SIABR: A Structured Intra-Attention Bidirectional Recurrent Deep Learning Method for Ultra-Accurate Terahertz Indoor Localization

Shukai Fan, Yongzhi Wu, *Graduate Student Member, IEEE*, Chong Han, *Member, IEEE*, and Xudong Wang, *Fellow, IEEE*

***Abstract*—High-accuracy localization technology has gained increasing attention in gesture and motion control and many diverse applications. Due to multi-path fading and blockage effects in indoor propagation, 0.1m-level precise localization is still challenging. Promising for 6G wireless communications, the Terahertz (THz) spectrum provides multi-GHz ultra-broad bandwidth. Applying the THz spectrum to indoor localization, the channel state information (CSI) of THz signals, including angle of arrival (AoA), received power, and delay, has unprecedented resolution that can be explored for positioning. In this paper, a Structured Intra-Attention Bidirectional Recurrent (SIABR) deep learning method is proposed to solve the CSI-based three-dimensional (3D) THz indoor localization problem with significantly improved accuracy. As a two-level structure, the features of individual multi-path rays are first analyzed in the recurrent neural network with the attention mechanism at the lower level. Furthermore, the upper-level residual network (ResNet) of the constructed SIABR network extracts hidden information to output the geometric coordinates. Simulation results demonstrate that the 3D localization accuracy in the metric of mean distance error is within 0.25m. The developed SIABR network has very fast convergence and is robust against THz indoor line-of-sight blockage, multi-path fading, channel sparsity and CSI estimation error.**

***Index Terms*—Terahertz communications, indoor localization, deep learning.**

I. Introduction

Fast developing mobile devices and wireless technology have brought diverse applications like autonomous driving, unmanned aerial vehicles, virtual/augmented reality, and gesture and motion control under the spotlight. On one hand, high-accuracy three-dimensional (3D) localization technology has never been strongly demanded as it is today [2]. On the other hand, the Terahertz (THz) band with ultra-broad bandwidth has drawn tremendous interests to support these promising applications in 5G and future 6G wireless systems [3], [4]. Therefore, a promising yet challenging research direction on THz localization is motivated and stringent.

Although the Global Positioning System (GPS) provides an excellent solution for outdoor localization, indoor localization still receives increasing interests to advanced accuracy in complex propagation environments [2], [5]–[12]. Different from the GPS mechanism where line-of-sight (LoS) transmission is mostly available, indoor localization involves complicated radio propagation phenomena, including multi-path fading, line-of-sight (LoS) blockage, among other effects. The problem evolves more challenging by incorporating the unique characteristics of THz propagation, including very high propagation attenuation, atmospheric molecular absorption, rough-surface reflection and scattering, and channel sparsity [13].

Specifically, since the spreading loss increases quadratically with the frequency and the atmospheric effects may attenuate the THz wave propagation, the propagation loss in the THz band becomes much higher than the lower frequency bands [14]. To compensate for this severe path loss, the massive multiple input multiple output (MIMO) and beamforming techniques are usually used to enable direction transmission with multiple antenna beams and high antenna gains [15]. Meanwhile, due to the ultra-high penetration loss, the THz signals can be obstructed by indoor blockages, such as static wall blockers and moving human bodies. Moreover, diffuse scattering in the THz band causes severe losses, which leads that the number of the dominant rays decreases and the sparsity of the THz channel arises [13]. In particular, the energy of the received THz signal is mainly concentrated in the LoS ray and several specular reflected rays.

As a result, THz non-line-of-sight (NLoS) propagation needs to be rigorously treated in indoor localization, in order to make full use of the information within several specular reflected rays. Existing solutions such as triangulation and radio map [6], [16], which use received signal strength indication (RSSI) for WiFi networks, are vulnerable to the LoS blockage and naturally have low positioning resolution at lower frequencies. Furthermore, current AoA based solutions [17], [18] struggle between high accuracy and specially modified experimental settings. Recently, rapidly rising artificial intelligence (AI) techniques are borrowed to solve the localization problem. The still limited efforts focus on 2D indoor environments and are not compatible with THz propagation peculiarities [5], [7], [9], [11], [19], due to the major differences between the THz spectrum and lower frequency bands in free-space attenuation, atmospheric molecular absorption, rough-surface reflection and scattering, and channel sparsity [13].

In this work, the objectives are to solve the 3D THz indoor localization problem and achieve ultra-accurate positioning, by

This work was in-part supported by National Key R&D Program of China under Project No. 2020YFB1805700, and Ministry of Education of China Joint Research Fund under Grant No. 6141A02033350. This paper was presented in part at the IEEE INFOCOM 2020 [1] *(Shukai Fan and Yongzhi Wu are co-first authors.) (Corresponding author: Chong Han.)*

Shukai Fan, Yongzhi Wu, Chong Han, and Xudong Wang are with University of Michigan–Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (E-mail:{kennyfan, yongzhi.wu, chong.han, wxudong}@sjtu.edu.cn).

leveraging deep learning networks. In particular, the channel state information (CSI) of THz indoor environment, including 3D angle of arrival (AoA), received power, and delay, are extracted and explored for localization. To address the THz indoor propagation challenges, we invent a structured intra-attention bidirectional recurrent (SIABR) deep learning method for ultra-accurate Terahertz indoor localization, which utilizes the CSI of THz wireless signals from multiple transmitters and estimates the 3D location of targeted receivers. According to the definition from [2], our proposed schema is device-based localization (DBL), where the user device uses three transmitters (Tx) to obtain its relative location. In the proposed SIABR method, two interconnected levels are designed in the neural network, in which the lower level consists of three base bidirectional long short-term memory (Bi-LSTM) networks that extract the THz NLoS ray features from three Tx cooperating with intra-attention mechanism, while the upper level integrates five blocks of residual network (ResNet). Remarkably, the mean bagging mechanism is implemented to reduce the variance of the estimated results.

Moreover, the proposed model is based on the extraordinary temporal and spatial resolution due to the ultra-broad bandwidth, and frequency- and distance-adaptive multi-path effects in the THz band. Compared to the mmWave and lower frequencies, the channel sparsity effect is much more severe in the THz band, due to the propagation attenuation, atmospheric molecular absorption, rough-surface reflection and scattering. Being incorporated in our proposed SIABR model, the intra-attention mechanism, upper-level residual learning and mean-bagging technique are designed to reinforce the robustness over the THz channel sparsity, static or dynamic blockage, limitation of dataset size and CSI estimation error.

The distinctive features of this work are summarized as:

- **We propose a novel paradigm for investigating the 3D THz indoor localization problem, including designing a complete schema to build a valuable and tailored THz localization database used for the supervised learning method**. The CSI dataset in the THz band is constructed based on a ray-tracing simulator, in consideration of the THz channel peculiarities. To the best of our knowledge, this is the first attempt on THz localization for beyond 5G wireless systems.
- **We develop the novel SIABR architecture for the THz localization, by constructing two interconnected levels**. The lower level consists of three base Bi-LSTM networks with an intra-attention mechanism that extract information from the THz NLoS ray features. By mutual interaction among features from different transmitters, the intra-attention mechanism is useful for the network to focus on the representative sequence segments with adaptive weights, and eventually grasp more comprehensive understanding of the indoor environment. The upper level integrates five blocks of ResNet to estimate the 3D location of the targeted receivers with ultra-high accuracy.
- **We conduct extensive performance evaluation of the SIABR method for 3D THz indoor localization in two representative environments**. Simulation results demonstrate the high efficiency and fast convergence of the proposed SIABR method in the THz multipath environment. Solid robustness of SIABR over THz channel sparsity, static or dynamic blockage, limitation of dataset size and estimation error is presented through the paper. Under these practical conditions, the localization accuracy in the metric of mean distance error achieves below 0.25m, which is highly competitive in contrast to the existing techniques from the literature.

The rest of this paper is organized as follows. Related work is surveyed in Section II. A framework of Terahertz indoor localization is depicted in Section III. Next, the system design of THz indoor localization is discussed in Section IV. We delineate the methodology of SIABR in Section V. Then, in-depth performance evaluation is described in Section VI. Finally, the conclusion is drawn in Section VII.

II. Related Work

For indoor localization, existing studies and experiments relate to ultra-wideband (UWB) [2], [5], [7], [9], [11], [19], [20] and millimeter-wave (mmWave) systems [8], [21]. However, these frequency bands suffer from crowded applications and high interference, which limit the accuracy of indoor localization to 1m, under practical 3D indoor environment with multi-path fading and blockage effects. By contrast, the THz band, specifically 95 GHz – 3 THz, has been approved by FCC for experimental use and reserved for future cellular systems [22]. This is aligned with the trend of requiring higher carrier frequencies in 5G and 6G wireless communications. Intrinsically, multi-GHz broad bandwidth, ultra-high temporal and spatial resolution of the THz signals bring advantages for localization in complex multi-path indoor environments.

To solve the indoor localization problem, methods can be classified into two main categories. On one hand, classic approaches use deterministic techniques, including triangulation [16], nearest neighbor in signal space (NNSS) [12] and radio map [6] based on the received signal strength (RSS) to estimate indoor positions. For RSS based non-learning methods, the estimation error cannot be reduced below 2m in indoor environments. Recent angle of arrival (AoA) based estimations such as ArrayTrack [17] and SpotFi [18] achieve a satisfying accuracy on median distance error. However, the performance of the localization strongly depends on the number of access points up to 6 and requires extra modification on them. Meanwhile, a fine-grained indoor fingerprinting system (FIFS) method is proposed to explore the channel status over all the subcarriers rather than using the RSS or AoA value, and uses the radio map to conduct location estimation [7]. Besides, while recent work in [23], [24] has explored some conventional THz localization methods with high estimation accuracy, they are not robust to a NLoS environment. In general, non-learning methods either struggle with low accuracy above 2.5m or the specially modified experimental setting.

On the other hand, with the uprising AI technologies, learning methods can be leveraged to solve the indoor location problem, including the studies of WiGEM leveraging gaussian mixture model (GMM) and expectation maximization [25], DeepFi using restricted Boltzmann machine

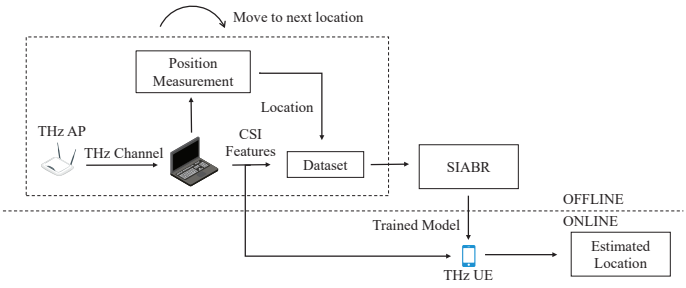


Figure 1: The system architecture of Tera-Loc.

(RBM) [5], BiLoc [19] and channel charting (CC) [24] based on auto-encoder, K-nearest neighbors (KNN) [10] and deep convolutional neural network (DCNN) [9], respectively. These studies extract the hidden features of CSI information and enhance the localization accuracy by at least 20%, in comparison to those non-learning methods. The highest accuracy, reported in [9], reduces the mean distance error to 2m in a complex lab environment and adopts the DCNN approach.

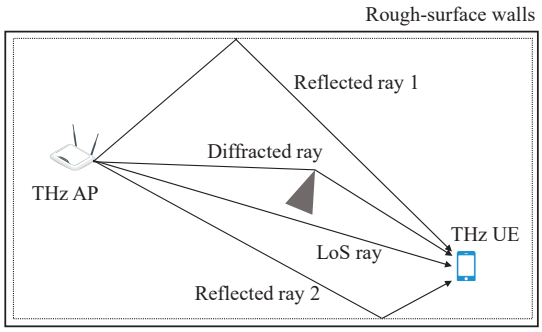
In summary, the literature studies on indoor localization mostly focus on microwave and mmWave frequency bands, while the THz localization problem in a 3D indoor environment with unique channel features and challenges is still untouched. Although the learning methods can substantially improve the accuracy over classic estimation approaches, ultra-accurate localization under practical indoor conditions that reduces the 3D distance error below 0.3m is still not reachable.

Hence, these aforementioned open problems and challenges motivate this study. In our preliminary and shorter version in [1], we invent a structured bidirectional LSTM (SBiLSTM) and improve the localization accuracy over the state-of-art solutions significantly. Remarkably, there are three major extensions to distinguish between our conference paper and this journal version. In Sec. III, we propose a novel paradigm for investigating the 3D THz indoor localization problem. The paradigm includes designing a complete schema to build a valuable and tailored THz localization database used for the supervised learning method, which is not formally covered in the conference paper. Besides, on the basis of our preliminary SBiLSTM structure, we further strengthen the deep learning network by introducing a new intra-attention mechanism, which enhances the ability to comprehensively extract the hidden patterns from the ray sequences and feed the upper level with less information loss. Furthermore, we also provide time complexity analysis for training and testing phases, and substantially more extensive performance evaluation under various practical indoor conditions, including THz multi-path channel sparsity, dynamic indoor blockage and influence of CSI estimation errors. In terms of localization accuracy, the proposed SIABR in this work achieves additional 7% improvement, compared to the SBiLSTM algorithm in [1].

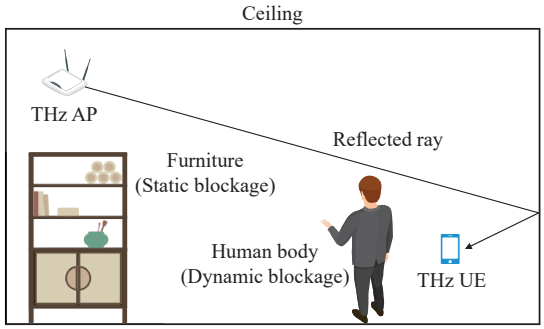
III. System Design of Terahertz Indoor Localization

A. System Framework

The overall framework of Terahertz localization (Tera-Loc) is described in Fig. 1, which is divided into two parts, namely, online and offline stages. At the offline stage, we first obtain the CSI features of the received signals at the user equipment (UE) from multiple THz access points (APs). To start with thousands of reference points, this transmission and reception process is repeated to enrich the size of the training dataset and improve the accuracy of the deep learning method. Meanwhile, the position of the UE is measured and transformed to 3D coordinates. With the dataset composed of these CSI features and locations, we train the proposed SIABIR neural network. Then at the online stage, by deploying the trained model on the CSI features of the UE, real-time localization is implemented to generate ultra-accurate 3D UE coordinates.



(a) The horizontal view of the multi-ray propagation between the THz access point and the user equipment.



(b) The vertical view of static and dynamic blockage, caused by furniture and human body.

Figure 2: The illustration of the indoor environment for the multi-ray propagation and blockage effects in the THz band.

B. Terahertz Channel

The THz propagation in the indoor environment has been characterized in [13], in which a multi-ray channel model is developed. In Fig. 2(a), a stationary indoor environment, there are N_{Ref} reflected rays and N_{Dif} diffracted rays, then the channel response of the multi-ray propagation is described as

$$\begin{aligned} h(\tau) =& \alpha_{\text{LoS}}\delta(\tau-\tau_{\text{LoS}})\mathbb{1}_{\text{LoS}} + \sum_{p=1}^{N_{\text{Ref}}} \alpha_{\text{Ref}}^{(p)}\delta(\tau-\tau_{\text{Ref}}^{(p)}) \\ &+ \sum_{q=1}^{N_{\text{Dif}}} \alpha_{\text{Dif}}^{(q)}\delta(\tau-\tau_{\text{Dif}}^{(q)}) \end{aligned} \tag{1}$$

where τ denotes the propagation delay, $\delta(\cdot)$ describes the unit impluse function, and $\mathbb{1}_{\text{LoS}}$ is the indicator function that is equal to 1 or 0 for the presence of LoS path or not. α_{LoS},

$\alpha_{\text{Ref}}^{(p)}$, $\alpha_{\text{Dif}}^{(q)}$ refers to the attenuation for the LoS path, p^{th} reflected path and q^{th} diffracted path, respectively. τ_{LoS}, $\tau_{\text{Ref}}^{(p)}$, $\tau_{\text{Dif}}^{(q)}$ represent the propagation delay for corresponding rays, and can be computed by the equation $\tau_n = \frac{r_n}{c}$, where r_n is the distance of the n^{th} path and c is the speed of the light.

In addition to the characteristic of the multi-ray propagation, THz signals are sensitive to the indoor blockage due to the high penetration loss. In Fig. 2(b), we describe the blockage effects in the indoor environment for the THz localization. On the one hand, the existence of the static blockage, such as walls and furniture, may result in the decrease of the number of the THz rays. On the other hand, some of the rays are likely to be suddenly blocked by the dynamic blockage, especially the human bodies in a crowded environment. This is a common phenomenon that may have a bad effect on the localization accuracy. Therefore, the robustness to the static and dynamic blockage of our method will be investigated in the performance evaluation. For localization, the main CSI features include AoA, propagation delay, and path loss, which can be obtained by the estimation algorithms as presented next.

C. Delay Estimation

The delay estimation algorithms have been widely investigated in the sensing and communication applications. Recently, attention has been drawn on joint communication and sensing (JCS) techniques [26], since they enable sharing of hardware and signal processing modules. The propagation delay can be estimated by the sensing algorithms from the received communication signals without sacrificing data rate.

For the multi-path delay estimation, a potential scheme in the communication system is the orthogonal frequency division multiplexing (OFDM) based JCS design [27], which is highly spectral-efficient and have strong robustness to frequency-selective channels and high MIMO compatibility. THz OFDM with hybrid beamforming and MIMO techniques is still of great significance for THz communications over frequency-selective channels [28]. With OFDM, different carriers experience beam-squint at THz frequencies, which can be mitigated by utilizing delay Vandermonde matrices [29].

We consider that a THz AP transmits a N-length OFDM symbol block $x[n], n = 0, 1, \cdots, N-1$, modulated over N subcarriers. The received block after OFDM demodulation is

$$y[n] = \sum_{l=1}^{L} \alpha_l e^{j\varphi_l} e^{-j2\pi\tau_l n\Delta f} x[n] + w[n], \tag{2}$$

where α_l and τ_l denote the attenuation and the propagation delay of l^{th} ray between the AP and the UE according to (1), φ_l describes the random phase distributed in $[0, 2\pi)$ and caused by the reflection that rotates the signal, Δf stands for the subcarrier spacing of OFDM waveform, and $w[n]$ represents the additive Gaussian noise. With the super-resolution sensing algorithms, such as the multiple signal classification (MUSIC) approach, the variance of the estimated delay $\hat{\tau}$ can be close to the Cramér-Rao lower bound (CRLB) [30]

$$\text{Var}\left[\hat{\tau}\right] \geqslant \frac{6}{\text{SNR} \cdot \left(N^2 - 1\right) NM} \left(\frac{1}{2\pi\Delta f}\right)^2 \tag{3}$$

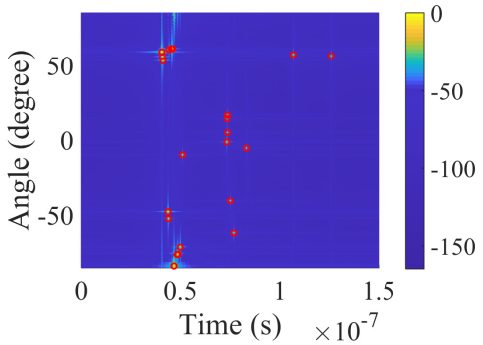


Figure 3: 2D estimation results of delay and AoA.

where SNR denotes the signal-to-noise ratio (SNR) of the ray power, and M describes the number of OFDM blocks. From (3) we learn that the delay estimation accuracy can be improved by increasing the number of subcarriers N and the subcarrier spacing Δf. Thanks to the ultra-broad bandwidth in the THz band, the estimation accuracy can achieve picosecond-level. For instance, when using 256 subcarriers and 5 MHz subcarrier spacing, the lower bound of the delay accuracy is about 6 picoseconds at the SNR of 10 dB.

D. AoA Estimation

Due to the sub-millimeter wavelength of the THz wave, an ultra-massive multiple-input multiple-output (UM-MIMO) system is allowed to be deployed at the THz APs and UEs. An array-of-subarrays (AoSA) hybrid beamforming structure is a promising technology owing to its high beamforming gain and reduced hardware complexity for THz wireless communications [31]. In this case, the AoA information at the UE is able to be accurately estimated.

We consider to employ uniform linear antenna arrays (ULA) for the azimuth AoA estimation, where the antenna elements is equally spaced at half wavelength. For an array with K antennas at both transmitter and receiver, the THz UM-MIMO channel matrix is given by

$$\mathbf{H} = \sum_{l=1}^{L} \alpha_l \delta(\tau - \tau_l) \mathbf{a}(\theta_{r,l}) \mathbf{a}^T(\theta_{t,l}) \tag{4}$$

where for the l^{th} path, τ_l is the propagation delay, $\theta_{r,l}$ and $\theta_{t,l}$ are the azimuth angle-of-arrival and the angle-of-departure, and the array response vector $\mathbf{a}(\theta)$ is given by

$$\mathbf{a}(\theta) = \left[1, e^{j\pi\sin(\theta)}, \ldots, e^{j\pi(K-1)\sin(\theta)}\right]^T \tag{5}$$

where θ is for either $\theta_{t,l}$ or $\theta_{r,l}$, and $(\cdot)^T$ denotes the transpose operation. Tailored for THz AoSA hybrid beamforming, a AoSA-MUSIC algorithm is proposed in [32] for the AoSA architecture, which can achieve a millidegree-level estimation accuracy. The CSI estimation techniques can be also extended to the spherical-wave channel model for the THz MIMO systems [33], which is more accurate than the uniform plane wave assumption.

The above parameters, delay and AoA, can be also estimated jointly for each path by developing 2D estimation algorithms.

In Fig. 3, we choose a sample in the dataset for the localization problem and show the joint delay and AoA estimation results. The real delays and AoAs for each ray are denoted by the red circles. The plot of delay-angle profile indicates that the features of all rays can be jointly and accurately estimated in wireless systems.

While the employment of the ultra-broad bandwidth and the use of large antenna arrays are able to provide accurate CSI features with less estimation error compared to the microwave band, it is still of great significance to study the influence on the localization accuracy caused by the bias when estimating the CSI features, especially the delay and AoA information. This type of error can be characterized by the estimation variance. It is predicted that the localization accuracy will be reduced if the variance of the estimation error becomes larger, owing to the loss of the information contained in the CSI features, which will be validated by the experiments.

IV. Ray-Tracing Channel Simulation of Tera-Loc System

A. Ray-Tracing Simulation

Ray-tracing techniques have been widely adopted to simulate the electromagnetic (EM) wave propagation, based on geometric-optics principles [13], [15]. As a complementary approach to the experimental measurement, a good balance between accuracy and computational complexity can be achieved by the ray-tracing simulation, which is favored for millimeter-wave and THz propagation studies in recent years [13], [15]. In this work, we generate the indoor environment and simulate the THz wave propagation, by using a commercial software *Wireless InSite*®. This software includes a suite of ray-tracing models and high-fidelity EM solvers for the analysis of site-specific radio wave propagation and wireless systems.

Specifically, the ray tracing simulations use the shoot-and-bouncing-ray (SBR) ray-tracing method. During the simulation, the number of the reflections is limited to 6, which covers a reasonable portion of rays that could be detected in the real scenarios. The maximum number of diffraction for the ray to reach the receivers is 1, considering the power of rays reduces significantly after once diffraction. With these constrains, the average number of rays for each receiver-transmitter pair is 24. Each ray corresponds to one group of CSI features including power, delay, elevation AoA and azimuth AoA. The carrier frequency is 100 GHz, and the bandwidth is 5 GHz. We build a 3D coordinate system, where O is regarded as the origin, x-y spans the horizontal plane, and the positive z-axis points vertically upwards to represent the height above the ground. In Fig. 4 and Fig. 5, two indoor environments are simulated for our analysis and described as follows.

1) Environment 1: The first environment in Fig. 4 is a hall, modeled as a cuboid with a size of 25m×25m×5m (length×width×height). It contains four cuboid pillars standing on the floor and below the ceiling with a size of 5m×5m×5m, whose centers are located at (7.5m, 9m), (17.5m, 9m), (7.5m, 18m), and (17.5m, 9m), respectively. The material of the ceiling, floor and walls, as well as the four pillars inside the hall, is concrete. This environment reveals the performance of the proposed model in a large and symmetrical hall environment with strong blockage effect.

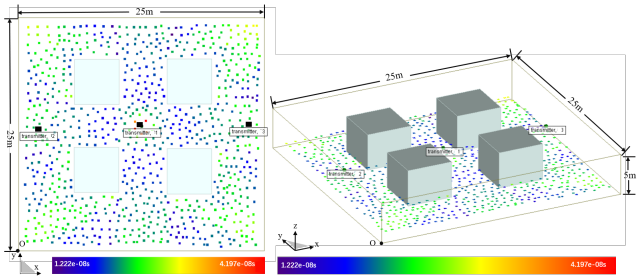


Figure 4: Environment 1. The left figure is the top view of the environment, in which the color points show the delay information of the signal from transmitter #1 to the receivers in the first layer with the height of 0.5m. The right figure is the 3D demonstration of the environment.

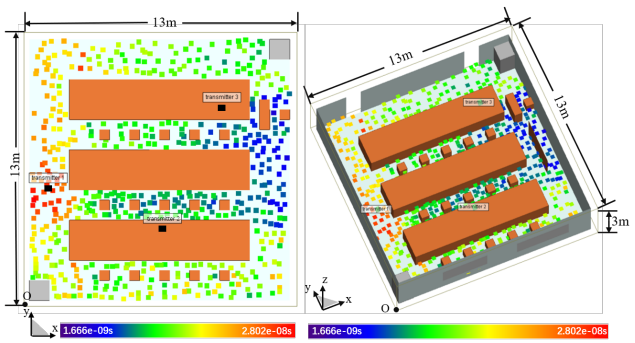
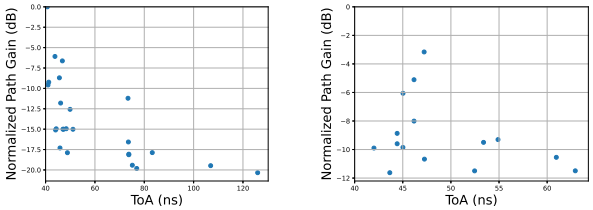


Figure 5: Environment 2. The left figure is the top view of the environment, in which the color points show the delay information of the signal from transmitter #1 to the receivers in the first layer with the height of 0.5m. The right figure is the 3D demonstration of the environment.

Three transmitters, represented by black points in the top view of Fig. 4, are respectively placed at (1.93m, 13.71m, 2.00m), (13.55m, 13.63m, 2.00m), (25.45m, 13.5m, 2.00m), from which a high coverage can be ensured in the indoor environment. As for the receivers, we choose 7 evenly-distributed horizontal layers with the heights ranging from 0.5m to 3.5m above the ground. The height difference between the neighboring two layers is 0.5m, which is chosen based on the trade off between none-oversized dataset and sufficient resolution on z-axis for common application. At each layer, receivers are randomly located outside the pillars. In this case, our training set can cover majority of the indoor environment.

2) Environment 2: The second environment in Fig. 5 is a laboratory, which is considered as a 13m×13m×3m cuboid model. There are three long tables in the middle and each of them has the size of 7.5m×1.5m×1.2m. Meanwhile, fifteen seats, whose size are identically 0.5m×0.5m×0.7m, are distributed between the tables. In the front of the lab, a board (2m×3m) is vertically placed, and a pair of desk (1.5m×0.5m×1.2m) and seat (0.5m×0.5m×0.7m) stands beside it. The tables, seats, and the board are made of wood. In addition, two air conditioners with a 1m×1m×2m rectangular shape, simplified as metal boxes, are installed in the two diagonal corners of the room. Moreover, two 1m-wide doors are



(a) LoS ray is present and the coherence bandwidth is 0.67 GHz. (b) LoS ray is absent and the coherence bandwidth is 0.10 GHz.

Figure 6: Normalized path gain of multiple rays versus propagation delay for LoS and NLoS cases in the simulated environment 2.

opened on the left side of the room, while there are two single-glazing windows on the right wall. This environment studies the proposed method in a laboratory with more materials, more polyhedral surfaces and less symmetry properties.

Similarly, three transmitters in Fig. 5 are selected, whose coordinates are (12.02m, 7.37m, 2.00m), (6.34m, 9.44m, 1.50m), (3.39m, 3.38m, 2.50m). Five layers of receivers are chosen, whose heights are 0.5m to 2.5m with a interval of 0.5m. The horizontal position of each receiver is randomly distributed in the lab, excluding the areas occupied by the furniture.

LoS Availability: Environment 1 contains four rectangular pillars in the hall, while environment 2 is more complicated involving with more obstacle objects and polyhedral surfaces. Therefore, the LoS probability varies between these two environments. To evaluate the LoS property of the receivers recorded in the dataset in both environments, $k_{LoS} = 0, 1, 2, 3$ indicates the total number of LoS ray for each receiver from the three transmitters. Given an integer threshold i, we denote $\gamma_{LoS}(i)$ as the ratio between the number of receivers having at least i LoS and the total number of receivers, which is listed in Table I for the two different environments.

Path gain versus propagation delay: Fig. 6 shows the normalized path gain of multiple rays as a function of the propagation delay for LoS and NLoS cases in the simulated environment 2. In Fig. 6(a), the rays between Tx#1 and a typical LoS Rx in environment 2 are illustrated, with the LoS ray in the upper left corner with the time-of-arrival (ToA) of 40 ns. With the path gain and delay spread, the coherence bandwidth of this Tx-Rx pair is 0.67 GHz. Moreover, Fig. 6(b) demonstrates the path gain of rays between a typical NLoS Rx and Tx#1 in environment 2, with absence of the LoS ray. The received energy is widely distributed among the NLoS rays, and the coherence bandwidth reduces substantially to 0.10 GHz. In both two cases, the coherence bandwidth is much smaller than the total used bandwidth, which causes the frequency-selective channels. Motivated by this, we propose the OFDM-based CSI feature estimation method for the TeraLoc framework in Sec. III.

B. CSI Signal Pre-processing

For the purpose of collecting data, large amount of receivers are placed under both environments. The labels are the $\{x, y, z\}$ coordinates of the receivers. From the three transmitters, each individual receiver detects 72 arrival rays in total. The features of rays, including the AoA, delay, power and index of the anchor transmitter, are recorded for training and testing the proposed network. The data structure is presented in Fig. 7. By contrast, for the network training purpose, 4000 and 2800 reference receiver points are used in environments 1 and 2, respectively. Correspondingly, environment 1 and 2 have 1.28 and 5.52 reference receiver points per cubic meter. In addition, another 500 points are reserved to test the accuracy of the proposed network in performance evaluation in Section VI.

TABLE I: Proportion of the receivers having LoS signal with respect to different k_{LoS} values for both environments.

	$\gamma_{LoS}(1)$	$\gamma_{LoS}(2)$	$\gamma_{LoS}(3)$
Env. 1	0.9175	0.2750	0.2475
Env. 2	0.9018	0.7619	0.7064

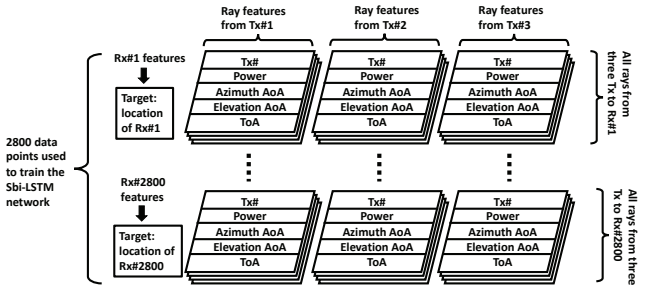


Figure 7: Data structure of the training set in the SIABR network.

In order to test the model robustness based on the SIABR, we further simulate the estimation error and background noise, by adding zero-mean Gaussian noise to all features in both environments, including AoA, power and delay.

Data normalization is implemented on the dataset, in order to lessen the computational burden of the network and increase the convergence ability. First, zero-score standardization, which linearly transforms the standard deviation to one and the mean to zero, is conducted for the features, including AoA values, delay values and power values. This zero-score standardization ensures that the effect of any one input is on the same scale as the others. Then, max-min normalization, applied on the location estimation targets, maps the coordinates, $\{$x, y, z$\}$, to $[-1, 1]$.

Moreover, we apply zero-padding on this dataset, for the following two reasons. First, the number of the rays between different pair of receiver and transmitter varies, causing different length of each network input, while zero-padding can unify their length. Second, the zeros padded to the end of the inputs help the network to identify where the ray sequence ends and ensure the model convergence. The number of rays from each transmitter is padded to 40, which ensures that no simulated rays are discarded. The final dimension of the input is 120×5, which indicates 120 rays in total after zero-padding and 5 features for each ray.

V. Structured Intra-Attention Bidirectional Recurrent Deep Learning Method

In this section, we delineate our proposed SIABR method in terms of the network structure and algorithms.

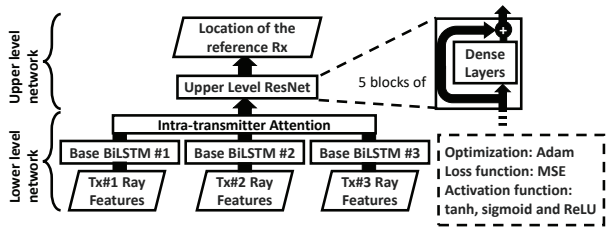


Figure 8: The proposed SIABR deep learning network for THz indoor localization.

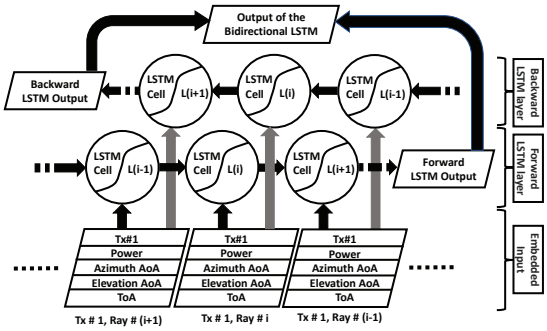


Figure 9: The base Bi-LSTM network for Transmitter #1.

A. Network Architecture

In order to address the problem of THz indoor localization, we develop a Structured Intra-Attention Bidirectional Recurrent network, which is a deep learning approach. As shown in Fig. 8, the whole network consists of the lower level and the upper level. The former part is composed of three parallel single layer Bi-LSTM neural networks with intra-attention mechanism [34]–[37], whose inputs gain from the information provided by the three transmitters. Then their outputs are concatenated and make up the inputs of the upper level network, which conducts the residual learning for the estimated location. The upper level network contains multiple residual blocks [38], as well as a fully connected layer used for outputting the targets. Experimenting to minimize the error and maximize the convergence efficiency, the number of residual blocks is set to 5. The technical details of each part are illustrated in the rest of this section.

B. Base Bi-LSTM Network

The *Bi-LSTM* model [34] is used at the lower level of our proposed network, where each Bi-LSTM network successfully extracts patterns from the ray information of one transmitter and significantly improves the accuracy of the location estimation. Fig. 9 is an illustration of one of the three Bi-LSTM at the lower level. We start with description of the classic recurrent layer, LSTM [35]. The main advantage of using LSTM is its ability in revealing the internal relation between rays. LSTM differs from fully-connected network and many other machine learning methods that flatten the 120×5 input matrices into 600 features and treat them equally. It maintains the structure of the input matrices, sequentially processes each rays and

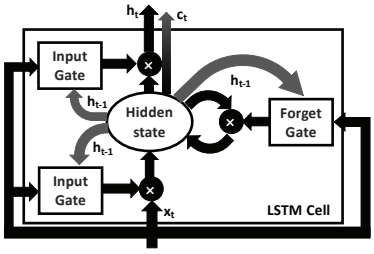


Figure 10: An LSTM unit.

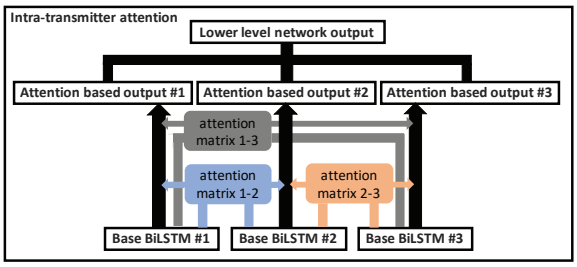


Figure 11: Intra-transmitter attention mechanism.

remembers the information inside the LSTM cell. Therefore, LSTM structurally captures the characters of the input matrices while reduces computational burden. In Fig. 10, a LSTM unit is composed of a memory cell, which is responsible for maintaining the information extracted from the former cell sequence. Three gates of a cell control the flow of the information among the LSTM units, including the input gate for the information inflow, the output gate for the outflow, and the forget gate for discarding values.

At timestep t, we assume that $f_t \in \mathbb{R}^h$ stands for the forget gate, $i_t \in \mathbb{R}^h$ describes the input gate, $o_t \in \mathbb{R}^h$ represents the output gate, $c_t \in \mathbb{R}^h$ denotes the memory cell in the unit, and $h_t \in \mathbb{R}^h$ stands for the hidden state. The superscript d refers to the number of the input features, i.e., the features of a ray in our problem, and h indicates the dimension of the hidden units. Then, the propagation of values in a LSTM unit contains

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f), \tag{6a}$$
$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i), \tag{6b}$$
$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o), \tag{6c}$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot \sigma_h(W_c x_t + U_c h_{t-1} + b_c), \tag{6d}$$
$$h_t = o_t \cdot \sigma_h(c_t) \tag{6e}$$

where $x_t \in \mathbb{R}^d$ is the input vector to the LSTM unit, and σ_g, σ_h are activation functions serving for the logic of forgetting and remembering in the LSTM gates and defined as

$$\sigma_g(x) = \frac{1}{1+e^{-x}}, \quad \sigma_h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

All $U \in \mathbb{R}^{h\times d}$s and $W \in \mathbb{R}^{h\times h}$s denote weight matrices and $b \in \mathbb{R}^h$s stands for the bias vector, which are learned during the training of the neuron network. At the lower level of our network, we adopt a Bi-LSTM approach [34], which is able to promote the accuracy for the indoor localization, though

applying the bidirectional property increases the complexity of the network. Naturally, in order to extract comprehensive information from the features of a ray, we need to consider it under the context of all rays. However, in classic LSTM, the information extracted from i^{th} ray only depends on the 1^{st} to $(i-1)^{\text{th}}$ rays so that only narrow part of the information can be remembered in LSTM cell. By contrast, with bidirectional structure in Fig. 9, when analyzing the i^{th} ray in the recurrent network, Bi-LSTM cells can remember both the $(i-1)$ rays before it and $(n-i)$ rays after it, where n indicates the number of total rays. Thus, using Bi-LSTM can make the network have a more comprehensive learning over all rays and achieve a better localization accuracy.

The traditional LSTM is forward LSTM, which sequentially accepts the rays in the input matrices from the first arrived ray to the ray with largest delay. By contrast, the backward LSTM accepts these rays in a reverse manner, from the last arrived ray to the ray with smallest delay. In our design, the adopted Bi-LSTM combines a forward LSTM and backward LSTM together. The input of a Bi-LSTM is described as $X = \{x_1, \cdots x_T\} \in \mathbb{R}^{T\times d}$, where T refers to the number of received rays for each receiver. Then, the output of a Bi-LSTM is calculated as

$$O = U_o C_F(X) + W_o C_B(X) + b_o \tag{8}$$

where O denotes the output neuron vector, C_F represents the neuron output for a forward LSTM that accept X in sequence $\{x_1 \cdots x_T\}$, while C_B is the neuron output for the backward LSTM, accepting X in sequence $\{x_T \cdots x_1\}$, and U_o, W_o, b_o are the trainable matrices.

C. Intra-transmitter Attention Mechanism

In addition to the Bi-LSTM applied at the lower level, we further propose an intra-transmitter attention mechanism aimed at enhancing the ability to comprehensively extract the hidden patterns from the ray sequences and feed the upper level with less information loss. The intra-transmitter attention mechanism is inspired by the intra-sentence attention mechanism [37], which was designed for natural language inference between short sentences. When using the bi-directional network, there is little information exchange among the three base networks and all sequences are treated independently. In other words, no information exchange happens before neuron weights forward propagate [34] into the upper level network. Information fusion of the three separate transmitters is only conducted non-sequentially at the upper-level network without sufficient interactions and augmentations. This motivates the utilization of the intra-transmitter attention mechanism, in which the adaptive weights assign different importance to different parts of the sequential inputs. By applying this mechanism, the base Bi-LSTM network can increase the information density and provide better location information for the upper level network. Intuitively, with information exchange at the lower level, three base Bi-LSTM model can generate more orthogonal features and further improve the localization accuracy. Thus, the information from the transmitters can be influenced in both upper- and the lower-level networks. Thanks to this kind of mutual information exchange in the early stage of neural network, each base Bi-LSTM is able to extract the information more accurately.

In our network, an information flow of this intra-transmitter attention mechanism is demonstrated in Fig. 11. We first absorb the sequential output of the three Bi-LSTM network at the lower level, and calculate the soft align attention matrix [36] between each two of them. The calculation of the attention matrix can be expressed as

$$E^{(mn)} = O_m^T O_n, \quad \text{for } m, n = 1, 2, 3 \tag{9}$$

where $O_m \in \mathbb{R}^{l_m \times h}$ and $O_n \in \mathbb{R}^{l_n \times h}$ represent the sequential output of the m^{th} and n^{th} base Bi-LSTM, respectively. $(\cdot)^T$ indicates the transpose operation. Moreover, l_m, l_n indicates the length of the signal sequence from transmitter m and n. In addition, h denotes the number of hidden neurons of the Bi-LSTM. Then, this matrix is used to find the weighted output sequence, given by

$$\widetilde{O}^i_{m(n)} = \sum_{i=1}^{l_m} \frac{\exp\left(e^{mn}_{ij}\right)}{\sum_k \exp\left(e^{mn}_{ik}\right)} O^j_n, \tag{10a}$$

$$\widetilde{O}^i_{n(m)} = \sum_{i=1}^{l_n} \frac{\exp\left(e^{nm}_{ij}\right)}{\sum_k \exp\left(e^{nm}_{ik}\right)} O^j_m \tag{10b}$$

where O^j_m is the j^{th} row of the sequential output $O_m \in \mathbb{R}^{l_m \times h}$. e^{mn}_{ij} indicates the i^{th} row, j^{th} column entry for attention matrix $E^{(mn)}$. The sequence O'_m is generated by concatenating O_m, attention-based $\widetilde{O}_m$ and the data augmentation mechanism. Based on the n^{th} attention, the m^{th} output of the base Bi-LSTM is defined as

$$O'_{m(n)} = [O_m; \widetilde{O}_{m(n)}; O_m - \widetilde{O}_{m(n)}; O_m \cdot \widetilde{O}_m] \tag{11}$$

Finally, the output of each lower level base network is

$$\begin{aligned} O'_1 &= [O'_{1(2)}; O'_{1(3)}] \\ O'_2 &= [O'_{2(2)}; O'_{2(3)}] \\ O'_3 &= [O'_{3(2)}; O'_{3(3)}] \end{aligned} \tag{12}$$

D. ResNet

We move upward to analyze the upper-level network, in which a residual neural network (ResNet) [38] plays the main role, which is essentially an artificial neural network (ANN) that utilizes skipped connections among layers and is able to learn the residual pattern between the prediction and the true values in the current network. Naturally, the deeper the network is, the more accurate model it can achieve. However, the gradient vanishing problem, which refers to the optimization gradients disappearing in the layers close to the input layers, is a common problem for deep neural networks. To address this problem, ResNet is a recently proposed method. In Fig. 8, ResNet in the upper-level network uses short-cuts between layers, which can, at least, form an identical mapping after the training and diminish the impact gradient vanishing problem.

Regarding $H(x)$ as the underlying target mapping, which is usually fitted by a non-residual network, the residual function is $F(x) = H(x) - x$. The weights in the residual blocks are trained to learn $F(x)$, which means they are trained to fill up the difference between the existed patterns and true underlying

patterns, instead of being trained to extract deeper patterns with direct linear transformations. By fitting $F(x)$ instead of $H(x)$, adding residual blocks can at least not worsen the performance of the network. Adopting the residual learning, the building blocks can be formally defined as

$$z_{i+1} = W_i \cdot a_i, z_{i+2} = W_{i+1} \cdot \sigma_r(z_{i+1}), a_{i+2} = \sigma_r(z_{i+2} + a_i) \tag{13}$$

where σ_r is the activation function ReLU, which ensures that the activated neurons are strictly positive and only residual can be learned, and is defined as

$$\sigma_r(x) = \max(0, x) \tag{14}$$

E. Design of the Upper-Level Network

The design of the upper level in our network is specially tailored for the indoor localization problem. Five blocks of ResNet are implemented for residual learning and improvement of accuracy with double layer skip each, while there is no additional skip weights added within the shortcuts.

In Fig. 8, the output of the lower level in network is a combination of neurons output by intra-attention mechanism based on three Bi-LSTM. Assume that $X^{(i)}$ denotes the rays and features from Tx #i, and $O'_i(X^{(1)}; X^{(2)}; X^{(3)}) \in \mathbb{R}^{8*h}$ indicates the attention-based output from the Base Bi-LSTM network #i. The output of the whole lower level network is presented by

$$o_{low} = \begin{bmatrix} O'_1(X^{(1)}; X^{(2)}; X^{(3)}) \\ O'_2(X^{(1)}; X^{(2)}; X^{(3)}) \\ O'_3(X^{(1)}; X^{(2)}; X^{(3)}) \end{bmatrix} \tag{15}$$

Then we can derive the expression of the upper level network as. By invoking the function $ResNet(\cdot)$ that refers to a residual block for residual learning, and incorporating the function $Dense(\cdot)$ that denotes a fully-connected layer used to transform the hidden neurons to the 3D coordinates, the output of the upper level network and equivalently, the final location estimate, $o_{up} \in \mathbb{R}^3$, is given by

$$o_{up} = \sigma_h(Dense^2 \cdot ResNet^5(o_{low})) \tag{16}$$

F. Loss function, Optimization Algorithm and Model Integration

For the problem of indoor localization, our objective is to minimize the distance error between the prediction and true values, so we select the Mean Square Error (MSE) as the loss function in our network, which is described as

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \|y_i - \hat{y_i}\|_2^2, \tag{17}$$

where n indicates the number of samples, y_i and $\hat{y_i}$ refer to the target and estimated values, and $\| \cdot \|_2$ denotes l_2-norm operation.

The optimizer used in our architecture is Adaptive Moment Estimation (Adam) [39], which can be regarded as a combination of Root Mean Square Propagation (RMSprop) and Stochastic Gradient Descent (SGD) with momentum [39]. It is the most popular algorithm recently because of its excellent ability of convergence for the increasing network depth nowadays and its remarkable efficiency comparing with normal SGD methods. The complete set of definition is

$$m_w^{(t+1)} = \beta_1 m_w^{(t)} + (1 - \beta_1) \nabla L^{(t)}, \tag{18a}$$

$$v_w^{(t+1)} = \beta_2 v_w^{(t)} + (1 - \beta_1)(\nabla L^{(t)})^2 \tag{18b}$$

with

$$\hat{m}_w = \frac{m_w^{t+1}}{1 - (\beta_1^{t+1})}, \quad \hat{v}_w = \frac{v_w^{t+1}}{1 - (\beta_2^{t+1})}, \\ w^{t+1} \leftarrow w^{(t)} - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w + \epsilon}} \tag{19}$$

where ϵ is a small scalar to prevent division by zero, β_1 and β_2 are the forgetting factors for first order gradients and the second moments. As we are performing an end-to-end optimization on the whole model, w represents any trainable weights in the lower-level and upper-level network. Note that the squaring and the square rooting operation here are element-wise. The algorithm of the adaptive moment estimation is stated in [1].

Furthermore, in order to achieve a better accuracy, two common techniques are used, i.e., k-fold cross validation and mean bagging. On one hand, the k-fold cross validation separates the training set into 10 slices. For each training process, one non-repetitive slice is reserved for the model evaluation, while the others are used for training. In this case, 10 training processes are conducted and generate 10 different models, from which the best model will be chosen. This technique aims to reduce the estimation bias and prevent over-fitting of the network. On the other hand, by averaging the output of multiple models, the mean bagging contributes to diminish the variance of the localization estimation, which is experimentally found the main obstacle of the proposed method. In order to combine these two techniques time-efficiently, instead of choosing only the best model, we choose the best $n = 5$ models from all models. These models are integrated using mean bagging, given by

$$O_{net} = \frac{o_{net}^{(0)} + o_{net}^{(1)} + \cdots + o_{net}^{(n-1)}}{n}, \tag{20}$$

where $o_{net}^{(i)}$ means the output of the model with i^{th} least MSE error in the validation.

G. Tree-structured Parzen Estimator Approach

Tree-structured Parzen Estimator (TPE) is a hyper-parameter optimizer [40], which optimizes those pre-training customized parameters and achieves the best performance within limited computational resources. Due to the high computation/time cost of one training process, hyper-parameter optimization tasks are required to generate the best parameters with the least number of training. Following the requirement, TPE is a novel method, of which the runtime is confined linearly by the size of observation set and the dimension of the optimizing hyper parameters space. In our framework, the SIABR neural network has many hyper-parameters, including the number of the output nodes for the base BiLSTM network and the number of the ResNet blocks for the residual learning.

In addition, all the learning methods, established for comparison purpose in Section VI, also apply the TPE method for hyper-parameter optimization.

Assume f to be a function modeled with Gaussian process, which takes the high dimensional hyper-parameters as input and the object function result (the mean square error in our validation) as an output. a represents the hyper-parameters in the model and b represents the performance of the model. b^* is the threshold value in observation set $\mathbf{H}$, which is usually chosen as a quantile value, as $p(b < b^*) = \gamma$. However, no specific evaluation for $p(b)$ is required in TPE approach. Similar to the Gaussian optimization approach, TPE uses the Expected Improvement (EI) algorithm to capture a balance between the regions where the mean function is close to or better than b and taking risk searching in the highly uncertain regions.

$$\begin{aligned} EI_{b^*}(a) &= \int_{-\infty}^{b^*} (b^* - b) p(b|a) dy \\ &= \int_{-\infty}^{b^*} (b^* - b) \frac{p(a|b)p(b)}{p(a)} db \end{aligned} \tag{21}$$

In the TPE method, $p(a|b)$ is defined as

$$p(a|y) = \begin{cases} l(a) & \text{if } b < b^* \\ g(a) & \text{if } b \geq b^* \end{cases} \tag{22}$$

where $l(a)$ denotes the estimated density formed with the observations $\{a^{(i)}\}$ satisfying $f(a^{(i)}) < b^*$, and $g(a)$ is the density formed by using the remaining observations. On each iteration, the algorithm returns the hyper-parameter group a^* with maximum value of $EI_{b^*}(u)$. Note that, $EI_{b^*}(a) \propto (\gamma + \frac{g(a)}{l(a)}(1-\gamma))^{-1}$, which is not necessary to be specifically calculated. The optimization logic is formed as the pseudo code in [1].

VI. Performance Evaluation and Analysis

In this section, the performance of the proposed SIABR method is extensively evaluated in the two different indoor environment described in Section IV-A, in contrast with both existing approaches and variations of the LSTM technique. The metrics of interest for comparison cover from convergence, localization accuracy, robustness to obstacles in NLoS environment, effect of limited training data, effect of reduced number of multi-path rays, effect of temporal blockage and effect of estimation errors. For the performance evaluation in the rest of the section, without explicit explanation, the dimensions of the training data sets are $4000 \times 120 \times 5$ for environment 1 and $2800 \times 120 \times 5$ for environment 2, i.e., 4000/2800 receiver locations, 120 multi-path arrival rays for each receiver with zero-padding introduced in Section IV-B, and 5 distinct features associated with individual arrival ray including elevation AoA, azimuth AoA, delay, received power, and index of the anchor transmitter. The experiment setting is listed in Tab. II. Instead of using MSE (17) for convenience of weights' update, the major performance metric of the experiments is the mean distance error (MDE), which is more realistic and defined as

TABLE II: Software and hardware setting for the experiments.

GPU	1 NVIDIA GEFORCE RTX 2080 Ti
CPU	9th Generation Intel Core i7 Processors
Memory size	64G
OS	CentOS 8
Python version	Python 3.7
Environment simulation	Wireless Insite
Deep learning framework	Tensorflow 1.15
CUDA/cuDNN version	CUDA 10.2 and cuDNN 7.6

Figure 12: The convergence evaluation of the SIABR model.

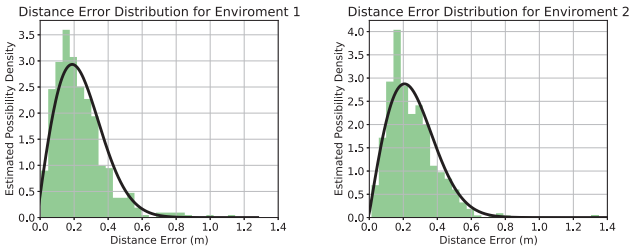
$$MDE = \frac{1}{n} \sum_{i=1}^{n} \|y_i - \hat{y}_i\|_2 \tag{23}$$

where y_i is the true location of a reference point and $\hat{y}_i$ indicates the estimated location.

Eight different localization techniques from the literature and intermediate products are selected for performance comparison with our proposed SIABR method, listed as triangulation, K-nearest neighbor (KNN) [41], auto-encoder [42], convolutional neural network (CNN), LSTM [35] with none of structuring, ResNet and bidirectional strategies, structured LSTM without bidirectionality, singular SBi-LSTM without model integration, and SBi-LSTM [1].

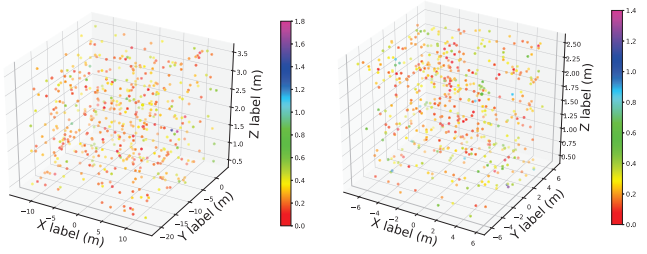
A. Convergence Evaluation of SIABR

Convergence and its speed are critical evaluation metrics for machine learning algorithms. The convergence of the proposed SIABR method in the training phase is demonstrated in Fig. 12. During the training phase of the network, evaluation sets are used to prevent over-fitting. In both indoor environments, the mean square error of the loss for the training data and evaluation data are plotted respectively. The loss follows an exponential decay in the log scale. The convergence proceeds with the number of epochs, where each epoch refers to one iteration of the training cycle. After 20 epochs, the loss reduces by at least two orders of magnitude, and hence, fast convergence of our proposed SIABR method is verified in both indoor environments.



(a) Environment 1 (b) Environment 2

Figure 13: Estimated probability distribution of localization errors.



(a) Environment 1 (b) Environment 2

Figure 14: The heat map for localization errors in both environments by using the proposed SIABR network.

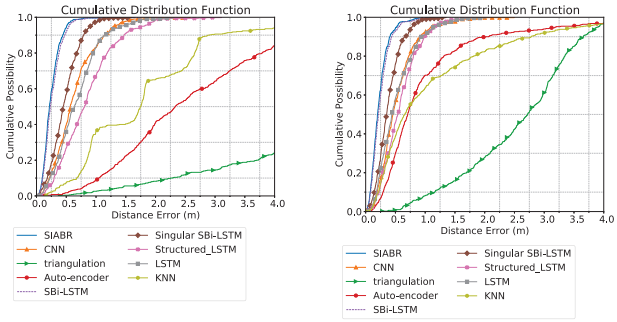
B. Localization Accuracy

The localization accuracy is evaluated based on the estimation error of the 3D distance. Fig. 13 shows the histogram and the fitted Rayleigh Distribution for distance errors, which is estimated by the proposed SIABR method. In environment 1, the mean distance error is 0.2431m and the standard deviation is 0.1703m. By contrast, the mean and standard deviation of the distance error are, respectively, 0.2312m and 0.1481m in environment 2. The proposed method achieves the localization accuracy under 0.25m within both environments, implying that the SIABR network is well applicable for localization in a simple, symmetrical hall as well as in a decorative room. The 3D heat map for the distance errors of location estimation in the two environments are shown in Fig. 14. The max distance error 1.7m in environment 1 and 1.3m in environment 2. Very few points separated in different corners have extremely large distance error. The proposed method has full coverage over both environments. According to our observation, while environment 1 is larger than environment 2, the two environments have negligible difference in terms of the mean distance error for localization accuracy. Therefore, the size of the indoor space mildly influences the performance of the model.

Next, eight different localization methods from the literature are compared with our proposed SIABR. For fair comparison, these methods are invoked to process the same ray tracing data under the two indoor environments. Table III summarizes the mean distance errors and standard deviations of different methods, while Fig. 15 shows the cumulative density function of the location estimation results of the methods. The proposed SIABR model has the best performance under the two indoor environments, in terms of the distance errors. From the state of art CNN deep learning method [9], SIABR improved about 70% of location estimation accuracy. The localization

TABLE III: The mean distance errors and its standard deviations for two environments. Unit in meter.

Model Structure	Env. 1 (mean, std.)	Env. 2 (mean, std.)
SIABR	**(0.2431, 0.1703)**	**(0.2312, 0.1481)**
SBi-LSTM	(0.2695, 0.1671)	(0.2480, 0.1645)
Singular SBi-LSTM	(0.4754, 0.2432)	(0.3814, 0.2091)
Triangulation	(5.5178, 2.9556)	(2.5524, 0.9315)
Auto-encoder	(3.3709, 2.6286)	(1.0420, 1.1024)
LSTM	(0.8465, 0.4749)	(0.5748, 0.3110)
Structured LSTM	(0.6851, 0.3880)	(0.5190, 0.3300)
KNN	(1.9620, 1.8169)	(1.1281, 1.1750)
CNN	(0.6350, 0.3620)	(0.5252, 0.3187)



(a) Environment 1 (b) Environment 2

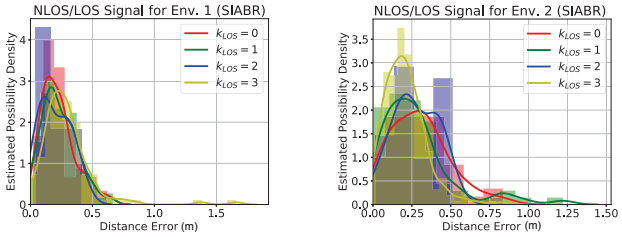
Figure 15: The CDF curve for distance errors in both environments with respect to different methods.

accuracy improvement of our network mainly contributes from the combination of residual learning in the upper level, the attention mechanism in the lower level and bidirectionality applied on the LSTM, as well as the model integration. By applying structuring and ResNet in the network, the accuracy of the structured LSTM improves about 15% from the classic LSTM. Furthermore, the accuracy of the Singular SBi-LSTM improves about 32% from the structured LSTM, by using bidirectionality and residual learning. Implementing model integration technique, the localization accuracy of SBi-LSTM method improves about 40% from its singular form. Finally, novel intra-attention mechanism introduced in this paper improves about 7% from our previous work by treating the input sequences with adaptive weights.

C. Robustness to Static Blockage

In real environments, varieties of blockages may have effect on the accuracy of indoor localization, including static blockage and dynamic blockage. In this section, the evaluation focuses on the static blockage, which should be captured by the SIABR during the offline training phase, such as pillars in environment 1 and furniture in environment 2. By observing distributions of mean distance error over different k_{LoS} values, We conduct the evaluation of the robustness to static blockage and study how the multi-path effect caused by static obstacles and complicated polyhedral surfaces influences the accuracy of indoor localization.

We introduce a LoS factor k_{LoS} in Section IV-A to distinguish different receiver points with different number of LoS rays due to the varying relative positions to blockages. In both environments, for the four groups of receivers with k_{LoS} equal



(a) Environment 1 (b) Environment 2

Figure 16: The estimated kernel density function for LoS/NLoS signal using SIABR method.

TABLE IV: The mean distance errors and their standard deviations with different k_{LoS} values. Unit in meter.

LoS factor	Env. 1 (mean, std.)	Env. 2 (mean, std.)
$k_{LoS} = 0$	(0.2260, 0.1248)	(0.2971, 0.1735)
$k_{LoS} = 1$	(0.2395, 0.1309)	(0.2603, 0.2045)
$k_{LoS} = 2$	(0.1981, 0.08827)	(0.2628, 0.1280)
$k_{LoS} = 3$	(0.2929, 0.1885)	(0.2126, 0.1315)

to 0, 1, 2, 3, we carry out their localization errors, described as the mean and standard deviation of corresponding probability density distribution in Table IV. Fig. 16 describes the distance error distributions with respect to different k_{LoS} values, in order to evaluate how the NLoS property impacts the accuracy of the location estimation.

In environment 1, there is no clear trend of change for the localization precision as k_{LoS} increases, i.e., more LoS signals from the transmitters for the receiver. This suggests that as a result of the simplicity and symmetrical property of the indoor structure in environment 1, the performance of our network is not impacted by the number of LoS rays with a specific trend. By contrast in environment 2, because of the complexity of the indoor structure, our SIABR performs worse in prediction accuracy when k_{LoS} decreases. We learn that the mean error of localization distance increases to 0.2971m as k_{LoS} reduces from 3 to 0 with an addition of 0.0845m. By comparison, the increments are 2.42m for triangulation, 0.716m for auto-encoder, 1.021m for KNN, 0.35m for LSTM, 0.33m for Structured LSTM 0.1063m for SBiLSTM and 0.23m for CNN, respectively. It indicates that our SIABR ensures strong robustness of the localization while considering the effect of NLoS phenomena.

D. Effect of the Dynamic Blockage

Before applying the proposed method in the real-world environments, one of the main obstacles that could significantly influence the final performance is the existence of dynamic blockage. In the practical indoor scenarios, moving human bodies or other objects can result in such dynamic blockages on some of the THz rays from the transmitters with a large blocking probability. This kind of blockage is able to introduce more severe attenuation, even fully-shield, on THz signals than on normal WiFi signals. Since the traditional triangulation method strongly depends on the accurate RSSI input, significant signal attenuation caused by the dynamic blockage may have a much worse influence on the accuracy of the indoor localization with this method. Moreover, according to [2], common fingerprinting methods are vulnerable to the slight environment change like dynamic blockage.

TABLE V: The mean distance errors with different ray blockage strategies in both environments. Unit in meter.

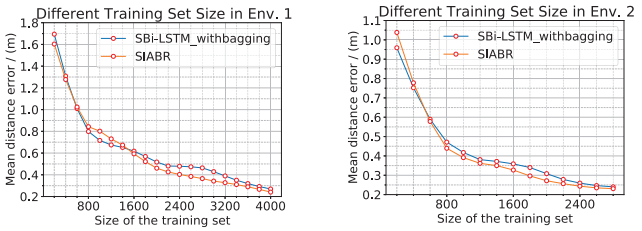
Blocked Rays	SIABR/SBiLSTM Env. 1	SIABR/SBiLSTM Env. 2
original	0.2431/0.2695	0.2312/0.2480
Rand. chosen 1	0.2451/0.2711	0.2343/0.2502
Rand. chosen 2	0.2646/0.2779	0.2398/0.2538
Rand. chosen 3	0.2759/0.2802	0.2450/0.2577
Rand. chosen 4	0.2794/0.2834	0.2582/0.2612
Rand. chosen 5	0.2830/0.2924	0.2602/0.2678
Top 1	0.3568/0.3423	0.3653/0.3019
Top 2	0.3150/0.4982	0.3425/0.4910
Top 3	0.5230/0.5458	1.2468/0.8169
Top 4	0.8420/0.7160	0.8614/1.0989
Top 5	1.2523/0.9153	0.9348/1.1783

As a result, we investigate the impact of the dynamic blockage on our proposed model. To mimic the dynamic blockage in the simulated environments, several rays of each sample are blocked based on the following two strategies, respectively. In the first situation, we randomly block n ray(s) from all three transmitters, corresponding to the first half of Table V, which is mimicking the people walking around the open area in the room. In the second situation, we randomly select one transmitter and block its first n ray(s) with the largest power, i.e., the top n power in the top power in Table V. This strategy reproduces the channel information in the room, when dynamic blockage is right in front of one certain transmitter.

Different from the static LoS blockage that is mentioned in Section IV, the dynamic blockage explored in this section is unpredictable and varies over time. In general, the LoS blockage is naturally embedded in the indoor environments and supposed to be captured by the proposed model during the training phase. However, the dynamic blockage is not captured during the offline training phase and unpredictable for the model. Therefore, it is a significant test on the robustness performance of the proposed model to introduce dynamic blockage experiments.

In addition, the difference between this subsection and Section VI-F is clarified here. During the experiments, both sections are trying to evaluate the performance of the proposed model by masking some rays, however, there are disparities between the logic behind them. In Section VI-F, rays are masked due to the large distinctions between the power of THz rays, while in this section rays are masked due to the temporal moving objects in the indoor environment.

From Table. V, it is observed that the proposed model is more sensitive towards the first strategy, intuitively implying that the localization accuracy of the SIABR strongly depends on the top-power rays from the transmitters. Though, drop of model performance during online testing phase is caused by dynamic blockage, the mean distance errors of both SIABR and SBiLSTM stay in a reasonable range in most cases. Nevertheless, to prevent top power of a transmitter rays being blocked in practical application, the positioning of the transmitter is critical. Locating transmitters at a open area with less probability of being blocked by moving objects could improve the online localization accuracy and reduce the risk of



(a) Environment 1 (b) Environment 2

Figure 17: Examination of the impact of limited dataset on accuracy. Here the performance indicator (y-axis) is the mean distance errors of the localization estimation on the test set.

malfunction for the proposed model. Furthermore, comparing with the SBiLSTM [1], there is no evidence that the robustness against temporal blockage of the SIABR model is worse.

E. Effect of Limited Training Data

For a real-life indoor localization problem, one main obstacle that the deep learning methods need to face is the limitation of the dataset. The exhaustive computation required to construct a 2800/4000 fingerprinting training set of the paper is idealized. In order to evaluate the proposed model under realistic scenarios, we test the network with different reduced sizes of the training set, i.e., different numbers of reference receiver points, in Fig. 17. The dataset size ranges from 100% to 10% of the original size, which causes the mean distance error to vary from 0.25m to 1.7m for environment 1 and 0.23m to 1m for environment 2. The results show that the proposed model is robust to the reduction of dataset size, which demonstrates feasibility in the realistic localization scenario with the limited dataset. Our proposed method, provided with limited number (i.e., less than 1000) of reference receiver points, still maintains effective for indoor localization.

The mean distance errors of the proposed model are within 1.85m in environment 1 and 1.10m in environment 2 for all sizes of dataset. Assigning proper hyper-parameters, the proposed network do not lose its ability to converge even with only 200 reference receiver points. The performance of the SIABR network improves steadily with the increase of the number of reference receiver points. Moreover, with the increasing of data in the the training set, there is no drastic drop and rise, implying that the model does not have sudden turn-on requirement of receiver numbers for these two environments. Therefore, we believe there is solid ground for us to conclude that the performance of our network is robust with limited dataset.

F. Effect of THz Channel Sparsity

Compared to the electromagnetic waves at the UWB and mmWave band, THz signals, used for the indoor localization, experience molecular absorption loss and reflection loss in addition to much higher spreading loss. Thus, there may be obvious distinctions between rays' power level of different paths absorbed by a receiver. In this case, only several rays with the strongest power for the receiver play major roles in predicting the location. Moreover, in real application, not all rays simulated by the ray tracing method can be detected,

TABLE VI: The mean distance errors with different reduced numbers of THz multi-path rays. Unit in meter.

Number of Reduced Rays	Env. 1	Env. 2
0 Lowest	0.2695	0.2480
10 Lowest	0.2933	0.2650
20 Lowest	0.2256	0.2545

which makes it interesting to investigate the effect of ray reduction for channel sparsity on THz localization. We delete 10/20 rays with lowest power among each transmitter-receiver pair before the zero-padding processing, aiming to simulate the impact of the ray reduction on the training data. These peculiarities of the THz channel are exploited in the proposed algorithms. The intra-attention mechanism, upper-level residual learning and mean-bagging technique are specially designed to solve the above problem, i.e. to minimize the impact caused by the channel sparsity effect.

The relation between the number of deleted rays and the accuracy of the model in both environments is demonstrated in Table VI. From test results, we learn that there are fluctuations on mean distance errors by deleting 10, or even 20 rays with lowest power. However, the mean distance errors still stay within 0.3m for both environments. The experimental results suggest that this model depends strongly on the first dozens of rays that each receiver receives to determine its location. Essentially, this phenomenon assures that the proposed method only acquires advantage from the high resolution in the THz band with ultra-broad bandwidth, while maintaining robust to the channel sparsity and the reduced number of multi-path.

G. Robustness to Estimation Error

Based on the ray-tracing techniques, our proposed model is trained and tested with idealized inputs from the simulated environments. Before applying the proposed method to practical applications, another part of the main obstacles that could significantly affect the final performances is the estimation error of the CSI features. In the real-world scenarios, the error mainly comes from the aforementioned inaccurate CSI estimation due to the thermal noise as well as the slight environmental changes, such as the shake of the mobile devices. Thanks to the ultra high bandwidth and frequency of THz signal, THz signals suffer less interference from background electromagnetic wave than UWB and mmWave band. However, the state-of-art highly accurate CSI estimation achieves up to 3.457% average difference ratio (ADR) [43] and still has a long way to go. Thus, it is necessary for us to provide robust indoor localization performance with noises in the simulated input.

Here, we apply zero-mean Gaussian distribution based random noise to the input based on the state-of-art CSI estimation [43], including power, time of arrival and angle of arrival. In Table VII, we evaluate the mean distance errors of our model based on different variance of Gaussian noise in both environments. The first three columns correspond to the variance of AoA, ToA and power. Moreover, we can observe that the estimation mean distance error of the proposed model apparently rises with increasing variance of the noise. However, all mean distance errors maintain within

TABLE VII: The mean distance errors with different ratio of Gaussian noise in both environments. Units in degree (AoA), second (ToA), dB (Power), and meter (mean distance errors).

AoA	ToA	Power	SIABIR Env. 1	SIABIR Env. 2
0	0	0	0.2431	0.2312
0.1	1e-9	0.1	0.2785	0.2444
0.2	2e-9	0.2	0.3011	0.2980
0.4	4e-9	0.4	0.3255	0.2745
0.8	8e-9	0.8	0.3356	0.3025
1.6	16e-9	1.6	0.4125	0.3508

a reasonable range, e.g., 0.5m. The performance drop reacts insensitively to reasonable amount noise. No malfunction of the proposed model is shown in the experiments. Therefore, the results suggest that the proposed SIABR model is not vulnerable towards the slight noises on the inputs and is robust to the inevitable errors during the THz CSI estimation. When the error of the estimated AoA, ToA and power are less than 0.1 degree, 1e-9, 0.1 dB, the localization accuracy variation is less than 5 cm.

H. Time Complexity Analysis

Besides the accuracy of the model, another key feature of an indoor localization method, is the time complexity in both training and testing phases. In the training phase, single SIABR model needs approximately 2 hours to converge in both environments, according to the experimental setting in Tab. II. When utilizing the mean-bagging technique in five SIABR models to improve accuracy, it takes approximately 10 hours to complete training. By contrast, a single CNN model [9] used for this localization problem consumes about 30 minutes for training, which corresponds to 2.5 hours to train 5 CNN models with mean begging. Furthermore, the SBiLSTM model in our preliminary work [1] consumes about one hour to train, and takes around 5 hours by applying mean bagging techniques on 5 SBiLSTM models.

In the testing phase, our proposed model takes less than 150 ms to generate a batch of 500 receivers' location with three sequences of input information from the transmitters, without any network pruning or compression. For comparison, the CNN model [9] and the SBiLSTM model [1] consume around 130 ms to generate the output location. Therefore, the SIABR improved accuracy comes at the cost of four times longer training complexity, and 15% increased testing time, which is acceptable in practical applications, since the rise in testing time is still in the same order of magnitude.

VII. Conclusion

In this paper, we have proposed a novel SIABR deep learning method for 3D THz indoor localization. The unprecedented high-resolution THz CSI, including 3D AoA, received power, and delay, are extracted and exploited. The two-level architecture is designed based on the SIABR method, in which the lower level extracts the THz NLoS ray features with the intra-attention mechanism, while the upper level integrates five blocks of ResNet to generate the 3D location of targeted receivers. Extensive simulation results have validated the accuracy, convergence and robustness against indoor NLoS environment. The mean distance error is within 0.25m, which is competitive compared to the state-of-the-art techniques, by considering the 3D indoor propagation with multi-path fading and dynamic blockage effects. Future work includes the adoption of THz real-world measurement data instead of simulation data, to evaluate the proposed SIABR deep learning method for THz indoor localization.

References

[1] S. Fan *et al.*, "Structured bidirectional lstm deep learning method for 3d terahertz indoor localization," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, July 2020.

[2] F. Zafari *et al.*, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 2019.

[3] I. F. Akyildiz *et al.*, "Terahertz band: Next frontier for wireless communications," *Elsevier Physical Communication*, vol. 12, pp. 16 – 32, Sept. 2014.

[4] I. F. Akyildiz *et al.*, "Combating the Distance Problem in the Millimeter Wave and Terahertz Frequency Bands," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 102–108, June 2018.

[5] X. Wang *et al.*, "Csi-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan 2017.

[6] M. Youssef *et al.*, "The Horus WLAN Location Determination System," in *Proc. of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, Seattle, Washington, June 2005.

[7] K. Wu *et al.*, "Csi-based indoor localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1300–1309, 2013.

[8] M. Vari *et al.*, "mmWaves RSSI indoor network localization," in *Proc. of IEEE ICC Workshop on Advances in Network Localization and Navigation*, June 2014.

[9] X. Wang *et al.*, "Deep convolutional neural networks for indoor localization with csi images," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 316–327, 2020.

[10] V. Otsason *et al.*, "Accurate GSM indoor localization," in *Proc. of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2005.

[11] H. Wymeersch *et al.*, "A Machine Learning Approach to Ranging Error Mitigation for UWB Localization," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1719–1728, June 2012.

[12] P. Bahl *et al.*, "RADAR: An in-building RF-based user location and tracking system," in *Proc. of IEEE International Conference on Computer & Communications (INFOCOM)*, 2000.

[13] C. Han *et al.*, "Multi-Ray Channel Modeling and Wideband Characterization for Wireless Communications in the Terahertz Band," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2402–2412, May 2015.

[14] J. M. Jornet *et al.*, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, 2011.

[15] C. Han *et al.*, "Propagation modeling for wireless communications in the terahertz band," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 96–101, June 2018.

[16] S. Sen *et al.*, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, June 2013.

[17] J. Xiong *et al.*, "Arraytrack: A fine-grained indoor location system," in *Proc. of the 10th USENIX Conference on Networked Systems Design and Implementation*, USA, 2013.

[18] M. Kotaru *et al.*, "Spotfi: Decimeter level localization using wifi," in *Proc. of the 2015 ACM Conference on Special Interest Group on Data Communication*, 2015.

[19] X. Wang *et al.*, "BiLoc: Bi-Modal Deep Learning for Indoor Localization With Commodity 5GHz WiFi," *IEEE Access*, vol. 5, no. 99, pp. 4209–4220, 2017.

[20] L. M. Ni *et al.*, "Landmarc: indoor location sensing using active rfid," in *Proc. of the First IEEE International Conference on Pervasive Computing and Communications (PerCom), 2003.*, 2003, pp. 407–415.

[21] A. Yassin *et al.*, "Mosaic: Simultaneous localization and environment mapping using mmwave without a-priori knowledge," *IEEE Access*, vol. 6, pp. 68 932–68 947, 2018.

[22] F. C. Commission, "FCC takes steps to open spectrum horizons for new services and technologies." [Online]. Available: https://docs.fcc.gov/public/attachments/DOC-356588A1.pdf

[23] H. Sarieddeen *et al.*, "Next generation terahertz communications: A rendezvous of sensing, imaging, and localization," *IEEE Communications Magazine*, vol. 58, no. 5, pp. 69–75, 2020.

[24] C. Studer *et al.*, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[25] A. Goswami *et al.*, "Wigem: A learning-based approach for indoor localization," in *Proc. of the Seventh Conference on emerging Networking Experiments and Technologies*, 2011.

[26] K. V. Mishra *et al.*, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 100–114, 2019.

[27] C. Sturm *et al.*, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proc. IEEE*, vol. 99, no. 7, pp. 1236–1259, July 2011.

[28] H. Yuan *et al.*, "Hybrid beamforming for terahertz multi-carrier systems over frequency selective fading," *IEEE Transactions on Communications*, 2020.

[29] S. M. Perera *et al.*, "Radix-2 self-recursive sparse factorizations of delay vandermonde matrices for wideband multi-beam antenna arrays," *IEEE Access*, vol. 8, pp. 25 498–25 508, 2020.

[30] D. Rife *et al.*, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 591–598, 1974.

[31] L. Yan *et al.*, "A dynamic array-of-subarrays architecture and hybrid precoding algorithms for terahertz wireless communications," *IEEE J. Sel. Areas Commun.*, 2020.

[32] Y. Chen *et al.*, "Millidegree-level direction-of-arrival (doa) estimation and tracking for terahertz wireless communications," in *Proc. of IEEE International Conference on Sensing, Communication and Networking (SECON)*, June 2020.

[33] Y.-H. Chen *et al.*, "Deep cnn-based spherical-wave channel estimation for terahertz ultra-massive mimo systems," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, 2020.

[34] M. Schuster *et al.*, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, pp. 2673–2681, 1997.

[35] S. Hochreiter *et al.*, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, Nov 1997.

[36] D. Bahdanau *et al.*, "Neural machine translation by jointly learning to align and translate," 2014.

[37] Q. Chen, X. Zhu, Z. Ling, S. Wei, and H. Jiang, "Enhancing and combining sequential and tree LSTM for natural language inference," 2016. [Online]. Available: http://arxiv.org/abs/1609.06038

[38] K. He *et al.*, "Deep Residual Learning for Image Recognition," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[39] D. P. Kingma *et al.*, "Adam: A Method for Stochastic Optimization," 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[40] J. S. Bergstra *et al.*, "Algorithms for Hyper-Parameter Optimization," in *Advances in Neural Information Processing Systems 24*, 2011, pp. 2546–2554.

[41] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[42] C.-Y. Liou *et al.*, "Autoencoder for words," *Neurocomputing*, vol. 139, pp. 84–96, 2014.

[43] C. Luo *et al.*, "Channel state information prediction for 5g wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, 2018.

Shukai Fan is a research student assistant at the Terahertz Wireless Communication Laboratory, UM-SJTU Joint Institute, Shanghai Jiao Tong University since 2018 and received his B.E degree in electrical and computer engineering August 2020. Meanwhile, he is also pursuing an M.S. degree in data science at the University of Michigan, Ann Arbor. His current research interests include indoor localization in the millimeter-wave and Terahertz band with deep learning methods.

Yongzhi Wu (S'19) received B.E degree in Electronic and Information Engineering from Huazhong University of Science and Technology in 2019. Since 2019, he is pursuing Ph.D. degree in the Terahertz Wireless Communication Laboratory, Shanghai Jiao Tong University. His research interests include Terahertz communications, joint communications and sensing.

Chong Han (M'16) received Ph.D. degree in Electrical and Computer Engineering from Georgia Institute of Technology, USA in 2016. He is currently a tenure-track Associate Professor with the Terahertz Wireless Communications (TWC) Laboratory, University of Michigan–Shanghai Jiao Tong University (UM-SJTU) Joint Institute, Shanghai Jiao Tong University, China. He is the recipient of 2018 Elsevier NanoComNet (Nano Communication Network Journal) Young Investigator Award, 2017 Shanghai Sailing Program 2017, and 2018 Shanghai ChenGuang Program. He is an editor with IEEE Open Journal of Vehicular Technology since 2020, an associate editor with IEEE Access since 2017, an editor with Elsevier Nano Communication Network journal since 2016, and is a TPC chair to organize multiple IEEE and ACM conferences and workshops. His research interests include Terahertz band and millimeter-wave communication networks, and electromagnetic nanonetworks. He is a member of the IEEE and ACM.

Xudong Wang (F'18) is a tenured professor with the UM-SJTU Joint Institute, Shanghai Jiao Tong University. He is also an affiliate faculty member with the Electrical Engineering Department at the University of Washington. Since he received the Ph.D. degree in Electrical and Computer Engineering from Georgia Institute of Technology in 2003, Dr. Wang has been working as a senior research engineer, senior network architect, and R&D manager in several companies. He has been actively involved in R&D, technology transfer, and commercialization of various wireless networking technologies. He holds a number of patents on wireless networking technologies and most of his inventions have been successfully transferred to products. Dr. Wang is an editor for IEEE Transactions on Mobile Computing, IEEE Transactions on Vehicular Technology, Elsevier Ad Hoc Networks, and China Communications. He was also a guest editor for several international journals. He was a general chair of 2017 IEEE 5G Summit in Shanghai and a TPC Co-Chair of the 32nd International Conference on Information Networking. He was the demo co-chair of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MOBIHOC 2006), a technical program co-chair of Wireless Internet Conference (WICON) 2007, and a general co-chair of WICON 2008. Dr. Wang is an IEEE Fellow (Communications Society) and was a voting member of IEEE 802.11 and 802.15 Standard Committees. His research interests include wireless communication networks (5G and beyond), distributed machine learning, and joint communications and sensing.