HIGH TECHNOLOGY LETTERS | Vol. 30 No. 1 | Mar. 2024 | pp. 61-67

doi:10.3772/j.issn.1006-6748.2024.01.007

Two-person device-free localization system based on ZigBee and transformer①

LIU Tianmeng(刘天蒙)[*,**], YANG Hai xiao[*,**], WU Hong②[*,**,***]

([*] School of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, P. R. China)

([**] Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Nankai University, Tianjin 300350, P. R. China)

([***] Engineering Research Center of Thin Film Optoelectronics Technology, Nankai University, Tianjin 300350, P. R. China)

Abstract

Most studies on device-free localization currently focus on single-person scenarios. This paper proposes a novel method for device-free localization that utilizes ZigBee received signal strength indication (RSSI) and a Transformer network structure. The method aims to address the limited research and low accuracy of two-person device-free localization. This paper first describes the construction of the sensor network used for collecting ZigBee RSSI. It then examines the format and features of ZigBee data packages. The algorithm design of this paper is then introduced. The box plot method is used to identify abnormal data points, and a neural network is used to establish the mapping model between ZigBee RSSI matrix and localization coordinates. This neural network includes a Transformer encoder layer as the encoder and a fully connected network as the decoder. The proposed method's classification accuracy was experimentally tested in an online test stage, resulting in an accuracy rate of 98.79%. In conclusion, the proposed two-person localization system is novel and has demonstrated high accuracy.

Key words: device-free localization, deep learning, ZigBee

0 Introduction

While satellite signals like Beidou and global positioning system (GPS) can achieve high localization accuracy outdoors, they are unable to penetrate indoor spaces due to obstructions such as walls and windows. Therefore, satellite signals are not practical for indoor localization.

Typical indoor localization signals include Bluetooth, Wi-Fi, and ZigBee. Indoor localization can be classified into two categories: localization with devices and device-free localization, depending on whether the testers need to use communication equipment. Localization with devices determines the position of communication devices to locate device wearers, which can be applied to cooperative goals. For example, customers can locate themselves in commercial centers by connecting their smartphones to Wi-Fi, and patients can locate themselves in medical facilities by wearing communication devices. Device-free localization is used for non-cooperative targets such as indoor intruder monitoring and prison personnel location monitoring. It predicts the location of testers based on the impact of the human body on the communication link. Microsoft[1] developed the RADAR indoor positioning device based on received signal strength indication (RSSI), with a positioning accuracy of 2 – 5 m. Xiao et al.[2] created the fine-grained indoor finger printing system, which was the first to use Wi-Fi channel state information (CSI) data for fingerprint matching position. While localization accuracy is high with devices, the target must actively collaborate with the communication device worn. Additionally, the device is costly and may cause discomfort to the user.

The concept of device-free localization was introduced by Youssef et al.[3] in 2007. Device-free localization leverages the fingerprint comparison technique. The fingerprint comparison procedure can be divided into the offline acquisition stage and online testing stage. During the offline acquisition stage, several sampling locations are set up in the experimental area to collect signal features received by the signal receiver when the tester is at different locations. A mapping

① Supported by the National Natural Science Foundation of China (No. U2031208, 61571244).

② To whom correspondence should be addressed. Email: wuhong@nankai.edu.cn.

Received on Feb. 6, 2023

model from the signal features to the tester's location is constructed using appropriate algorithms. During an online test, when a target enters an experimental region, the gathered signal attributes and the mapping model created during the offline phase are utilized to predict the individual's location. By analyzing signal intensity changes in wireless networks, Seifeldin et al. [4] confirmed the feasibility of using Wi-Fi signals for device-free localization systems. Chiang et al. [5] developed a fuzzy support vector machine approach and used it with device-free localization techniques. Wang et al. [6] utilized deep learning to construct a localization system without a device and extract behavior recognition features. Dang et al. [7] designed a two-person location system for CSI signals based on Wi-Fi, although the experimental area was small, and the accuracy was low. Yang and Wu[8] created a single-person indoor locating system using deep learning and ZigBee RSSI.

In scenarios where location services are necessary, there may be more than one user present. Common two-person localization scenarios include a prison cell with two inmates, a nursing care room with two elderly residents, and so on, all of which require ongoing observation. To expand the application potential of device-free localization, this study extends the common one-person location to a two-person location. In the two-person localization with devices scenario, there is no interference between the two communication devices. Therefore, the position calculations of the communication equipment are carried out in two independent systems, and the position calculations of the two users are independent of each other. Device-free localization is based on how individuals can impact a signal link. Therefore, in the case of two individuals, each individual will have an impact on a shared signal link. Investigating device-free localization for two persons is more challenging because the solution to localization is not an independent problem.

1 ZigBee sensor network and signal analysis

ZigBee is a short-range, low-power wireless communication technology that is based on the IEEE 802.15.4 standard. It is also known as Purple Bee. Devices that use ZigBee technology are compact, inexpensive, and power-efficient. This chapter introduces the construction of a ZigBee sensor network and the analysis of ZigBee data packages.

1.1 ZigBee sensor network

In this paper, a wireless sensor network (WSN) is established to collect ZigBee RSSI. To maximize the number of wireless links and enhance the characteristics of indoor localization signals, the mesh topology is utilized. In the mesh topology, every two router nodes communicate with each other, and the router node simultaneously sends information to the coordinator node for information summary. Fig. 1 illustrates the network topology, which includes a coordinator node and multiple router nodes.

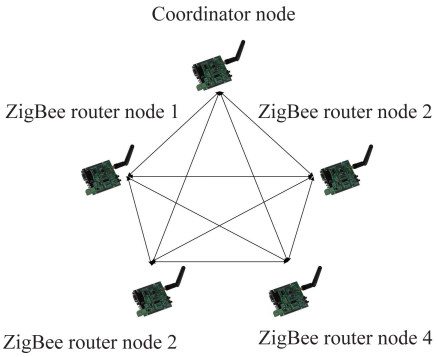

Fig. 1 ZigBee mesh topology

The steps involved in data acquisition can be categorized into the following groups.

(1) Start the system up.

(2) Setup system tasks: establish network connectivity between router nodes and coordinator nodes.

(3) After the network is constructed, the router node receives the data in the package, which contains the RSSI that is needed for this study.

(4) The coordinator node receives data packages from the router node on a regular basis.

The coordinator node is responsible for launching and setting up the entire sensor network. In a ZigBee sensor network with M router nodes, the number of two-way wireless links between routers is $(M(M-1))/2$ pairs. In this study's case, there are 45 pairs of two-way wireless links connecting the 10 router nodes. Each router node should be set up to broadcast data packages to all other router nodes every second while simultaneously receiving data packages from other router nodes. The coordinator node consolidates the data packages sent by each router node before importing them into the computer through the universal serial bus (USB) connection.

1.2 Signal analysis

This paper employed a total of 11 ZigBee nodes, consisting of one coordinator node and ten router nodes. In each cycle, 10 router nodes can connect with one another, producing an RSSI matrix with 10 rows and 10 columns. The RSSI that the ith router

node received from the *j*th router node is represented by the value of column *j* in row *i*. The value is 0 if *i* and *j* are equal.

Signal processing and mathematical modeling processes become more challenging if the signal variation is too significant over a short period of time. In this section, two groups of experiments were conducted under static conditions of the human body. In experiment 1, two ZigBee routing nodes were installed in the test area. In experiment 2, the transmitter and receiver were placed at a location consistent with the two routing nodes in experiment 1 and the ZigBee transmission cycle.

The receiver is a host with an Intel 5300 network adapter, and the transmitter is a 2.4 GHz antenna with Wi-Fi. The Fig. 2 depicts how a Wi-Fi link's CSI amplitude and RSSI value fluctuate from router node No. 1 to router node No. 2. The ordinate and abscissa, both in dBm, represent the RSSI and the signal's sample period, respectively. In comparison to CSI, the signal from ZigBee is more stable and has a lower variance.

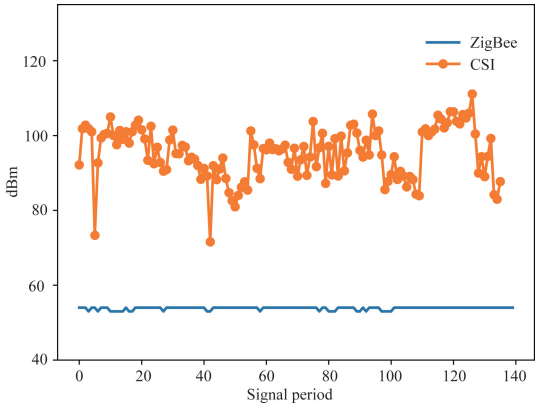

Fig. 2 The stationarity comparison between ZigBee and CSI

To input ZigBee RSSI into the neural network, data preprocessing is required. After creating the ZigBee Dataset class, the ZigBee RSSI matrix and its related location label are added to a sample. The DataLoader function is used to define the batch size and create the DataLoader file. In this study, the batch size is set to 4, meaning that each training batch contains four samples fed into the neural network. DataLoader is an iterable object that a neural network's cyclic function can iterate over. DataLoader is an iterable object that a neural network's cyclic function can read.

2 The algorithm

This chapter comprises two sections: data preprocessing and algorithm model design. The data preprocessing section uses the box plot analysis method, while the algorithm model section utilizes a deep learning model based on Transformer encoder.

2.1 Data preprocessing

Defects in ZigBee node devices can lead to the receiver collecting anomalous data, which can reduce the accuracy of the positioning model. This section applies the box plot approach to identify outliers.

A box plot is a statistical diagram that depicts the dispersion of a data set. Its box-like shape inspired its name. In box plot analysis, the four primary indicators are the lower quartile, upper quartile, lower bound, and upper bound. The quartile L represents the point below which one-fourth of the samples' data falls. The quartile H represents the point above which one-fourth of the data in all samples falls. The top bound is set at $H + 1.5IQR$, and the lower bound is set at $L - 1.5IQR$, where IQR is the difference between the upper and lower quartiles.

Outliers refer to data points that lie outside the upper and lower boundaries. The following graph illustrates the experimental data containing outliers and the results obtained using the box plot method, using a set of ZigBee RSSI as an example. The black dots represent outlier data, while the data within the 'box' represent normal data. Fig. 3 demonstrates how the box plot method can reliably identify unusual data points.

Outliers require to be processed. In this paper, outliers are replaced with the mean value of the training set's concentrated signals in the same signal link. It is worth noting that this approach is only employed in the training set for this study. Since the positioning algorithm must operate in real-time in the test environment, obtaining the average RSSI value is impossible. Modifying the test data can also have adverse effects.

2.2 The deep learning model

This paper uses a deep learning model even though the tabular data from the ZigBee RSSI is appropriate for classical machine learning modeling.

Deep learning generally performs worse on tabular data compared with classical machine learning methods like XGBoost, LightGBM, and support vector machines (SVM). However, deep learning demonstrates its benefits in processing non-tabular data, such as text, image, and speech. The primary reason for choosing deep learning for modeling in this study is that the two-person localization problem can be abstracted into a multi-label classification problem, which traditional machine learning cannot handle directly. Traditional

machine learning requires tedious label transformation steps, and the performance is poor.

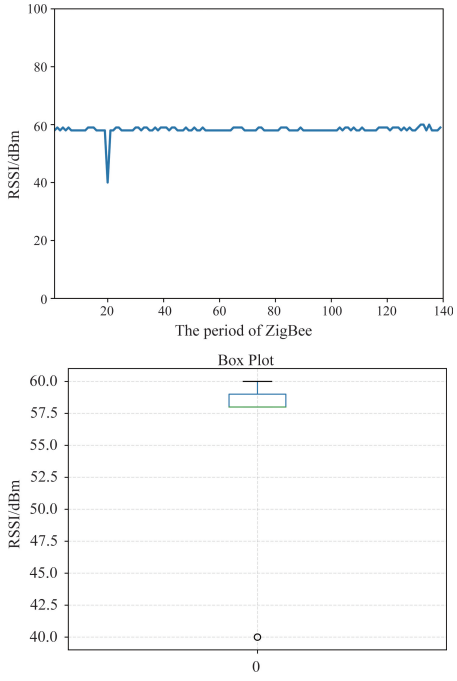

Fig. 3 Box plot method identifies abnormal data

Multi-label classification problems can be more effectively addressed with deep learning, as it involves matrix operations. In this study, the two-person localization problem's label is defined as a 12-dimensional, one-dimensional vector. Deep learning can produce the output vector of label type without the need for label type conversion.

ZigBee packages use 10×10 RSSI matrices. This study ingeniously maps a 10×10 matrix to two locations and uses it as an analogy to model multi-label classification of a sentence with 10 words and 10 embedding dimensions. This approach addresses the challenge of text multi-label classification in natural language processing. Furthermore, the model incorporates the Transformer network, which has gained popularity in recent years for natural language processing applications.

In their paper 'Attention is All You Need' [9], Google introduced a self-attention-based neural network structure called Transformer. Initially developed to solve machine translation problems, it has since been extensively employed in natural language processing pre-training models such as Bert, GPT-3, and so on. The Transformer network separates encoders and decoders. The encoder consists of several encoder blocks, including a self-attention layer, a residual and normalization layer, and a feedforward neural network layer. The components of an encoder block are further explained in this section.

2.2.1 Self attention layer

Take the ZigBee packages used in this paper as an example, the encoding method of self-attention layer is as follows: set the ith column data of ZigBee RSSI matrix $\boldsymbol{A}$ as $\boldsymbol{A}_i$, the dimension of $\boldsymbol{A}_i$ is 1×10. Randomly initialize the learnable matrix $\boldsymbol{W}_Q$, $\boldsymbol{W}_K$, and $\boldsymbol{W}_V$ with three dimensions of 10×10, and dot multiply $\boldsymbol{A}_i$ with three matrices respectively to obtain three matrices $\boldsymbol{Q}_i$, $\boldsymbol{K}_i$, and $\boldsymbol{V}_i$. The flow chart of the above steps and the changes of matrix dimensions are shown in the Fig. 4.

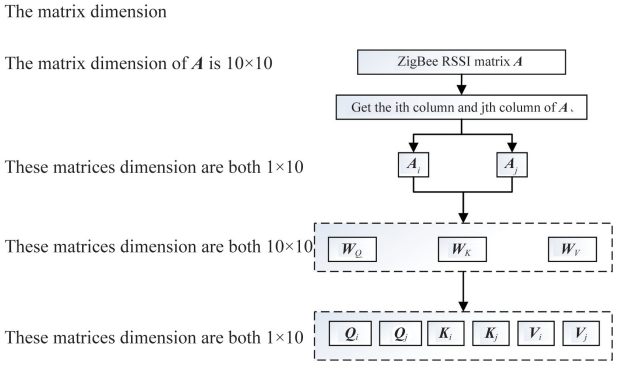

Fig. 4 The self attention layer

The $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V}$ matrices are used to calculate the attention score. Take $Score_{i,j}$, the calculation of $\boldsymbol{A}_j$' s attention score for $\boldsymbol{A}_i$, as an example. The calculation formula is as follows.

$$Score_{i,j} = \frac{Q_i K_j^{\mathrm{T}}}{\sqrt{d_{k_j}}} \tag{1}$$

After calculating the $Score_{i,j}$ of all columns for $\boldsymbol{A}_i$, the softmax function is used to calculate all the attention score values to get the corresponding weight of each column, multiplied the weight by the $\boldsymbol{V}$ matrix corresponding to each column, and added the weight to get the coded output of the current column. It can be expressed as

$$t_i = \sum_{j=1}^{10} \mathrm{softmax}(Score_{i,j}) V_j \tag{2}$$

where, t_i is the ith column of the target matrix output through the self-attention coding layer.

2.2.2 Residuals and layer normalization

The purpose of adding layer normalization is to accelerate the training speed and enhance the stability of training. Layer normalization normalizes the outputs of a layer across the features, and it has been shown to be effective in reducing the internal covariate shift problem, which can slow down training.

2.2.3 Feedforward neural network layer

The feedforward neural network layer in the Transformer network comprises two fully connected neural network layers, and the activation function used is the rectified linear unit (ReLU) function. As the output layer in the Transformer network, its output dimension is the same as the original input dimension, which is 10×10.

In this paper, the training set's batch size is set to 4. The neural network topology depicted in Table 1 is constructed in this paper after experimental validation.

Table 1 The detail of network

The index of network	Network structure
1	Transformer encoder layer
2	Linear(10,128) + Tanh
3	Linear(128,10) + ReLU
4	Linear(10,12) + Tanh
Dimension transfer	$4 \times 10 \times 12$ matrix to $4 \times 12 \times 10$
5	Linear(10,1) + Sigmoid
Dimension transfer	$4 \times 12 \times 1$ matrix to $4 \times 1 \times 12$
6	BCELoss

The neural network topology depicted in the following table is constructed in this paper after experimental validation. The linear layer is a fully connected layer, and the activation functions used are Tanh, ReLU, and Sigmoid. The torch permute function is used for dimension transfer.

The loss function applied to multi-label classification is the Binary Cross-Entropy (BCE) Loss. Suppose the number of categories is n, y_i is the real label of the ith category, and x_i is the probability of the ith category output by the model. The expression for BCELoss is

$$BCELoss(x,y) = -\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i \ln x_i \tag{3}$$

After calculating the loss value, the back propagation of the gradient will update the parameters of the network.

3 Experiment

The experiment was conducted in a classroom in the building of the College of Electronic Information and Optical Engineering at Nankai University, Tianjin. The experiment site was 5.10 m × 7.86 m, and the area was divided into 12 uniform rectangles. It is worth noting that the premise of the study is that there are two people indoor, and the system determines the two locations with the highest probability based on the output of the neural network, regardless of whether there is one or more people in the room. If the assessment is negative, the system outputs 'there are not two individuals in the residence'.

During the training set collection stage, the two experimental persons were stationed at 66 different locations, as depicted in Fig. 5. For each experiment, two people stand at different star symbols in the middle, and data were collected for a minute at each location. During the test set collection stage, the two testers stood at the training data collection point, or on the left or right side of the training data collection point. The test set has 198 types of stations, which is three times the number of station types in the training set, with 10 s of data collected for each station.

The training set and the test set were collected at different time periods to evaluate the model's generalization performance.

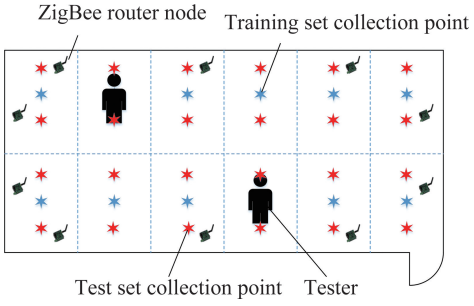

Fig. 5 Experimental environment

The Fig. 6 depicts the change in the loss value on the training set as the number of epochs increases. For this study, the number of epochs is set to 15.

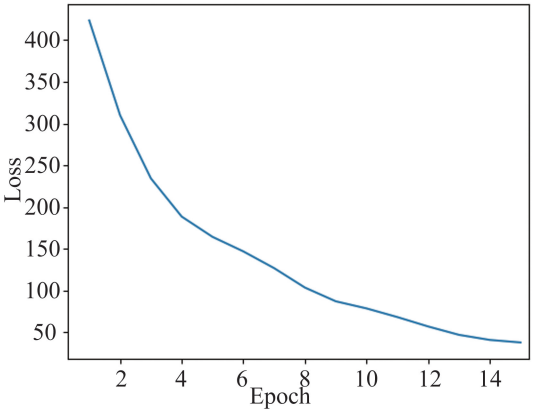

Fig. 6 The loss on the training set

Since the objective of this study is two-person localization, the Euclidean distance formula cannot be directly used to calculate the error. Let loc_pred_1 and loc_pred_2 be the predicted positions, and loc_real_1 and loc_real_2 be the correct positions. These four parameters are two-dimensional coordinates in the form of (x, y). The error in two-person positioning is defined using the following formula, where $dis(\cdot)$ represents the Euclidean distance formula:

$$error = \min(dis(loc\_pred_1, loc\_real_1) + dis(loc\_pred_2, loc\_real_2), dis(loc\_pred_1, loc\_real_2) + dis(loc\_pred_2, loc\_real_1)) \quad (4)$$

Lastly, the model's performance is validated using the test set. The output layer of the neural network created in this paper produces a one-dimensional vector with a length of 12. The 12 values in the vector represent the likelihood that a person will occupy each position, with values ranging from 0 to 1. If the output vector's first two values are 0.9 and the final ten values are 0, the likelihood of people in the model's predicted areas 1 and 2 is 0.9, and the probability of humans in the remaining regions is 0. Setting a threshold value of 0.5 generates the desired output, with values greater than 0.5 indicating the expected presence of people at the location and values less than 0.5 indicating the absence of such predictions.

Table 2 Comparison of different algorithm

Algorithm	Accuracy/%	Error/m
Transformer	98.79	0.575
LSTM	93.21	0.737
GRU	92.26	0.778
LightGBM	88.83	0.969
SVM	91.52	0.826

The classification accuracy is 98.79% when using location accuracy as the assessment metric for the anticipated location. The error is 0.575 meters when using the two-person positioning error stated in this paper as the evaluation index. In this study, the Transformer network is used as an encoder in the neural network structure. In sequential modeling, long short-term memory (LSTM) and gate recurrent unit (GRU) are also typically employed as encoders of neural networks. Fully connected neural networks are frequently used as decoders in sequence modeling. This study compares the experimental results obtained using the encoders Transformer, LSTM, and GRU, respectively. Additionally, the accuracy of SVM and LightGBM is compared as a sample machine learning algorithm. The statistical table of error and the cumulative distribution function (CDF) of the error diagram are displayed.

The algorithm described in this paper outperforms other algorithms, including conventional machine learning techniques, as demonstrated by the CDF diagram and table.

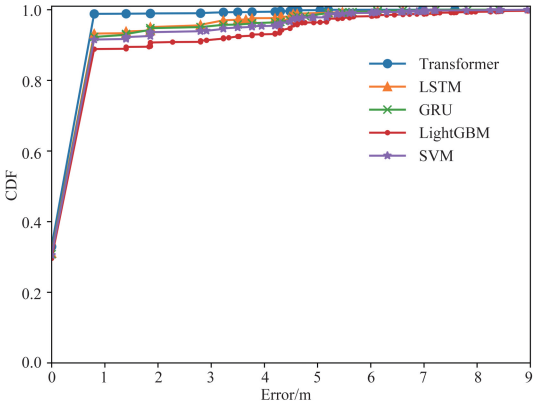

Fig. 7 The CDF of error diagram

This algorithm focuses on two-person DFL and achieves high accuracy, but there are still limitations. The algorithm cannot determine the position of a person when there is no one or only one person in the room. Future research will investigate this issue further.

4 Conclusion

This paper proposes a design for a device-free localization system in a two-person situation. ZigBee data packages are processed and input into a structure that incorporates a Transformer encoder and a fully connected neural network, using the text multi-label classification of natural language processing. Experimental results show that the two-person positioning accuracy is 98.79% with a positioning error of 0.575 m, demonstrating high accuracy.

References

[1] BAHL P, PADMANABHAN V N. RADAR: an in-building RF-based user location and tracking system[C]//The 19th Joint Conference of the IEEE Computer Communications Societies. Tel Aviv, Israel: IEEE, 2000:775-784.

[2] XIAO J, WU K S, YI Y W, et al. FIFS:fine-grained indoor fingerprinting system[C]//2012 21st International Conference on Computer Communications and Networks. Munich, Germany: ICCCN, 2012:1-7.

[3] YOUSSEF M, MAH M, AGRAWALA A. Challenges: device-free passive localization for wireless environments [C]//Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking.

Montréal, Canada: Association for Computing Machinery, 2007: 222-229.
[4] SEIFELDIN M, SAEED A, KOSBA A E, et al. Nuzzer: a large-scale device-free passive localization system for wireless environments[J]. IEEE Transactions on Mobile Computing, 2012, 12(7):1321-1334.
[5] CHIANG Y Y, HSU W H, YEH S C, et al. Fuzzy support vector machines for device-free localization[J]. IEEE International Instrumentation and Measurement Technology Conference. Graz, Austria: IEEE, 2012: 2169-2172.
[6] JIE W, XIAO Z, GAO Q, et al. Device-free wireless localization and activity recognition: a deep learning approach[J]. IEEE Transactions on Vehicular Technology, 2017, 66(7):6258-6267.
[7] DANG X C, CAO Y, HAO Z J. A CSI-based double positioning method[J]. Chinese Journal of Sensors and Actuators, 2019, 32(11): 1700-1705.
[8] YANG M G, WU H. Deep learning approach for device-free localisation based on Internet of things[J]. Electronics Letters, 2020, 56(11):575-577.
[9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 8(1):8-15.

LIU Tianmeng, born in 1997. He is currently pursuing his M. S. degree at Nankai University. He received his B. S. degree in Harbin Institute of Technology in 2019. His current research interests include device-free localization and natural language processing.