Toward 5G NR High-Precision Indoor Positioning via Channel Frequency Response: A New Paradigm and Dataset Generation Method

Kaixuan Gao[ID], Huiqiang Wang[ID], Hongwu Lv[ID], and Wenxue Liu

Abstract—Location-based services (LBSs) provide necessary infrastructure for daily life, from bicycle sharing to nursing care. In contrast to traditional positioning methods such as Wi-Fi, Bluetooth, and ultra-wideband (UWB), fifth-generation (5G) networking is defined as a paradigm of *integrated sensing and communication* (ISAC). With its advantages of wide-range coverage and indoor-outdoor integration, 5G is promising for high-precision positioning in indoor and urban canyon environments. However, 5G location studies face great obstacles due to the lack of commercialized 5G ISAC base stations that support positioning functions as well as publicly available datasets. In this paper, we first propose a dataset generation method, the Multilevel Feature Synthesis Method (Multilevel-FSM), to obtain positioning features. In particular, the features of a multiple-input multiple-output (MIMO) channel are flattened into a single image to increase the information density and improve feature expression, and data augmentation is performed to provide stronger robustness to noise. Subsequently, we devise a specially designed deep learning positioning method, Multipath Res-Inception (MPRI), trained on the proposed dataset to enhance positioning accuracy. Finally, the results of extensive experiments conducted in two typical 5G scenarios (indoors and urban canyon) show that Multilevel-FSM and MPRI outperform state-of-the-art works in accuracy, time overhead and robustness to noise.

Index Terms—5G, high-precision positioning, dataset generation, deep learning.

I. Introduction

HIGH-PRECISION location-based services (LBSs) have become a necessary infrastructure, similar to water, electricity, and gas, facilitating bicycle sharing, unmanned workshops, nursing care services and other scenarios [1]–[4]. Benefiting from the widespread existence of satellite networks, such as the Global Positioning System (GPS), LBSs have become a reality in open outdoor scenarios [5], [6]. However, LBSs still face enormous challenges in high-rise urban

canyons and indoor scenarios, where people spend more than 80% of their time. Although Wi-Fi [7], Bluetooth [5], [8], ultra-wideband (UWB) [9] and other technologies have been widely adopted for indoor positioning, city-wide coverage is still difficult to achieve due to their limited coverage and additional deployment costs. Fortunately, 5G technology as specified by the 3rd Generation Partnership Project (3GPP) Release 16 (R16) standard [10], as a new paradigm of *integrated sensing and communication* (ISAC), provides potential support for such scenarios [11], [12]. Specifically, 5G base stations (gNodeBs, or gNBs) can support ISAC natively via a joint signal processing framework, making full use of spectrum resources. Additionally, in 3GPP R16, 6 positioning schemes are listed to reach metre- or even sub-metre- level precision. Hence, 5G New Radio (NR) with ISAC is expected to usher in a new era of high-precision LBSs.

It is well known that high-quality public datasets are a major driver of research in a variety of fields, such as computer vision and wireless perception. The demand for publicly available high-quality datasets is even more urgent in the field of 5G positioning, because most carrier-grade commercial gNBs have only communication capabilities, without positioning support. Although some of the main manufacturers (e.g., ZTE, Huawei) have developed 5G ISAC gNB prototypes, they have not yet been commercialized. At present, people have no alternative but to develop their own dedicated ISAC gNBs. This results in costly real-world data collection and the exclusivity of these precious non-public data. Consequently, as a kind of fallback solution, researchers often turn to synthetic datasets for 5G NR positioning studies.

Unfortunately, the existing methods for synthesizing positioning datasets produce either insufficiently authentic or coarse-grained data. Fan *et al.* [13], [14] proposed a dataset generation method for positioning oriented towards the terahertz spectrum, providing user equipment devices (UEs) with channel state information (CSI), including the angle of arrival (AoA), received power, and delay. However, under practical conditions, it is impossible for UEs to obtain such ray information accurately, let alone estimate the specific parameters of all multipath components in an indoor environment [2]. Therefore, the rationality and authenticity of this work are greatly reduced. As reported in [15], Bell Labs proposed a method of generating reference signal received power (RSRP) values through simulated synthesis, which is similar to the idea of using received signal strength indicator (RSSI) data

as a reference in traditional fingerprint localization [2], [16], [17]. Accordingly, RSRP fingerprint datasets can be employed for deep learning (DL)-based positioning tasks. Unfortunately, the RSRP value is a coarse-grained measure that is sensitive to noise and thus unsuitable for high-precision indoor positioning.

As seen from the above discussion, there is still a lack of high-quality datasets for 5G indoor high-precision positioning studies. Such a dataset should meet three requirements: 1) The position-related features embedded in the data should be detailed and prominent. 2) The data characteristics should be close to those of the real environment. 3) The dataset should be universal, general in purpose and not limited to a specific research method.

Therefore, we propose a fine-grained dataset generation paradigm for 5G NR indoor positioning, named the Multilevel Feature Synthesis Method (Multilevel-FSM). The data of position-related features, i.e., the channel frequency response (CFR) matrices, are synthesized strictly in accordance with the 3GPP R16 specifications in terms of signal design, the air interface and the multilevel processing chain. Therefore, our method can provide sample data to support 5G high-precision positioning experiments. Moreover, the process of our dataset generation method is simple and suitable to be ported between different scenarios or devices working in different frequency bands. Furthermore, the positioning performance of several machine learning (ML)-based methods trained on the proposed dataset has been examined, with the results showing high accuracy. Finally, a data augmentation method for positioning feature data is also proposed to achieve stronger robustness to noise.

The main contributions of this article are as follows:

- **We propose a novel paradigm for constructing a valuable and general-purpose 5G high-precision localization dataset, including Multilevel-FSM, to provide fundamental support for 5G ISAC studies.** Different from the datasets containing overly detailed CSI used in previous works [13] and [14], our work is fully compliant with the 3GPP R16 standards, and the CSI data is generated in a manner more consistent with the principle of real 5G communication. To the best of our knowledge, this is the first attempt to study dataset construction with a focus on 5G ISAC positioning.

- **We present a novel feature synthesis method using CFR to achieve better resolution and robustness.** Instead of using the RSRP [15], which is a single value, the CFR matrices are flattened into a CFR image to exhaustively characterize the features of 5G multiple-input multiple-output (MIMO) channels. Furthermore, the subtle differences between adjacent antennas are simultaneously reflected in a single CFR image, thereby enhancing the information density. This improved approach has not been previously reported.

- **We develop a DL-enhanced method, Multipath Res-Inception (MPRI), for high-precision ISAC positioning by constructing a series of specially sized filters and a multipath feature propagation structure.** The former aims to be more suitable for extracting

position-related features in the frequency domain and capturing the subtle channel differences between adjacent antennas. The latter enables our MPRI method to go deeper and converge faster.

- **Extensive positioning experiments have been conducted in typical 5G positioning scenarios, including an indoor office and an urban canyon.** The results demonstrate the fast convergence and noise robustness of the proposed MPRI method. Specifically, several existing DL methods (e.g., GoogLeNet [18], ResNet-18 [19], MobileNet-v2 [20] and Xception [21]) can achieve sub-metre precision. By contrast, MPRI achieves better precision values of 0.28 m and 0.20 m (indoors and outdoors, respectively), a lower time overhead, and stronger robustness in environments with a low signal-to-noise ratio (SNR).

The remainder of this paper is organized as follows. Related work is summarized in Section II. In Section III, the paradigm for generating a 5G NR high-precision positioning dataset is presented. In section IV, the details of the MPRI DL method are discussed. Our experimental design and results analysis are provided in Section V. Conclusions are drawn in Section VI.

II. RELATED WORK

A. Datasets for 5G NR Positioning

There are two approaches for obtaining 5G NR positioning data: actual measurements and synthesis methods [22].

For actual measurements, Bell Labs [23] developed a prototype architecture of a channel sounder capable of measuring CSI that supports millimetre-wave (mm-wave) bands. The prototype was designed to generate a dataset to evaluate the performance of a massive MIMO system. Regrettably, the study mainly targeted mm-wave channel estimation rather than 5G NR positioning. Consequently, the study characteristics were far from the specifications defined by 3GPP in certain key aspects related to positioning, e.g., the signal design, experimental scenarios and equipment parameters. The work in [24] introduced an ML-based approach for inferring UE locations using CSI data collected by a Huawei carrier-grade 5G test bed.

Unfortunately, the equipment required for the experiments mentioned above is extremely expensive, and no corresponding dataset has been made publicly available. Therefore, as the next best alternative, researchers may turn to simulation methods to collect research data. For simulation-based dataset synthesis, companies such as ZTE [25], [26] and Huawei [27], [28] and institutions such as Harbin Engineering University (HEU) [29] and Beijing University of Posts and Telecommunications (BUPT) [25] have jointly carried out a large number of positioning experiments. The results were made public during the course of formulating the R16 standard based on classic deterministic techniques, including time difference of arrival (TDoA) [30] and AoA [31]. Different from the traditional methods mentioned above, a special radio map dataset [32] has been used to simulate a positioning scenario in the Internet of Vehicles. That study utilized multipath signals to achieve improved accuracy, converting the multipath effect into an

advantage rather than a hindrance, which is enlightening to our work. However, the article assumed an enormous vehicle antenna, which limits the applicability of the data, making it a dedicated dataset.

Moreover, Bell Labs [15] proposed a *general* 5G NR positioning data synthesis method. Based on the synthesized data, a deep neural network (DNN) was used for positioning, achieving a positioning accuracy of 1.4 m. The RSRP value was used as fingerprint data in that paper. However, as a single value representing signal quality, the RSRP varies greatly in the presence of noise. This results in a coarse-grained dataset, making positioning performance unstable.

In [13] and [14], a *fine-grained* dataset construction method for high-frequency signals was proposed. Data generated by commercial software, *Wireless Insite®*, were used for DNN training to achieve a positioning accuracy of 0.25 m. However, the study assumed that the ray information of the received signal between each transmitter–receiver pair was *known* to the UEs and treated this ray information as input data for the DNN. This assumption is overly strong. In fact, the received signal components are delayed and scaled replicas from different points within the same source [2], [33]–[35], and there could be performance limits in terms of ray-tracking accuracy [2], [34]. Therefore, a UE acting as a receiver can obtain only the superposition of the multipath components [2] rather than accurately distinguishing the information of the 40 rays mentioned in the paper, let alone the detailed parameters of all these rays. That is, a UE in a multipath environment has *no* ability to obtain such fine-grained data for either training or positioning. Hence, the dataset generation method proposed in that study clearly violates real-world logic, and its rationality is greatly reduced.

In summary, there is still a lack of practical, fine-grained and general-purpose datasets for 5G NR positioning research. This motivates us to think about how to design a method of generating such datasets so as to ensure authenticity when simulating the actual data collection process at UEs.

B. ML Methods for High-Precision Positioning

Due to the non-line-of-sight (NLoS) effect in indoor scenarios, large positioning errors are incurred with traditional positioning methods [36]. Different from ranging-based localization approaches, ML methods learn rich features from multipath channel data to estimate UE locations, thereby achieving better accuracy and robustness in rich multipath environments [3], [37].

DNNs have been widely applied for Wi-Fi localization, for example, in [38]–[40]. By learning position-related features from captured Wi-Fi CSI data, metre-level positioning accuracy is achieved. Although Wi-Fi technology is significantly different from 5G NR in some aspects [41], this approach provides inspiration for our work.

In [15], Bell Labs utilized ML to capture 5G channel characteristics and achieved an average positioning error of 1.4 m on an RSRP value dataset. However, the neural network used was relatively primitive, with a limited fitting ability, and failed to achieve sub-metre positioning accuracy.

For further improvement, a two-step DNN training procedure was proposed in [42], with training on simulated line-of-sight (LoS) data in the first step and fine-tuning on measured NLoS positions in the second step. This work reduced the complexity of the actual measurements required to some extent, but the NLoS positioning accuracy was limited.

Recently, Fan *et al.* [13], [14] proposed a neural network with a two-level structure, whose claimed accuracy outperformed other similar methods. In those studies, ray information was employed as input data for training and positioning. However, there is a logical paradox in these papers: the neural network relies on the detailed information of the multipath components (which is assumed to be known) for positioning, but that very information could already be directly used in equation solving to accomplish UE localization (exactly as done in a TDoA or AoA algorithm), thus making such a neural network unnecessary.

In summary, actual measured datasets from internet service provider (ISP)-level 5G NR equipment are currently *not* available, and there is still a lack of universal fine-grained dataset generation methods. Therefore, high-quality dataset construction could clearly be a bottleneck for ML-based positioning studies that rely on high-quality data. These obstacles prompted us to conduct the research presented in this paper.

III. MULTILEVEL-FSM: A DATASET GENERATION METHOD TO SUPPORT 5G NR POSITIONING

In this section, a dataset generation paradigm for 5G NR positioning, named Multilevel-FSM, is proposed. The elements in the dataset are synthesized based on measured environmental data, which are fine-grained data that can actually be obtained by UEs. In addition, the synthesis process strictly complies with the 3GPP R16 technical specifications, making the data generation consistent with the principles of 5G NR communication. Our Multilevel-FSM can be divided into the following main components:

1) Channel modelling at the physical (PHY) level: Modelling and electromagnetic analysis of the physical scenario is carried out with the aid of the ray-tracing method to obtain the ray information at each point. Then, a *multipath channel filter* is constructed based on this ray information, packaging the complex multipath propagation information into a black box for the UE.

2) CFR estimation at the system level: Using the *multipath channel filters* generated for channel modelling, system-level simulation of the 5G NR air interface conforming to the 3GPP standard [43] is run to estimate CFR matrices for each UE. Key factors related to the CFR measurement capabilities of the UEs, such as signal design, bandwidth allocation and parameter settings, are carefully configured following the R16 technical specifications to ensure that the synthesized CFR results serving as the dataset are generated in a manner close to the real 5G communication.

3) Dataset construction and CFR image augmentation at the application level: The raw CFR data are preprocessed into CFR images and then augmented to increase the information density and robustness to noise, strengthening the advantages of our dataset in terms of feature expression.
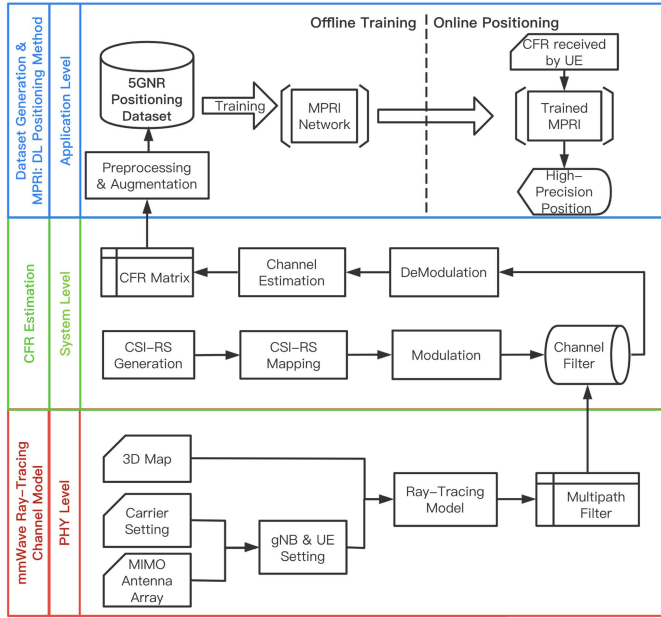
Fig. 1.   Architectural overview of our 5G NR positioning method.

A. Architectural Overview of Our 5G Positioning Method

In Fig. 1, we abstract our 5G NR positioning method into a three-layer structure, consisting of the PHY level, the system level, and the application level. The first two levels refer to the synthesis of position-related feature data, and the highest layer corresponds to the dataset generation and positioning processes.

At the PHY level, the ray-tracing method is used to model a multipath channel and to obtain *ray information*, including the AoA, delay, path loss (PL), number of rays, time offset and frequency offset, at each point. There can be a conflict that, this ray information is clear and knowable to us, but in contrast, it is unknown and unknowable to UEs. Specifically, only the results of multipath superposition can be observed by a UE, while the individual rays (multipath components) are obscured.

To resolve this conflict of information asymmetry, we introduce a multipath channel filter, $Filter_{multipath}$, into the processing chain. The workflow of this $Filter_{multipath}$ is defined as follows: First, according to the obtained *ray information* mentioned in the last paragraph, a set of copies of the original signal is created for each propagating ray. Second, each copy is transformed based on its *ray information*, to complete attenuation, delay, and offset as a multipath component. Finally, all the *multipath components* are superimposed to form a *received signal*. $Filter_{multipath}$ represents a black-box filter for a multipath fading channel defined by a set of propagation rays from a gNB to a UE. This filter can provide the UE with the channel characteristics and avoid the UE requiring knowledge of *overly detailed* ray information.

Therefore, our approach has two advantages: 1) We introduce $Filter_{multipath}$ to furnish a UE with the results of multipath superposition, rather than the details of each ray. This provides a correction for the overly strong hypothesis in [13] and [14]. 2) The ray-tracing method, rather than a

statistical model, is utilized to model multipath channels. This method is a mature technique for capturing the fine-grained parameters of a channel.

At the system level, we generate CSI reference signals (CSI-RSs) as pilots for each gNB. The 5G signals are generated, modulated, and then filtered by $Filter_{multipath}$. Finally, the CFR matrices are estimated via the received pilot signals at the UEs, as shown at the system level in Fig. 1. The parameters of the whole process, such as frame structure, reference signal design and carrier parameters, are set following the 3GPP R16 specifications.

It can be observed that the CFR matrix is a fine-grained and position-related measure, showing marked differences at different locations. This feature is particularly useful for high-precision positioning. Furthermore, the channel correlations (e.g., the embedded channel differences) derived from a MIMO antenna array are also beneficial for positioning accuracy. They yield more effective information, which is beneficial for further improving the positioning accuracy. In this respect, our method has two advantages: 1) For characterizing the CSI, the CFR matrix provides better resolution and robustness than does the RSRP value used in [15]. 2) The subtle differences between MIMO channels are used to capture embedded features and to improve the positioning accuracy compared with that achieved in the literature when adjacent channels are disregarded [13]–[15].

At the application level, data preprocessing and augmentation are conducted. The CFR matrices of MIMO channels obtained by the UEs are intercepted, down-sampled, visualized and flattened to generate CFR images. Finally, after data augmentation, the CFR images are stored as a dataset with enhanced feature expression. That is, the proposed dataset can be employed in 5G NR positioning research to address the problem of dataset shortfalls.

B. Ray-Tracing-Based Channel Model

In the PHY layer, we build a ray-tracing model for a 5G channel to represent the signal propagation path between each gNB–UE pair. Accordingly, the power of the received signal at the UE can be determined as follows [44]:

$$P^{3D}(d)$$
$$= \frac{P_t\lambda^2}{(4\pi d)^2}\Big| \sum_{k=-m}^{m} \cos\theta_k \sqrt{G_{tx}G_{rx}}$$
$$\times (\Gamma^{ceiling}(\theta_k))^{\left|\left[\frac{k+1}{2}\right]\right|}(\Gamma^{floor}(\theta_k))^{\left|\left[\frac{k}{2}\right]\right|} \prod_{i=0}^{N} I_t(\theta_k) \tag{1}$$

where θ_k is the elevation of the k-th ray; G_{tx} and G_{rx} are the gains of the transmitting antenna and the receiving antenna, respectively; and the wavelength is λ. After N reflections or transmissions from or through walls between the transmitting antenna and the receiving antenna, the length of the propagation path is D. Γ and I are the reflection and interaction coefficients, respectively, along the path, and $P^{3D}(d)$ represents the receiving power on the path.

Our ray-tracing model follows the terms of the 3GPP specification TR 38.901, setting the maximum order of reflection

on a path without diffraction as 2 for reasons of simplicity and simulation speed [45]. Only LoS and specular reflection paths are considered in this work to alleviate the computational burden. This simplicity is reasonable because diffraction and refraction paths contribute less to multipath superposition than reflection paths do, which has been discussed in detail in [34], [46]–[48]. Additionally, there are no millimetre- or centimetre-sized objects in our scenarios. Therefore, scattering, based only upon small objects, is omitted [45]. In summary, the parameters of each path, which cannot be suitably provided by statistical channel models, can be calculated by our model with a good balance between the simulation speed and accuracy.

In this work, the gNBs and UEs are each considered to be equipped with a uniform rectangular array (URA) with $k \times k$ elements. The inter-element spacing is set to 0.5λ in both the vertical and horizontal directions. There are $n_{tx} \times n_{rx}$ MIMO channels in the proposed ray-tracing-based channel model, considering a gNB with n_{tx} elements and a UE with n_{rx} elements.

C. CFR Estimation

In this subsection, we describe how to simulate the generation of a 5G NR reference signal, signal fading in multipath channels, and CSI estimation at a UE, following the 3GPP R16 specifications.

1) Generating a CSI-RS at a gNB: The CSI-RS [43] is one of the essential reference signals in 5G NR and is designed for four purposes: time-frequency tracking, CSI measurement, L1-RSRP measurement, and mobility management. As our main focus is 5G NR positioning, we make use of the CSI-RS for CSI estimation. Therefore, our method is intrinsically different from the direct usage of ray information to represent channels in [13] and [14].

The CSI-RS has two modes: a zero-power (ZP) mode and a non-zero-power (NZP) mode [43]. In the ZP mode, the CSI-RS does not transmit any power on its resource elements in the uplink direction, while the NZP-mode CSI-RS is mainly used in the downlink direction. The signal sequence $r(m)$ is created as follows:

$$r(m) = \frac{1}{\sqrt{2}}(1 - 2 \cdot c(2m)) + j\frac{1}{\sqrt{2}}(1 - 2 \cdot c(2m+1)) \quad (2)$$

where the pseudo-random sequence $c(i)$ is given in clause 5.2 of TS38.211 [43], initialized as:

$$c_{init} = (2^{10}(N_{sym}^{sl} n_{s,f}^{\mu} + l + 1)(2n_{ID} + 1)n_{ID}) \bmod 2^{31} \quad (3)$$

at the start of each orthogonal frequency-division multiplexing (OFDM) symbol, where $n_{s,f}^{\mu}$ is the number of slots within a radio frame, l is the number of OFDM symbols in a slot, and n_{ID} is given by higher-layer parameters. We map $r(m)$ to resource elements $(k, l)_{(p,\mu)}$ as follows:

$$a_{k,l}^{(p,u)} = \beta_{CSIRS} w_f(k') \cdot w_t(l') \cdot r_{l,n_{s,f}}(m') \quad (4)$$

2) Conducting Channel Estimation at a UE: The CSI-RS, as a pilot signal, is adopted to obtain CFR information for non-blind channel estimation. The following 5 steps are required to fully characterize the individual channels.

Step 1: We utilize a temporal linear filter, $Filter_{multipath}$, to model 5G multipath channels and describe the superposition of multipath components for a complex rich-multipath environment [45], [49]–[51]:

$$h(\tau) = \sum_{n=1}^{N} a_n e^{-j\theta_n} \delta(\tau - \tau_n) \quad (5)$$

where a_n, θ_n, and τ_n are the amplitude, phase and time delay on the n-th path, respectively; $\delta(\cdot)$ is the Dirac delta function; and N represents the total number of paths in the environment. In addition, $h(\tau)$ is the channel impulse response (CIR) of this multipath channel, which consists of the delay on multiple paths multiplied by the amplitude and phase. Therefore, the received signal spectrum $Rx(f)$ can be expressed as the product of the transmitted signal spectrum $Tx(f)$ and the CFR $H(f)$, corresponding to the convolution of the two in the time domain. Because the CIR calculation is nontrivial, we estimate the channel by means of frequency multiplication and then denoise it using the inverse Fourier transform.

Step 2: After OFDM demodulation and time offset correction of the received signal, the received resource grid, $rxGrid$, can be obtained by the UE. The CFR is the Fourier transform of the CIR.

Step 3: The raw CFR matrices of all subcarriers for each OFDM symbol and each antenna can be expressed in the form of a k-dimensional column vector:

$$H_{raw} = (rxGrid)_{k \times 1} / (refGrid)_{k \times 1} \quad (6)$$

where k is the number of subcarriers and $refGrid$ represents the known pilot.

Step 4: We perform interpolation in the frequency domain on both sides of H_{raw} and use inverse Fourier transformation to convert the information into the CIR for noise reduction. Then, the denoised CIR can be obtained as follows:

$$H_{denoised} = H_{raw}W \quad (7)$$

where W is the raised-cosine window in the time domain for CIR denoising.

Step 5: Finally, the CFR H is acquired after Fourier transformation. Below, $\mathscr{F}^{-1}(\cdot)$ denotes the inverse Fourier transform:

$$H = \mathscr{F}^{-1}(H_{denoised}) \quad (8)$$

The CFR results are observed to be position-related, fine-grained and highly distinguishable. Fig. 2a shows the CFR matrices for the signals received at a UE from 4 gNBs at different locations. Fig. 2b shows the CFR matrices observed by 4 UEs with a spacing of only 0.5 m from the same gNB. All CFR observations show good discriminability, indicating that the CFR-based approach is promising for high-precision positioning.

D. Multilevel-FSM to Support 5G NR Positioning

The CFR estimation results presented above are used as the raw data to construct the dataset. We perform data pre-processing to establish the desired data structure and introduce
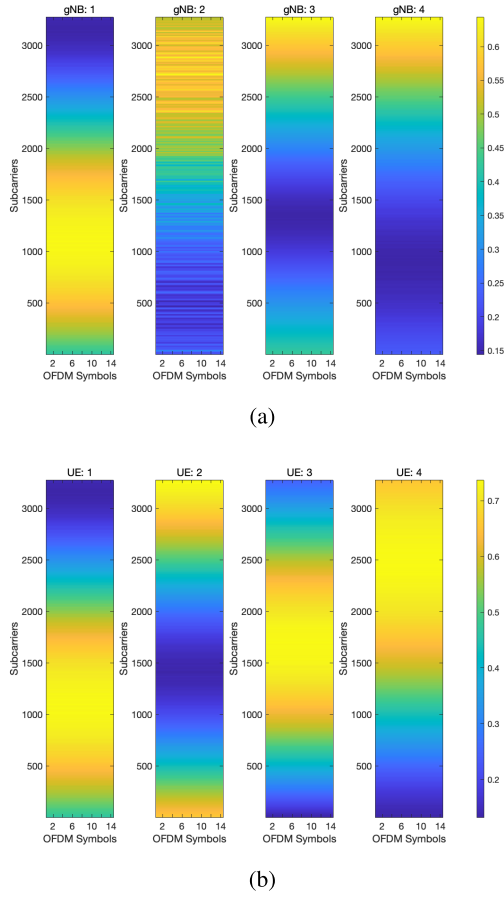
(a)

(b)

Fig. 2. CFR observations received (a) at a single UE from 4 gNBs and (b) at 4 UEs from one base station.

data augmentation to reduce redundancy and enhance the information density.

Specifically, the following procedures are applied for data preprocessing.

1) Truncation in the time domain: The CSI in an indoor environment remains essentially the same within a single subframe in the time domain. Based on this, we clip the results of CFR estimation to a single OFDM symbol for each sample to reduce data redundancy.

2) Down-sampling in the frequency domain: The subcarriers over the full bandwidth are down-sampled in the frequency domain. Thus, the amount of data used as input to the neural network is reduced while preserving the continuous variation characteristics of the data as much as possible.

3) MIMO CFR imaging and flattening: The CFR observations on adjacent physical antennas show subtle differences, which contain position-related features such as channel differences and AoAs. The high-dimensional CFR observations are flattened into a two-dimensional CFR image for each pair of transmitting and receiving antennas in the MIMO system. Thus, the subtle differences between adjacent antennas can be simultaneously reflected in a single CFR image, thereby increasing the information density.

4) Data structuring: Each CFR image is labelled with its real location and stored in our dataset. The estimation results for different positions compose an $N_s \times N_p \times N_g$
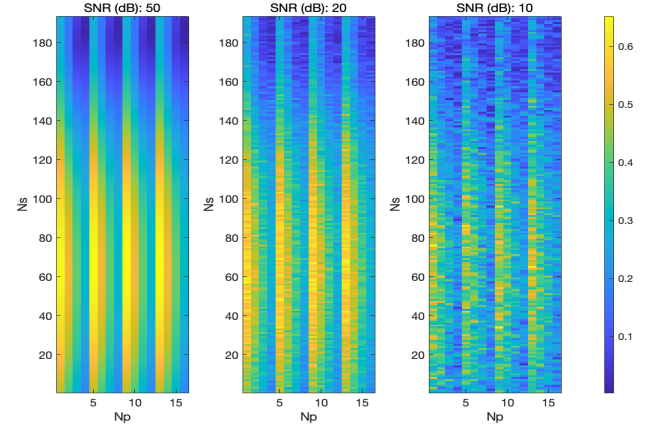

Fig. 3. CFR images obtained by adding noise at different SNR levels (50 dB, 20 dB and 10 dB), where $N_s = 193$, $N_p = 16$ and $N_g = 1$.

three-dimensional matrix, where N_s is the number of samples, N_p is the number of MIMO channels, and N_g is the number of gNBs. For example, the 5G wireless channels established between a UE and each base station are described as two-dimensional matrices ($N_s \times N_p$). Considering a UE receiving signals from N_g base stations, its CFR image is a 3-D matrix with depth N_g.

The leftmost subfigure of Fig. 3 presents an example of a CFR image that describes the position-related features of the channel between gNB1 and UE1.

Finally, we propose a data augmentation method for CFR image data. For each gNB–UE pair, noise is added to each $Filter_{multipath}$ at s SNR levels to simulate the actual communication conditions. As shown in Fig. 3, by adding noise in this way, the number of CFR images at the same UE location is increased by a factor of s. The advantage is that position-related features can be displayed simultaneously under a variety of noise interference conditions. This can be helpful to allow DL-based positioning methods to learn the essential mapping of "signal to location" under interference from noise. Consequently, the feature representation and generalization ability of the proposed dataset is improved.

Note that in our CFR image augmentation method, the noise is added directly to $Filter_{multipath}$ for the RF link in the PHY layer rather than being added to the channel estimation results, as in references [13] and [14]. This makes our dataset more consistent with real channel propagation behaviour and the logic of wireless communication systems.

E. ISAC Discussion

Because of the 5G's native communication–positioning-integration, the proposed Multilevel-FSM supports ISAC research and impose *NO added computational burden* on a 5G NR system.

Moreover, the CSI-RS used for channel estimation is one of the 5G reference signals defined in R16, also in R17. All these reference signals and the user communication data occupy different resource blocks (RBs) and thus are independent of each other in terms of spectrum resources. This means that communication and sensing functions can coexist within a time-frequency band.

However, the problem is also regarded as a balancing problem; that is, denser reference signal deployment can improve the quality of channel estimation and positioning accuracy at the cost of user communication bandwidth. In this regard, our Multilevel-FSM adopts RB mapping principles in compliance with TS38.211 [43] for a reasonable balance.

IV. MPRI: DL METHOD FOR NR INDOOR POSITIONING

A. Motivation

We tested the dataset generated by Multilevel-FSM using a variety of classic DNN models with proven high performance, such as GoogLeNet [18], ResNet-18 [19], and MobileNet-v2 [20], and all of them achieved sub-metre precision (see Section V). However, there are still two weaknesses: 1) The network scale is too large for deployment on UEs, and training takes too much time. 2) The positioning accuracy of classic DL models is difficult to further improve. Therefore, we attempted to analyse and overcome these weaknesses from the source, namely, the dataset itself.

First, a CFR image generated by Multilevel-FSM has an elongated shape, different from that of a normal rectangular image. Each column of pixels in a CFR image (a 2D $N_s \times N_p$ image) corresponds to a single channel, while the width corresponds to the number of channels, and N_s is typically much larger than N_p, with a ratio of more than 200:1. Therefore, the convolution kernel needs to capture the features in CFR images mainly in the vertical direction, which is not considered by classic DL models.

Second, the physical meaning of a CFR image is mainly concentrated in the frequency direction (vertical direction). This phenomenon is different from the typical information distribution in natural images. Moreover, because of the existence of noise, the feature map in the frequency domain becomes discontinuous, which hinders feature extraction. In contrast, traditional DNN models are designed to identify objects in photographs using square convolution filters, which are not suitable for CFR images.

Additionally, the subtle differences between adjacent antennas contain additional position-related features (such as channel differences) and may therefore be adopted to improve the positioning accuracy. These small differences should be extracted while simultaneously capturing the characteristics in the frequency domain. Until now, this issue has also not been considered in classic DNN models.

Based on the above three considerations, we propose a DL network structure named MPRI to achieve improved positioning accuracy.

B. Architectural Design of the MPRI Network

As shown in Fig. 4, the MPRI network is composed of an input module, an MPRI module and an output module.

The input module includes an image-input layer and a basic convolution block. The latter consists of a convolution (Conv) layer, a batch normalization (BN) layer, a rectified linear unit (ReLU) layer and a pooling layer. It is able to extract position-related features from CFR images and convert them into abstract initial representations for higher layers.

The MPRI module is the main body of the proposed network for inferring UE locations. It includes an upper half for providing rich positioning features, and a lower half for feature-learning and location estimation. The core idea of our design includes multipath Inception and Residual Sharing: The former enables MPRI to independently determine an appropriate range for feature extraction, providing diverse features, while the latter ensures that the depth of the MPRI network can be increased while maintaining fast convergence properties. Note that these two design ideas run through the entire module.

The output module combines the results obtained on the two paths of the MPRI module and passes them to a basic convolution block to extract mixed features, followed by a dropout layer before the final fully connected layer to avoid over-fitting.

C. Design of the Upper Half of the MPRI Module

The upper half of the MPRI is designed to ensure the diversity of extracted features, considering the special information distribution law of CFR images as mentioned in Section IV-A.

The richer the extracted features, the more beneficial the DNN is in improving the positioning accuracy in the CFR-to-location mapping process. Hence, we adopt four parallel paths with different structures to extract features of varying ranges, and use *Depth Concatenation* at the outlet layer. By doing so, a single input tensor can produce output features at diverse scales with the help of convolution kernels of various sizes, corresponding to different receptive fields. Moreover, the *Depth Concatenation Layer* at 256 channels is used at the outlet of the upper half instead of the widely used *Addition Layer*, so that these 2D output features can be reserved along the third dimension (the channel dimension). This ensures the diversity of features during propagation.

Other detailed designs of the upper half of the MPRI module can be summarized as follows.

1) Convolution kernels of different sizes are used on the four parallel paths, including 1×1 filters, a 3×3 filter and a pair of $5 \times 1 + 1 \times 3$ filters. The advantage is that MPRI has the ability to automatically select the convolution kernel of the appropriate size for feature extraction. An $n \times 1$ convolution kernel better focuses on the CFR features in the frequency domain, reducing the number of parameters and increasing the training speed. A larger convolution kernel can capture the channel differences between adjacent antennas more effectively, from which to infer implicit position-related features, such as AoA information.

2) For the second path, we utilize a combination of 5×1 and 1×3 convolution kernels instead of a single 5×3 convolution kernel. This decreases the number of parameters to 53.3% of the original number while leaving the size of the receptive field unchanged. At the same time, the nonlinear fitting ability of MPRI can be enhanced by deeper layers. Therefore, our design endows MPRI with stable positioning accuracy in complex positioning scenarios and greatly improves training efficiency.

3) The 256 channels finally converging to the Depth Concatenation Layer are composed of the following parts:
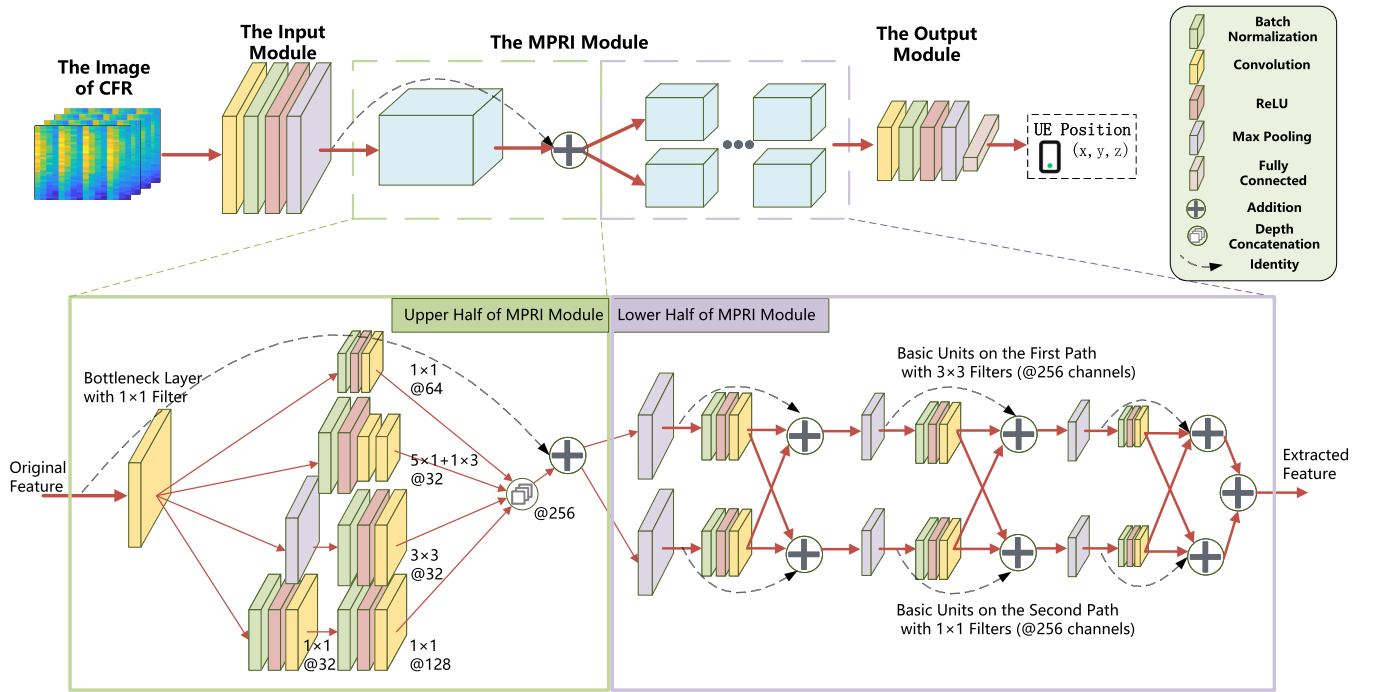
Fig. 4.   Architecture of the MPRI network, consisting of an input module, an MPRI module and an output module.

64 channels from path 1, 32 from path 2, 32 from path 3, and 128 from the path 4. Among them, a bottleneck strategy (32 channels first, then 128 channels) is adopted at path 4 to reduce memory requirements.

4) At the entrance to this module, we introduce a bottleneck to reduce the computational effort. Meanwhile, identity mapping is added from the entrance to the exit of the module to speed up convergence.

D. Design of the Lower Half of the MPRI Module

The lower half of the MPRI module completes the feature-to-coordinate mapping through dual-path parallel convolution blocks.

One disadvantage of traditional DNNs is the fixed parameter design, which makes it impossible for DNNs to determine the most suitable convolution kernel size in different task scenarios (e.g., indoor and outdoor). To solve this problem, we construct two basic units of parallel feature propagation paths with cross-connections, and then stack this dual-path structure three times. In this design, the output of each convolutional block at the same depth is allowed to be directly exchanged between the two paths. These cross-connections enable the MPRI network to flexibly select a more appropriate receptive fields, and thus improve the positioning precision.

Other detailed designs of this structure can be described in terms of the following three points.

1) Residual connections are added to each basic unit to guarantee that a deeper MPRI network has better positioning accuracy and convergence speed. Six such basic units (2 units per layer, 3 layers in total) in the lower half of the MPRI module are equipped with residual connections.

2) The convolution kernels on the feature propagation paths are of different sizes. This provides two advantages. First,

MPRI has the power to capture position-related features (such as channel differences) on adjacent channels by means of receptive fields of different sizes. Second, MPRI achieves better robustness to noise. In real-world scenarios, the continuity of CFR images in the frequency domain is disturbed by noise, which is unfavourable for feature extraction. However, with the help of multi-size receptive fields, the MPRI network can still effectively capture the CSI features of MIMO channels from such discontinuous changes.

3) In contrast to the classic Inception model, we use an *addition layer* instead of a *depth concatenation layer*, which greatly reduces memory consumption and speeds up the training process, allowing MPRI to be deployed on performance-limited UEs.

As discussed above, the MPRI module serves as the backbone for the whole network, which is composed of 63 layers, and the number of parameters is approximately 2.6M. This is far smaller than the scale of similar networks (144 layers and 7M parameters for GoogLeNet [18], 71 layers and 11.7M parameters for ResNet-18 [19], and 170 layers and 22.9M parameters for Xception [21]). Note that the activation layer is omitted since it does not contain any parameters to be trained.

V. EXPERIMENTAL STUDY AND RESULTS ANALYSIS

A. Experimental Settings

A server equipped with an Intel Xeon E5-2626 v2 CPU and a GeForce GTX 1080 GPU was used in our simulation. A Nikon DTM-352C Total Station was utilized for site surveying and positioning-point labelling.

Two typical 5G NR positioning scenarios were selected for our experiments, as shown in Figs. 5 and 6.

Scenario 1 is an indoor office hall of a new building of Chinese Academy of Sciences (CAS) in Beijing, China. Five

TABLE I
EXPERIMENTAL SCENARIO PARAMETERS

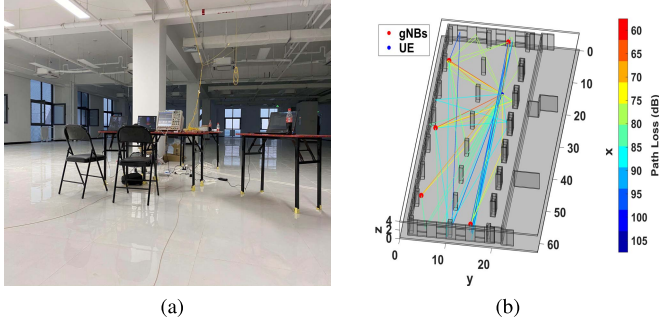| Scenario	Size	UE Interval @Layers	Number of UEs	Height of gNBs	Number of gNBs
1	60.5 m × 20 m	0.5 m@2	3876	2.4 m	5
2	61 m × 33 m	0.5 m@2	5150	40 m	4

Fig. 5. A typical indoor positioning scenario in an office hall: (a) photograph and (b) electromagnetic simulation results.

Fig. 6. A typical outdoor positioning scenario in an urban canyon: (a) photograph and (b) electromagnetic simulation results.

ISAC gNBs operate at 3.5 GHz with 100 MHz bandwidth and 40 W power. These five gNBs are suspended on plastic holders 2.4 m above the ground. For simulation, there is a random floating height of 0.1 m to prevent coplanarity. The UE, as a receiver, is mounted on a marked liftable cart 1.2 metres above the ground to simulate a 1.8-m-tall person holding a mobile phone. The second scenario is a typical urban canyon environment, which makes it difficult for a UE to access sufficient satellites for positioning. The UEs are located on a low-rise platform between two high-rise buildings. Parameters specific to the two scenarios are detailed in Table I.

Based on the proposed Multilevel-FSM, datasets were generated for the two scenarios. **First**, a cellular channel physical simulation toolkit was developed as described in Section III-B. The reliability of our tool was verified through comparison with simulation results from the commercial software *Wireless Insite®*. Ray tracing was performed for each gNB–UE pair to produce multipath channel filters. According to TR38.901 [45], for indoor scenario, we adopted both 2-order reflection and 1-order penetration; For outdoor scenario, the maximum order of reflection on a path without diffraction is set to 2. The carrier frequency is 3.5 GHz and 40 GHz for indoor and outdoor scenarios, belonging to the FR1 band and FR2 band respectively. The signal propagation paths between one of the gNB–UE pairs in each scenario are visualized in Figs. 5b and 6b. Ray-tracing results show an average of 53.9 rays and 21.6 rays between each gNB–UE pair for indoor and outdoor, as the rich multipath scenarios. **Second**, we built 4×4 MIMO channels for 5G NR with 100 MHz of bandwidth resources. The NZP-mode CSI-RS was adopted as the pilot signal in the downlink direction, in accordance with Section III-C. **Third**, the received signal at the UE was used to conduct channel estimation, including gNB searching, signal demodulation, time-offset correction, etc. Then, the CFR matrices were stored as the raw estimation results. **Fourth**, matrix preprocessing was further performed for dataset construction. The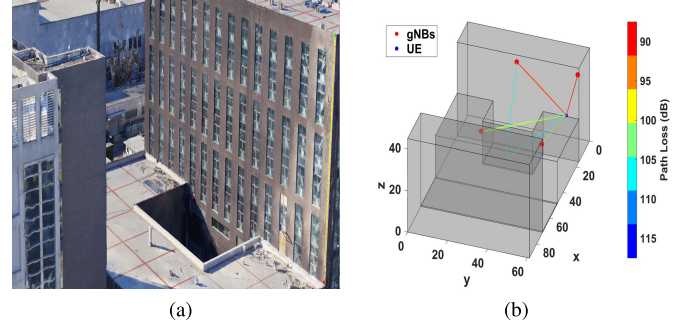 CFR matrices for all 16 channels in the MIMO system were compressed, down-sampled and then flattened into a single image to increase the information density as described in Section III-D. Next, data were collected at the same location with three different SNRs to construct an enhanced dataset to achieve robustness to noise.

In this way, we obtained a dataset composed of CFR matrices with dimensions of $N_s \times N_c \times N_g \times N_u$, where N_s is the number of samples, N_c is the number of MIMO channels, N_g is the number of gNBs and N_u is the number of samples at all UE positions. For instance, the provided dataset of scenario 1 contains 14448 CFR images, of which 11559 are used for training and 2889 for testing, with a segmentation ratio of 0.8. The dataset of scenario 2 contains 11628 CFR-images, which are also divided into a training set and a test set at a ratio of 0.8. In all training processes, we used exactly the same parameters for an objective comparison. The learning rate is set to 1e-3 during 400 training epochs with a periodical learning-rate drop schedule of 50/0.3 (epoch/drop-factor). At the beginning and end of each epoch, the test set is used to evaluate the current validation accuracy. Accordingly, the training accuracy is also recorded at the completion of each iteration. The analysis of the training process is given in Section V-C.

The used instances of the datasets generated by our Multilevel-FSM can be downloaded for free at IEEE DataPort https://dx.doi.org/10.21227/jsat-pb50 [52].

B. Metrics

The root mean square error (RMSE) and mean error ($MeanErr$), which are commonly used to evaluate positioning performance, are defined as follows:

$$MeanErr = \frac{1}{N_{UE}} \sum_{i=1}^{N_{UE}} \left(\left\| S_i - \hat{S}_i \right\|_2 \right) \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N_{UE}} \sum_{i=1}^{N_{UE}} (S_i - \hat{S}_i)^2} \quad (10)$$

where $\hat{S}_i$ denotes the estimated position of the i-th UE, S_i is the actual position of the i-th UE, N_{UE} is the number of UEs and $\|\cdot\|_2$ denotes the Euclidean distance.

Additionally, a widely used metric, the Structural Similarity (SSIM), is introduced for assessing a generated CFR image with the corresponding sampled CFR image. Different from
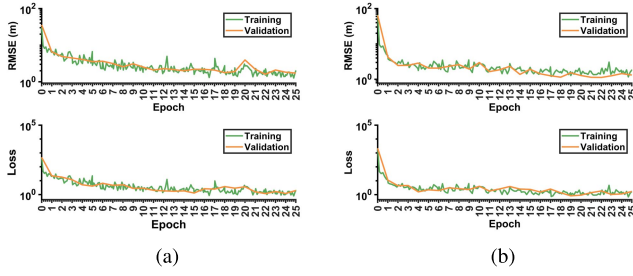
Fig. 7. Convergence behaviour of the MPRI network for (a) the indoor office scenario and (b) the urban canyon scenario.

Fig. 8. Convergence comparison of the first 50 training epochs of 5 DNNs.

the coarse-grained index mean squared error (MSE), the SSMI considers three characteristics of an image: luminance, contrast and structure, matched better to the perceived quality [53]. The simplified version of the SSIM is adopted as follows:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \quad (11)$$

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (12)$$

where x and y are two nonnegative image signals for comparison and where μ_x is the mean intensity of luminance. The square root of variance serves as an estimate of the signal contrast, given by

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{\frac{1}{2}} \quad (13)$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y). \quad (14)$$

In addition, C_1 and C_2, the regularization constants for the luminance and contrast terms, respectively, are set according to [53].

C. Convergence Analysis

A convergence test is essential to evaluate the performance and structure of a DNN. The RMSE and loss on both the training set and the test set are shown in Fig. 7. During the first 25 epochs of training, the RMSE decreases from 63.97 to 1.45 in the urban canyon scenario and from 35.80 to 1.83 in the indoor scenario. Notably, the curve on the test set is always close to that on the training set during the entire training process, indicating that no over-fitting occurred.

We introduce other four classic outstanding DNNs for comparison, namely, GoogLeNet [18], ResNet-18 [19], and improved versions thereof, such as MobileNet-v2 [20] and Xception [21]. As shown in Fig. 8, the convergence of MPRI is significantly faster than that of MobileNet-V2, and close to the performance of ResNet and Xception. However, considering fewer high peaks and fluctuations, MPRI has advantages in the stability of rapid convergence.

In summary, fast convergence of the MPRI network has been proven on the Multilevel-FSM datasets for both environments.
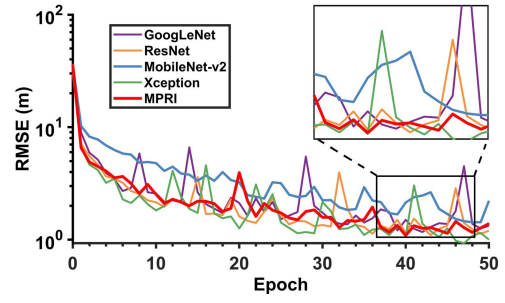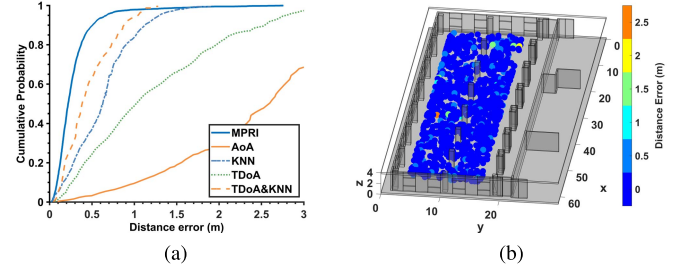

Fig. 9. Positioning performance in scenario 1, as reflected by (a) the cumulative distribution function (CDF) curve and (b) the visualized distance error.
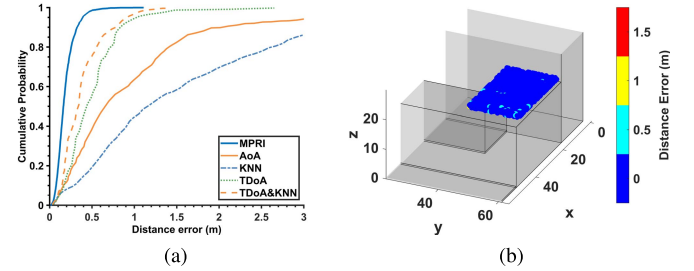

Fig. 10. Positioning performance in scenario 2, as reflected by (a) the CDF curve and (b) the visualized distance error.

D. Positioning Accuracy

The positioning performance of MPRI in the two scenarios is shown in Figs. 9 and 10.

1) Exp. 1: Indoor Scenario: In scenario 1, the mean error is 0.28 m, with 66% of the points having an error of less than 0.28 m and 90% having an error of less than 0.51 m. We chose four other methods, namely, TDoA, AoA [54], K-nearest neighbours (KNN) and TDoA-AoA, for comparison. The results show that the corresponding mean errors of these methods are 0.52 m, 4.43 m, 0.77 m and 0.54 m, respectively. Hence, the MPRI positioning method improves the performance by up to a factor of 15, showing notable advantages.

2) Exp. 2: Urban Canyon Scenario: In scenario 2, the mean error is 0.204 m, with an error of less than 0.23 m for 66% of the points and less than 0.36 m for 90% of the points. For comparison, the mean errors of the other four location algorithms are 0.56 m, 1.01 m, 1.57 m and 0.40 m, indicating that the location estimation accuracy of MPRI is improved by up to a factor of approximately 6.6.

3) Exp. 3: Ablation Experiment for the MPRI: To validate the performance of the MPRI network, we conduct an ablation
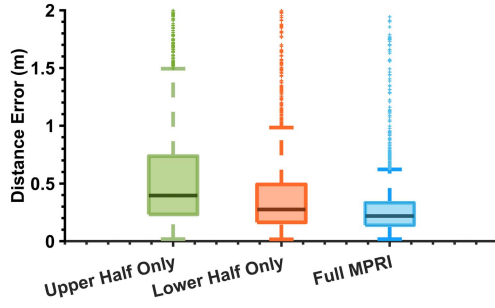
Fig. 11. Ablation experiment for the MPRI network.

experiment in scenario 1. First, the control group is defined as a 'full' version of the MPRI network. Next, we set up a new MPRI but disabled the Lower Half as the experimental group 1. Correspondingly, an MPRI with the Upper Half Module disabled is considered as the experimental group 2.

From the result, as shown in Fig. 11, we observed that the control group ('full' version) achieves the best precision of 0.28 m, while that of the control group 1 and 2 were 0.64 m and 0.41 m respectively. The performance has been improved by 56.09% and 31.93%. Additionally, the standard deviation of the control group is 0.26 m, leading the other two groups (0.85 m and 0.47 m). The result demonstrates that the proposed MPRI with the Upper and Lower Half modules has advantages in improving positioning accuracy and reducing dispersion.

In summary, the proposed method is suitable and, indeed, shows significant advantages for both complex indoor environments and large-scale outdoor scenarios.

E. Robustness to Noise

To check the *robustness to noise* and the *applicability of our data augmentation method*, we defined two experiments in this section.

1) Exp. 1: Robustness to Noise With/Without Our Data Augmentation Method: In the experimental group, noise at *three* different SNRs (50 dB, 20 dB and 10 dB) was added to $Filter_{multipath}$ to augment the CFR image dataset. Next, this *augmented dataset* was adopted for MPRI training. In the blank control group, only *one* level of noise, at an SNR of 50 dB, was added for dataset generation, as done in most studies. Then, MPRI was trained on this *normal dataset* for comparison. To ensure objective experiments, the MPRI networks representing the experimental group and the blank control group were trained under exactly the same conditions.

In the validation phase, positioning experiments were performed with 9 different SNRs (ranging from 10 dB to 50 dB, in steps of 5 dB) for comparison, as shown in Fig. 12. Despite being disturbed by the same levels of noise, the blank control group showed a significant increase in the error at low SNR levels compared with the more stable accuracy in the experimental group. With our augmented dataset, $MeanErr$ was reduced by 72.26%, 58.82% and 45.94% in the cases of low SNRs of 10 dB, 15 dB and 20 dB, respectively. At higher SNRs (30-50 dB), the experimental group still showed an approximately 34.48% improvement in accuracy. This indicates that the proposed data augmentation method can improve the robustness to noise of a DNN.
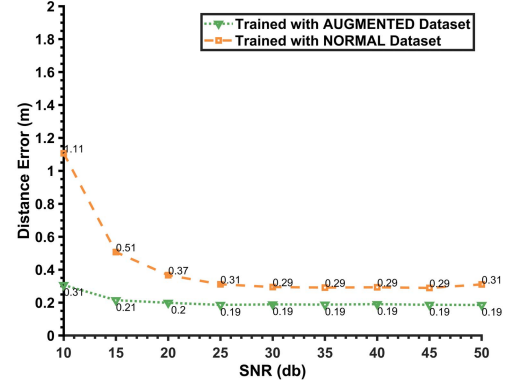

Fig. 12. $MeanErr$ comparison. Triangles denote the results of MPRI trained on an *augmented dataset*; squares denote the results of MPRI trained on a *normal dataset*.
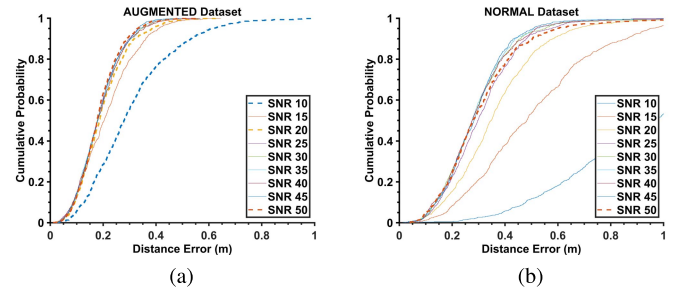

Fig. 13. Positioning performance (represented by the CDF) of MPRI trained on (a) an *augmented dataset* and (b) a *normal dataset* (right). Noise at nine SNR levels is used for testing, and a thick line represents an SNR that is also used for training.

For further analysis, the cumulative distribution function (CDF) curves in Fig. 13 show the results in greater detail. We plot the positioning performance of the experimental group at 9 different SNR levels in Fig. 13a, where the thick lines represent SNR levels that were also used for training. Similarly, the results for the blank control group are shown in Fig. 13b.

In the experimental group, the cumulative probability of $MeanErr$ reached 98% at 0.5 m in 8 out of 9 noise environments (all except SNR = 10 db). Additionally, MPRI still converged in low-SNR environments (10 db and 15 db) with a $MeanErr$ within 0.31 m.

As observed in the comparison, the CDF curves in Fig. 13a are more concentrated, while the curves in Fig. 13b are more dispersed. This demonstrates that MPRI achieves robustness to noise during training on noisy CFR images. Most importantly, through training with noise at several SNRs (e.g., 10 dB, 20 dB and 50 dB), noise robustness with respect to *additional SNRs* (e.g., from 10 dB to 50 dB) can be achieved.

2) Exp. 2: Comparison of Noise Robustness Between 5 DNNs: To validate the DNNs' native robustness to noise, a mutual-control verification is conducted among 5 DNNs. Two representative factors are taken into account in this experiment: positioning accuracy and standard deviation of positioning error. All the tested DNNs follow the principle of "training in a single SNR environment and verification in a multiple SNR environment". Figure 14 shows the results.
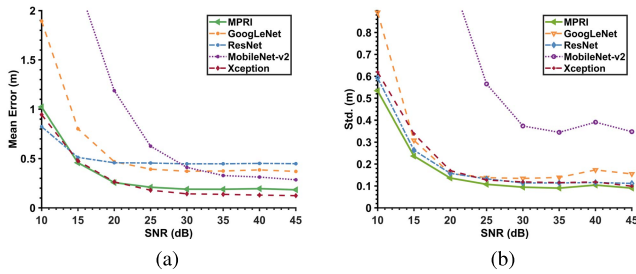
Fig. 14.   Mean error (left) and Std. (right) comparison of 5 DNNs, which are trained at *a single SNR level* and tested at *8 SNR levels*.
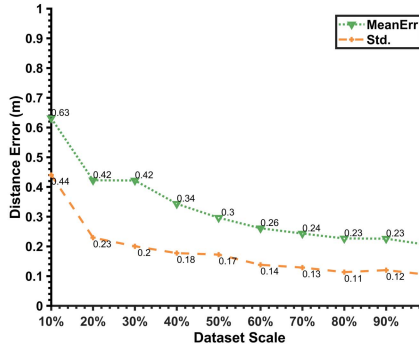

Fig. 15.   Impact of dataset scale. The horizontal axis reports the scale relative to the original dataset adopted for MPRI training.
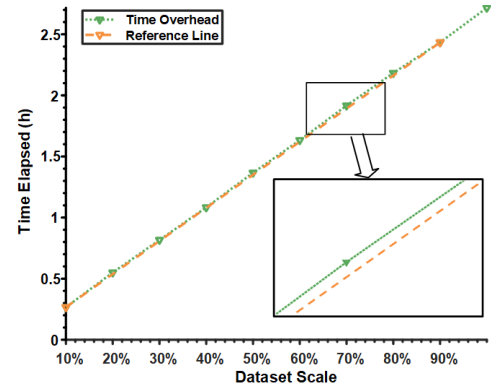

Fig. 16.   Training time overhead compared with a straight reference line.
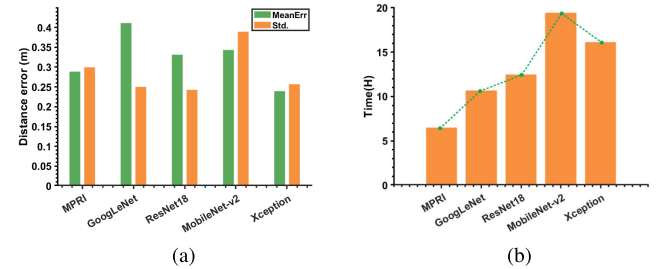

Fig. 17.   Comparisons of MPRI with various classic DL models in terms of (a) distance error and (b) time overhead.
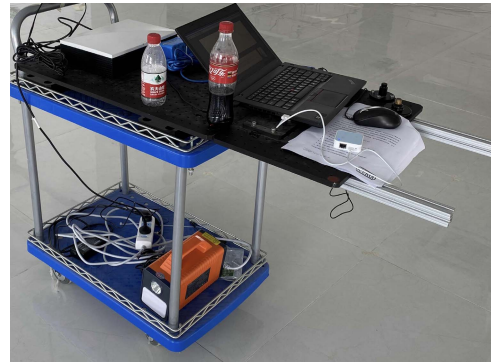

Fig. 18.   A gNB prototype on a cart that is collecting data.

It is observed in Fig. 14a that in the low-SNR cases, the positioning errors of MobileNet-V2 and GoogLeNet increase rapidly, showing that they are sensitive to noise. In contrast, MPRI and Xception present better positioning accuracy. The latter has advantages in the high SNR cases, while MPRI is more accurate in the 15-20 dB area. More importantly, the standard deviation of MPRI is superior to others in all tested SNR cases, as shown in Fig. 14b. The MPRI's smaller standard deviation shows a lower degree of positioning dispersion, indicating that the positioning performance is more robust in a variety of SNR levels.

Experiments in this section show that the proposed MPRI and data enhancement method can provide advantages in robustness to noise. This indicates that our DL-based positioning method can adapt to noisy environments for positioning tasks, in addition to the ideal laboratory environment. This enhancement has not been reported in other datasets in the positioning field.

F. Robustness to Dataset Scale

The impact of the dataset scale on the robustness of MPRI is analysed in this section. In practice, the enormous cost of large-scale data collection and the consequent increase in time overhead are unacceptable. Therefore, we tested the robustness of MPRI with respect to the scale of the dataset generated by Multilevel-FSM. The control group was defined as an MPRI network trained on the full-size original dataset (3876 CFR images). Correspondingly, experimental groups 1-9 were defined as MPRI networks trained using 10-90% of the original dataset (with a step size of 10%).

As shown in Fig. 15, the $MeanErr$ of experimental group 1 was 0.6 m, with a standard deviation of 0.4 m. The

$MeanErr$ of the control group was 0.2 m, corresponding to improvements of 67.57% and 53.08% compared with experimental groups 1 and 5, respectively.

As demonstrated by experimental groups 2-9, the positioning accuracy continued to improve as more data were added to the training process. In experimental group 9, $MeanErr$ reached 0.22 m, and the standard deviation was 0.12 m, close to the performance of the control group. It is worth noting that even when only 300 points were used for training, sub-metre accuracy could still be achieved. Therefore, Multilevel-FSM dataset generation and the MPRI network are already effective at a small dataset scale, and the performance can be further improved by expanding the dataset scale.

G. Time Overhead

In this section, we examine the time consumption of the proposed MPRI method, including the training time and the refresh rate for real-time positioning.

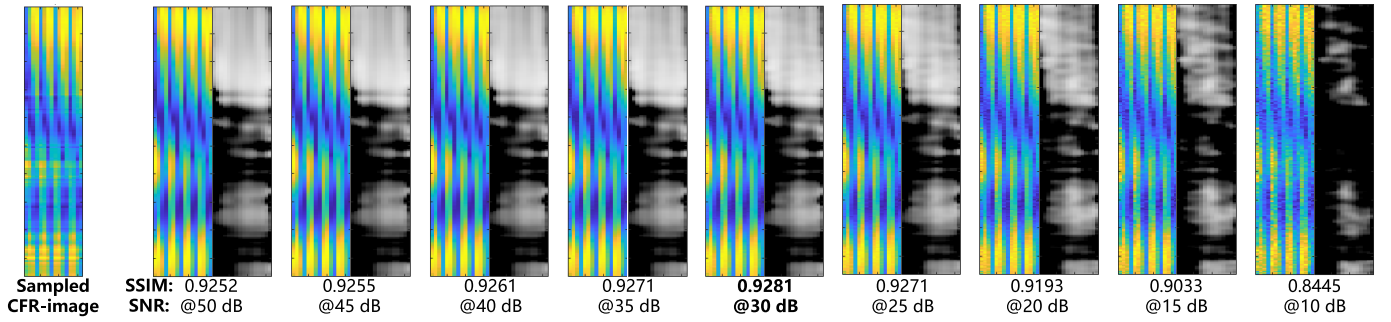| Sampled CFR-image	SSIM:	0.9252	0.9255	0.9261	0.9271	**0.9281**	0.9271	0.9193	0.9033	0.8445
	SNR:	@50 dB	@45 dB	@40 dB	@35 dB	**@30 dB**	@25 dB	@20 dB	@15 dB	@10 dB

Fig. 19. Comparison of the sampled CFR image with the synthetic images. The real CFR image sample is displayed on the far left, and the following 9 groups are composed of generated CFR images and SSIM thermal maps.

Fig. 16 shows the training time overheads of MPRI on the datasets of different scales described in Section V.F, with the number of training epochs set to 400 in each case. We find that the change in time consumption is basically linear, with training on 300 data points taking approximately 16 minutes and training on 3101 data points taking approximately 2 hours and 43 minutes.

We also tested the speed of the neural network in calculating a UE's three-dimensional location for real-time positioning. We gave the MPRI network 1,000 positioning tasks to calculate the "thousand-location positioning time" (TPT). When this test was repeated 200 times, the average TPT was 9.76 s, indicating that the time required for each task was less than 0.01 seconds. These results show that with the MPRI method, a UE can achieve a positioning refresh rate of more than 100 Hz, which is much higher than the usual GPS refresh rate (5-10 Hz) and sufficient to meet the real-time positioning needs for medium-speed moving objects.

H. Comparison With Classic DL Methods

A comparative analysis (as mentioned in Section IV.A) of the proposed MPRI network and several classic image recognition networks is presented here. The test group is the proposed MPRI network. The control group includes GoogLeNet, ResNet-18, MobileNet-v2, and Xception. The experiments were all carried out using the exact same dataset and parameters.

As shown in Fig. 17a, the positioning accuracy of the MPRI network is better than 0.3 m, being 29.76%, 12.68%, and 15.78% higher than those of the classic networks GoogLeNet [18], ResNet-18 [19], and MobileNet v2 [20], respectively. Xception [21] slightly outperforms MPRI at the cost of a time complexity that is higher by a factor of two. In terms of time overhead, as shown in Fig. 17b, the training time overhead of the MPRI network is only 6.55 h, representing savings of 38.8%, 47.81%, 66.41% and 59.48% compared with GoogLeNet, ResNet-18, MobileNet v2 and Xception, respectively. Thus, the training speed is significantly increased.

Therefore, the weaknesses noted in Section IV.A are successfully addressed by MPRI, indicating that MPRI is more compatible with the proposed CFR image dataset. While greatly reducing training time consumption, MPRI also improves the positioning accuracy by 15-29%, thereby demonstrating better practical applicability.

I. Comparison With Sampled CFR Images

In this section, a verification of the similarity of synthetic data is presented compared with a field measurements in the scenario 1, indoor office hall of CAS.

For detailed information on scenario size and other information, please refer to the experimental settings in Section V-A. As shown in Fig. 18, signal acquisition is carried out when the cart is stationary, and 50 sets of sampling data are collected at each sampling point. To compare the synthesized CFR image with the corresponding sampled CFR image, we randomly select a point, #17. Subsequently, we defined the control group as the values of each pixel averaged over all 50 samples. The experimental groups 1-9 were defined as the CFR images generated with 9 different SNRs (ranging from 10 dB to 50 dB, in steps of 5 dB). As shown in Fig. 19, the averaged CFR image sample is presented on the left, and 9 sets of generated CFR image are evaluated in SSIM value with an SSIM heatmap. The brightness of each pixel in the heatmap is directly proportional to the local similarity. The best consistency is reached at 30 dB (experimental group 5) with a mean SSIM of 0.928, which is a good SNR ratio in the laboratory environment.

Therefore, it can be observed that the CFR image generated by Mutilevel-FSM is close to the real sampled CFR image in SSIM.

VI. CONCLUSION AND FUTURE WORK

As a crucial foundation for 5G ISAC high-precision positioning, high-quality datasets for this task are urgently needed. A new paradigm for generating 5G datasets for ISAC high-precision positioning, named Multilevel-FSM, is proposed in this paper. 1) First, we have built a multilevel architecture for dataset generation, with the aim of providing UEs with sufficient but not overly detailed CSI. The generation and positioning processes follow the R16 standard and thus are able to provide fine-grained data in a manner more consistent with the real 5G NR communication principle. 2) The key to our dataset generation method is to flatten the features of MIMO channels into a single CFR image so that the subtle differences in features among adjacent antennas can be simultaneously embedded. Consequently, the information density of the resulting CFR images is enhanced. 3) Furthermore, we design a data augmentation method for Multilevel-FSM to achieve better robustness to noise for DL-based positioning. Simulation results show that the positioning accuracy improves

even under noise at an unknown SNR level. 4) Finally, we develop a novel DNN model, named MPRI, to improve the accuracy of 5G NR positioning. Our experimental results confirm that the training time for MPRI is significantly reduced compared with that for classic DNNs, while the positioning accuracy is simultaneously improved by approximately 15-29%. Thus, the superiority of MPRI compared to other state-of-the-art methods has been definitively demonstrated.

This work is our first research attempt focusing on datasets intended to support 5G ISAC high-precision positioning. Further efforts involving more extensive experiments will be necessary to refine the proposed method to accommodate dynamic scenarios and achieve better feature expression.

REFERENCES

[1] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 3rd Quart., 2017, doi: 10.1109/COMST.2019.2911558.

[2] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–32, Nov. 2013, doi: 10.1145/2543581.2543592.

[3] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J. Oh, "Semisupervised deep reinforcement learning in support of IoT and smart city services," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 624–635, Apr. 2018.

[4] P. Chhikara, R. Tekchandani, N. Kumar, V. Chamola, and M. Guizani, "DCNN-GA: A deep neural net architecture for navigation of UAV in indoor environment," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4448–4460, Mar. 2021.

[5] F. Zafari, I. Papapanagiotou, and K. Christidis, "Microlocation for Internet-of-Things-equipped smart buildings," *IEEE Internet Things J.*, vol. 3, no. 1, pp. 96–112, Feb. 2016.

[6] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, and C. Fischione, "A survey of enabling technologies for network localization, tracking, and navigation," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3607–3644, 4th Quart., 2018.

[7] F. Dou, J. Lu, T. Xu, C.-H. Huang, and J. Bi, "A bisection reinforcement learning approach to 3-D indoor localization," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6519–6535, Apr. 2021.

[8] Y. Li, X. Hu, Y. Zhuang, Z. Gao, P. Zhang, and N. El-Sheimy, "Deep reinforcement learning (DRL): Another perspective for unsupervised wireless localization," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6279–6287, Jul. 2020.

[9] Y.-Y. Chen, S.-P. Huang, T.-W. Wu, W.-T. Tsai, C.-Y. Liou, and S.-G. Mao, "UWB system for indoor positioning and tracking with arbitrary target orientation, optimal anchor location, and adaptive NLOS mitigation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9304–9314, Sep. 2020.

[10] *Study on NR Positioning Support*, Standard 38.855, 03-28, 3GPP, Tech. Rep., 2019. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDeta%ils.aspx?specificationId=3501

[11] F. Liu, W. Yuan, J. Yuan, J. A. Zhang, Z. Fei, and J. Zhou, "Radar-communication spectrum sharing and integration: Overview and prospect," *J. Radars*, vol. 10, Nov. 2021, Art. no. R20113, doi: 10.12000/JR20113.

[12] H. Wang, K. Gao, and H. Lv, "A survey of high-precision localization and the prospect of future evolution," *J. Commun.*, vol. 42, pp. 198–210, 2021.

[13] S. Fan, Y. Wu, C. Han, and X. Wang, "A structured bidirectional LSTM deep learning method for 3D terahertz indoor localization," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Jul. 2020, pp. 2381–2390.

[14] S. Fan, Y. Wu, C. Han, and X. Wang, "SIABR: A structured intra-attention bidirectional recurrent deep learning method for ultra-accurate terahertz indoor localization," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2226–2240, Jul. 2021.

[15] M. M. Butt, A. Rao, and D. Yoon, "RF fingerprinting and deep learning assisted UE positioning in 5G," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–7.

[16] J. Sangthong, J. Thongkam, and S. Promwong, "Indoor wireless sensor network localization using RSSI based weighting algorithm method," in *Proc. 6th Int. Conf. Eng., Appl. Sci. Technol. (ICEAST)*, Jul. 2020, pp. 1–4.

[17] J. Luo, Z. Zhang, C. Wang, C. Liu, and D. Xiao, "Indoor multifloor localization method based on WiFi fingerprints and LDA," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5225–5234, Sep. 2019.

[18] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[22] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw. (Mobicom)*, 2012, pp. 269–280, doi: 10.1145/2348543.2348578.

[23] M. Arnold, J. Hoydis, and S. T. Brink, "Novel massive MIMO channel sounding data applied to deep learning-based indoor positioning," in *Proc. SCC 12th Int. ITG Conf. Syst., Commun. Coding*, Feb. 2019, pp. 1–6.

[24] A. Decurninge *et al.*, "CSI-based outdoor localization for massive MIMO: Experiments with a learning approach," in *Proc. 15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2018, pp. 1–6.

[25] *System-Level Performance Evaluation of Co-Band TBS for TR 38.855*, document BUPT, ZTE and CAICT, Tech. Rep., 2019. [Online]. Available: https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_96/Docs/R1-1902698.zip

[26] *System-Level Evaluations on Rat-Dependent Positioning*, document R1-1901781, ZTE and Sanechips, Tech. Rep., 2019. [Online]. Available: https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_96/Docs/R1-1901781.zip

[27] *Performance Evaluations for NR Positioning*, document R1-1901577, Huawei and HiSilicon, Tech. Rep., 2019. [Online]. Available: https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_96/Docs/R1-1901577.zip

[28] *Remaining Issues on DL Based Positioning*, document R1-1901574, Huawei and HiSilicon, Tech. Rep., 2019. [Online]. Available: https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_96/Docs/R1-1901574.zip

[29] C. D. C. Shijun and W. Huiqiang, "5G oriented high-precision fusion positioning architecture and key technologies," *ZTE Technol. J.*, vol. 99, pp. 1–9, Jan. 2018.

[30] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2013, pp. 249–262, doi: 10.1145/2462456.2464463.

[31] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. USENIX Symp. (NSDI)*. Chicago, IL, USA, Apr. 2013, pp. 71–84. [Online]. Available: https://www.usenix.org/conference/nsdi13/technical-sessions/presentatio n/xiong

[32] H. Kim, K. Granstrom, L. Gao, G. Battistelli, S. Kim, and H. Wymeersch, "5G mmWave cooperative positioning and mapping using multi-model PHD filter and map fusion," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3782–3795, Jun. 2020.

[33] J. Lee, I. Song, H. Kwon, and S. Ro Lee, "Low-complexity estimation of 2D DOA for coherently distributed sources," *Signal Process.*, vol. 83, no. 8, pp. 1789–1802, Aug. 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168403001038

[34] F. Wen, J. Kulmer, K. Witrisal, and H. Wymeersch, "5G positioning and mapping with diffuse multipath," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1164–1174, Feb. 2021.

[35] A. Zoubir and Y. Wang, "Efficient DSPE algorithm for estimating the angular parameters of coherently distributed sources," *Signal Process.*, vol. 88, no. 4, pp. 1071–1078, Apr. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168407003799

[36] C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu, and Y. Bo, "UWB NLOS/LOS classification using deep learning method," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2226–2230, Oct. 2020.

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

[38] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with CSI images," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 316–327, Jan. 2020.

[39] X. Wang, X. Wang, and S. Mao, "Indoor fingerprinting with bimodal CSI tensors: A deep residual sharing learning approach," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4498–4513, Mar. 2021.

[40] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[41] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 227–236, Jan. 2020.

[42] M. Arnold, S. Dorner, S. Cammerer, and S. Ten Brink, "On deep learning-based massive MIMO indoor user localization," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[43] *Physical Channels and Modulation*, Standard 38.211, 3GPP, Technical Specification, 2021.

[44] N.-R. Jeon, C.-H. Lee, N.-G. Kang, and S.-C. Kim, "Performance of channel prediction using 3D ray-tracing scheme compared to conventional 2D scheme," in *Proc. Asia–Pacific Conf. Commun.*, Aug. 2006, pp. 1–6.

[45] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, Standard 38.901, 3GPP, Tech. Rep., 2021.

[46] R. Mendrzik, F. Meyer, G. Bauch, and M. Z. Win, "Enabling situational awareness in millimeter wave massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1196–1211, Sep. 2019.

[47] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1822–1835, Mar. 2018.

[48] H. Wymeersch *et al.*, "5G mm wave downlink vehicular positioning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 206–212.

[49] R. Ziemer and W. Tranter, *Principles of Communications: Systems, Modulation, and Noise*. Hoboken, NJ, USA: Wiley, 2002. [Online]. Available: https://books.google.com.sg/books?id=6R0fAQAAIAAJ

[50] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Towards omnidirectional passive human detection," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 3057–3065.

[51] G. Calcev *et al.*, "A wideband spatial channel model for system-wide simulations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 389–403, Mar. 2007.

[52] K. Gao, H. Wang, and H. Lv, "CSI dataset towards 5G NR high-precision positioning," *IEEE Dataport*, 2021. [Online]. Available: https://ieee-dataport.org/open-access/csi-dataset-towards-5g-nr-high-precision-positioning, doi: 10.21227/jsat-pb50.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[54] H.-J. Shao, X.-P. Zhang, and Z. Wang, "Efficient closed-form algorithms for AOA based self-localization of sensor nodes using auxiliary variables," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2580–2594, May 2014.

Kaixuan Gao received the B.E. degree in computer science and technology in June 2018. He is currently pursuing the Ph.D. degree with Harbin Engineering University (HEU). His current research interests include indoor localization, communication and navigation (CaN)-integrated networks, AI, and future networks.

Huiqiang Wang received the M.E. and Ph.D. degrees from Harbin Engineering University (HEU) in 1985 and 2005, respectively. He is now engaged in teaching and research as a Professor and a Post-Doctoral Advisor with HEU. His research interests include CaN-integrated networks, future and cognitive networks, and autonomic computing.

Hongwu Lv received the B.A. and Ph.D. degrees from Harbin Engineering University (HEU) in 2006 and 2011, respectively. He is currently an Associate Professor with the College of Computer Science and Technology, HEU. His research interests include the study of indoor localization, network security, formal modelling, and performance evaluation.

Wenxue Liu received the bachelor's and Ph.D. degrees from the University of Chinese Academy of Sciences, China, in 2013 and 2019, respectively. He is a Senior Researcher with the Institute of Microelectronics of Chinese Academy of Sciences. His current research interests include GNSS application in time synchronization and BDS/5G joint positioning technology.