where the RIS adopts $\boldsymbol{C}^*$ with $K = 10$ and $K = 50$. In the semantic segmentation results, the human, suitcase, laptop, and bottle are labeled by red, white, blue, and green colors, respectively.

[2]We use photos to show the ground truth to facilitate the comparison between the segmentation results and real scenes. The photos can be used to get the labeled point clouds in the training data set. After training, the MetaSketch can perform semantic segmentation based on only the received RF signals, which ensures privacy protection.

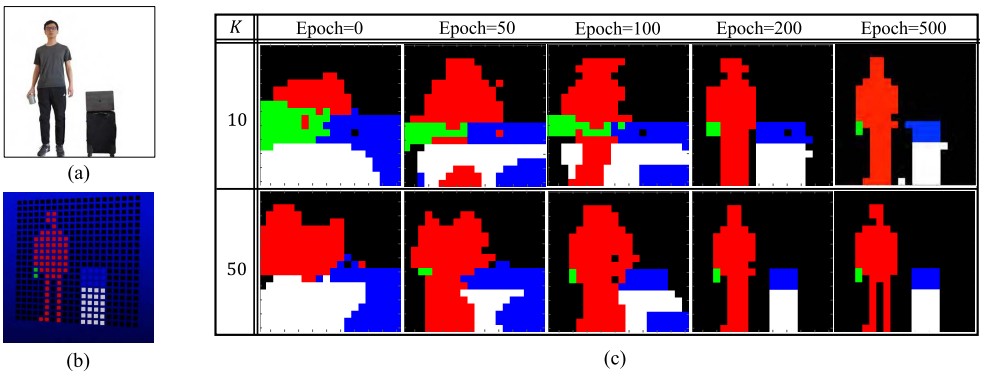


Fig. 17. (a) and (b) are the image of the human and objects and the ground truth labeled point cloud, respectively, and (c) is the semantic segmentation results in different training epochs given RIS adopting the optimized $\boldsymbol{C}^*$ with $K = 10$ and $K = 50$.

In Fig. 16, comparing the cases where $K = 10$ and $K = 50$, we can observe that increasing the number of configurations in a cycle enables the MetaSketch to obtain the labeled point cloud closer to the ground truth. In Fig. 17 (c), it can be seen that when K is larger, the training process is faster, as the semantic segmentation result gets close to the ground truth in an earlier epoch. Besides, in Fig. 16, it can also be seen that using the optimized configuration matrix improves the semantic segmentation accuracy. When $K = 50$, the improvement due to using the optimized configuration matrix is smaller than that when $K = 10$. Nevertheless, since a large K results in a long duration of the data collection phase, a small K is preferable, where adopting the optimized configuration matrix is necessary and important.

VIII. Conclusion

In this paper, we have presented MetaSketch, an RIS-based RF-sensing system, to perform semantic segmentation based on RF signals. We have designed MetaSketch to have three component modules, i.e., a radio environment reconfiguration module, a point cloud extraction module, and a semantic segmentation module. MetaSketch can actively modify the radio environment according to the configurations of the RIS and generate favorable propagation channels for sensing purposes. To optimize its performance, we have proposed a configuration matrix optimization algorithm for the radio environment reconfiguration. By using the point cloud extraction module, MetaSketch can extract the spatial reflection coefficient point cloud, which can be processed by its semantic segmentation module for semantic recognition and labeling.

Our results have shown that, firstly, the RIS-based radio environment reconfiguration module with the proposed algorithm can generate measurement matrices with low AMC, which can promote the accuracy of point cloud extraction. Secondly, after training, MetaSketch can label semantic meanings of the spatial blocks with an average error rate of $\leq 1\%$, given the setup of a human, a suitcase, a laptop, and a bottle in a 1.6 m^3 indoor space. Thirdly, optimizing the measurement matrix can reduce the number of training epochs and measurements required to obtain accurate semantic segmentation results.

Appendix
Calculation of Measurement Matrix

Given configuration matrix $\boldsymbol{C}$, we now calculate the corresponding measurement matrix $\boldsymbol{H}$. Based on ray-tracing technique [53], we first calculate channel gain matrix $\boldsymbol{A}$, where the elements indicate the channel gains of the radio paths from the Tx to Rx via the L RIS groups in N_s different states and the M spatial blocks. Specifically, $\boldsymbol{A}$ is a $(LN_s) \times M$ matrix. Based on [54], given $l \in [1, L]$, $i \in [1, N_s]$, and $m \in [1, M]$, the element of $\boldsymbol{A}$ can be expressed as

$$(\boldsymbol{A})_{N_s(l-1)+i,m} = \sum_{n \in \mathcal{N}_l} \Big(\frac{\lambda \cdot r_{n,m}(\hat{s}_i) \cdot \sqrt{g_{T,n} g_{R,m}}}{4\pi d_n^{\mathrm{T,M}} \cdot d_{n,m}^{\mathrm{M,S}} \cdot d_m^{\mathrm{S,R}}} \cdot e^{-j2\pi (d_n^{\mathrm{T,M}} + d_{n,m}^{\mathrm{M,S}} + d_m^{\mathrm{S,R}})/\lambda} \Big), \quad (12)$$

where $r_{n,m}(\hat{s}_i)$ denotes the reflection coefficient of the n-th RIS element in the l-th group with state $\hat{s}_i$ for the incident signals towards the m-th spatial block, $g_{T,n}$ is the gain of the Tx antenna towards the n-th RIS element, $g_{R,m}$ is the gain of the Rx antenna towards the m-th spatial block, and $d_n^{\mathrm{T,M}}$, $d_{n,m}^{\mathrm{M,S}}$, and $d_m^{\mathrm{S,R}}$ are the distance from the Tx antenna to the n-th RIS element, from the n-th RIS element in the l-th group to the m-th spatial block, and from the m-th spatial block to the Rx antenna, respectively. Here, $r_{n,m}(\hat{s}_i)$ is calculated by using the CST software [51]. Besides, $g_{T,n}$ and $g_{R,m}$ are obtained from the data-sheets of Tx and Rx antennas, respectively.

We then transform $\boldsymbol{C}$ to a $K \times (LN_s)$-dim zero-one matrix $\boldsymbol{D}$. Given $\forall k \in [1, K]$, $l \in [1, L]$, and $i \in [1, N_s]$, the element of $\boldsymbol{D}$ can be expressed as

$$(\boldsymbol{D})_{k, N_s \cdot (l-1)+i} = \begin{cases} 1, & \text{if } (\boldsymbol{C})_{k,l} = i, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Therefore, $\boldsymbol{H}$ can be calculated by $\boldsymbol{H} = \boldsymbol{D}\boldsymbol{A}$. We denote process of calculating $\boldsymbol{H}$ from $\boldsymbol{C}$ by the function $\boldsymbol{H} = \boldsymbol{g}(\boldsymbol{C})$.

References

[1] L. Shapiro, *Computer Vision and Image Processing*. San Diego, CA, USA: Academic Press, 1992.
[2] D. Zhang, J. Wang, J. Jang, J. Zhang, and S. Kumar, "On the feasibility of Wi-Fi based material sensing," in *Proc. ACM MobiCom*, Los Cabos, Mexico, Oct. 2019, pp. 1–16.

[3] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," in *Proc. ACM MobiCom*, New Delhi, India, Oct. 2018, pp. 289–304.

[4] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "CrossSense: Towards cross-site and large-scale WiFi sensing," in *Proc. ACM MobiCom*, New Delhi, India, Oct. 2018, pp. 305–320.

[5] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, "Enabling identification and behavioral sensing in homes using radio reflections," in *Proc. ACM CHI*, Glasgow, Scotland U.K., May 2019, pp. 1–13.

[6] M. Zhao *et al.*, "Through-wall human pose estimation using radio signals," in *Proc. IEEE CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 7356–7365.

[7] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Trans. Graph.*, vol. 34, no. 6, p. 219, 2015.

[8] M. Zhao *et al.*, "RF-based 3D skeletons," in *Proc. ACM SIGCOMM*, New York, NY, USA, Aug. 2018, pp. 267–281.

[9] A. Pedross-Engel *et al.*, "Orthogonal coded active illumination for millimeter wave, massive-MIMO computational imaging with metasurface antennas," *IEEE Trans. Comput. Imag.*, vol. 4, no. 2, pp. 184–193, Jun. 2018.

[10] N. Honma, D. Sasakawa, N. Shiraki, T. Nakayama, and S. Iizuka, "Human monitoring using MIMO radar," in *Proc. IEEE Int. Workshop Electromagn., Appl. Student Innov. Competition (iWEM)*, Nagoya, Japan, Aug. 2018, pp. 1–2.

[11] Z. Li *et al.*, "Programmable radio environments with large arrays of inexpensive antennas," *GetMobile, Mobile Comput. Commun.*, vol. 23, no. 3, pp. 23–27, Jan. 2020.

[12] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–20, May 2019.

[13] C. Huang, R. Mo, and Y. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, Jun. 2020.

[14] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.

[15] J. Hu *et al.*, "Reconfigurable intelligent surface based RF sensing: Design, optimization, and implementation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2700–2716, Nov. 2020.

[16] H. Zhang, B. Di, L. Song, and Z. Han, *Reconfigurable Intelligent Surface-Empowered 6G*. Cham, Switzerland: Springer, 2021.

[17] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing improvement via deep learning configured intelligent metasurfaces," in *Proc. IEEE SPAWC*, Cannes, France, Jul. 2019, pp. 1–5.

[18] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in *Proc. USENIX NSDI*, Oakland, CA, USA, May 2015, pp. 279–292.

[19] N. Patwari and J. Wilson, "RF sensor networks for device-free localization: Measurements, models, and algorithms," *Proc. IEEE*, vol. 98, no. 11, pp. 1961–1973, Nov. 2010.

[20] B. Kellogg, V. Talla, and S. Gollakota, "Bringing gesture recognition to all devices," in *Proc. USENIX NSDI*, Seattle, WA, USA, Apr. 2014, pp. 303–316.

[21] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 907–920, Apr. 2014.

[22] J. Lien *et al.*, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, p. 142, Jul. 2016.

[23] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[24] Y. Tian, G.-H. Lee, H. He, C.-Y. Hsu, and D. Katabi, "RF-based fall monitoring using convolutional neural networks," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–24, Sep. 2018.

[25] J. N. Gollub *et al.*, "Large metasurface aperture for millimeter wave computational imaging at the human-scale," *Sci. Rep.*, vol. 7, p. 42650, Feb. 2017.

[26] T. Zhou *et al.*, "Short-range wireless localization based on meta-aperture assisted compressed sensing," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2516–2524, Jul. 2017.

[27] J. Yao, Z. Zhang, X. Shao, C. Huang, C. Zhong, and X. Chen, "Concentrative intelligent reflecting surface aided computational imaging via fast block sparse Bayesian learning," in *Proc. IEEE VTC*, Helsinki, Finland, Apr. 2021, pp. 1–6.

[28] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "MetaLocalization: Reconfigurable intelligent surface aided multi-user wireless indoor localization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7743–7757, Dec. 2021.

[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2015, pp. 1–9.

[31] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Jan. 2017.

[32] D. Mehta *et al.*, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, p. 44, 2017.

[33] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3d classification and segmentation," in *Proc. IEEE CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 652–660.

[34] T. J. Cui, D. R. Smith, and R. Liu, *Metamaterials*. Boston, MA, USA: Springer, 2010.

[35] M. A. El Mossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.

[36] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.

[37] Z. Han, H. Li, and W. Yin, *Compressive Sensing for Wireless Networks*. New York, NY, USA: Cambridge Univ. Press, 2013.

[38] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, May 2006.

[39] J. Dai, K. He, and J. Sun, "BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 1635–1643.

[40] L. Li *et al.*, "Machine-learning reprogrammable metasurface imager," *Nature Commu.*, vol. 10, no. 1, pp. 1–8, 2019.

[41] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[42] R. M. Lewis, A. Shepherd, and V. Torczon, "Implementing generating set search methods for linearly constrained minimization," *SIAM J. Sci. Comput.*, vol. 29, no. 6, pp. 2507–2530, Jan. 2007.

[43] D. E. Goldberg, *Genetic Algorithms*. London, U.K.: Pearson Education India, 2006.

[44] R. M. Lewis and V. Torczon, "Pattern search methods for linearly constrained minimization," *SIAM J. Optim.*, vol. 10, no. 3, pp. 917–941, 2000.

[45] R. R. Sharapov and A. V. Lapshin, "Convergence of genetic algorithms," *Pattern Recognit. Image Anal.*, vol. 16, no. 1, pp. 392–397, Jul. 2006.

[46] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Mar. 2021.

[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[48] J. Hu, H. Zhang, K. Bian, M. D. Renzo, Z. Han, and L. Song, "MetaSensing: Intelligent metasurface assisted RF 3D sensing by deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2182–2197, Jul. 2021.

[49] H. Zhang *et al.*, "MetaRadar: Indoor localization by reconfigurable metamaterials," *IEEE Trans. Mobile Comput.*, early access, Dec. 14, 2020, doi: 10.1109/TMC.2020.3044603.

[50] I. Goodfellow, Y. Bengio, and A. Courvile, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[51] F. Hirtenfelder, "Effective antenna simulations using CST MICROWAVE STUDIO®," in *Proc. INICA*, Munich, Germany, Mar. 2007, p. 239.

[52] E. Blossom, "Gnu radio: Tools for exploring the radio frequency spectrum," *J. Linux*, vol. 2004, no. 122, p. 4, Jun. 2004.

[53] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

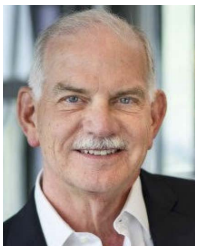
[54] W. Tang *et al.*, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, Jan. 2021.

Jingzhi Hu (Graduate Student Member, IEEE) received the B.S. degree in electronic engineering from Peking University, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Electronics. His main research interest includes metamaterial-aided passive RF sensing techniques for the Internet of Things. He served as a TPC Member of IEEE/CIC ICCC in 2017 and 2018.

Hongliang Zhang (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Electrical Engineering and Computer Science, Peking University, in 2014 and 2019, respectively. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA. He is currently a Post-Doctoral Associate with the Department of Electrical and Computer Engineering, Princeton University, NJ, USA. His current research interests include reconfigurable intelligent surfaces, aerial access networks, optimization theory, and game theory. He received the Best Doctoral Thesis Award from the Chinese Institute of Electronics in 2019. He is an Exemplary Reviewer for IEEE TRANSACTIONS ON COMMUNICATIONS in 2020. He was also a recipient of the 2021 IEEE ComSoc Heinrich Hertz Award for Best Communications Letters and the 2021 IEEE ComSoc Asia-Pacific Outstanding Paper Award. He has served as a TPC Member of many IEEE conferences, such as GLOBECOM, ICC, and WCNC. He is currently an Editor for IEEE COMMUNICATIONS LETTERS, *IET Communications*, and *Frontiers in Signal Processing*. He has also served as a Guest Editor for several journals, such as IEEE INTERNET OF THINGS JOURNAL and *Journal of Communications and Networks*.

Kaigui Bian (Senior Member, IEEE) received the B.S. degree in computer science from Peking University, Beijing, China, in 2005, and the Ph.D. degree in computer engineering from Virginia Tech in 2011. He was a Visiting Young Faculty with Microsoft Research Asia in 2013. His research interests include wireless networking and mobile computing. He received the best paper awards of international conferences, such as IEEE ICC 2015, ICCSE 2017, and BIGCOM 2018; the Best Student Paper Award of IEEE DSC 2018; and the IEEE Communication Society Asia-Pacific Board (APB) Outstanding Young Researcher Award in 2018. He currently serves as an Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE ACCESS, and an organizing committee member as well as a technical program committee member of many international conferences.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was a Research and Development Engineer at JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an Assistant Professor at Boise State University, ID, USA. He is currently a John and Rebecca Moores Professor with the Department of Electrical and Computer Engineering and the Department of Computer Science, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He received the NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS) in 2016, and several best paper awards in IEEE conferences. He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018, and has been an AAAS Fellow since 2019 and an ACM Distinguished Member since 2019. He is 1% highly cited researcher since 2017 according to Web of Science. He is also the winner of the 2021 IEEE Kiyo Tomiyasu Award, for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."



H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana–Champaign. Since 1990, he has been on the faculty at Princeton, where he is currently the Michael Henry Strater University Professor. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other universities, including most recently at Berkeley and Cambridge. His research interests are in the areas of information theory, machine learning and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the forthcoming book *Machine Learning and Wireless Communications* (Cambridge University Press). Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences and is a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. He received the IEEE Alexander Graham Bell Medal in 2017.

Lingyang Song (Fellow, IEEE) received the Ph.D. degree from the University of York, U.K., in 2007, where he received the K. M. Stott Prize for excellent research. He worked as a Research Fellow at the University of Oslo, Norway, until rejoining Philips Research, U.K., in March 2008. In May 2009, he joined the Department of Electronics, School of Electronics Engineering and Computer Science, Peking University, where he is currently a Boya Distinguished Professor. His main research interests include wireless communication and networks, signal processing, and machine learning. He was a recipient of the IEEE Leonard G. Abraham Prize in 2016 and the IEEE Asia Pacific (AP) Young Researcher Award in 2012. He has been an IEEE Distinguished Lecturer since 2015.