IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 September 2022, accepted 14 October 2022, date of publication 19 October 2022, date of current version 26 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3215504

RESEARCH ARTICLE

TIPS: Transformer Based Indoor Positioning System Using Both CSI and DoA of WiFi Signal

ZHONGFENG ZHANG, HONGXIN DU, SEUNGWON CHOI, (Member, IEEE), AND SUNG HO CHO, (Member, IEEE)

Department of Electronic Engineering, Hanyang University, Seoul 04763, South Korea

Corresponding author: Seungwon Choi (choi@dsplab.hanyang.ac.kr)

This work was supported by the National Research Foundation (NRF), South Korea, under Grant NRF-2022R1A2C2008783.

ABSTRACT In a channel state information (CSI) based indoor positioning system, the positioning performance becomes susceptible to multipath fading effects especially in non-line-of-sight environments. We propose a transformer-based indoor positioning system (TIPS) to address this challenge. The proposed TIPS utilizes a self-attention mechanism to process the continuous WiFi CSI observed from predetermined routes as fingerprints in a given indoor environment. Each route is then considered a sentence, whereas the position along the route is treated as a word in terms of natural language processing. Consequently, the problem of predicting the position with the fingerprints can then be considered the task of predicting the current word with previous words, which can be efficiently solved using the proposed TIPS. In order to fully exploit the relations among positions, we propose embedding the information of the direction of arrival (DoA) on top of the collected CSI as inputs to the TIPS. Thus, the transformer of the proposed TIPS can better capture the dependencies of the positions in the route and significantly boost positioning accuracy. To exhibit the superiority of the proposed TIPS in a radio frequency (RF) environment, we demonstrate a hardware implementation of an RF testbed consisting of an emulator of WiFi access point and user equipment. Through extensive computer simulations and experimental tests, it is demonstrated that the proposed TIPS can reduce the positioning error down to 20 cm, which is a significant improvement compared to the current state-of-the-art models.

INDEX TERMS CSI, DoA, indoor positioning system, transformer.

I. INTRODUCTION

The accurate positioning of mobile user equipment (UE) is of great help in providing location-related services such as navigation, virtual reality, and motion tracking [1], [2], [3]. It is also beneficial to smart buildings [4], the Internet of Things [5], and machine type communication [6]. A global positioning system (GPS) enables accurate localization in the outdoor environment, whereas the GPS signals degrade significantly in an indoor environment [7], resulting in poor indoor positioning accuracy. Therefore, the GPS is not applicable in the robot or drone navigation cases, which require a high degree of centimeter-level positioning precision. The prevalent indoor positioning technologies include ultra-wideband, Bluetooth, ZigBee, WiFi, geomagnetic [8], [9], [10]. Among these technologies, WiFi has been extensively researched over the last few decades owing to its widespread deployment and low cost. As a state-of-the-art technology, WiFi fingerprinting indoor localization systems (IPS) have been largely studied for both localization [11], [12], [13], [14], tracking [15], and activity recognition [16], [17] applications. The radio frequency (RF) characteristics of WiFi signals at each location are unique due to their different propagation paths, and the RF characteristics can be considered unique fingerprints. With all the fingerprints of locations collected and stored in the database

The associate editor coordinating the review of this manuscript and approving it for publication was Zihuai Lin.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

beforehand, accurate localization can be achieved by comparing the received WiFi signal with the data in the database. The construction of a WiFi-based IPS comprises two phases. One is the offline phase, during which the fingerprints (RF characteristics) of the WiFi signal at the reference points (RP) are collected to construct the database for the IPS. The other is the online phase, during which the fingerprints available in the database are compared with the new fingerprints of the WiFi signal received at the test positions.

The fingerprinting-based IPS requires a site survey process where a radio map is created by measuring the signal characteristics of the predetermined positions in the indoor environment. By comparing the signal received at a RP and the one stored in the radio map, the position with the fingerprint that has the highest correlation with the fingerprint of the test position is determined as the predicted position of the UE. A WiFi-based IPS can employ either the received system strength indicator (RSSI) or channel state information (CSI) as fingerprints for reference positions. The RSSI is a single value representing the received power level, which can fluctuate significantly in both line-of-sight (LOS) and non-line-of-sight (NLOS) environments [18]. By contrast, CSI provides fine-grained signal information provided by multiple subcarriers (SCs) in orthogonal frequency-division multiplexing (OFDM) symbols available in WiFi signals. In the presence of the multipath effect, CSI is more stable than RSSI [19], [20], [21]. Consequently, CSI-based IPSs have gained momentum in recent years.

Recently, deep learning (DL) based algorithms are gaining a keen interest in the application of IPS. Some related works [22], [23], [24], [25], [26] have shown that the DL algorithms such as dense neural network (DNN), convolutional neural network (CNN), long short-term memory (LSTM), etc. can help the IPS better capture the correlation between the fingerprint and corresponding position, which consequently enhances the accuracy compared to the conventional methods [27], [28]. Although the DL algorithms commonly used in IPS can reduce the positioning error down to 1-2 meter, it is still challenging for the IPS to reach centimeter level accuracy considering many adverse indoor signal environments involving the signal instabilities caused by multipath fading effects. In this paper, we claim that the centimeter level accuracy can be achieved by adopting direction of arrival (DoA) as well as CSI as the fingerprints of the radio map because the DoA of the WiFi signal has been proven to be stable in [26]. In order to integrate the two heterogeneous fingerprints, DoA and CSI, we adopt the transformer neural network [29], which converts the indoor positioning task into the problem of language processing.

The main contributions of this paper are as follows.

- We propose a transformer-based IPS with an extremely high positioning accuracy. To the best of our knowledge, this paper is the first to apply the self-attention mechanism provided by a transformer to the indoor positioning problem utilizing the WiFi signals.
- We propose utilizing both the CSI extracted and the DoA estimated from the received WiFi signal by embedding them together during the preprocessing stage. Then, the embedded data are fed to the proposed transformer based indoor positioning system (TIPS) to learn the relationship between the positions and their corresponding CSI & DoA.
- We demonstrate the superiority of the proposed TIPS through extensive computer simulations, utilizing the ray-tracing technique to simulate an indoor environment with rich multipath propagation. Then, we evaluate and analyze the performance of the TIPS trained on different types of datasets, i.e., CSI-only dataset, DoA-only dataset, and CSI & DoA dataset, as well as the impact of the input batch size.
- We implemented an RF testbed using multiple universal software radio peripherals (USRPs) and a central processing unit (CPU) to emulate a single access point (AP) and UE. The RF experiments were conducted to verify the high accuracy of the proposed TIPS by comparing it with state-of-the-art solutions.

The remainder of this paper is organized as follows. Section II describes the related works on the fingerprinting-based IPS. Section III briefly explains the CSI and DoA

TABLE 1. Acronyms and corresponding explanations.

Acronym	Explanation
AP	Access point
CNN	Convolutional neural network
CPU	Central processing unit
CSI	Channel state information
DCNN	Dimensional deep convolutional neural network
DL	Deep learning
DNN	Dense neural network
DoA	Direction of arrival
EVD	Eigenvalue decomposition
GPS	Global positioning system
GPT	Generative pretrained transformer
GPU	Graphics processing unit
IPS	Indoor positioning system
KNN	K-nearest neighbors
LOS	Line of sight
LSTM	Long short-term memory
MUSIC	Multiple signal classification
NLOS	Non line of sight
NLP	Natural language processing
OFDM	Orthogonal frequency-division multiplexing
RF	Radio frequency
RNN	Recurrent neural network
RP	Reference point
RSSI	Received system strength indicator
Rx	Receive
SC	Subcarrier
SCM	Sample covariance matrix
SDR	Software-defined radio
SNR	Signal-to-noise ratio
SVD	Singular value decomposition
TIPS	Transformer-based indoor positioning system
Tx	Transmit
UE	User equipment
UHD	USRP hardware driver
ULA	Uniform linear array
VRP	Virtual reference point
USRP	Universal software radio peripheral

that are utilized as fingerprints. Section IV introduces the proposed TIPS and details the fingerprinting technique. Section V presents the transformer model adopted for TIPS and describes the CSI and DoA preprocessing processes. Section VI explains the detailed training process and demonstrates the performance of TIPS under various scenarios through computer simulations. Section VII shows the experimental results with RF signals and verifies the superior performance of TIPS by comparing it with other positioning solutions. Finally, Section VIII concludes the paper.

II. RELATED WORKS

This sections describes the previous works on the fingerprinting-based IPS using WiFi signal. The fingerprinting-based IPS employs matching algorithms to estimate the position of the target. Prior to applying the algorithms, a site survey process is carried out. During the site survey process, a radio map consisting of the signal characteristics of WiFi signal of each predetermined locations is created.

In [28], the authors adopt an fingerprinting-based IPS utilizing a probabilistic algorithm to model the RSSI received from an AP as random variable over time and space. With a radio map consisting the information of the RSSI at different locations, the Horus system estimates the location of the UE by calculating the probability distribution of the measured RSSI. However, the performance is limited by the instability of the RSSI due to multipath effects in indoor environment. In [27], the authors calculate the time-reversal resonance strength and Euclidean distance between the target location and the RP using multidimensional scaling analysis. Then, an optimized kNN algorithm is used to predict the locations. In [30], a passive radio map is utilized to address the localization problem by matching the CSI anomalies to the fingerprint database via a probabilistic algorithm. In [22], Wang et al. propose DeepFi, a DL solution, to solve the localization problem. A greedy learning algorithm is employed to train the proposed DL network to reduce complexity. Subsequently, a probabilistic method with a radial basis function is used to estimate the location. However, the greedy learning algorithm does not guarantee optimized weights for the networks. Gao et al. in [31] adopt a DL network to estimate the location and activity of a person using radio images transformed from CSI measurements from multiple channels. In [32], a partially connected neural network is proposed to make the best use of both the amplitude and phase of the CSI given in a phasor format. However, the proposed networks may suffer from slow training and instability when dealing with complex nonlinear problems. In [33], the authors propose an autoencoder-based indoor positioning method, which utilizes data of reduced dimension by an autoencoder from the data collected data by smartphones. Chen et al. in [34] propose transforming the CSI amplitude into a time-frequency matrix, which is treated as image data and used to train a CNN-based model for localization. In [24], Wang et al. propose to estimate the DoA with the phase difference of WiFi signal. The DoA values of a given location are more robust than that of the phase due to the stability of the phase difference. Then, they adopt the two dimensional convolutional neural network (2DCNN) to process the DoA images to exploit the time-frequency features. In [23], the authors propose utilizing the software-defined radio to capture WiFi beacon frames passively. A feed-forward neural network and one dimensional convolutional neural network (1DCNN) deep learning models are adopted to use full SCs collected from the SDR. In [26], the authors propose a deep residual sharing learning based IPS which uses the 2DCNN to exploit both frequency and time features in bimodal CSI data. In [25], Zhang et al. propose utilizing the trajectory CSI to enhance the robustness of the instability of RF signals in the indoor environment. They employ an 1DCNN to extract the spatial information from the trajectory CSI. Then, an LSTM is used to further extract temporal information from the spatial information. Both the spacial and temporal information are used to enhance the robustness and accuracy of the IPS.

TABLE 2. Accuracy comparison for representative IPS.

Ref.	Year	Features	Algorithm	Positioning Accuracy
[28]	2005	RSS-based	Probabilistic	1.4m
[27]	2017	CSI-based	kNN	1.23m
[22]	2017	CSI-based	DNN	1.8m
[34]	2017	CSI-based	2DCNN	1.37m
[24]	2018	CSI-based	2DCNN	1.79m
[23]	2019	CSI-based	1DCNN	0.99m
[32]	2020	CSI-based	DNN	1.74m
[26]	2020	CSI-based	2DCNN	1.55m
[25]	2021	CSI-based	1DCNN-LSTM	0.9m

III. CSI AND DoA PRELIMINARY

This section describes the two components, CSI and DoA, used in this paper to construct the fingerprinting-based IPS. To explain how to combine these two different components, a brief introduction to the multiple signal classification (MUSIC) algorithm and the spatial smoothing method will be provided for estimating the DoA.

A. CHANNEL STATE INFORMATION

The CSI of the WiFi signal can be extracted from the SCs of IEEE 802.11 ac based OFDM symbols. The fine-grained characteristics of the wireless channel are then available in the extracted CSI. Various CSI observations can be reflected at different positions in an indoor environment due to the RF front-end impairment between the AP and UE. This property makes CSI an ideal choice as fingerprints for fingerprinting-based radio maps.

In an OFDM system, the received signal can be expressed as

$$\mathbf{r} = \mathbf{h} \cdot \mathbf{t} + \mathbf{n}, \tag{1}$$

where $\mathbf{t}$ and $\mathbf{r}$ denote the signal vectors from the transmit (Tx) and receive (Rx) signals, respectively. $\mathbf{n}$ denotes the additive

white Gaussian noise vector and $\mathbf{h}$ denotes the channel vector that carries the CSI.

The ith SC channel h_i is a complex-valued quantity that can be written as

$$h_i = |h_i| \, e^{j\angle h_i}, \quad i = 1, \ldots, N_s, \tag{2}$$

where $|h_i|$ and $\angle h_i$ are the magnitude and phase of the channel, respectively, for the ith SC. N_s is the number of SCs in an OFDM symbol. Note that we only use the magnitude of channel h_i because the phase information $\angle h_i$ is unstable owing to the random jitters and noise caused by the imperfect hardware of the RF transceiver [22].

B. DIRECTION OF ARRIVAL

We consider a typical far-field signal scenario, in which signals are transmitted from several sources. The impinging signal received by an M-element uniform linear array (ULA) can be expressed as

$$\mathbf{y}(t_i) = \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k(t_i) + \mathbf{n}(t_i), \quad i = 1, \ldots, T, \tag{3}$$

where K is the number of far-field sources, T is the number of snapshots, and the source signal $s_k(t_i) \in \mathbb{C}$ is received at snapshot t_i from the angle θ_k. The array steering vector $\mathbf{a}(\theta_k)$ is given by

$$a_m(\theta_k) = e^{-jm(2\pi\lambda)d \sin\theta_k}, \tag{4}$$

where $\lambda = c_0/f_c$ is the wavelength of the signal with the SC frequency being f_c, and c_0 being the speed of light. The distance between antenna elements of the ULA is d. The objective of the DoA is to estimate the angle θ_k.

In this paper, we use the MUSIC algorithm to estimate the DoA of the impinging signals. The MUSIC algorithm [35] is a subspace-based algorithm that provides super-resolution with lower computational complexity compared to other DoA techniques [36]. The essential idea behind the MUSIC algorithm is to conduct eigenvalue decomposition (EVD) or singular value decomposition (SVD) on the sample covariance matrix (SCM) of the received signal to acquire the signal and noise subspaces that are orthogonal to each other. The two orthogonal subspaces are used to construct a pseudo spectrum with the largest peaks corresponding to the DoA. SCM can be written as

$$\mathbf{R}_{\mathbf{xx}} = \frac{1}{T} \sum_{i=1}^{T} \mathbf{y}(t_i) \mathbf{y}(t_i)^H, \tag{5}$$

In an indoor environment, multipath signal propagation makes the impinging signals coherent, of which the SCM is a singular matrix. In such a case, the MUSIC algorithm fails; therefore, it is necessary to apply smoothing technologies to SCM to obtain a non-singular matrix. Spatial smoothing [37] is performed by splitting the ULA into L subarrays, and the SCM can be written as

$$\mathbf{R}^s_{\mathbf{xx}} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{R}^{(l)}_{\mathbf{xx}}. \tag{6}$$

The rank of $\mathbf{R}^s_{\mathbf{xx}}$ is the sum of the ranks of signal the space and noise subspace. We can construct the MUSIC pseudo-spectrum by taking EVD or SVD on the SCM, as in (5).

$$P(\theta) = \frac{1}{\mathbf{a}^H(\theta) \mathbf{V}_n \mathbf{V}_n^H \mathbf{a}^H(\theta)}, \tag{7}$$

where $\mathbf{a}$ is the steering vector, as in (4), and $\mathbf{V}_n$ is a matrix spanning the noise subspace.

The K largest peaks in the result of (7) correspond to the DoAs of the signals impinging on the ULA from K sources. The DoAs estimated using the MUSIC algorithm are the inputs to the proposed model, which are detailed in Section IV.

IV. SYSTEM MODEL

This section describes the proposed TIPS architecture. The fingerprinting technique, including the online and offline phases of constructing an IPS, is explained.

A. SYSTEM OVERVIEW

Figure 1 shows the architectural overview of the proposed TIPS. The TIPS comprises three main modules: WiFi AP, preprocessor, and DL model. The WiFi AP equipped with an M-element ULA receives WiFi uplink signals from a single UE. The objective of the preprocessor is to extract the CSI and estimate the DoA from the received WiFi signal. The extracted CSI and estimated DoA corresponding to each position are stored in the database. The data in the database are then used as the input to the DL model for the training process. The training process is performed using a graphics processing unit (GPU) for a speed-up operation. The output of the DL model is the prediction of the UE's current position.

B. FINGERPRINTING TECHNIQUE

The proposed TIPS based on fingerprints has two phases: offline phase and online phase.

During the offline phase, the WiFi signal from predetermined routes is measured. As shown in Figure 1, there are four predetermined routes 1, 2, 3, and 4. We adopted the data collection method proposed in [25] to collect the WiFi signals continuously along each route. The collected WiFi signal is preprocessed in the preprocessor to extract the CSI and estimate the DoA using equations (2) and (7), respectively, as briefly explained in Section II. Subsequently, the CSI and DoA corresponding to all to-be-estimated positions along each route are stored in the database. The CSI and DoA, along with the label (i.e., the correct position corresponding to the CSI and DoA), are fed as input to the transformer. More details regarding the training process are provided in Section V-C. After the training is complete, the trained model, including the neural network structure and its optimized weights, is available during the online phase.

During the online phase, the test data are collected and preprocessed in the preprocessor to extract the CSI and estimate the DoA, similar to the collection and preprocessing of training data during the offline phase. Subsequently, the

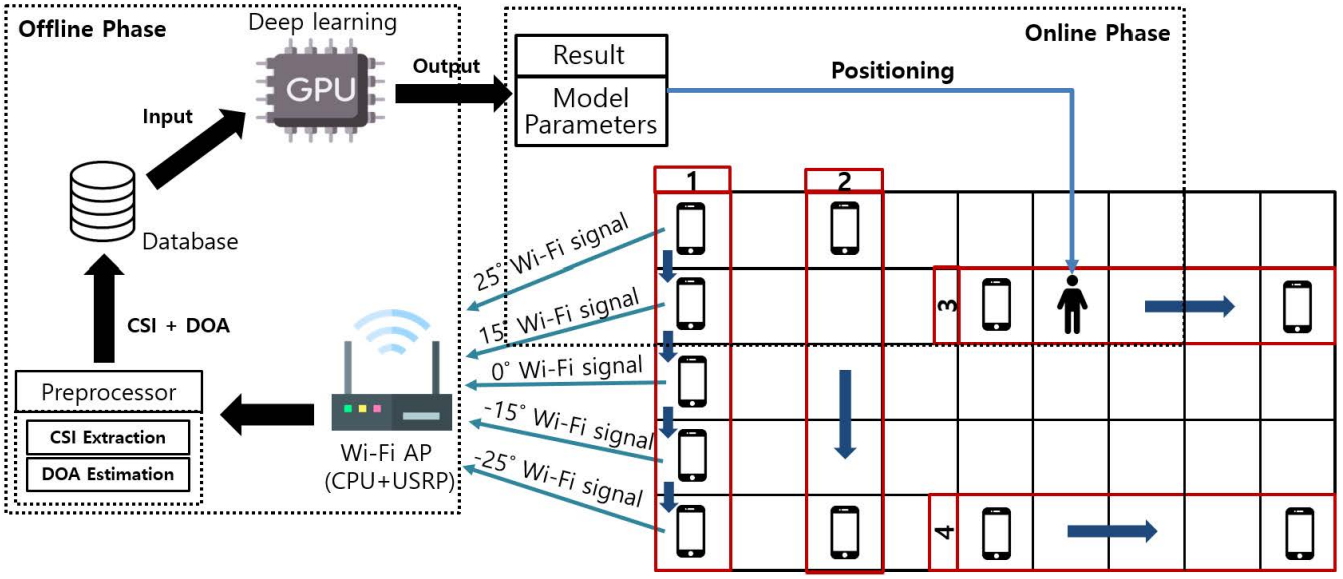

FIGURE 1. The proposed TIPS system overview.

transformer model, which has been trained during the offline phase, is employed to predict the exact position corresponding to the current test data input. The model is evaluated in terms of positioning accuracy obtained from the difference between the predicted and actual positions.

V. TRANSFORMER FOR IPS

This section explains the transformer DL model for the IPS in detail. First, the CSI and DoA data processing for the input data to the transformer is introduced. Then, the adopted transformer structure is presented. Finally, we demonstrate the application of the transformer model to an IPS.

A. TRANSFORMER MODEL

The objective of the desired model is to predict the current position by capturing the long-term dependencies from a continuous WiFi signal collected from predetermined routes. To achieve this goal, we adopt a transformer-based neural network model, which was initially introduced for the task of natural language processing (NLP) [29]. The original transformer has an encoder and decoder structure; however, in contrast to the original transformer model, we used the generative pretrained transformer (GPT) model [38], which has only the decoder part. GPT is an autoregressive model that predicts words at the current time step based on the words generated from the previous time steps. The following words are then predicted based on the output of the current time step in addition to the previously generated words. In particular, GPT is advantageous in the applications that predict sequential outputs based on previous predictions, which we believe is suitable for achieving the objective of predicting the next position based on previous positions. A simplified blog diagram of the transformer model is shown in Figure 2.

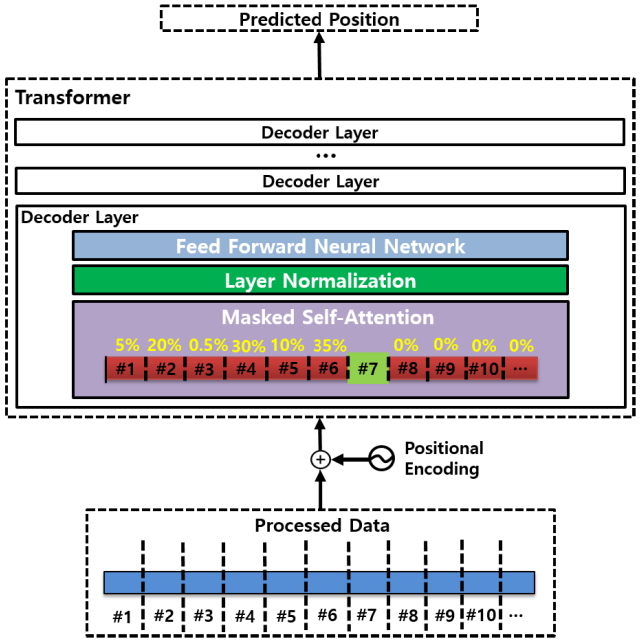

FIGURE 2. Adopted transformer model.

The self-attention mechanism used by the transformer models discards sequential operations in favor of parallel computation, which is entirely different from recurrent neural networks (RNNs) [39]. Consequently, the transformer model is unaware of the order of the input sequence. In other words, the positional information denoted as #1, #2, . . ., as shown in Figure 2, is lost. To compensate for the loss of positional information, position-dependent signals are generated based on the index of each sample of the input sequence through positional encoding. The generated

position-dependent signals are added to the corresponding samples of the processed data sequence, $\mathbf{X}_p$, which is explained in detail in Section IV-B.

The transformer consists of several decoder layers. In each decoder layer, a self-attention layer, normalization layer, and feedforward network are included.

In the self-attention layer, a masked self-attention mechanism is employed to prevent the prediction of the current position (e,g., #7), from attending to future positions. Instead of attending to future positions, the model only attends to past positions, for example, from #1 to #6. The decoder produces the joint probability of the current position as the product of the conditional probabilities of the previous positions.

$$p(\mathbf{y}_n) = \prod_{i=1}^{n-1} p(\mathbf{x}_p^i | \mathbf{x}_p^1, \ldots, \mathbf{x}_p^{i-1}). \tag{8}$$

The normalization layer [40] normalizes the distribution of the intermediate layers and stabilizes the gradients of loss; therefore, faster training and better generalization can be achieved. Finally, the feedforward network transforms the attention vector, i.e., the output of the self-attention layer, into a nonlinear representation that suits the input of the next decoder layer.

B. DATA PREPROCESSING

The input to the original transformer is a sequence of a sentence consisting of a fixed number of words. However, we modified the model for the IPS application to make it compatible with the inputs of CSI and DoA continuously collected from predetermined routes.

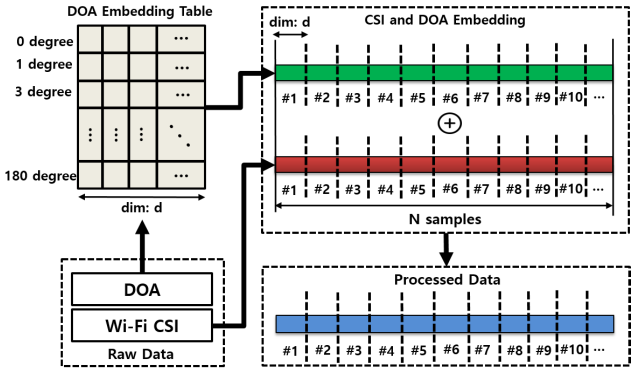

FIGURE 3. Data processing of embedding the CSI and DoA.

The CSI can be considered as a multivariate time series data, $\mathbf{X}_{\text{csi}} \in \mathbb{R}^{W\times d} = [\mathbf{x}_{\text{csi}}^1, \mathbf{x}_{\text{csi}}^2, \ldots, \mathbf{x}_{\text{csi}}^W]^T$. Each feature vector corresponding to a single time step t is $\mathbf{x}_{\text{csi}}^t \in \mathbb{R}^d$, where d is the number of variables in one vector. The raw DoA can also be considered as time series data but with only a single integer value for a single time step, $\mathbf{X}_{\text{doa}} \in \mathbb{Z}^{W\times 1} = [x_{\text{doa}}^1, x_{\text{doa}}^2, \ldots, x_{\text{doa}}^W]^T$ falls within the range from $-90°$ to $90°$. However, we add $90°$ to the raw DoA so that it falls within the range from $0°$ to $180°$ because the embedding layer in the neural network takes only a positive integer as input. The DoA is then linearly projected onto a d-dimensional vector space via an embedding table, as shown in Figure 3. The embedded vector can be expressed as:

$$\mathbf{u}_{\text{doa}}^t = \mathbf{W}_e \mathbf{x}_{\text{doa}}^t + \mathbf{b}_e, \quad t = 1, \ldots, W, \tag{9}$$

where $\mathbf{W}_e \in \mathbb{R}^{d\times 1}$ and $\mathbf{b}_e \in \mathbb{R}^d$ are trainable parameters optimized during the training of the neural network. $\mathbf{u}_{\text{doa}}^t \in \mathbb{R}^d, t = 0, \ldots, W$ is the embedded output, which corresponds to a word vector in the case of an NLP transformer. The embedded data $\mathbf{U}_{\text{doa}} \in \mathbb{R}^{W\times d} = [\mathbf{u}_{\text{doa}}^1, \mathbf{u}_{\text{doa}}^2, \ldots, \mathbf{u}_{\text{doa}}^W]^T$ are added to $\mathbf{X}_{\text{CSI}}$ to form the processed data $\mathbf{X}_{\text{p}} \in \mathbb{R}^{W\times d} = [\mathbf{x}_{\text{p}}^1, \mathbf{x}_{\text{p}}^2, \ldots, \mathbf{x}_{\text{p}}^W]$

VI. COMPUTER SIMULATIONS

This section introduces a modeled indoor propagation environment for computer simulations performed using MATLAB. First, we explain how the datasets are prepared to train the transformer model shown in Section IV. A detailed explanation of the transformer training process is provided. Finally, the performance of the proposed TIPS in various scenarios is evaluated and analyzed.

A. INDOOR PROPAGATION ENVIRONMENT

Figure 4 shows the 8 m $\times$ 5 m indoor propagation environment for which the simulations are carried out. The blue points and red points in the figure denote the RPs and APs, respectively.

The indoor propagation environment employs four APs, each of which is equipped with 8-antenna element ULA. The channel bandwidth is set to 20 MHz, corresponding to 56 SCs for each OFDM symbol of IEEE 802.11ac. CSI is generated using the ray-tracing technique [41] implemented in MATLAB. The ray-tracing technique can help build multipath channel models for indoor environments.

In our simulations, a UE with a single antenna transmits WiFi OFDM signals from all the predetermined RPs. Each AP extracts the CSI and estimates the DoA from the received WiFi signals as fingerprints. Predetermined RPs are uniformly distributed with a spacing of 10 cm. The total number of RPs is $79 \times 49 = 3871$ points, with 49 points along the x-axis and 79 points along the y-axis. Note that the number of the RPs used in the simulations is not the same as shown in Figure 4.

We set the parameters of the ray propagation model to consider only the LOS and first-order reflection. The number of rays is of size $N_{\text{AP}} \times N_{\text{UE}}$, where N_{RP} is the number of APs, and N_{UE} is the number of RPs. Hence, the size of the rays generated using the ray-tracing technique is 4×3871 in our simulation. A visualization of the rays at an arbitrary RP is shown in Figure 5. From the figure, three APs have LOS propagation to the RP, whereas the signal emitted from the remaining AP only reaches the RP via reflection without LOS propagation. The different colors denote different path losses in decibels (dB).

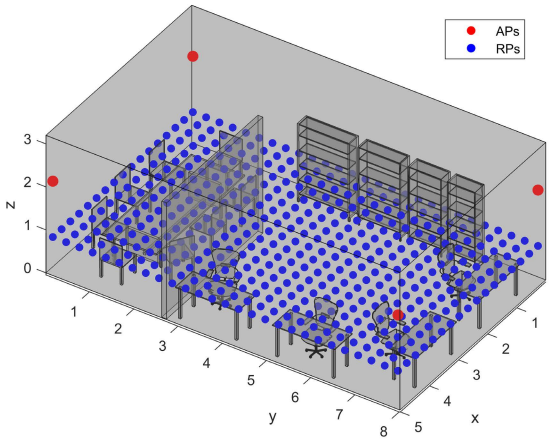

FIGURE 4. Indoor propagation environment.

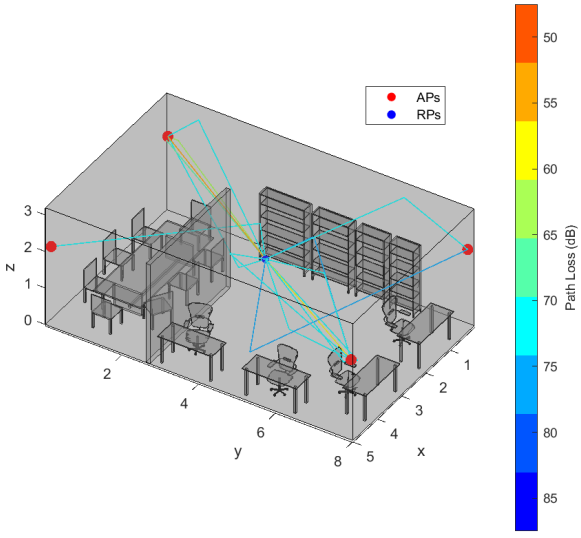

FIGURE 5. Visualization of rays received at an arbitrary RP.

B. PREPARATION OF DATASETS

The four APs receive 802.11 ac packets transmitted from a single UE at each of the RPs in the modeled indoor environment shown in Figure 4. The CSI extracted from the received WiFi signal is a complex-valued variable including both the magnitude and phase for 56 SCs as shown in (2), of which we only use the magnitude. Consequently, the data size of the CSI is $N_s \times N_{\text{tx-rx}} \times N_{\text{AP}} \times N_{\text{UE}}$, where N_s is the number of SCs, $N_{\text{tx-rx}}$ is the number of Tx-Rx antenna pairs. Therefore, in our simulation, the size of the dataset is $56 \times 8 \times 4 \times 3871$. As explained in Section IV-B, the DoA estimated from the received WiFi signal is a real-valued variable ranging from 0° to 180°. Consequently, the data size of the DoA is $1 \times N_{\text{AP}} \times N_{\text{UE}}$, i.e., $1 \times 4 \times 3871$ in our simulation.

To extensively evaluate the performance of our transformer model under various SNR scenarios, we generate WiFi signals under SNR scenarios from 0 to 30 dB. For each SNR, we generate 100 WiFi signal samples, which are then divided into 80 training data samples and 20 validation data samples. In addition, 20 WiFi signal samples are generated as test data samples.

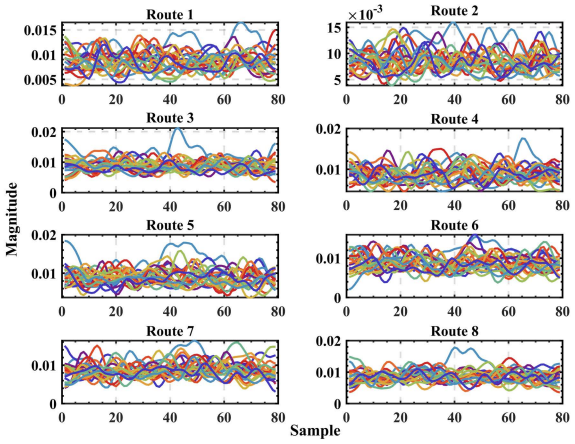

FIGURE 6. The CSI of 79 RPs along eight arbitrary routes.

We divide the 3871 RPs into 49 routes, along each of which there are 79 RPs. Each route can be considered a *sentence* in terms of NLP, whereas the CSI measured at each RP along the route can be considered an individual *word*. Figure 6 shows the CSI measured at 79 RPs along eight arbitrary routes from a total of 49 routes. The lines with different colors in each subplot denote the CSI of each SC. To match the data size of the CSI with that of the input of the transformer, we combine the dimensions of N_s, $N_{\text{tx-rx}}$, and N_{AP} into one dimension of size 1792, i.e., 56 (number of SCs) × 8 (number of Rx antennas) × 4 (number of APs). The CSI dataset is then reshaped to a size of $79 \times 49 \times 1792$. Likewise, the DoA dataset is reshaped to $79 \times 49 \times 4$. Table 3 lists the datasets corresponding to the environmental parameters.

C. TRAINING PROCESS

The processed CSI and DoA, as shown in Figure 3, are first mapped into the corresponding d_{model}-dimensional input embeddings via the embedding lookup table. The processed data are then fed into the transformer model. CSI and DoA can be considered fingerprints of RPs. The label of the fingerprint is $c_i, i \in \mathbb{Z}, i = 0, 1, \ldots, 3870$ corresponding to the position. The position-prediction task is to estimate the probability for a given position (or a sequence of positions) based on the transformer outputs for the previous positions. The probability estimation is calculated by putting the output of the transformer model into a linear layer followed by a log-softmax function. Then, a cross-entropy loss for updating the model's weights via backpropagation is calculated. We use the Adam optimizer [42] with a learning rate of 0.0001 to accelerate the gradient descent algorithm. The number of transformer layers l_{tr} is set to three, and the number of attention heads per layer h_{tr} is set to four.

TABLE 3. Detailed dataset information.

Property	Value
SNR (dB)	[0,30]
Number of APs	4
Number of routes	49
Number of RPs per route	79
Number of Tx antennas	1
Number of Rx antennas	8
Number of SCs per OFDM symbol	56
Dimension of CSI dataset	$79 \times 49 \times 1792$
Dimension of DoA dataset	$79 \times 49 \times 4$
Number of Training data samples	80
Number of Validation data samples	20
Number of Test data samples	40

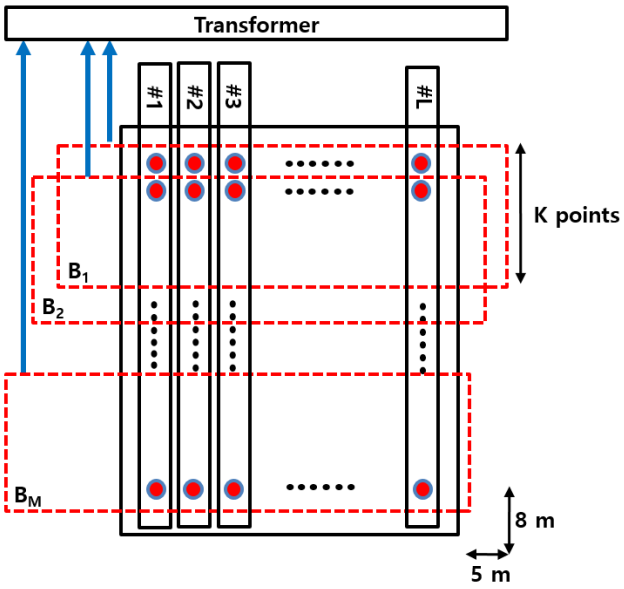

FIGURE 7. Transformer training process.

All the routes can be trained together simultaneously by enabling the parallel processing due to the fact the training for each route is independent of that for the other routes. As shown in Figure 7, the input data of size $K \times L$ in the first batch B_1 are fed to the transformer to be trained at once, where K is the number of samples trained in a route at once and L is the number of routes. K is 40, and L is 49 in our simulation. After the first batch is trained, the samples at positions from 2 to $K+1$ of all routes of the next batch B_2 are fed to the transformer to be trained. The training continues until the training for the last batch B_M is completed. Note that the data at the same positions on different routes can be trained together without interference because the training for each route is independent of one another. We may consider changing the sample size K to adjust the sequence length. With a larger K, the model can capture longer-term dependencies among positions and vice versa. Table 4 summarizes the hyper-parameter settings for our transformer model.

TABLE 4. Hyper-parameter settings.

Hyper-parameter	Value
Dimension of input embeddings	1792
Input data size $K \times L$	40×49
Number of batches	40
Number of labels	3871
Number of transformer layer	3
Number of attention heads per layer	4
Number of epochs	200
Optimization algorithm	Adam
Learning rate	0.0001
Dropout rate	0.2
Loss function	Cross-entropy

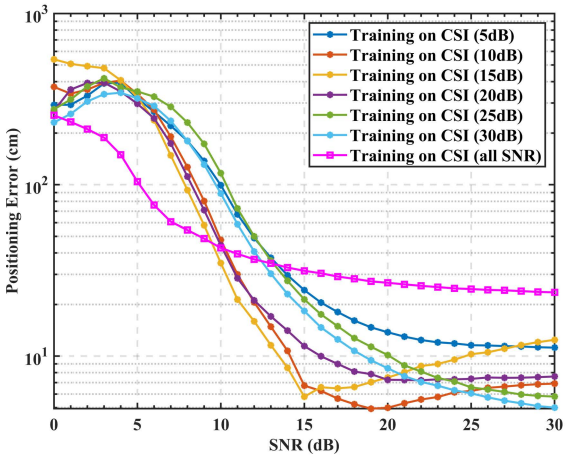

FIGURE 8. Positioning error of models trained on datasets including only CSI under different SNRs.

D. PERFORMANCE OF THE PROPOSED TIPS TRAINED ON CSI-ONLY DATASET

We conducted experiments to investigate the performance of TIPS as a function of the SNRs, while the training datasets are generated only with the CSI of WiFi signals. First, our model is trained with the CSI generated under six different SNR values, i.e., 5, 10, 15, 20, 25, and 30 dB. Then, our transformer model is trained with CSI in an environment where the SNR value varies between 0 dB and 30 dB.

Figure 8 compares the positioning error in centimeters as a function of the SNR for the seven different models. As the SNR increases, the positioning error decreases for the models trained on datasets of 5 dB, 25 dB, 30 dB, and all SNRs. The models trained on datasets with 10 dB and 15 dB SNR exhibit relatively superior performance when SNR varies from 10 dB to 24 dB. However, it can also be observed that the performance superiority of the model trained on a dataset of 15 dB (10 dB) is slightly reduced when the SNR becomes greater than 15 dB (20 dB). This observation implies that the transformer model trained in a noisy environment does not necessarily exhibit a better performance in a noise-free environment.

It is also observed that, for lower SNRs, i.e., from 0-9 dB, the model trained on datasets of all SNRs shows the best performance because the model is better aware of the data

generated from the lower SNRs, which are included in the training dataset. However, this model appears to perform worse than the other models for higher SNRs, i.e., from 14-30 dB, due to the erroneous information caused by the heavily noisy training samples. The models trained on datasets with an SNR of 25 or 30 dB exhibit superior performance than other models for the high SNR range, i.e., from 25-30 dB.

E. PERFORMANCE OF THE TIPS TRAINED ON DoA-ONLY DATASET

We conducted experiments to investigate the performance of TIPS as a function of the SNRs, while the training datasets are composed of only the DoA of WiFi signals. First, our transformer model is trained with the DoA estimated under five different values for SNR, i.e., 10, 15, 20, 25, and 30 dB. Subsequently, our model is trained with the DoA estimated in an environment where the SNR value varies between 0 and 30 dB.

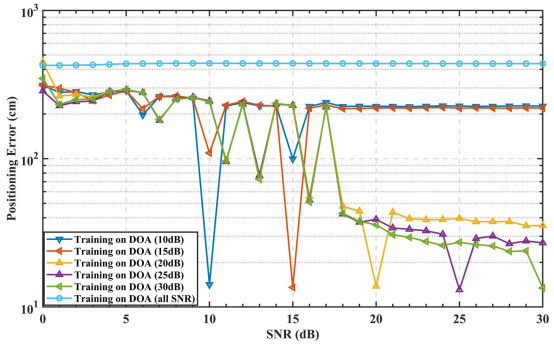

FIGURE 9. Positioning error of models trained on datasets including only DoA under different SNRs.

Figure 9 compares the positioning error in centimeters as a function of the SNR for the six different models. As the SNR increases, the positioning error does not decrease in the same manner as when the training dataset is the CSI-only dataset. It can be observed that the models trained on DoA-only datasets perform excellently when the SNR of the test data is the same as that of the training data. However, the models exhibit significant performance degradation when the SNR of the test dataset does not match that of the training dataset.

The model trained on the dataset generated under all SNRs exhibit the worst performance. This is a substantial limitation concerning the ability of the model to learn the similarities of the DoA under different SNRs. Consequently, the DoA-only datasets prove to be of little help in training the model.

F. PERFORMANCE OF THE TIPS TRAINED ON BOTH CSI AND DoA

We conducted experiments to investigate the performance of TIPS as a function of SNRs, while the training datasets are composed of CSI & DoA of WiFi signals. We compared the performance of the models trained on DoA-only datasets and those trained on CSI & DoA datasets under five different SNRs, i.e., 10, 15, 20, 25, and 30 dB.

Figure 10 compares the positioning error in centimeters as a function of the SNR for ten different models. The dotted lines denote the models trained on DoA-only datasets, and the solid lines denote the models trained on the CSI & DoA datasets. It can be observed that the models trained on CSI & DoA datasets have a much lower positioning error than those trained on DoA-only datasets when the SNR of the test data is the same as that of the training data. However, the models show significant performance degradation which bears a strong resemblance to the models trained on DoA-only datasets. This indicates that the DoA plays a predominant role during the training process compared with the CSI when the WiFi signal is generated under a single SNR.

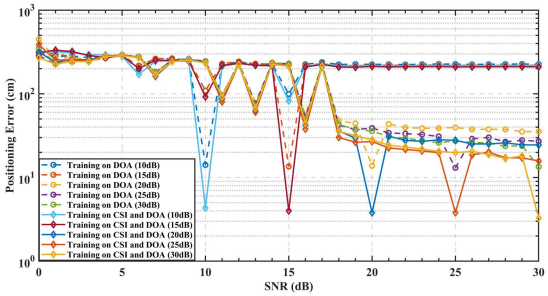

FIGURE 10. Positioning error of models trained on CSI & DoA datasets under different SNRs.

It has been observed that the models trained on CSI-only datasets exhibit more consistent and better performance as a function of SNR than those trained on DoA-only datasets. We compare the performance of the models trained on CSI-only datasets and the model trained on CSI & DoA datasets, where the SNR varies between 0 and 30 dB.

Figure 11 compares the positioning error in centimeters as a function of the SNR for the eight different models. The model trained on CSI & DoA datasets, where the SNR varies between 0 and 30 dB, exhibits the best performance. This indicates that the model can learn more beneficial information from both the CSI & DoA datasets, where the SNR has a wider range than the CSI-only dataset.

G. IMPACT OF BATCH SIZE K

The performance of the proposed transformer model was observed with various values for the batch size K, which should be determined on account of a trade-off between the dependencies among the positions and the time for training the model.

The transformer model can attend to more fingerprints of previous positions with a larger K, corresponding to longer-term dependencies among positions. Thus, the model can selectively pay more attention (by assigning a larger significance probability) to positions where multipath fading has less severe impacts. The transformer model is more likely to

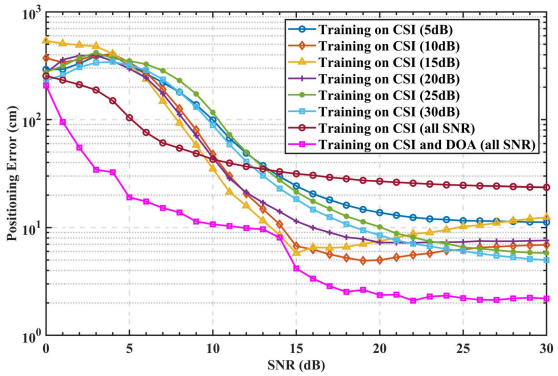

FIGURE 11. Positioning error comparison for the models trained on the CSI-only and CSI & DoA datasets.

correctly predict the subsequent position with more exposure to effective fingerprints. However, longer dependencies may lead to a longer time required for the model to converge. Furthermore, if the value of K is too large, the model performance may be degraded unless the transformer model can handle the large size of the long input data sequence.

In contrast, short-term dependencies among positions, which correspond to a small K, can accelerate the training speed; however, the performance may not be satisfactory due to the lack of effective fingerprints to which the transformer can attend for predicting the next position.

Table 5 shows the training time and training loss of the proposed TIPS for various values of the batch size K. As the batch size K increases, the training time increases, and the training loss reduces because the model can enhance its ability to capture the long-term dependencies of the input sequence. However, when the batch size K exceeds 40, the decreasing rate of the training loss becomes significantly small, but the increasing rate of the training time remains almost the same. Hence, considering the training loss and the cost of the training time, we conclude that a batch size K equal to 40 is an optimal value for our model.

H. TRAINING TIME COMPARISON

The training time of the model is observed for different datasets, i.e., the CSI-only dataset and CSI & DoA dataset. To be specific, we measured the training time and training loss to evaluate the training time provided by the CSI-only and CSI & DoA datasets.

Table 6 shows the training loss and training time for two different training datasets, i.e., the CSI-only dataset and CSI & DoA dataset. The model trained on CSI-only has 0.07 training loss with 170-second training time, whereas the model trained on CSI & DoA dataset has only 0.02 training loss with only 65-second training time. In other words, the model trained on CSI & DoA dataset converges much faster than the model trained on the CSI-only dataset. The intuition behind the fast training with CSI & DoA dataset is that the embedding of CSI and DoA can help the transformer model more effectively and optimally allocate attention to the previous positions where the fingerprints can be utilized for predicting the current position.

TABLE 5. Model's training time and training loss for different batch size K.

Batch size K	Training time (seconds)	Training loss
1	58.5	11.31
5	60	1.85
10	62	1.43
20	83	1.18
30	103.5	1.11
40	119	1.08
50	130.5	1.07
60	140	1.07
70	150	1.06

TABLE 6. Comparison of training time and loss of models training on different types of datasets.

Training dataset	Training time (second)	Training loss
CSI-only	170	0.07
CSI & DoA	65	0.02

I. IMPACT OF ANTENNA ARRAY SIZE AND SIGNAL BANDWIDTH

Having noticed that the IPS performance might seriously be affected by the antenna array size as well as the bandwidth of the WiFi signal, we summarise in this subsection the impact of antenna array size and signal bandwidth obtained from various computer simulations.

To evaluate the actual effects of antenna array size and signal bandwidth on the performance of our proposed TIPS, we generated the datasets of both the CSI and DoA under different antenna array size and signal bandwidth. The model is trained for 200 epochs with the batch size K being equal to 40. Figure 12 shows the cumulative probability over the distance error in centimeters under four different cases of antenna array size and signal bandwidth. As the figure illustrates, the positioning error decreases as the signal bandwidth increases with the 2-dimensional antenna array size being fixed to 4×4. The positioning error is within 20 cm for signal bandwidth 80MHz, 40MHz, and 20MHz with a probability of 90%, 70%, and 60%,respectively. It can be observed that the larger signal bandwidth contributes to the more accurate positioning. It is because the richer position-related CSI from the wider signal bandwidth can result in the better performance. Note that, as the bandwidth of the WiFi signal gets larger, more SCs of the OFDM symbol can be utilized for constructing the radio map. On the other hand, as the size of the 2-dimensional antenna array decreases from 4×4 to 2×2 with the bandwidth being fixed to 20MHz, the performance of the IPS is degraded. With antenna size of 2×2, the probability of holding the positioning error within 20 cm decreases to 50% compared to the case of

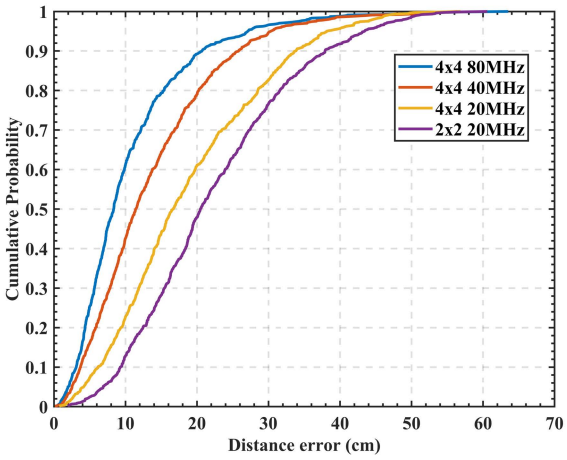

FIGURE 12. Predetermined routes in the indoor RF environment.

4×4 antenna array that provides the probability of 60%. The main reason is that the larger antenna size can lead to the better DoA resolution, which results in the more accurate positioning.

VII. EXPERIMENTAL RESULTS WITH RF SIGNALS

This section introduces the RF experiments conducted using a testbed implemented for the proposed TIPS. The RF WiFi signals are collected along the predetermined routes and preprocessed to extract CSI and DoA. As explained in Section IV-B, the preprocessed CSI and DoA are then embedded together to prepare for model training. We observed the impact of data scaling on model training and compared the performance of the proposed TIPS and other state-of-the-art indoor positioning methods.

A. RF TESTBED

We implemented an RF testbed for the proposed TIPS using USRPs and a CPU. USRPs, which are frequently used for SDR communication applications, are reconfigurable RF devices consisting of an RF front end, an analog-to-digital converter, and a frequency down-converter.

Figure 13 shows the implemented RF testbed consisting of five USRPs. Among the five USRPs, we used four USRPs (Ettus X310 [43]), each of which provides two Rx channels, to implement a single AP. Each channel is connected to an omnidirectional antenna (VERT2450 [44]). The eight antennas constitute an 8-element ULA for the DoA estimation. For UE, we used a single USRP (Ettus X310) with a single Tx channel connected to an omnidirectional antenna (VERT2450) as the RF transceiver. The host PC executes the USRP hardware driver (UHD) using MATLAB to control all the five USRPs, including four USRPs for the AP and one USRP for the UE. An Ethernet switch connects all the USRPs with the host PC via 10 gigabit Ethernet cables for data packet transmission. For time synchronization among Rx channels at the AP, we used OctoClock (CDA-2990 [45]) to synchronize

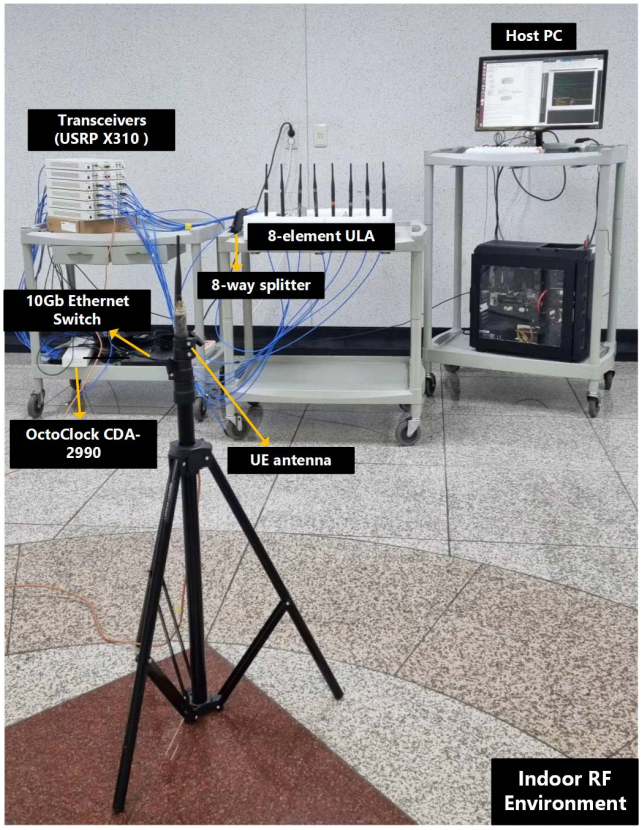

FIGURE 13. The RF testbed located in an indoor RF environment.

the eight Rx channels of the AP to the common timing source provided by the OctoClock.

To acquire an accurate DoA estimation, it is critical to align the phase among the eight Rx channels of the AP. Due to the random phase offset incurred at each channel of the USRPs [46], the DoA estimation becomes infeasible. To tackle the phase offset problem, we implemented phase calibration for the alignment of the phase among USRP channels before performing DoA estimation. For phase calibration, we utilized an 8-way splitter to distribute the sinusoidal calibration signal to each Rx channel of the AP with RF cables of identical length. We estimate and compensate for the phase difference among the eight Rx channels using the received sinusoidal signal.

B. PREPARATION OF DATASETS

To evaluate the performance of the proposed TIPS in an indoor RF environment, we predetermined 16 routes, as illustrated in Figure 14. Each route consists of 480 uniformly distributed virtual RPs (VRPs) with a spacing of 1 cm. Note that the spacing can be set to different values for different resolutions.

The UE continuously transmits the 802.11ac WiFi signal generated by MATLAB while moving along each of the 16 routes. The WiFi signal emitted from the moving UE is received by the AP located at a fixed location. The AP then

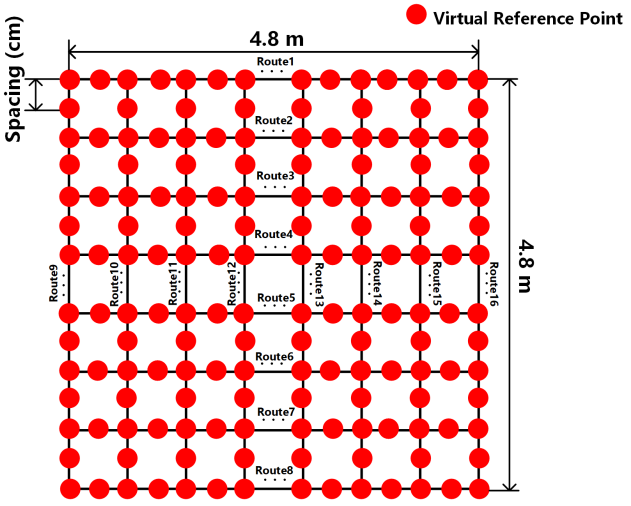

FIGURE 14. Predetermined routes in the indoor RF environment.

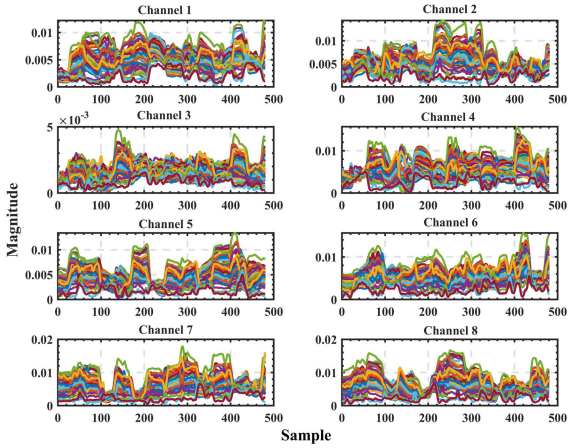

FIGURE 15. CSI of 480 VRPs along a single route measured from eight Rx channels of AP.

extracts the CSI and estimates the DoA from the received WiFi signal. Consequently, the CSI dataset has a size of $480 \times 16 \times 448$, where 480 is the number of VRPs, 16 is the total number of routes, and 448 is the dimension of the variable for each VRP determined by the number of SCs and channels ($448 = 56 \times 8$). For each VRP, only a single DoA value is estimated, and the DoA dataset has a size of $480 \times 16 \times 1$.

Figure 15 shows the CSI measured by the eight Rx channels of the AP at 480 VRPs along a single route. The lines with different colors in each subplot correspond to the CSI of each SC.

C. EXPERIMENTAL PROCEDURES

To further explain the application of the proposed TIPS in the indoor RF environment, Figure 16 shows the flowchart of the experimental procedures. As shown in the flowchart, the multiple USRPs are phase aligned through phase calibration before the reception of the WiFi signal. Then, the received WiFi signals at all 480 VRPs are through the process of resampling and denoising so that the extracted WiFi CSI of each VRP should become of the same sample length with less noise. With the preprocessed WiFi signal, we employ the MUSIC algorithm to estimate the DoA of the WiFi signal at each VRP. Both the WiFi CSI and DoA are embedded as explained in Section V-B to form the dataset of 480 VRPs. We perform the data scaling on the dataset in order to standardize the dataset. The standardized dataset is then used to train the adopted transformer model. Note that the training process is detailed in Section VI-C.

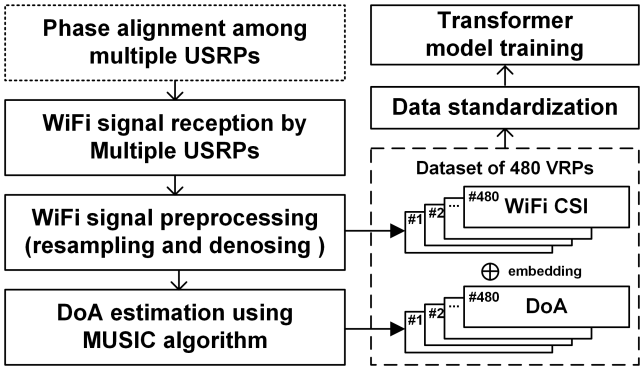

FIGURE 16. Flowchart of the experimental procedures.

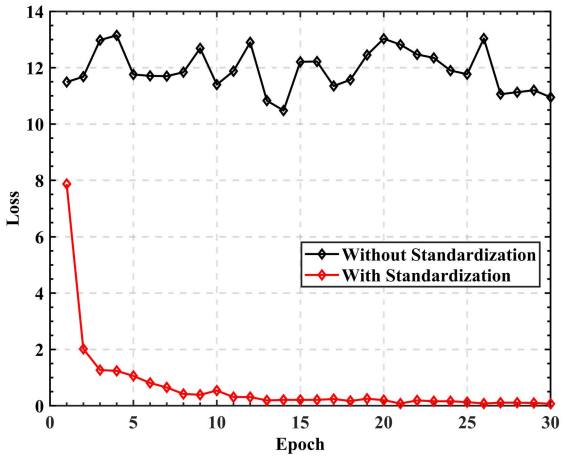

FIGURE 17. Model's training loss over epochs.

D. IMPACT OF DATA SCALING

We observed the impact of data standardization on the training process of the transformer model. Our model is trained with two different datasets; one is composed of raw data, and the other is composed of data standardized to have zero mean and unit variance. As shown in Figure 17, the model trained on the dataset without standardization fails to converge and maintains a high training loss throughout 30 epochs. In contrast, the model trained on the standardized dataset converges within ten epochs and exhibits a significantly low training loss. This implies that data standardization is essential for the training process of the proposed TIPS.

E. PERFORMANCE COMPARISON WITH EXISTING METHODS

We compare the performance of the proposed TIPS with that of state-of-the-art solutions, i.e., 1DCNN-LSTM [25], ConFi [34], DeepFi [22], and Horus [28]. 1DCNN-LSTM utilizes a CNN and LSTM neural structure to extract spatial and temporal information from the trajectory CSI. ConFi employs a CNN with CSI input collected at multiple time instances and antennas while considering the input as an image. DeepFi uses a deep neural network with the CSI amplitude as its input. Horus estimates the target position using a probabilistic model with the signal strength vector from multiple APs.

Figure 18 shows the cumulative probability over the distance error in meters for all state-of-the-art methods and the proposed TIPS. It can be observed that the proposed TIPS significantly outperforms 1DCNN-LSTM, ConFi, DeepFi, and Horus with a probability of 70%, 85%, and 100% within a 10 cm, 12 cm, and 20 cm distant error, respectively.

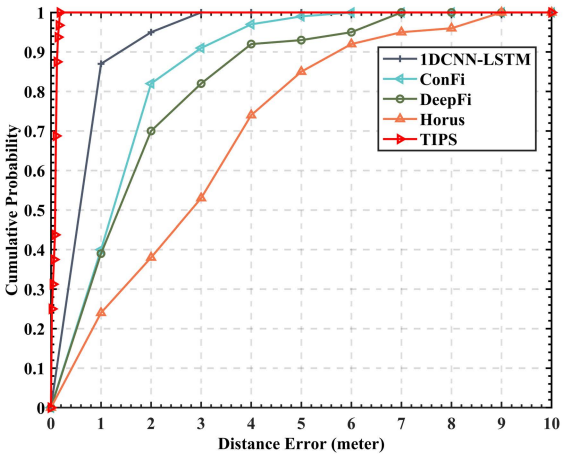

FIGURE 18. Positioning accuracy of the proposed TIPS and existing methods.

VIII. CONCLUSION

This paper proposed TIPS, a transformer-based indoor positioning system that uses WiFi signals. First, we extract the CSI and estimate the DoA from the Wi-Fi signal transmitted from predetermined positions. Subsequently, the extracted CSI and estimated DoA are preprocessed and embedded to be fed to the transformer model to learn the correlations between the fingerprints and their corresponding positions. Using the attention mechanism, the transformer model can predict the current locations by effectively attending to the fingerprints of the previous position, which significantly boosts positioning accuracy. We evaluated the superior performance of TIPS through extensive MATLAB simulations under a modeled multipath propagation environment. The TIPS trained on the CSI & DoA dataset were demonstrated to achieve the highest positioning accuracy with a short training time. To evaluate the performance of TIPS in a real-world RF environment, we implemented an RF testbed that included a single AP and UE with five USRPs. The results show that TIPS exhibits the highest positioning accuracy compared to state-of-the-art indoor positioning methods, with a probability of 100% within a 20 cm distance error.

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.
[2] L. Barreto, A. Amaral, and T. Pereira, "Industry 4.0 implications in logistics: An overview," *Proc. Manuf.*, vol. 13, pp. 1245–1252, Jan. 2017.
[3] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017.
[4] S. Subedi and J.-Y. Pyun, "A survey of smartphone-based indoor positioning system using RF-based wireless technologies," *Sensors*, vol. 20, no. 24, p. 7230, Dec. 2020.
[5] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, and A. Bagheri, "A survey on indoor positioning systems for IoT-based applications," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7680–7699, May 2022.
[6] S. Horsmanheimo, S. Lembo, L. Tuomimaki, S. Huilla, P. Honkamaa, M. Laukkanen, and P. Kemppi, "Indoor positioning platform to support 5G location based services," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2019, pp. 1–6.
[7] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1327–1346, 2nd Quart., 2016.
[8] P. Kriz, F. Maly, and T. Kozel, "Improving indoor localization using Bluetooth low energy beacons," *Mobile Inf. Syst.*, vol. 2016, pp. 1–11, Mar. 2016.
[9] X.-H. Li and T. Zhang, "Research on improved UWB localization algorithm in NLOS environment," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City (ICITBS)*, Jan. 2018, pp. 707–711.
[10] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.
[11] Y. He, Y. Chen, Y. Hu, and B. Zeng, "WiFi vision: Sensing, recognition, and detection with commodity MIMO-OFDM WiFi," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8296–8317, Sep. 2020.
[12] Y. Duan, K.-Y. Lam, V. C. S. Lee, W. Nie, K. Liu, H. Li, and C. J. Xue, "Data rate fingerprinting: A WLAN-based indoor positioning technique for passive localization," *IEEE Sensors J.*, vol. 19, no. 15, pp. 6517–6529, Aug. 2019.
[13] X. Dang, X. Tang, Z. Hao, and Y. Liu, "A device-free indoor localization method using CSI with Wi-Fi signals," *Sensors*, vol. 19, no. 14, p. 3233, Jul. 2019.
[14] Y. Yin, C. Song, M. Li, and Q. Niu, "A CSI-based indoor fingerprinting localization with model integration approach," *Sensors*, vol. 19, no. 13, p. 2998, Jul. 2019.
[15] S. Shi, S. Sigg, L. Chen, and Y. Ji, "Accurate location tracking from CSI-based passive device-free probabilistic fingerprinting," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5217–5230, Jun. 2018.
[16] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint activity recognition and indoor localization with WiFi fingerprints," *IEEE Access*, vol. 7, pp. 80058–80068, 2019.
[17] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-LSTM for human activity recognition using WiFi signal," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4628–4641, Mar. 2021.
[18] E. Ben Hamida and G. Chelius, "Investigating the impact of human activity on the performance of wireless networks—An experimental approach," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Networks (WoWMoM)*, Jun. 2010, pp. 1–8.
[19] W. Liu, Q. Cheng, Z. Deng, H. Chen, X. Fu, X. Zheng, S. Zheng, C. Chen, and S. Wang, "Survey on CSI-based indoor positioning systems and recent advances," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2019, pp. 1–8.
[20] Y. Wang, C. Xiu, X. Zhang, and D. Yang, "WiFi indoor localization with CSI fingerprinting-based random forest," *Sensors*, vol. 18, no. 9, p. 2869, Sep. 2018.

[21] L. Tang, Z. Zhang, Y. Zhao, T. Feng, W.-C. Wong, and H. K. Garg, "A comparison of WiFi-based indoor positioning methods," in *Proc. 13th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Dec. 2019, pp. 1–6.
[22] 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. 2016.
[23] E. Schmidt, D. Inupakutika, R. Mundlamuri, and D. Akopian, "SDR-fi: Deep-learning-based indoor positioning via software-defined radio," *IEEE Access*, vol. 7, pp. 145784–145797, 2019.
[24] 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.
[25] Z. Zhang, M. Lee, and S. Choi, "Deep-learning-based Wi-Fi indoor positioning system using continuous CSI of trajectories," *Sensors*, vol. 21, no. 17, p. 5776, Aug. 2021.
[26] 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.
[27] Q. Song, S. Guo, X. Liu, and Y. Yang, "CSI amplitude fingerprinting-based NB-IoT indoor localization," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1494–1504, Jun. 2017.
[28] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proc. 3rd Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2005, pp. 205–218.
[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*.
[30] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "Pilot: Passive device-free indoor localization using channel state information," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst.*, Jul. 2013, pp. 236–245.
[31] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10346–10356, Nov. 2017.
[32] M. Kim, C. Kim, D. Han, and J.-K.-K. Rhee, "CompFi: Partially connected neural network using complex CSI data for indoor localization," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.
[33] Z. E. Khatab, A. Hajihoseini, and S. A. Ghorashi, "A fingerprint method for indoor localization using autoencoder based deep extreme learning machine," *IEEE Sensors Lett.*, vol. 2, no. 1, pp. 1–4, Mar. 2018.
[34] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18066–18074, 2017.
[35] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[36] S. Chen, X. Li, and Z. Shao, "Study on the performance of DOA estimation algorithms," in *Proc. IEEE Int. Conf. Commun. Problem-Solving (ICCP)*, Oct. 2015, pp. 475–477.
[37] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 806–811, Apr. 1985.
[38] A. Radford, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.
[39] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.
[40] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.
[41] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[43] X300/X310. (2021). *X300/x310 Ettus Knowledge Base*. Accessed: May 25, 2022. [Online]. Available: https://kb.ettus.com/index.php?title=X300/X310&oldid=5092
[44] VERT 2450. (2016). *Antennas ettus Knowledge Base* Accessed: May 25, 2022. [Online]. Available: https://kb.ettus.com/index.php?title=Antennas&oldid=3098
[45] OctoClock CDA-2990. (2020). *Octoclock CDA-2990 Ettus Knowledge Base*. Accessed: May 25, 2022. [Online]. Available: https://kb.ettus.com/index.php?title=OctoClock_CDA-2990&oldid=4916
[46] H. Fu, S. Abeywickrama, C. Yuen, and M. Zhang, "A robust phase-ambiguity-immune DOA estimation scheme for antenna array," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6686–6696, Jul. 2019.

ZHONGFENG ZHANG received the B.S. degree in electronic engineering from Yanbian University, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electronics and Computer Engineering, Hanyang University, Seoul, South Korea. His current research interests include beamforming, indoor positioning systems, wireless communications, mmWave communication, and MIMO systems.

HONGXIN DU received the B.S. degree in electronic engineering from Yanbian University, in 2020. He is currently pursuing the M.S. degree with the Department of Electronics and Computer Engineering, Hanyang University, Seoul, South Korea. His current research interests include DoA, beamforming, wireless communications, and DL.

SEUNGWON CHOI (Member, IEEE) received the M.S. degree in computer engineering and the Ph.D. degree in electrical engineering from Syracuse University, Syracuse, NY, USA, in 1985 and 1988, respectively. From 1988 to 1989, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Syracuse University. He joined Hanyang University, Seoul, South Korea, as an Assistant Professor, in 1992, where he is currently a Professor with the School of Electrical and Computer Engineering. His research interests include software reconfiguration, and digital communications with a recent focus on the implementation of various kinds of MIMO systems for mobile communication systems.

SUNG HO CHO (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from The University of Utah, Salt Lake City, UT, USA, in 1989. From 1989 to 1992, he was a Senior Member of Technical Staff with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. He joined the Department of Electronic Engineering, Hanyang University, Seoul, South Korea, in 1992, where he is currently a Full Professor. He has been the Director of the Radar Computing Laboratory, since 2010. His research interests include applied signal processing, machine learning, radar computing, digital health, and radar-based smart space.