The following publication Y. Wang, I. W. -H. Ho, S. Zhang and Y. Wang, "Intelligent Reflecting Surface Enabled Fingerprinting-Based Localization With Deep Reinforcement Learning," in IEEE Transactions on Vehicular Technology, vol. 72, no. 10, pp. 13162-13172, Oct. 2023 is available at https://dx.doi.org/10.1109/TVT.2023.3275581.

# Intelligent Reflecting Surface enabled Fingerprinting-based Localization with Deep Reinforcement Learning

Yuhao Wang, Ivan Wang-Hei Ho, *Senior Member, IEEE*, Shuowen Zhang, *Member, IEEE* and Yuhong Wang

***Abstract*—Intelligent reflecting surface (IRS) is considered a promising solution to manipulate the radio frequency transmission environment in the sixth-generation (6G) wireless systems. However, little attention was received by IRS-aided localization techniques. Among range-free wireless localization strategies, received signal strength indicator (RSSI) fingerprinting-based technique is preferred since it can be easily accessed. Inspired by these and the tremendous success of deep reinforcement learning (DRL), we propose an IRS-enabled fingerprinting-based localization methodology with the aid of DRL. Specifically, we firstly propose an IRS-enabled fingerprinting-based localization system. In this system, RSSI lists are created by periodic IRS configurations and pre-collected as database. When a request of localization from a receiver is sent to the server, the database is compared with the online-measured RSSI data to identify the best receiver position estimate using the nearest neighbor algorithm. In addition, we develop a DRL-based IRS configuration selector to identify the most qualified IRS configurations so as to minimize the localization error. We also propose a communication protocol for the operation of the proposed localization methodology. Extensive simulation under different circumstances have been conducted and the results indicate that the localization accuracy scales with the number of IRS configurations. With the aid of DRL, the localization accuracy is further boosted by more than $40\%$ as compared with previous work.**

***Index Terms*—Intelligent reflecting surface, Localization, Deep reinforcement learning, Fingerprinting.**

## I. Introduction

THE worldwide roll-out of the fifth-generation (5G) mobile networks leads to a rapid proliferation of base stations and antennas. Though superior user experience has been unlocked by 5G, the existing 5G technology also has various drawbacks such as high energy consumption and expensive hardware. Fortunately, beyond 5G, a new way to transform transmission has been proposed to address these challenges termed as the intelligent reflecting surface (IRS) [1]. An IRS is usually made of a great quantity of almost passive reconfigurable units [2], where each unit can reflect or refract the incident signal with designed phase shifts. Among conventional wireless communication research, a communication channel is regarded as an uncontrollable physical transmission medium. With the new IRS technology, the communication channel can be manipulated to create smart radio environments. Additionally, the IRS can be flexibly coated on facades of buildings so that they can be easily implemented in practice. It is worth noting that various names have been proposed for this new technology such as large intelligent surfaces, software-controlled meta-surfaces, software-defined surfaces, reconfigurable intelligent surfaces, and intelligent reflecting surfaces [3].

With the increasing interest in accurate localization of autonomous vehicles and simultaneous localization and mapping (SLAM) of complicated environments, reliable localization service provided by the 5G wireless system has received more attention. Accurate geographical information not only could enhance next-generation vehicular technologies but also enable vehicles to sense and connect with the road infrastructure ubiquitously. Moreover, the operation of a large number of Internet of things (IoT) applications relies on the provision of location information of mobile devices, especially with the explosive growth of IoT devices in both densely urban and unpopulated remote areas [4]. Traditionally, wireless-based localization techniques can be categorized according to the utilized data type such as Direction of Arrival (DoA), Angle of Arrival (AoA), Time of Arrival (ToA), Received Signal Strength (RSS), and Channel State Information (CSI). RSS-based techniques are widely employed because RSS information can be accessed easily and does not need a strictly synchronized system. However, unfavorable RSS distribution (i.e., similar RSS values among adjacent locations) can degrade the localization accuracy. Now, RSS-based localization techniques can be further developed with the aid of IRS-enhanced next-generation wireless communication systems.

The sixth-generation (6G) IRS-enhanced technologies bring more possibilities to localization research because it enables us to customize the radio environment to create a more distinguishable RSS distribution. The reader is referred to the next section for details. However, to rapidly configure a large amount of IRS units requires high computation power and is difficult to implement in practical scenarios. Recent advances in data-driven techniques such as deep learning, reinforcement learning, and multi-task learning make this problem feasible [5]. A large and growing body of literature investigated the deep learning architecture for IRS and they

Yuhao Wang, Ivan Wang-Hei Ho and Shuowen Zhang are with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong (e-mail: yuhao1995.wang@connect.polyu.hk, ivanwh.ho@polyu.edu.hk).

Yuhong. Wang is with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong (e-mail: yuhong.wang@polyu.edu.hk)

This work was supported in part by the Research Impact Fund (Project No.R5007-18) established under the University Grant Committee (UGC) of the Hong Kong Special Administrative Region (HKSAR), China.

Manuscript received April 19, 2020; revised August 26, 2020.

are well surveyed in [6]. Nevertheless, training a supervised deep learning model needs enormous labeled training data and usually, they are difficult to obtain. Motivated by the success stories of reinforcement learning (RL) in solving the complicated problem, some studies have already employed deep reinforcement learning (DRL) to configure IRS. As compared to supervised learning, DRL does not require the training data to be labeled and can improve the overall storage efficiency.

The reconfigurability of IRS provides us with more flexibility to create received signal strength indicator (RSSI) fingerprints. Compared with conventional trilateration-based techniques with multiple anchor points, only one transmitter and one IRS are needed in the RSS-based localization system. Moreover, DRL can reduce the complexity of IRS configuration significantly. Motivated by these advantages, in this study, we leverage DRL techniques and the diversity offered by IRS-enhanced radio environments to perform fingerprinting-based localization. More easily differentiable radio maps are generated with DRL-selected IRS configurations for localization purposes. This study has potentials to be integrated with IRS-assisted communication to achieve integrated sensing and communication (ISAC). The major contributions of this paper are as follows.

- We propose an IRS-enabled fingerprinting-based localization methodology and design a communication protocol to coordinate the operation of the proposed methodology in the localization process. Compared with previous fingerprinting-based localization methodologies, the proposed one can perform localization with a single transmitter and an IRS, which is more cost-effective in a future scenario where IRS has already been deployed on large scale.
- We propose a research framework based on DRL to assist the IRS configuration of the proposed IRS-enabled fingerprinting-based localization methodology. The proposed DRL-assisted framework can tackle the localization accuracy.
- We optimize the IRS configuration selection problem with the aid of DRL and achieve a $40\%$ localization performance boost as compared with previous work.

The rest of this paper is organized as follows: We introduce and review related works regarding IRS, IRS-aided localization, and DRL in section II; We propose a novel DRL-based IRS-enabled fingerprinting localization framework as well as its protocol to localize mobile devices in section III; We conduct extensive simulation of the proposed scenario to evaluate the performance in section IV; Finally, section V concludes the whole paper.

## II. Related Work

Three principal elements are considered and reviewed in our proposed methodology: IRS, localization technique aided by the wireless communication system, and deep reinforcement learning.

### A. What is IRS and Why IRS

An IRS is usually made of low-cost, compact, and quasi-passive electronics mirrors in the form of a planer sheet [7], [8]. Each element enables amplitude/phase change of the input signal by setting different load impedance. By tuning each reflecting element electrically and digitally, it could reflect or steer the signal in a tailored direction so as to optimize the connectivity or bypass the obstacles. Thus, an IRS enables a Virtual-Line-Of-Sight (VLOS) to propagate the signal wave. In case of Line-Of-Sight (LOS) blockages, IRS-aided wireless networks can improve communication accuracy and continuity significantly. The reader is referred to [1] for the details of the IRS hardware architecture.

Practically, discrete phase shift and reflection amplitude are considered because it is difficult to tune the hardware continuously. Researchers always divide the phase shift $2\pi$ into several equal subsets to implement discrete phase shift such as $0$ and $\pi$ for 1-bit hardware design and half $\pi$ interval for 2-bit hardware. For the amplitude control, 1-bit means two-level amplitude attenuation (i.e., 1 for total reflecting and 0 for fully absorbing) [9]. IRS channel usually is considered as a cascaded channel that combines three elements: two complex channel coefficients: the channel between the transmitter and IRS, the channel between the IRS and receiver, and the complex reflection coefficient. In the tutorial of [10], the baseband signal model of IRS-aided wireless communication is expressed as $y(t) = \mathbf{h}_r^H \boldsymbol{\Theta} \mathbf{g} x(t)$, where $x(t)$ denotes the equivalent complex-valued baseband transmit signal, $\mathbf{h}_r^H$ denotes the narrow-band frequency-flat channel matrix between the transmitter and IRS, $\boldsymbol{\Theta}$ is the diagonal complex reflecting matrix (i.e., $\beta_i e^{j\alpha_i}$ where $\beta_i$ and $\alpha_i$ are the amplitude reflection coefficient and phase-shift reflection coefficient respectively), and $\mathbf{g}$ denotes the channel between the IRS and receiver.

The passive reflecting characteristic of IRS endows it with numerous advantages than Massive Multi-Input Multi-Output (MIMO) and backscatter technologies such as low energy consumption, high spectral efficiency with full-duplex mode, and less self-interference. In addition, the implementation of IRS in wireless localization system can also reduce the number of APs and radio frequency anchors. This is an economical solution for the construction of next-generation mobile network system. Typical applications of IRS include indoor dead zone coverage extension, outdoor signal enhancement at the cell edge, assisting hybrid satellite-terrestrial relay networks in a link blockage scenario [11], and enhancing vehicle-to-everything (V2X) communications in downtown areas. It is especially suitable when the LOS is blocked. Other extended applications involve communication security (e.g., mitigating the eavesdropper signal) localization, and positioning with mobile network [12].

### B. Localization with IRS-aided Wireless Network

The research of localization and positioning with wireless communications can be categorized as follows: DoA, AoA, ToA, RSS, and CSI. AoA and ToA based techniques usually have large errors and need a strictly synchronized system

TABLE I
VARIOUS TECHNIQUES USED IN WIRELESS LOCALIZATION

| Ref. Example | Techniques | Advantages | Disadvantages |
|---|---|---|---|
| [13], [14] | CSI fingerprinting-based | insensitive to multipath effect, accurate | few chips supported, environment noise and dynamics |
| [18] | RSSI fingerprinting-based | infrastructure support | LOS only, multiple APs needed |
| [19] | UWB+ToA | low power consumption | prone to interference, strictly synchronized |
| [20], [21] | AoA, DoA | does not require clock synchronization | prone to interference, small coverage |
| | IRS+RSSI | work with both LOS and NLOS | IRS configurations |

which is difficult to be perfect. Meanwhile, the receiver is difficult to distinguish the signal source (i.e., IRS or transmitter). On the other hand, since CSI is difficult to be extracted from conventional Wi-Fi devices and only a few Wi-Fi chips, such as Intel 5300, are available [13], [14]. The RSS-based technique thus is widely applied due to its good compatibility and easy accessibility. In Table I, the advantages and disadvantages of various techniques of wireless localization are compared. However, the RSS-based localization technique also has its drawbacks such as instability and inaccuracy under the None-Line-Of-Sight (NLOS) situation, and thus we propose the IRS-aided localization scenario. IRS-aided terminal localization is a promising technique to solve the modern wireless localization problem. It can play an important role in the Global Navigation Satellite System (GNSS) denied environments but less attention was received. [15] proposed a signal model which is suitable for both near-field and far-field scenarios, and employed the Cramer-Rao lower bound to assess the localization performance. The IRS-aided terminal localization can be either active or passive. Most of the recent works conducted were active, and few of them focused on passive localization. In [15], the authors considered a three dimensional active localization scenario. Besides, this paper also considered both synchronous and asynchronous systems. In MIMO radar research, researchers use IRS to strengthen the received power and improve the target detection performance. They also validated the performance with the Cramer-Rao bound [16]. The authors in [17] proposed a method to identify the position of the receiver via configuring the IRS to generate more distinguished RSS patterns. Another interesting work was conducted by Nguyen et al. [7]. They employed machine learning methods for wireless fingerprinting localization aided by IRS. They first generated a database of fingerprints with different periodical IRS configurations. When the mobile terminal sends a request signal, the IRS will repeat the configuration and the terminal will collect the data. Then the data is used for comparing with the database utilizing different machine learning algorithms.

In conventional wireless localization research, the generation of multiple RSS fingerprints, and the deployment of multiple access points are time-consuming and costly. In this paper, we propose to generate more distinguished RSS fingerprints to improve the localization accuracy with the assistance of IRS and machine learning techniques. Compared with the existing works discussed above, this work not only utilizes IRS to improve the NLOS situations in RSS-based localization but also introduces the use of both discrete and continuous IRS phase shift to tackle the challenges in practical scenarios.

### C. Deep Reinforcement Learning and IRS

Sequential decision-making in an uncertain environment is a core topic in machine learning [22]. Reinforcement learning, a branch of machine learning, addresses this problem well. In its framework, an agent can learn to achieve optimal status by means of interacting within an environment and gathering experience for decision making based on the obtained rewards. Q-learning is a well-known model-free reinforcement learning model [23] but it can hardly deal with a large amount of data due to the size of Q-table. Additionally, only discrete states and actions are compatible with Q-table. To ride this out, [24] developed Deep Q-Network (DQN) that combines deep neural network (DNN) and Q-learning together. In DQN, the Q-table is replaced by DNN which makes infinite spaces possible for both action and state. However, the action spaces and state spaces still have to be discrete. The policy gradient (PG) algorithm is a policy-based algorithm and it can deal with continuous spaces. Nevertheless, it can only update the network after an episode is done, which has a slow convergence rate [25]. An elegant solution to this problem is the deep deterministic policy gradient (DDPG) algorithm proposed by [26]. This algorithm is based on the actor-critic network which can interact with continuous action spaces and state spaces.

There have already been various success stories of reinforcement learning in plentiful research areas, especially in sophisticated optimization and control problems. Motivated by this, some studies have already attempted to employ reinforcement learning to configure IRS so as to optimize the IRS-aided [2] and multi-user communication [27], enhance secure communication [28], assist anti-jamming communication [29], and design passive beamforming [30]. How to configure massive IRS passive elements is always a great challenge. Training beam and reflecting coefficient matrices is extremely time-consuming. [31] proposed a deep reinforcement learning-based algorithm for predicting such matrices with minimal training overhead. To the best of our knowledge, this paper is the first work that exploits DRL for IRS-enabled fingerprinting-based localization.

## III. RESEARCH FRAMEWORK AND SYSTEM MODEL

Motivated and inspired by existing works, this work proposes a DRL-based framework to simplify the solution for the IRS configuration selection problem in IRS-enabled fingerprinting-based localization. First, we propose the general framework for IRS-enabled fingerprinting-based localization with DRL. Then, we propose a practical protocol under this framework. Finally, the system models of IRS localization and DRL-based methodology are described in detail.

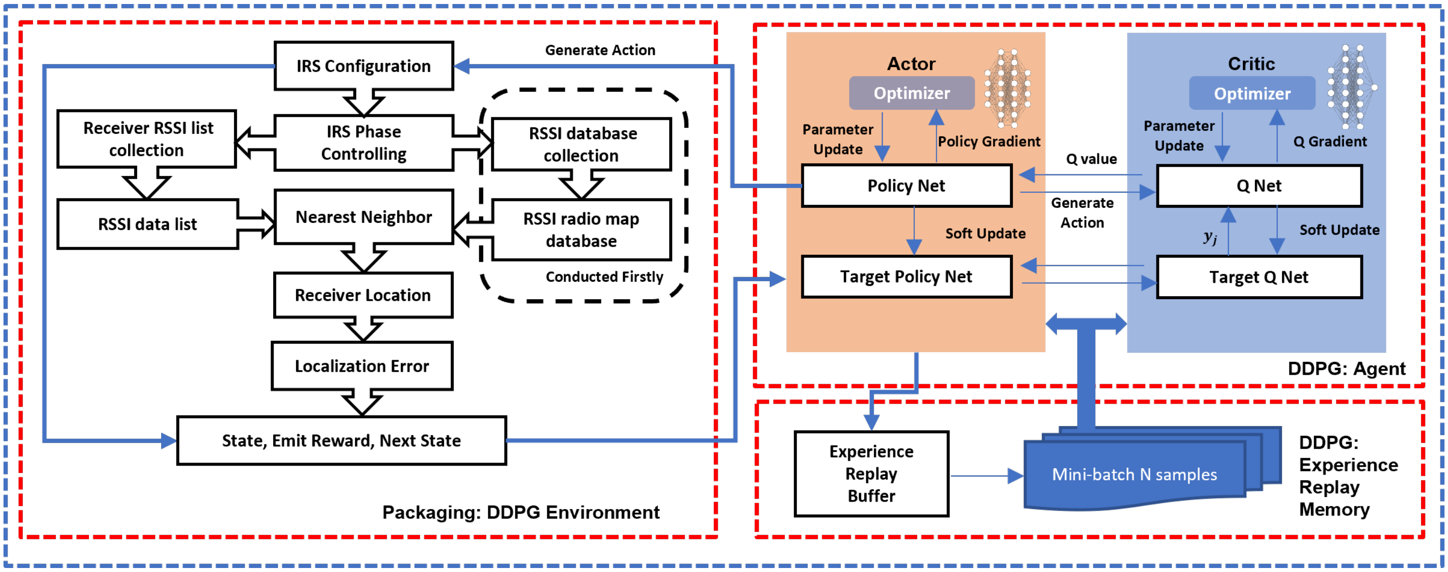



Fig. 1. The research framework of DDPG assisting IRS localization.

### *A. Problem Statement and Research Framework*

The research framework of the proposed IRS-enabled fingerprinting-based localization with DRL is illustrated in Fig. 1. As illustrated on the left, we first propose an IRS-enabled fingerprinting-based localization system. We leverage the reconfiguration capability of the IRS to create multiple RSSI lists instead of conventional trilateration techniques which utilize multiple anchor points. Then, as shown on the right part, we employ the DRL algorithm to select the qualified IRS configurations to achieve high localization precision. Specifically, instead of randomly or uniformly setting the phase shift (configuration) of each IRS element in a localization process, we use a DRL-based IRS configuration selector to identify the best combinations of IRS configurations. Note that for different application scenarios, the IRS configurations need to be re-selected to obtain the most suitable IRS configurations. After selection, the specific IRS configurations are stored on the server. When a request for localization is sent to the server, the server will use this set of IRS configurations to perform the localization and forward the result to the mobile user. Therefore, the arithmetic requirements on the IRS and server are not stringent since the IRS configurations are trained offline.

DRL usually consists of two essential parts: the agent and the environment which are demonstrated in Fig. 1. The left part of the framework is the developed DRL environment where our IRS-enabled fingerprinting-based localization algorithm is packaged. The right part of the framework is the DRL agent which is responsible for identifying the optimized combinations of IRS configurations through interacting with the DRL environment. The interaction between the environment and the agent can be regarded as a Markov Decision Process.

### *B. DRL Environment: System Model of IRS-enabled Fingerprinting-based Localization*

*1) Operating Principle:* Conventional fingerprinting-based localization methodologies utilize multiple APs to create Received Signal Strength Indicator (RSSI) list and match the data with pre-collected database to obtain the location [32]. In IRS-aided wireless communication system, only one AP and one IRS surface are needed for generating multiple fingerprints. Although single user is illustrated in the following description, the proposed methodology scales to serve multiple users simultaneously.

For the purpose of illustration, let us consider a downlink transmission from a multi-antenna AP to a single-antenna mobile device via an IRS. We assume that the direct link between the AP and the mobile device is totally blocked, and the IRS is deployed so that the channel between the AP and the IRS and that between the IRS and the mobile device are both LOS. Specifically, as shown on the left side of Fig. 1, an AP sends constant signals towards the IRS. The server prepares a set of different IRS configurations in advance and each IRS configuration is denoted by $\mathbf{C}$ (i.e., sets of phase shifts). The IRS applies different configurations at different time slots periodically, which generates a vector of distinguished RSSI for each location denoted by $\mathbf{r}$. The collection of all sets of IRS configurations is denoted by $\mathbf{\Phi}$ with a length of $N$. This RSSI vector can be regarded as the online fingerprint for each location. In the next step, the fingerprints of all locations are collected, which serves as the offline RSSI vector fingerprint database denoted by $\hat{\mathcal{R}}$, and this database will be stored in the backend server. The offline fingerprint of each location in the database is denoted by $\hat{\mathbf{r}}$. Next, during the localization process, the transmit beamforming and IRS configurations are the same as in the offline training phase. The mobile receiver located at $(x, y)$ (the true position) collects its online RSSI vector fingerprint $\mathbf{r}$ and sends it to the server. The server compares this fingerprint $\mathbf{r}$ against pre-collected fingerprints database $\hat{\mathcal{R}}$ and estimates the mobile device's location $(\hat{x}, \hat{y})$ using the nearest neighbor algorithm. Note that the proposed methodology in this paper can be directly extended to the setup with multiple APs operating at orthogonal frequencies, where the RSSI fingerprint will be represented by a matrix (consisting of the RSSI vectors for each AP) and the number of feature points of the fingerprint will increase. Based on this, the localization error is anticipated to decrease as the number of APs increases. Moreover, it will also be interesting to consider the scenario with multiple multi-antenna APs operating at the same frequency. In this case, the AP precoding, IRS configurations, and receiver signal processing need to be jointly designed for minimizing the localization error, which is left as our future work.

In the proposed methodology, we consider both continuous and discrete phase shifts at the IRS. Discrete phase shift at each reflecting element is chosen from a set of quantized values (i.e., 0, $2\pi/b$, $4\pi/b$, ..., $2(b-1)\pi/b$), where $b$ denotes the number of quantization bits. Note that such discrete phase shifts are easier to implement in practice (the reader is referred to [33] for the system using 1- bit phase shift). Since the aim of using IRS is to achieve total reflection with its main beam angle, we thus simply consider one set IRS configuration to a desired reflection angle $\theta_r$. The collection of all IRS configurations at different reflection angles is represented by the vector $\mathbf{\Phi}$. The step can also reduce the burden of the DRL training process since instead of training each IRS element for different configurations, we just need to train the whole IRS surface with desired reflection angles. After this, the phase shift of each IRS element can be restored according to (13) in [34]. For example, let us assume the number of IRS configurations is equal to ten. Then $N = 10$ and $\mathbf{\Phi}$ is the list of the simplified ten desired reflection angles $\mathbf{\Phi} = \{\theta_{r1} \dots \theta_{r10}\}$. $\mathbf{C}$ refers to all sets of phase shifts of each IRS element after the IRS restoration. In case of discrete phase shifts, we can set the local phase shift to closely approximate the calculated phase shift. It is worth noting that the accuracy of the approximation is related to the size of the IRS element. The smaller the IRS element, the more accurate the phase shift can be approximated. Specifically, in each IRS configuration, we select one phase shift value from the above sets for each reflecting element. For example, to generate the configuration of 1-bit IRS discrete elements, we set the value of the phase shift reflection coefficient $\alpha_i$ of each element to 0 or $\pi$. For 2-bit IRS, the value would be 0, $\pi/2$, $\pi$ and $3\pi/2$. For the reflection amplitude of each element, we set it to the maximum value of one for maximum signal reflection.

*2) Localization Protocol:* In this subsection, we propose a protocol for the IRS-aided localization as shown in Fig. 2. Let $N$ denote the number of IRS configurations which is equal to the length of $\mathbf{\Phi}$. In each localization cycle, the AP firstly uses several time slots to broadcast the location information calculated in the previous cycle, while the IRS configures its phase shifts according to the pre-designed pattern over the $n$ time slots. The mobile device receives the signal reflected by the IRS and collects the list of RSSI vectors, which is then fed back to the AP and the backend server. The server then calculates the mobile device location at the end of the cycle and broadcasts this location information at the beginning of the next cycle.

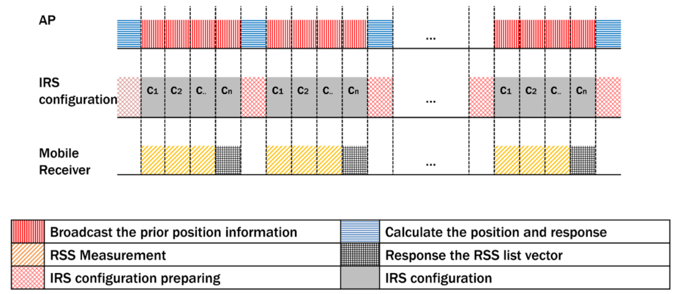



Fig. 2. The protocol for IRS-enabled fingerprinting-based localization

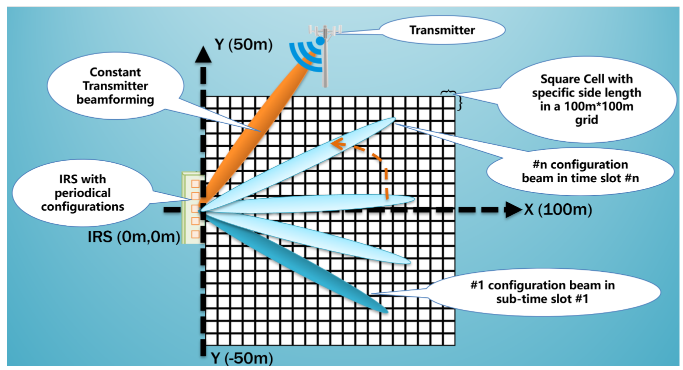



Fig. 3. Illustration of the proposed IRS-aided localization scenario

In Fig. 3, we illustrate the system model for the proposed IRS-aided localization protocol. For example, let us consider the scenario of locating a mobile device on a 100 m $\times$ 100 m two-dimensional plane. The main lobe of the transmit beamforming is illustrated with the orange beam, while the reflected beams via the IRS at different time slots are illustrated using the blue beams. It can be observed that the main lobe of the beam at each time slot is different, which results in different RSSI fingerprints at the mobile device over multiple time slots.

*3) Localization Accuracy:* The localization accuracy can be implied by the localization error. In some cases where two or more nearest neighbors are found with the same distance, we use an average function to get sole output to avoid program error. In the proposed IRS-aided localization methodology, the final estimate is obtained via averaging over all possible locations as identified by the model. Thus, the mean localization error is calculated by averaging all possible locations. Let us assume that there are $l$ locations that are found in a localization process. The ground truth location is denoted by $(x, y)$ and the $l$ estimated locations are denoted by $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ where $\hat{\mathbf{x}} = \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_l\}$ and $\hat{\mathbf{y}} = \{\hat{y}_1, \hat{y}_2, ..., \hat{y}_l\}$. The probability of each possible location is $\xi_i$ where $i = 1, 2, ..., l$. Then the averaged location estimate $(\bar{\hat{x}}, \bar{\hat{y}})$ is calculated though the following equation:

$$\bar{\hat{x}} = \sum_{i=1}^{l} \xi_i \hat{x}_i, \tag{1}$$

$$\bar{\hat{y}} = \sum_{i=1}^{l} \xi_i \hat{y}_i. \tag{2}$$

The mean localization error is calculated though the following equation:

$$\varepsilon_{(x,\hat{x})} = \sqrt{(\bar{\hat{x}} - x)^2 + (\bar{\hat{y}} - y)^2} \tag{3}$$

or we can also employed the root mean squared error (RMSE):

$$\varepsilon_{(x,\hat{x})} = \sqrt{\frac{1}{l} \sum_{i=1}^{l} \xi_i [(\hat{x}_i - x)^2 + (\hat{y}_i - y)^2]} \tag{4}$$

### *C. DRL Agent: IRS Localization with Deep Deterministic Policy Gradient Algorithm*

According to the discussion in related works, the DQN algorithm cannot deal with continuous IRS configurations and the PG algorithm has a slow convergence rate. Thus, we employ the DDPG algorithm here. The DDPG algorithm combines the excellence of both the DQN and PG algorithms. It is a model-free off-policy actor-critic algorithm that can deal with continuous action spaces. A DDPG usually consists of two main parts: the agent and the environment. As shown in Fig. 1, the environment is the packaged IRS-enabled fingerprinting-based localization algorithm that we discussed in the last subsection. It can conduct the action received from the agent and emit the Award which is used to evaluate the performance of the action under the current State. Meanwhile, it also updates the State to the Next State. The agent has four deep neural networks: Actor network, Critic network, target Actor network, and target Critic network. The Actor employs a deterministic policy gradient network to choose an action from a continuous action space. The Critic employs a DQN to measure the performance of the chosen action. An experience replay buffer is also employed here to reduce the correlation of different training samples. Just as how DQN operates to update its network, the Actor and Critic networks also need two target networks (that share the same structure with its corresponding evaluation networks) to update the parameters. The soft update algorithm is used to update parameters:

$$\boldsymbol{\omega}_{i'} = \tau \boldsymbol{\omega}_i + (1 - \tau) \boldsymbol{\omega}_{i'} \tag{5}$$

where $\boldsymbol{\omega}$ denotes the parameters of the DNN in the agent and $\tau \ll 1$ denotes the update parameter. Soft update can make sure that the parameters are updated slowly and increase the stability of the training process [26]. Moreover, we add a random noise sample $\mathcal{N}$ to the actor policy since the exploration problem can be treated independently in DDPG. The DDPG aims to maximize the reward and it is configured as follows:

- **Reward**: The aim of DRL is to select the most qualified IRS configurations with the highest localization precision (i.e., minimum localization error $\varepsilon$). In related works [2], [31], [35], the reward functions in an IRS study are usually defined as the corresponding objectives such as SNR, quality of service, and the ratio of data rate to energy consumption. Therefore, in the DRL training process, to maximize the accumulated rewards $r_t$, the reward function is defined as the opposite of the localization error $-\varepsilon$:

$$r_t = -\varepsilon_{(x,\hat{x}),t}. \tag{6}$$

- **Action Space**: The Action at time $t$ is defined as the superposition of the updated IRS configurations under current channel states and a random noise sample $\mathcal{N}$ for the network exploration. The updated IRS configurations are the output of the agent with the input State $S_t$ at time $t$. Thus the Action vector $A_t$ at time $t$ is:

$$A_t = [\theta_{r1}^t, ..., \theta_{rN}^t] + \mathcal{N}. \tag{7}$$

- **State Space**: The State vector $S_t$ at time $t$ is defined as the IRS configurations and reward at the previous step $(t-1)$ which is

$$S_t = [r_{t-1}, \theta_{r1}^{t-1}, ..., \theta_{rN}^{t-1}]. \tag{8}$$

**Algorithm 1:** The IRS Localization with DRL

**Input:** The learning rate $\rho$; The discount factor $\lambda$; The soft update coefficient $\tau$; The experience replay $\mathcal{M}$ and its capacity $\mathcal{C}$. The length of IRS configurations $N$. The mini batch size $B_M$. Randomly initialize the parameters of critic and actor network as well as the target critic and target actor network. Empty the experience replay buffer $\mathcal{M}$.

**Output:** The optimal IRS configurations $\mathbf{\Phi}^*$ and the minimized localization error of IRS-enabled fingerprinting-based localization system $\varepsilon_{min}$.

```
for episode i = 1, ..., K do
    Initialize the IRS localization system;
    Randomly choose the initial set of IRS
      configurations Φ^(0) and obtain the reward r_0 as
      the current state;
    Initialize a random process N;
    for step t = 1, ..., T do
        Action a_t = μ(s_t; ω) + N; Reform a_t into IRS
          configurations and calculate the localization
          error (i.e., reward r_t);
        Obtain the Next State S_{t+1} = [r_t, θ^t_{r1}, ..., θ^t_{rN}];
        Store the experience {S_t, A_t, r_t, S_{t+1}} into the
          experience replay M;
        Sample a B_M-size minibatch
          {S_j, A_j, r_j, S_{j+1}}(j = 1, ..., B_M) from the
          experinece replay M;
        Calculate the target Q value;
        Update the critic network by minimizing the
          loss function and stochastic gradient descent
          algorithm;
        Update the actor network utilizing sampled
          policy gradient;
        Soft Update the two target networks according
          to Equation (3);
        Update the State to the Next State;
```

**DRL-based Workflow**: As shown in Fig. 1, the overall workflow is as follows: the Agent generates action (i.e., IRS configurations) and feeds it to the DDPG environment. After conducting the localization algorithm, the Environment emits the reward and the Next State to feedback to the agent for updating the DDPG networks. Then the Agent will generate a new action of IRS configuration for the next cycle until the maximum reward is observed. Specifically, prior to commencing the algorithm, an initialization stage is conducted. Four DNNs are generated: the actor network with parameter $\boldsymbol{\omega}_{\mu}$, the critic network with parameter $\boldsymbol{\omega}_{Q}$, the target actor network with parameter $\boldsymbol{\omega}_{\mu'}$ and the target critic network with parameter $\boldsymbol{\omega}_{Q'}$. These parameters are initialized with normally distributed random numbers. Moreover, an experience replay memory $\mathcal{M}$ is initialized with capacity $\mathcal{C}$ as well. It is used to store the experience during training. The IRS configurations are initialized randomly at the beginning of each episode. During each episode, firstly, feed the state $S_t$ at time $t$ as input to actor network and obtain the corresponding action $A_t$ at time $t$. Then, the generated action flows to the DDPG environment and emits the reward $r_t$ and the Next State $S_{t+1}$. Once finished, the algorithm will store $\{S_t, A_t, r_t, S_{t+1}\}$ as one transition into $\mathcal{M}$. The critic network samples a $B_M$ size mini-batch $\{S_j, A_j, r_j, S_{j+1}\}(j = 1, ..., B_M)$ from the experience replay $\mathcal{M}$ to calculate the target Q value $y_j$ according to the following equation:

$$y_j = \begin{cases} r_j, & j = B_M \\ r_j + \lambda Q'\left(s_{j+1}, \mu'\left(s_{j+1}; \boldsymbol{\omega}_{\mu'}\right); \boldsymbol{\omega}_{q'}\right), & j < B_M. \end{cases} \tag{9}$$

The loss function of critic network is defined as:

$$L\left(\boldsymbol{\omega}_q\right) = \frac{1}{B_M}\sum_{j=1}^{B_M}\left(y_j - Q\left(s_j, a_j; \boldsymbol{\omega}_q\right)\right)^2 \tag{10}$$

Then, the critic network can be updated with the stochastic gradient descent algorithm and using policy gradient to update the actor network with ascent factor according to the equation below:

$$\Delta_{\boldsymbol{\omega}_\mu} = \frac{1}{B_M}\sum_{j=1}^{B_M}\left(\nabla_a Q\left(s_j, \mu\left(s_j; \boldsymbol{\omega}_\mu\right); \boldsymbol{\omega}_q\right) \mid \nabla_{\boldsymbol{\omega}_\mu}\mu\left(s_j; \boldsymbol{\omega}_\mu\right)\right). \tag{11}$$

Finally, the two target networks are updated utilizing soft update according (3). The whole algorithm is shown in Algorithm 1.

## IV. Simulation and Results

In this section, we conduct several simulations to demonstrate the performance of the proposed IRS-enabled fingerprinting-based localization with DRL. The first subsection presents the detailed configurations of the simulation system. The second subsection demonstrates and analyses the results.

### A. Configuration of Simulation Systems

As shown in Fig. 3, the simulation setup of the IRS localization system includes a transmitter, an IRS, and receivers.

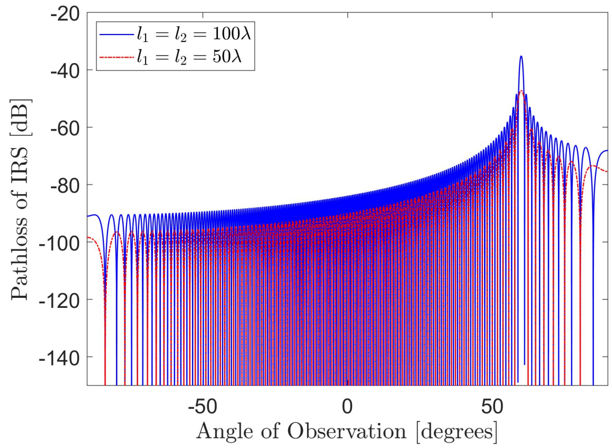



Fig. 4. An example of IRS phase-shift reflection coefficient with configuration of $\theta_r = \pi/3$.

The AP-IRS and IRS-device channels are assumed to be LOS channels. We ignore the signal reflected by the IRS more than once. The path loss of IRS is modeled with (18) in [34]. Fig. 4 shows how the pathloss of IRS reflected path varies with the angle of observation when $\theta_r = \pi/3$ and the angle of incident signal is $\pi/6$. Two different IRS sizes are considered and shown in different colors. The blue curve has a larger size and a narrower main beam-width. When the observation angle is equal to $\theta_r$, the maximum pathloss is achieved.

Practically, the desired local IRS configurations are discretized by separating the IRS surface into sub-wavelength elements. The smaller the elements, the more precise the local configurations can be approximated. In Fig. 5, we illustrate how to configure the IRS element with four different types of IRS (i.e., continuous IRS, 1-Bit discrete IRS, 2-Bit discrete IRS, and 3-Bit discrete IRS) where $y$ denotes the location of IRS element along the direction of the y-axis of the IRS plane as shown in Fig. 3. The readers is referred to [34] for the details of the IRS configuration model. For example, if the size of the IRS element is half of the wavelength $\lambda$, then the centroid position of each IRS element is integer multiple of half of the wavelength (i.e., $0.5\lambda$, $\lambda$, $1.5\lambda$, etc.). The values of the horizontal axis thus can be regarded as $y/\lambda = 0.5, 1, 1.5, ...$ in Fig. 5, and the local phase shift $\alpha_i$ of each IRS element corresponds to the value of the vertical axis. We only show an IRS with a size of 20 times the wavelength due to the limited figure size. The phase-shift of a larger IRS size can be implied in the same way.

In addition, a zero-mean Gaussian noise is added to the RSSI with a standard deviation of 3 dBm. The IRS is configured with maximum signal reflection. Fig. 6 gives four examples of the generated RSSI database of four different IRS configurations in form of radio maps. These radio maps are generated with IRS configurations of $\theta_r = \pi/6, \pi/4, \pi/6, -\pi/3$. The frequency is 5 GHz. The size of the whole IRS planer surface is 200 times the wavelength on both sides. Let us consider a region of 100 m x 100 m with 1 m$^2$ grids. The transmitter is located 50 meters away from the localization plane and the IRS is located at the origin (0,0) of the

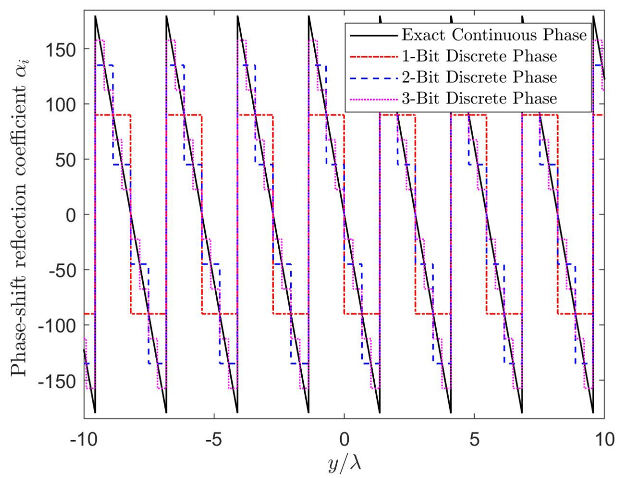



Fig. 5. An example of pathloss of IRS reflected path with configuration of $\theta_r = \pi/3$.

coordinate system.

The DRL methodology for IRS localization is configured as follows: all neural networks are three-layer fully connected DNN. The Actor and Critic networks use Adam optimizer for parameter updating. The Actor network is structured with $N$ (i.e., the number of IRS configurations) output neurons and $N + 1$ input neurons. The Critic network contains one output neuron and $2N + 1$ input neurons. The hidden layer of all networks has 50 neurons. The first two layers are followed by ReLU activation while the output layers use the tanh function to guarantee the gradient. Moreover, the size of experience replay $\mathcal{M}$ is set to 200,000 and the mini-batch size $B_M$ is set to 16. We configure the learning rate of the Actor and Critic networks to $5 \times 10^{-5}$ and $5 \times 10^{-4}$ respectively. The soft update rate $\tau$ is set to 0.02. The gamma value is 0.99.

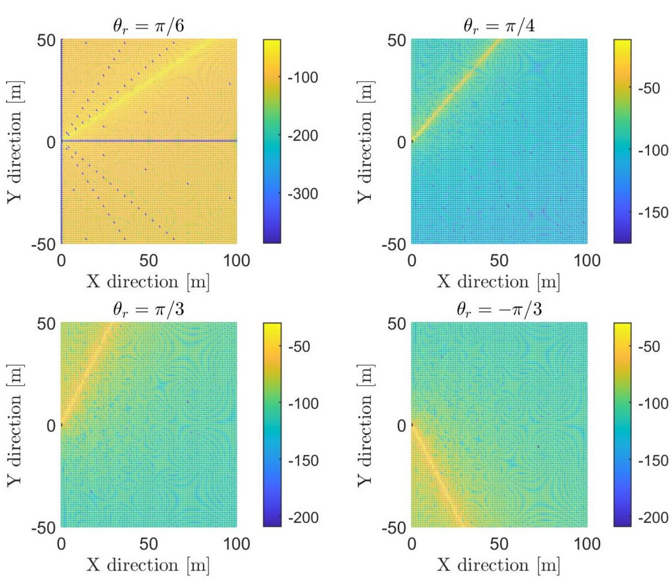



Fig. 6. Examples of generated radio maps with different IRS configurations in dB.

## B. Simulation Results

In the proposed methodology, each IRS configuration will generate a unique radio map that contains the RSSI of each cell in the grid. Generally, the localization accuracy will increase with the number of IRS configurations due to the larger difference among adjacent locations. In the following simulation, we examine the impact of the number of IRS configurations on the localization error. Additionally, to evaluate the quality of the proposed scheme, we implement random configurations and uniform configurations to the IRS as benchmarks. Continuous phase-shift means that the phase shift can be of any value within the limitation and this can increase the distinguishability of each location's fingerprint significantly. The discrete phase shift of each IRS element can be restored according to (13) in [34].

Fig. 7 presents the localization errors versus different numbers of IRS configurations. Random IRS configuration means we choose the IRS configurations randomly to perform localization. Uniform IRS configuration means that we choose a linear spaced vector of IRS configurations $\theta_r$ from $-\pi/2$ to $\pi/2$ with a size of $N$ and conduct the localization. For example, if the number of configurations is 5, then the uniform IRS configurations are $-\pi/2$, $-\pi/4$, 0, $\pi/4$, $\pi/2$. The DDPG IRS means that the IRS localization is assisted by DDPG to select the best combinations of IRS configurations to perform the localization. All of these localization errors are average localization errors over 5000 random noise realizations. Each realization picks a cell randomly from the considered region of $100 \times 100$ grid and performs the localization with the selected IRS configurations. The channel model we assumed is the same as (18) in [34]. In this reference, the author proposed an IRS-reflecting signal propagation and path loss model. The AP does not have a line of sight to any cell and this is assumed in the proposed DRL environment. In our study, the localization capability is enabled by the RSSI fingerprint distinguishability generated by IRS configurations. Thus, even though the AP has the line of sight to some cells, the DRL is still expected to select the optimized IRS configurations to create fingerprints with significant distinguishability .

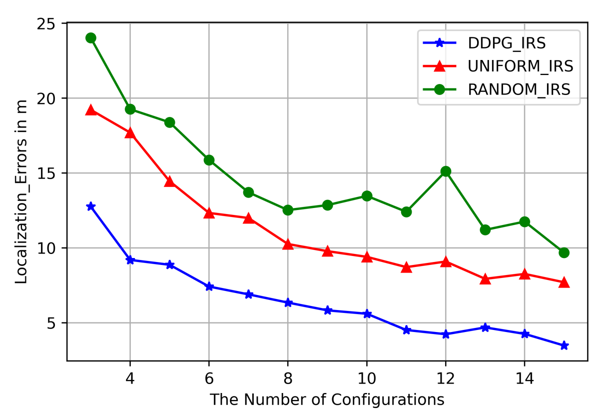



Fig. 7. The localization error v.s. The number of IRS configurations.

In Fig. 7, it can be seen that as the number of IRS configurations increases, the trends of all three error lines decrease even though some small fluctuations exist on all three lines and the differences among the different configuration methodologies are clear. The DDPG IRS always performs the best among these methodologies since the DDPG can select the most efficient IRS configurations to generate the most differentiable radio maps for localization. The random IRS configuration scenario has the largest errors since randomly chosen IRS configurations cannot generate a differentiable RSSI database for localization. When the number of IRS configurations is 12. The most significant fluctuation happens on the random IRS line when the number of IRS configurations is 12. This occurs for two reasons. Firstly, all the results are averaged over 5000 random noise realizations. The randomness of both the IRS selection process and noise realization can lead to such significant fluctuation. Secondly, the added zero-mean Gaussian noise could introduce the fluctuation to the localization error as well and the proposed model is not deterministic. The red line in the figure is the localization error with uniformly chosen IRS configurations, which is used to compare the performance of the DDPG and random IRS. It is always located between the other two methodologies. This means that if we choose the IRS configurations uniformly, we still can obtain more differentiable radio maps to conduct localization than the random method. Moreover, when the number of IRS configurations is larger than ten, an acceptable localization service with an error of less than 5 meters can be obtained. It is also observed that when the number of configurations is beyond 13, the performance improvement almost starts to saturate, and only marginal performance gain can be achieved. It is worth noting that as the number of IRS configurations increases, the delay of the localization protocol increases since each configuration takes a time slot. Hence this is a fundamental trade-off between localization accuracy and efficiency. The running time analysis is shown in Fig. 8. The CPU for DRL training is Intel(R) Core(TM) i7-9750H with 2.60 GHz with 32.0 GB RAM and it is accelerated by the NVIDIA GeForce GTX 1650 GPU. All the configuration sets are trained with 100 steps and 50 episodes. Fig. 8 demonstrates the amount of time needed for training different numbers of IRS configurations. The running time increases with the rising number of IRS configurations.

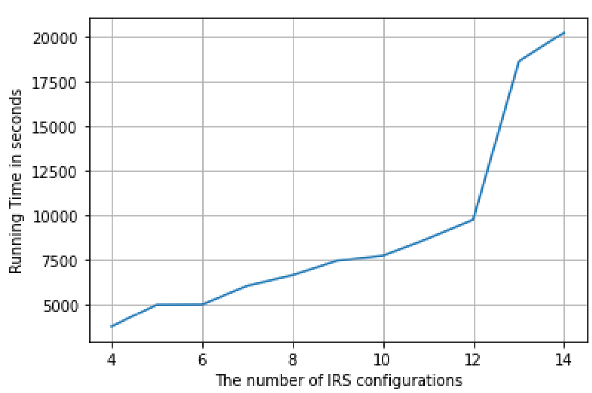



Fig. 8. The running time in seconds for different numbers of IRS configurations.

TABLE II
COMPARISON OF RESULTS USING DIFFERENT METHODOLOGIES

| **Number of IRS Configurations** | **4** | **6** | **8** | **10** |
|---|---|---|---|---|
| **kNN-FS L400(Approximate) in m** | 3.2 | 2.5 | 1.95 | 2 |
| **DDPG-IRS in m** | 1.723 | 1.01 | 0.929 | 0.543 |

To evaluate the localization performance of our proposed methodology, we also compare the results with the localization methodology proposed in [7] which is the most similar work to ours. We try to configure our settings in the same way to the best of our knowledge. For example, we change the number of grids from 10,000 to 400. In Table II, we compare our DRL-based methodology with the kNN method presented in [7]. We test the DRL-based methodology with the number of IRS configurations from 4 to 15. When the number of IRS configurations is 4, DRL-based localization can achieve a 1.723 m error while the result in [7] is over 3 m. When the number of IRS configurations is 10, the proposed method can achieve an error of 0.5428 m while the error of the benchmark is over 2 m. This comparison indicates that the proposed DRL-based methodology can achieve better performance and increase the accuracy over $40\%$ in IRS-enabled fingerprinting-based localization.

## V. CONCLUSION

With the growing interest of IRS, future 6G systems will not only provide ubiquitous and reliable communications but also deliver many location-based information. In this study, we have had a throughout review of IRS, IRS-aided localization methodology, and IRS with DRL. Then, we have proposed a practical IRS-aided localization methodology and developed a DRL-based framework to optimize the IRS configuration selection problem. Moreover, we have proposed a protocol for the implementation of the methodology. In the simulation, we have investigated the impact of the number of IRS configurations on localization errors and compared the results with previous work. Overall, our proposed DRL-based methodology can achieve a better performance with over $40\%$ improvement. It can be concluded from the simulation results that the localization error can be reduced by increasing the number of IRS configurations.

In future studies, the practical challenges of establishing the radio map database and real-time configuration of DRL need to be further addressed. We plan to develop a practical IRS-RSSI collection system to interact with the server so that the DRL training process can be performed automatically in practice. The initial idea is to integrate the RSSI collector and mobility capability into a single board computer (e.g., Raspberry Pi) and let it cooperate with the server for the fingerprint collection and DRL training. Although a simulated channel model is used to train the DRL selector in this paper, in real practice, the empirical data or a hybrid of both real and simulated data can be employed to train the selector. In our prior work [36], we exploited empirical channel data collected on campus and simulated data to train a city-model-aware DNN to model the wireless communication channel. The results indicate that such model can significantly improve the channel estimation performance. Therefore, this work can be further extended and the algorithm can be verified in such a digital twin environment that includes channel data from both the simulated and real worlds in the future.

## REFERENCES


[1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2019.

[2] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for miso communication systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 745–749, 2020.

[3] E. Björnson, H. Wymeersch, B. Matthiesen, P. Popovski, L. Sanguinetti, and E. de Carvalho, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *arXiv preprint arXiv:2102.00742*, 2021.

[4] Z. Lin, M. Lin, T. De Cola, J.-B. Wang, W.-P. Zhu, and J. Cheng, "Supporting iot with rate-splitting multiple access in satellite and aerial-integrated networks," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 123–11 134, 2021.

[5] J. Zhang, Q. Su, B. Tang, C. Wang, and Y. Li, "Dpsnet: Multitask learning using geometry reasoning for scene depth and semantics," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[6] A. M. Elbir and K. V. Mishra, "A survey of deep learning architectures for intelligent reflecting surfaces," *arXiv preprint arXiv:2009.02540*, 2020.

[7] C. L. Nguyen, O. Georgiou, G. Gradoni, and M. Di Renzo, "Wireless fingerprinting localization in smart environments using reconfigurable intelligent surfaces," *IEEE access*, vol. 9, pp. 135 526–135 541, 2021.

[8] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Metalocalization: Reconfigurable intelligent surface aided multi-user wireless indoor localization," *IEEE Transactions on Wireless Communications*, 2021.

[9] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, 2019.

[10] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Transactions on Communications*, 2021.

[11] Z. Lin, H. Niu, K. An, Y. Wang, G. Zheng, S. Chatzinotas, and Y. Hu, "Refracting ris aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Transactions on Aerospace and Electronic Systems*, 2022.

[12] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE access*, vol. 7, pp. 116 753–116 773, 2019.

[13] J. M. Rocamora, I. W.-H. Ho, W.-M. Mak, and A. P.-T. Lau, "Survey of csi fingerprinting-based indoor positioning and mobility tracking systems," *IET Signal Processing*, vol. 14, no. 7, pp. 407–419, 2020.

[14] D. Khan and I. W.-H. Ho, "Deep learning of csi for efficient device-free human activity recognition," *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)*, pp. 19–24, 2021.

[15] A. Elzanaty, A. Guerra, F. Guidi, and M.-S. Alouini, "Reconfigurable intelligent surfaces for localization: Position and orientation error bounds," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5386–5402, 2021.

[16] W. Lu, B. Deng, Q. Fang, X. Wen, and S. Peng, "Intelligent reflecting surface-enhanced target detection in mimo radar," *IEEE Sensors Letters*, vol. 5, no. 2, pp. 1–4, 2021.

[17] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Towards ubiquitous positioning by leveraging reconfigurable intelligent surface," *IEEE Communications Letters*, vol. 25, no. 1, pp. 284–288, 2020.

[18] S. Boonsriwai and A. Apavatjrut, "Indoor wifi localization on mobile devices," *2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 1–5, 2013.

[19] J. Xiao, Z. Liu, Y. Yang, D. Liu, and X. Han, "Comparison and analysis of indoor wireless positioning techniques," *2011 International conference on computer science and service system (CSSS)*, pp. 293–296, 2011.

[20] T. L. N. Nguyen and Y. Shin, "A new approach for positioning based on aoa measurements," *2013 International Conference on Computing, Management and Telecommunications (ComManTel)*, pp. 208–211, 2013.

[21] H. Zheng, Z. Shi, C. Zhou, M. Haardt, and J. Chen, "Coupled coarray tensor cpd for doa estimation with coprime l-shaped array," *IEEE Signal Processing Letters*, vol. 28, pp. 1545–1549, 2021.

[22] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *arXiv preprint arXiv:1811.12560*, 2018.

[23] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[24] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[25] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," *International conference on machine learning*, pp. 387–395, 2014.

[26] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[27] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser miso systems exploiting deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, 2020.

[28] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 375–388, 2020.

[29] H. Yang, Z. Xiong, J. Zhao, D. Niyato, Q. Wu, H. V. Poor, and M. Tornatore, "Intelligent reflecting surface assisted anti-jamming communications: A fast reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1963–1974, 2020.

[30] L. Wang, K. Wang, C. Pan, W. Xu, and N. Aslam, "Joint trajectory and passive beamforming design for intelligent reflecting surface-aided uav communications: A deep reinforcement learning approach," *arXiv preprint arXiv:2007.08380*, 2020.

[31] A. Taha, Y. Zhang, F. B. Mismar, and A. Alkhateeb, "Deep reinforcement learning for intelligent reflecting surfaces: Towards standalone operation," *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2020.

[32] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless rssi fingerprinting localization," *Signal Processing*, vol. 131, pp. 235–244, 2017.

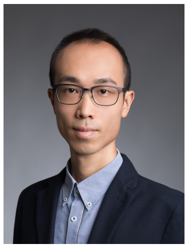
[33] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light: Science & Applications*, vol. 3, no. 10, pp. e218–e218, 2014.

[34] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, 2019.

[35] G. Lee, M. Jung, A. T. Z. Kasgari, W. Saad, and M. Bennis, "Deep reinforcement learning for energy-efficient networking with reconfigurable intelligent surfaces," *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2020.

[36] C. Ding and I. W.-H. Ho, "Digital-twin-enabled city-model-aware deep learning for dynamic channel estimation in urban vehicular environments," *IEEE Transactions on Green Communications and Networking*, 2022.


**Yuhao Wang** was born in Shuangyashan, Heilongjiang, China. He received the B.Eng. degree in traffic engineering from Southeast University, China, in 2018 and the M.S. degree in transport from Imperial College London in 2019. He is currently working toward the Ph.D. degree with The Hong Kong Polytechnic University, Kowloon, Hong Kong, focusing on intelligent transport system and smart pavement. His work on a smart, integrated road pavement and drainage system for storm-water storage, deicing, dust suppression, and cooling received the Gold Medal with the Organizer's Choice Award in the International Invention Innovation Competition in Canada (iCAN) in 2020.



**Ivan Wang-Hei Ho** (M'10–SM'18) received the B.Eng. and M.Phil. degrees in information engineering from The Chinese University of Hong Kong, Hong Kong, in 2004 and 2006, respectively, and the Ph.D. degree in electrical and electronic engineering from the Imperial College London, London, U.K., in 2010. He was a Research Intern with the IBM Thomas J. Watson Research Center, Hawthorne, NY, USA, and a Postdoctoral Research Associate with the System Engineering Initiative, Imperial College London. In 2010, he cofounded P2 Mobile Technologies Ltd., where he was the Chief Research and Development Engineer. He is currently an Associate Professor with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong. His research interests include wireless communications and networking, specifically in vehicular networks, intelligent transportation systems (ITS), and Internet of things (IoT). He primarily invented the MeshRanger series wireless mesh embedded system, which received the Silver Award in Best Ubiquitous Networking at the Hong Kong ICT Awards 2012. His work on indoor positioning and IoT also received a number of awards, including the Gold Medal at the International Trade Fair Ideas and Inventions New Products (iENA) in 2019, the Gold Medal with the Organizer's Choice Award at the International Invention Innovation Competition in Canada (iCAN) in 2020, and the Gold Medal at the International Exhibition of Inventions Geneva in 2021. He is currently an Associate Editor for the IEEE Access and IEEE Transactions on Circuit and Systems II.

**Shuowen Zhang** (S'14-M'18) received the B.Eng. degree in information engineering from Chien-Shiung Wu Honors College, Southeast University, Nanjing, China in June 2013, and the Ph.D. degree from NUS Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore, in January 2018 under the NGS scholarship. From 2017 to 2020, she was a Research Fellow at the Department of Electrical and Computer Engineering, National University of Singapore. She is currently a Research Assistant Professor at the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University. Her current research interests include intelligent reflecting surface aided communications, UAV communications, multiuser multiple-input multiple-output (MIMO) communications, and optimization methods.

**Yuhong Wang** received his M.Eng. and B.Eng. degree at Tongji University, Shanghai, China in 1996, and the Ph.D. and MSc in civil engineering at the University of Kentucky, the USA in 2003 and 2001, respectively. He is currently a Professor with the Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong. His research interest focuses on the new generation of urban infrastructure, which includes how to make future cities cleaner and more environmentally friendly, more resistant to floods, better serve urban residents, smarter, and how to promote biodiversity in the urban environment.