

> TIM-24-05298R1<

TransAoA: Transformer-based Angle of Arrival Estimation for BLE Indoor Localization

Wei Wu, *member, IEEE*, Didi Zhou, Leidi Shen, Zhiheng Zhao, *senior member, IEEE*, Congbo Li, *senior member, IEEE*, and George Huang, *Follow, IEEE*

Abstract—Bluetooth Low Energy (BLE) technology, characterized by its low energy consumption, cost-effectiveness, and scalability, has gained prominence as a viable solution for indoor localization within industrial contexts. However, the dynamic nature of industrial environments poses considerable challenges to the accuracy of BLE-based indoor positioning systems (IPSs), particularly those dependent on signal strength for localization. Accordingly, this paper proposes a novel method framework TransAoA that leverages the Transformer deep learning architecture to enhance Angle of Arrival (AoA) estimation for BLE indoor positioning. First, a data filtering method is designed to eliminate low-quality In-phase and Quadrature (I/Q) samples affected by noise. Second, a specialized feature extraction method is developed to distill multiple informative features from I/Q samples prior to the prediction model to enable rapid convergence and improve accuracy. Third, the Transformer-based AoA estimation model is constructed to establish a mapping relationship between angles (azimuth and elevation) and the combined I/Q samples and features. Fourth, several BLE anchors collaborate to localize targets using a least squares approach, and a self-adjusting calibration mechanism is devised to bolster the long-term robustness and stability of the IPS. Finally, experiments are conducted in a lab that simulates industrial conditions to verify the effectiveness of the framework. By comparison, the TransAoA shows superiority over existing benchmark methods, achieving estimation errors within 5 degrees for 98.85% of azimuth and 99.97% of elevation measurements.

Index Terms—Indoor localization, Bluetooth low energy, Angle of Arrival, Deep learning, Transformer.

I. INTRODUCTION

THE significance of location-based services (LBS) and spatial-temporal traceability has been highlighted in Industry 4.0 for enhancing operational and management efficiency among activities such as rapid object-picking, stock-

taking, and strategic asset allocation [1,2]. Furthermore, they have also been expected to assume a more vital role in the context of Industry 5.0 [3], which prioritizes human-centric management, sustainability, and production resilience, especially in human safety monitoring, product lifecycle supervision, and real-time decision-making [4-6]. Correspondingly, indoor positioning systems (IPS) serve as a foundation to generate spatial-temporal information necessary for these service offerings. In most industrial applications, stakeholders favor an IPS that is not only accurate but also cost-efficient and easy-to-deploy. Bluetooth Low Energy (BLE) technology, featuring low energy consumption, affordability, and high scalability, emerges as a prime candidate to fulfill those indoor localization requirements [7]. However, the intricate and dynamic industrial environment, replete with noise pollution, signal interference, and physical obstructions, can severely impair BLE signal integrity, exacerbate multipath fading and, thus, challenge the accuracy, robustness, and stability of the positioning system.

In the early stages, proximity detection, trilateration and fingerprinting, mainly relying on the Received Signal Strength Indicator (RSSI), were the predominant techniques employed in BLE-based location [8,9]. However, the reliability of RSSI values is compromised by their propensity to significant fluctuations due to signal reflection, refraction, and interference, even in the absence of any movement within the space. The advancement in localization performance was limited until the release of BLE 5.1 in 2019, which introduced a direction-finding capability predicated on Angle of Arrival (AoA) estimation [10]. AoA measures the angle at which the signal from a transmitter arrives at a receiver. This enhancement marked a departure from the limitations of RSSI-based positioning. AoA-based techniques are inherently less vulnerable to multipath interference, as they concentrate on the

This work was supported in part by China Postdoctoral Science Foundation under Grant 2023M730406; in part by the Natural Science Foundation Project of Chongqing under Grant CSTB2024NSCQ-MSX0561; in part by the National Natural Science Foundation of China under Grant 52305557; in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515011930; in part by Open Fund of State Key Laboratory of Intelligent Manufacturing Equipment and Technology under Grant IMETKF2024022; in part by the China Postdoctoral Science Foundation under Grant 2022M712394; in part by Hong Kong RGC TRS Project under Grant T32-707/22-N; in part by the Research Impact Fund under Grant R7036-22. (Corresponding author: Zhiheng Zhao & Wei Wu).

W. Wu and C. Li are with the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China (e-mail: bravew@cqu.edu.cn; congboli@cqu.edu.cn).

D. Zhou is with the School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: 1222077429@njupt.edu.cn).

L. Shen is with the Department of Industrial and Manufacturing Systems Engineering, the University of Hong Kong, Hong Kong 999077, China (e-mail: sldc@hku.hk).

Z. Zhao is with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong 999077, China, with Research Institute for Advanced Manufacturing, The Hong Kong Polytechnic University, Hong Kong 999077, with China State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430072, China (e-mail: zhiheng.zhao@polyu.edu.hk).

G. Huang is with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong 999077, China (e-mail: gq.huang@polyu.edu.hk).

direct signal path from a transmitter to a receiver. The utilization of AoA enables the attainment of more granular location insights, thereby broadening the scope of potential applications. Consequently, the adoption of AoA-based BLE indoor localization has garnered increasing interest within both academic research and industrial implementation [11,12], signaling a shift towards more sophisticated and comprehensive indoor positioning solutions.

The AoA estimation primarily relies on the analysis of In-phase and Quadrature (I/Q) data. A plethora of algorithms, including MUSIC [13], PDDA [14], ISSS [15], ESPRIT [16], and maximum likelihood estimation [17], have been developed to refine AoA estimation in triangulation-based localization. Despite their contributions, these algorithms exhibit limitations in handling interference, environmental dynamics, and systematic inaccuracies stemming from antenna array configurations and radio frequency switching processes, which can compromise AoA measurement accuracy. Moreover, the computational complexity of some algorithms may impede real-time applications and increase energy consumptions. In contrast, data-driven models demonstrate a profound capacity to assimilate complex environmental interactions and signal propagation phenomena, thereby enhancing the fidelity and adaptability of AoA estimation for various scenarios [18]. Different machine learning methods exemplify such approaches. While studies have demonstrated the advantages of learning methods [19-21], there remains significant room for improvement given the rapid advancements in artificial intelligence techniques. Besides, the majority of extant studies often focus on positioning accuracy as the primary metric of performance [22-24]. However, the response time, robustness, and stability of IPS should also be crucial considerations in algorithm design.

Accordingly, this paper proposes a method framework for BLE-based indoor positioning, which integrates raw data preprocessing, feature extraction, AoA estimation, location prediction, and self-calibration to enhance system performance. This framework, termed TransAoA, employs the Transformer deep learning architecture for AoA computation, leveraging the multi-head attention mechanism to capture long-range dependencies with parallelizable computation and bolster the model's resilience to noise and outliers in data. A tailored feature extraction approach is devised to process raw I/Q samples, capturing a comprehensive set of features, involving temporal, spectral, energy, envelope, phase difference, and I/Q component features, to facilitate AoA estimation. The least squares approach is applied for target localization using multiple anchors, with a moving average technique designed to refine prediction accuracy. Additionally, a self-adjusting calibration mechanism is introduced to ensure the localization model remains responsive to system and environmental variations. The efficacy of TransAoA is validated through experiments in a signal-rich laboratory setting that emulates industrial conditions, and the results are benchmarked against existing methods to demonstrate its superiority.

Therefore, the main contributions of this study are fourfold:

- 1) An AoA-based indoor localization method framework, TransAoA, is developed to improve the overall performance of BLE-based IPSs, which can serve as a valuable reference model for industry practitioners to adopt, enabling rapid and effective implementation of high-accuracy indoor localization solutions.

- 2) A quality filtering technique is designed to exclude substandard I/Q samples based on the quality analysis that mainly considers signal amplitude and phase difference to elevate the accuracy and reliability of localization.

- 3) A specialized feature extraction method is proposed to extract multi-domain features from I/Q data, which reinforces the generalization ability of the AoA estimation model and also expedites its training convergence.

- 4) A Transformer-based deep learning model, equipped with a multi-head attention mechanism, is constructed to capture the intricate relationship between I/Q data and AoA, significantly enhancing the precision and robustness of AoA predictions.

The remainder of this paper is structured as follows. Section II provides a literature review of BLE-based indoor localization, focusing on AoA techniques and machine learning applications. An overview of the TransAoA method framework is introduced in Section III. This section is followed by an elaboration of the BLE-based indoor positioning methodology in Section IV. Section V dwells on the experimental setup and analyzes the results. The paper concludes with Section VI, summarizing the findings and pointing out several potential directions of future research.

II. RELATED WORK

Traditional BLE-based positioning approaches, relying on theoretical and empirical formulas to estimate location from parameters like RSSI [25], time of flight (ToF) [26], and phase difference [27], are often susceptible to performance degradation due to multipath propagation and environmental dynamics. To overcome these limitations, researchers have explored the integration of machine learning (ML) techniques [28]. For instance, Philip *et al.* [29] demonstrated the superior localization accuracy of k-nearest neighbor (k-NN) compared to Support Vector Regression (SVR) and Decision Tree (DT) using BLE 5.1. Kriz *et al.* developed a weighted k-NN in signal space algorithm [30] to further enhance the accuracy. Kamal *et al.* [31] proposed a method combining recursive continuous wavelet transform for feature extraction with the extreme gradient boosting machine for location classification. However, despite these advancements, ML techniques for BLE-based positioning still face challenges. Their reliance on linear assumptions or hand-crafted features extracted from raw BLE signals can limit their ability to capture complex spatial relationships. Therefore, ML-based approaches may struggle to fully address the complexities and dynamics of indoor environments.

Many studies have verified that deep learning is such a promising technique in capturing the complex relationships between signal propagation effects and informative factors, leading to more accurate estimation. For example, Cui *et al.* [32] designed a feedforward neural network (FNN) with a

> TIM-24-05298R1<

random vector functional link and demonstrated its superiority over several traditional ML methods in terms of both localization accuracy and generalization ability. Babakhani *et al.* [33] explored the use of recurrent neural networks (RNNs) in dynamic indoor environments, achieving notable improvements in tracking moving objects. Turgut and Kakisim [34] proposed a hybrid model combining convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to bolster accuracy and robustness in complex indoor settings. Talla-Chumpitaz *et al.* [35] converted RSSI samples into image representations and applied CNN for location classification, with an evolutionary algorithm used to find the best combination of results. Ho *et al.* [36] presented a hybrid learning approach that incorporated unsupervised learning, supervised learning, and genetic algorithms, drawing on classical ML techniques and neural networks to refine distance estimation accuracy. Despite great strides made in BLE-based indoor localization through these models, a research gap persists in the integration of advanced artificial intelligence frameworks, such as Transformers [37-39].

Unlike traditional neural networks, Transformers can handle sequential data more efficiently and capture long-range dependencies using multi-head attention that enables the model to simultaneously focus on information from various representation subspaces at any positions [40]. This method can potentially offer superior performance in dynamic and intricate indoor environments and also promote computational efficiency due to parallelization. Therefore, this work aims to bridge this gap by applying a Transformer framework to AoA estimation, leveraging its capability to process high-dimensional I/Q samples. To the best of our knowledge, this study represents a pioneering effort to establish a comprehensive methodology framework for deep learning-enabled AoA localization, taking advantage of enhanced accuracy, robustness and cost-effectiveness.

III. TRANSAOA FRAMEWORK

This section provides an introduction to the TransAoA framework that combines AoA and deep learning techniques to realize indoor localization based on BLE technology. This framework mainly consists of five parts, from hardware configuration to localization, as shown in Fig. 1. Furthermore, the effects of key factors on positioning performance are discussed here.

A. Hardware Configuration

In AoA-based IPSs, the precision and dependability of localization are profoundly impacted by the hardware configuration, particularly the arrangement of antenna arrays [41]. Thereinto, the form of antenna arrays, the polarization of antennas, and the spacings between antennas emerge as pivotal elements that necessitate meticulous consideration to enhance AoA estimation performance. It is important to note that a set of I/Q data comprises two segments corresponding to signal reception phases, namely reference and sampling periods. During the reference period, one antenna is designated as a

reference to gather eight pairs of I/Q samples at a 1μs frequency, followed by the sampling period, during which antennas are switched in sequence to capture signals. Consequently, the form of antenna arrays is instrumental in defining the directionality and spatial resolution of AoA estimation. Common forms include linear, planar (e.g., L-shape and rectangular), and circular arrays. Furthermore, antenna polarization, which denotes the orientation of the electromagnetic wave's electric field vector in space, is crucial for optimizing signal reception quality. In practice, linear and circular polarization techniques are employed. Lastly, in line with AoA estimation theory, antenna spacing is critical to acquire phase difference, and a proper one can help prevent grating lobes while ensuring a compact array size. This spacing is recommended to be approximately half the wavelength.

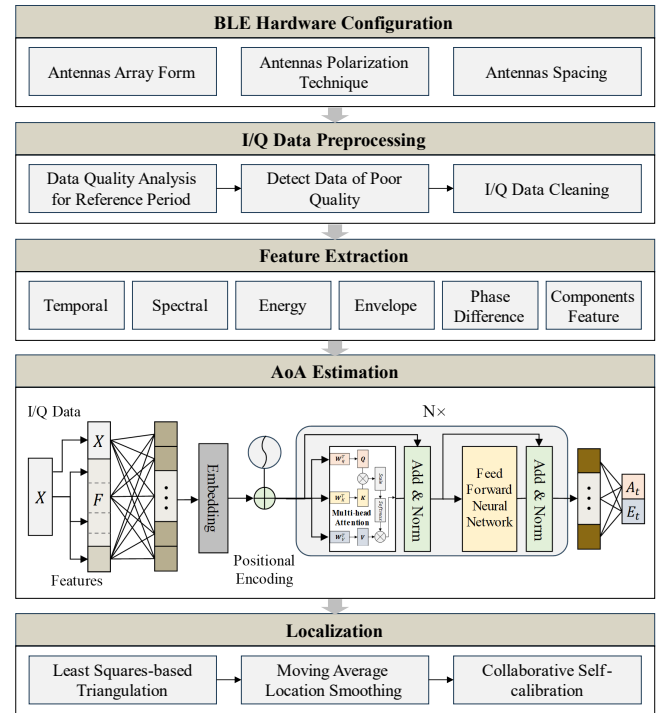


Fig. 1. The framework of TransAoA.

B. I/Q Data Preprocessing

In addition to environmental influences, the hardware components of an AoA receiver inherently affect the quality of I/Q values. The conversion of BLE signals into I/Q values involves a series of components, including the antenna, radio frequency (RF) switch, local oscillator (LO), 90-degree phase shifter, low-pass filter, and analog-to-digital converter (ADC). Each of these components has the potential to introduce noise and bias into the results. However, low-quality data compromises the effectiveness and efficiency of AoA estimation, necessitating the implementation of data quality analysis to identify and eliminate flawed data. Given that the reference period characterizes a continuous sampling process, quality analysis can be conducted based on the data acquired during this period. When eight pairs of normal data are plotted

on a rectangular coordinate system, each pair ideally forms a 90-degree angle with its adjacent sets and is arranged in a counter-clockwise manner. Theoretically, values within the same quadrant should overlap. Nonetheless, due to variations in amplitude, frequency, and phase, such perfect alignment is often not achieved. It is, therefore, imperative to examine the overall distribution of the data and the variations within each pair to evaluate the quality of each data set. Should the distribution pattern diverge from the anticipated norm or if the variations are excessive, the data set will be deemed invalid and excluded. Furthermore, concerning the input to the AoA estimation neural network, the segment corresponding to the reference period will also be omitted for each set of I/Q sample within this framework to enhance the generalizability of the prediction model.

C. Feature Extraction

Feature extraction in this context refers to the process of transforming raw I/Q samples into a comprehensive set of features that can be effectively used by a deep learning model. Previous research predominantly leveraged neural networks' inherent ability to autonomously identify and derive features from raw data, which, however, imposed constraints on further improvements in predictive performance and interpretability of the outcomes. In response to these limitations, this study introduces a tailored feature extraction approach to fortify AoA estimation accuracy. This approach involves the extraction of multiple attributes from I/Q samples, encompassing temporal, spectral, energy, envelope, phase difference, and I/Q component features. Specifically, statistical attributes such as mean, variance, skewness, and kurtosis of the I/Q samples are employed to summarize the data distribution with respect to the temporal domain, while Fourier transforms are utilized to delineate frequency characteristics. The energy metric is indicative of the signal strength or power level, calculated through the summation of the squares of I and Q components. Due to the computational efficiency and robustness, the absolute value of a complex number is adopted as an envelope feature to depict signal shape changes. Additionally, this method incorporates features based on the phase difference between consecutive samples and the individual I/Q components. This feature extraction method is expected to significantly enhance computational efficiency, interpretability, and predictive accuracy.

D. AoA Estimation

Experimental evidence suggests that incorporating a composite input vector of I/Q samples and additional features into the AoA prediction model improves accuracy. Initially, this amalgamated input vector is subjected to normalization to mitigate the impact of scale discrepancies. Subsequently, a fully-connected neural network (FCNN) is employed to integrate information and partially transform features. This is followed by embedding the transformed features into a higher-dimensional space, enabling the model to discern more intricate representations. Upon integration into the Transformer model, positional encodings are appended to the embeddings to compensate for the lack of inherent sequential data processing

regarding the model. These encodings furnish the model with spatial information of each feature in a sequence, thereby facilitating the subsequent implementation of the attention mechanism. The architecture of the Transformer typically encompasses both encoder and decoder components. However, in this context, only the encoder is utilized, with an FCNN supplanting the decoder to accomplish regression towards azimuth and elevation angles. The encoder consists of multiple layers of multi-head self-attention, FCNN, layer normalization, and residual connection. The quantity of encoder layers is adjustable, contingent on the intricacy of the task. The multi-head attention mechanism empowers the model to prioritize different segments of the I/Q feature sequence during AoA estimation, concentrating on the most relevant features. Consequently, this approach substantially augments the accuracy and dependability of AoA estimation.

E. Localization

Based on the acquired AoA values, the least squares (LS) method [42] is applied to calculate the spatial coordinates of the target by minimizing an error function related to pertinent angles. To enhance positioning accuracy and robustness, three BLE anchors are strategically deployed to work in concert. Moreover, the ultimate location determination is refined through the averaging of multiple sequential prediction outcomes, including historical data, facilitated by a moving time window. This approach ensures smoother and more precise estimations. To accommodate environmental variations and sustain long-term positioning accuracy, the system leverages fixed BLE tags at predetermined locations as calibration points, enabling online self-calibration of the BLE positioning model. Specifically, the system adaptively modifies the regularization target of the positioning model, iteratively refining it to attain a global optimum within the updated constraints. If the adjustments in regularization fail to satisfy the positioning criteria, the positioning model would take dynamical expansion, with appropriate modifications to the network structure. Through iterative calibration and model updates, the system can maintain optimal performance amidst evolving environmental conditions.

IV. METHODOLOGY

This section elaborates on the mathematical models underpinning the methods involved within the TransAoA framework. The initial focus is on the theory of I/Q sampling, which lays the foundation for subsequent data quality analysis and multi-feature construction. Then, the section delves into the implementation of AoA-based localization using deep learning techniques.

A. I/Q Sampling and Quality Analysis

I/Q sampling is a technique for capturing and processing continuous-time analog RF signals to extract their I and Q components, then effectively capturing both amplitude and phase information. Fig. 2 provides a simplified circuit diagram illustrating the key components enabling I/Q sampling. The process involves receiving the BLE RF signal, converting it to

an analog signal, and multiplying it with a local oscillator signal in two paths with a 90° phase shifter. Ideally, the LO signal frequency aligns with the RF carrier frequency, resulting in an upper sideband frequency and a lower sideband frequency that matches the constant tone extension (CTE) baseband frequency. Subsequently, low-pass filters eliminate the high-frequency upper sideband signals, allowing only the signals at the CTE baseband frequency to pass. An ADC digitizes the signals to produce I and Q samples.

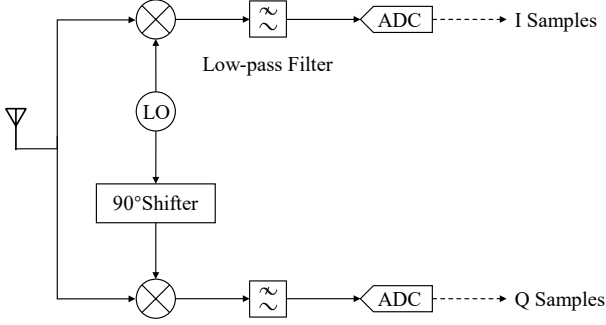


Fig. 2. A simplified circuit for BLE I/Q sampling.

Several factors can influence the quality of I/Q sampling, leading to potential errors. A significant factor is the carrier frequency offset (CFO) between the LO signal and the BLE carrier signal, which can cause a systematic residual and deviate the final samples from the intended baseband frequency. Another issue is quantization error, which occurs when the signal amplitude is too low, reducing the signal-to-noise ratio (SNR) and potentially leading to inaccuracies in ADC sampling outcomes. Additionally, random noise from various electrical components can alter the amplitude and phase of the signal, introducing cumulative errors in I/Q samples that are difficult to detect or cancel. Therefore, it necessitates the data quality analysis and discard substandard samples to maintain AoA estimation efficiency.

Like traditional algorithms to address CFO by processing the I/Q samples from the reference period, this study proposes a quality filtering method predicated on the reference samples. This approach establishes two criteria for quality assessment, taking into account the signal's amplitude and phase difference. First, low amplitude is identified as indicative of error, as explained above. Let the I/Q samples in the reference period be denoted as $X_t = \{(i_{t,1}, q_{t,1}), (i_{t,2}, q_{t,2}), \dots, (i_{t,8}, q_{t,8})\}$. The amplitude of each sample vector can be expressed as:

$$A_{t,n} = \sqrt{i_{t,n}^2 + q_{t,n}^2}, n = 1, 2, \dots, 8 \quad (1)$$

A threshold value $T_{amplitude}$ is settled to enable the filtering process:

$$a_{X_t} = \begin{cases} 0 & \max A_{t,n} \leq T_{amplitude} \\ 1 & \max A_{t,n} > T_{amplitude} \end{cases} \quad (2)$$

If a_{X_t} equals zero, it needs to discard this set of data due to insufficient quality. Pertaining to the second criterion, given that phase difference is intrinsically linked to the AoA estimation, sample quality is appraised through this differential.

Specifically, each I/Q sample pair should be converted into a phase angle $\theta_{t,i}$, computed by:

$$\theta_{t,n} = \arctan\left(\frac{q_{t,n}}{i_{t,n}}\right), \theta_{t,n} \in (-\pi, \pi] \quad (3)$$

As a signal propagates space, its phase evolves continuously. However, phase measurements are typically confined to a finite range, introducing potential ambiguity in the true phase difference. Phase unwrapping is thus imperative for generating a seamless phase profile. In this study, with phase rotation assumed to be counterclockwise, the phase for each sample is updated as:

$$\theta'_{t,n} = \begin{cases} \theta_{t,n} & \text{if } \theta_{t,n} \leq \theta'_{t,n-1} \\ \theta_{t,n} - 2\pi p & \text{if } \theta_{t,n} > \theta'_{t,n-1} \end{cases} \quad (4)$$

Here, the p represents the number of complete cycles traversed by the signal. To exclude anomalous phase differences, upper and lower bounds are defined, ensuring that significant deviations from the expected phase difference, ideally $\pi/2$, are filtered out:

$$a_{X_t} = \begin{cases} 1 & T_{lower} \leq |\theta'_{t,n} - \theta'_{t,n-1}| \leq T_{upper} \\ 0 & \text{else} \end{cases} \quad (5)$$

By applying these criteria, the dataset is refined based on the quality of the I/Q samples collected during the reference period to elevate the reliability of subsequent prediction.

B. Multi-feature Extraction

Prior to the AoA estimation model development, a variety of features is constructed from the I/Q data within the sampling period to expedite model convergence during training and promote prediction accuracy. Each I/Q sample pair can be denoted as $X_t = \{(i_{t,1}, q_{t,1}), (i_{t,2}, q_{t,2}), \dots, (i_{t,n}, q_{t,n})\}$. Given the I and Q components originating from a complex signal, each unit in the I/Q vector is transformed into a complex representation, namely $X_t = \{i_{t,1} + j \cdot q_{t,1}, i_{t,2} + j \cdot q_{t,2}, \dots, i_{t,n} + j \cdot q_{t,n}\}$. In this study, six distinct feature groups are extracted from the complex I/Q vectors.

1) Temporal feature: This category concerns statistical attributes of the signal. The real and imaginary parts of the mean of single I/Q vector constitute two features, which are formulated by:

$$F_1 = \frac{1}{N} \sum_{k=1}^N (i_k + j \cdot q_k), f_1 = \text{Re}(F_1), f_2 = \text{Im}(F_1) \quad (6)$$

The standard deviation of I/Q vector is given by:

$$f_3 = F_2 = \sqrt{\frac{1}{N-1} \sum_{k=1}^N ((i_k + j \cdot q_k) - F_1)^2} \quad (7)$$

Additionally, skewness and kurtosis are used to depict the shape of a probability distribution, which provide information about the asymmetry and peakedness of the distribution, respectively. These two features are defined as:

$$F_3 = \frac{1}{NF_2^3} \sum_{k=1}^N ((i_k + j \cdot q_k) - F_1)^3, f_4 = \text{Re}(F_3), f_5 = \text{Im}(F_3) \quad (8)$$

$$F_4 = \frac{1}{NF_2^4} \sum_{k=1}^N ((i_k + j \cdot q_k) - F_1)^4, \quad f_6 = \text{Re}(F_3), f_7 = \text{Im}(F_3) \quad (9)$$

2) Spectral feature: These features reveal the energy distribution across different frequencies within the signal. This study applies a Fourier transform to the I/Q vectors to obtain the frequency spectrum, as modeled by:

$$F_5 = \sum_{k=0}^{N-1} (i_k + j \cdot q_k) \cdot e^{-j2\pi f_k / N} \quad (10)$$

Hereinto, the mean and standard deviation of both the magnitude and phase of the frequency-transformed I/Q vector provide four spectral features: f_8 , f_9 , f_{10} , and f_{11} .

3) Energy feature: This feature is a measure of the total energy or power contained within a signal. This feature provides a straightforward metric for signal intensity, making it widely used in signal processing and machine learning. It is calculated by:

$$f_{12} = F_6 = \sum_{k=1}^N (i_k^2 + q_k^2) \quad (11)$$

4) Envelope feature: The envelope captures the amplitude variations of a signal over time and is calculated by taking the absolute value of each set of I/Q samples as:

$$F_7 = |i_k + j \cdot q_k| = \sqrt{i_k^2 + q_k^2} \quad (12)$$

The mean, standard deviation, maximum, and minimum values of the absolute vector serve as four envelope features: f_{13} , f_{14} , f_{15} , and f_{16} .

5) Phase difference feature: These features have a close relation with AoA prediction in line with the theory. The mean and standard deviation of a vector of phase difference give birth to f_{17} and f_{18} . This difference is estimated by the following formula:

$$F_8 = \arctan(q_{k+1} / i_{k+1}) - \arctan(q_k / i_k) \quad (13)$$

6) I/Q component feature: The separate mean and standard deviation of the in-phase and quadrature components are used to produce features f_{19} , f_{20} , f_{21} , and f_{22} .

In total, twenty-two features are crafted and then integrated with the I/Q sample vector to comprise the input for the AoA estimation neural network, which is denoted as

$$X'_t = \{i_{t,1}, q_{t,1}, \dots, i_{t,n}, q_{t,n}, f_{t,1}, \dots, f_{t,22}\}.$$

C. Transformer model for AoA Estimation

It should be clarified that each single numerical value of I and Q data is a scalar. A sequence of I/Q scalars collected within one round of reference period and sampling period is parallelly combined with these feature values to construct a vector as the input to AoA estimation. In the Transformer architecture, embeddings are pivotal for transforming input tokens into vectors of continuous values that the model can process. In the feature input, simplified as $X' = \{x_1, x_2, \dots, x_{n+22}\}$, each data point could be mapped to a high-dimensional matrix with rows and columns through an embedding function

$Em(x_k) = \{e_{k,1}, e_{k,2}, \dots, e_{k,m}\}$. The embedding matrix is characterized by $m * (n + 22)$ dimensions.

Subsequently, due to the Transformer's inability to inherently process sequential data, it entails the integration of positional encodings with each point in the sequence of input embeddings. For each position k and dimension m in the embedding, the positional encoding $PE(k, m)$ is defined as follows:

$$\begin{cases} PE(k, 2i) = \sin\left(\frac{k}{s^{2i/m}}\right) \\ PE(k, 2i+1) = \cos\left(\frac{k}{s^{2i/m}}\right) \end{cases} \quad (14)$$

The s represents a specific scalar. These equations alternate between sine and cosine functions for even and odd indices, respectively, allowing the model to distinguish between different positions. The final input representation for each data point in the sequence is obtained by consolidating its input embedding with the positional encoding.

Next, these representations are fed into encoder layers, the number of which is dependable. An encoder layer primarily consists of two sub-layers, a multi-head attention function and a FCNN. The multi-head attention function features three key elements, including matrices of queries Q , keys K , and values V . It computes the dot products of the query with all keys and applies a softmax function to derive the weights on the values. The output matrix is given by:

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (15)$$

The d_k denotes the dimension of queries and keys. To mitigate the issue of the dot products growing large in magnitude, which leads to the softmax function entering regions with extremely small gradients, the dot products are scaled by $1/\sqrt{d_k}$. In contrast to single-head attention, multi-head attention enables the model to concurrently attend to various aspects at different positions, fostering a more nuanced understanding of the relationships between data points and enhancing parallelism. The final results are acquired by a concatenation of all the heads, as depicted by:

$$MH(Q, K, V) = \text{Concat}\left(\text{Attn}(QW_1^Q, KW_1^K, VW_1^V), \dots, \text{Attn}(QW_h^Q, KW_h^K, VW_h^V)\right) \quad (16)$$

Here, the projection matrices $W_h^Q \in \mathbb{R}^{m \times d_k}$, $W_h^K \in \mathbb{R}^{m \times d_k}$, $W_h^V \in \mathbb{R}^{m \times d_v}$ are allocated to queries, keys and values, respectively, and h represents the number of attention heads.

In addition to the attention sub-layer, each encoder layer contains a FCNN, which is applied to each position separately and identically. The FCNN is aimed to map the input dimension from one space to a higher-dimensional space, and then back to the original dimension or another appropriate dimension through two linear transformations with a ReLU activation in between. Hence, this process facilitates nonlinear transformation, enabling the model to capture more complex feature relationships and thereby augmenting its expressive capability. The FCNN is structured as follows:

> TIM-24-05298R1<

$$\bar{X}_{l+1} = \bar{W}_2^l \cdot \max(0, \bar{W}_1^l \cdot \bar{X}_l + \bar{b}_1^l) + \bar{b}_2^l \quad (17)$$

The $\bar{W}_1^l$ and $\bar{W}_2^l$ are weight matrices assigned to the input at the l^{th} layer for two linear shifts, with $\bar{b}_1^l$ and $\bar{b}_2^l$ as the corresponding bias matrices. Besides those two sub-layers, each sub-layer is enhanced with residual connections and layer normalization to foster model convergence and stabilize training.

To optimize model parameters, it is essential to identify an appropriate loss function. The mean square error (MSE) is utilized as the loss function in response to the application of the FCNN. The adaptive moment estimation (Adam) algorithm is employed for optimization due to its efficacy across various parameter initializations. This algorithm aims to iteratively update parameters in the direction of gradient descent to minimize loss.

D. Triangulation

This study aims to achieve indoor localization through triangulation that leverages the AoA values garnered from signals transmitted between the target and multiple BLE anchors. By acquiring AoA measurements from a minimum of three BLE anchors, the spatial coordinates $P_t = (x_t, y_t, z_t)$ of the target at time t are determined by the minimization of an error function. This function quantifies the discrepancy between the measured and predicted angles. Mathematically, the LS problem can be articulated as:

$$\text{Min} : \sum_{i=1}^n \left[\left(\theta_{t,i} - \hat{\theta}_i(x_t, y_t, z_t) \right)^2 + \left(\varphi_{t,i} - \hat{\varphi}_i(x_t, y_t, z_t) \right)^2 \right] \quad (18)$$

where n denotes the number of BLE anchors. The $\theta_{t,i}$ and $\varphi_{t,i}$ are the measured elevation and azimuth angles from the i^{th} BLE anchor to the target at time t , respectively. Conversely, the $\hat{\theta}_i(x_t, y_t, z_t)$ and $\hat{\varphi}_i(x_t, y_t, z_t)$ are the predicted elevation and azimuth angles grounded on the estimated position of target relative to the i^{th} anchor. The goal is to ascertain the values of x , y , and z that minimize this sum, thereby providing the best estimate of the target's position based on the AoA measurements.

Furthermore, a moving average technique is applied to smooth out short-term fluctuations and underscore longer-term trends in location data. This approach involves averaging a sequence of consecutive position estimates to yield a single and smoothed location point, as illustrated below:

$$\bar{P}_t = \frac{P_t + P_{t-1} + \dots + P_{t-w+1}}{T_w} \quad (19)$$

The T_w signifies the length of the time window for averaging. This method proves particularly useful when dealing with noisy measurements or when the target's position is subject to rapid changes, engendering accurate and reliable location estimates.

V. EXPERIMENT AND EVALUATION

To validate and evaluate the efficacy of the TransAoA method, experiments were conducted in a laboratory that is settled to

mimic an industrial environment, characterized by the presence of numerous signals that could potentially cause interference. Besides, a comparative analysis was conducted to demonstrate the superiority of the proposed method relative to other benchmark approaches.

A. Experimental Settings

The laboratory environment incorporates various physical objects and personnel to simulate potential sources of signal reflection and multipath interference, which are common in industrial environments, as displayed in Fig. 3. The layout is designed to introduce challenges typical of industrial settings, such as non-line-of-sight (NLOS) conditions and dynamic obstacles. The designated experimental area measures 7.2m by 5.4m within the laboratory, avoiding fixed arrangements, as depicted in Fig. 4. Below are detailed the hardware configuration and the establishment of datasets.

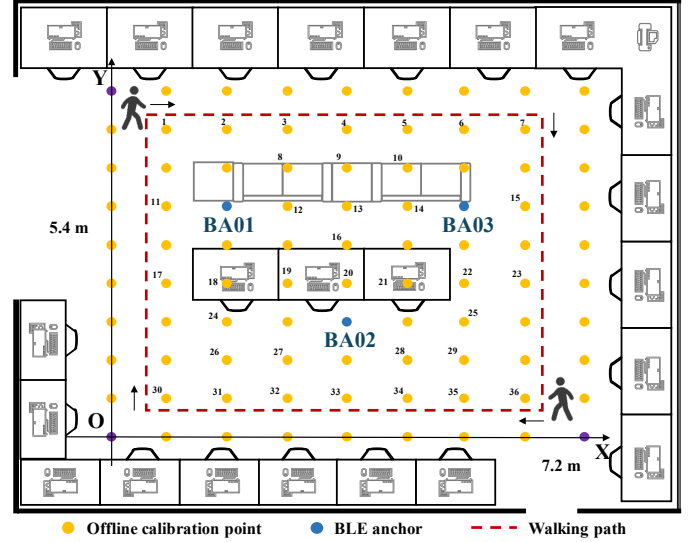


Fig. 3. The floor plan of experimental settings.

1) Hardware configuration: The experimental setup utilizes BLE tags based on the BRD4184B board from Silicon Labs, configured to periodically transmit data packets with CTE via data channels excluding 37, 38, and 39, as specified by the BLE 5.1 protocol. For signal reception, a dual-polarized antenna array radio board, the BRD4185A, also from Silicon Labs, is employed. This board features a rectangular antenna array in a 4×4 configuration, with an inter-antenna spacing of 0.036m. Moreover, it is equipped with both horizontal and vertical antenna polarizations to facilitate the measurement of radiated powers across fundamental frequencies and harmonics. An antenna board interfaces with a gateway composed of a Raspberry Pi 3B+, which provides the necessary computing capabilities for I/Q sample collection, data preprocessing, and transmission. Three anchors are mounted on tripods at a height of two meters within the experimental area, as shown in Fig. 3 and 4. Data transmission from the gateway to a server is accomplished via the Message Queuing Telemetry Transport (MQTT) protocol, which is well-suited for industrial applications due to its effectiveness in environments with constrained resources, low bandwidth, high latency, or

> TIM-24-05298R1<

unreliable networks. Lastly, the Transformer model for AoA estimation is enforced using PyTorch, and the training process is executed on a workstation furnished with an Intel(R) Xeon(R) Gold 6230R CPU @ 2.10GHz, featuring 8 cores.



Fig. 4. The deployment of BLE indoor positioning system.

2) Dataset establishment: The experimental datasets are categorized into two primary groups. The first group involves collecting data from fixed points where BLE tags are mounted at a height of one meter and remain stationary throughout the data collection period. The second group gathers data from

mobile BLE tags. An experimenter carries these tags and walks along a predefined path within the experimental area. For the fixed points, the experimental setup includes 9×10 points along both the x and y axes. However, data collection focuses on a total of 36 points to support the exploration of neural network generalization, as illustrated in Fig. 4. Each of these points contributes 3000 data entries, encompassing I/Q samples, RSSI, channel information, transmitter and anchor identifications, and timestamps. The I/Q samples comprise 8 pairs of I/Q data for the reference period and 74 pairs of dually-polarized I/Q data for the sampling period, reflecting cyclic sampling across 16 antennas. Overall, the dataset, exceeding 100,000 entries, is segmented into training, validation, and testing subsets for AoA estimation, following an 8:1:1 ratio. Regarding the dynamic target localization, an experimenter carrying a BLE tag walks a predetermined route three times at a normal pace, yielding 9557 data entries. Concurrently, three anchors work in unison to localize the moving target in real-time using individual AoA estimation models.

B. Experimental Results and Analysis

In this study, several prominent deep learning techniques for angulation are employed, including FNN, CNN, and LSTM, to perform a comparative analysis with the proposed TransAoA method. Additionally, the PDDA approach, which has demonstrated superiority over the traditional MUSIC algorithm, is included for comparison. The impact of I/Q data quality filtering and specific multi-feature extraction on AoA prediction accuracy is further investigated in the following experiments.

First, the TransAoA is benchmarked against other four methods, involving data quality filtering and feature extraction

TABLE I
RESULTS OF AOA ESTIMATION WITH QUALITY FILTERING AND FEATURE EXTRACTION

Method	Azimuth					Elevation					TTPP (ms)	PT (ms)
	MSE	MAE	Within 5°C	Within 10°C	Within 20°C	MSE	MAE	Within 5°C	Within 10°C	Within 20°C		
PDDA+QF	114.62	9.78	5.52%	62.74%	95.58%	43.91	5.95	39.81%	89.92%	100%	\	857
FNN+QF+FE	11.81	2.52	89.48%	98.85%	99.91%	1.51	0.94	99.66%	99.97%	100%	2.41	1
CNN+QF+FE	27.82	3.70	74.78%	93.88%	99.63%	2.09	1.08	98.79%	99.97%	100%	2.16	52
LSTM+QF+FE	6.51	1.91	95.52%	99.63%	100%	0.80	0.65	99.91%	100%	100%	1.89	19
TransAoA	3.62	1.35	98.85%	99.91%	99.97%	0.61	0.53	99.97%	100%	100%	1.53	31

Note: QF: quality filtering; FE: feature extraction; TTPP: training time per parameter; PT: prediction time.

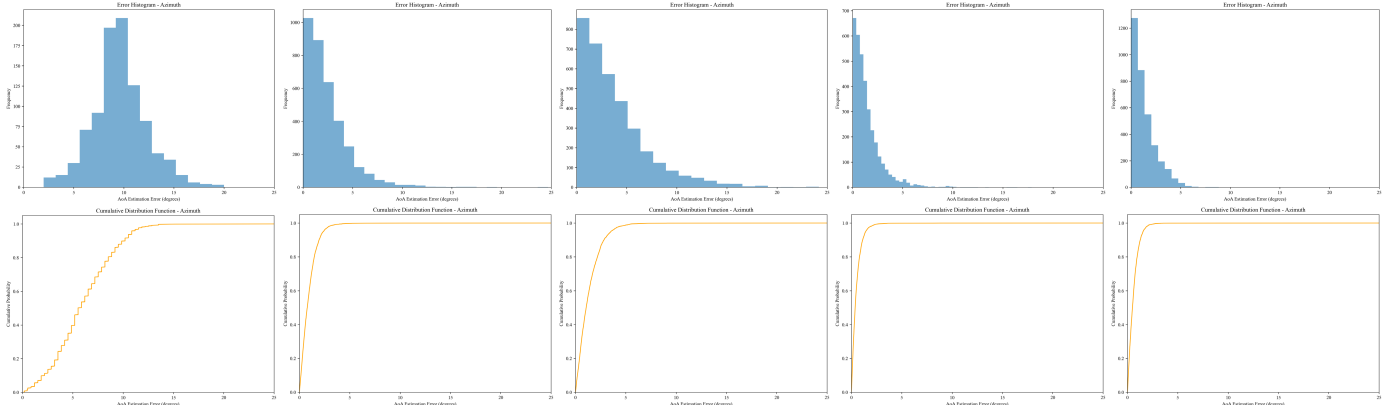


Fig. 5. The error histogram and CDF of azimuth estimation.

> TIM-24-05298R1<

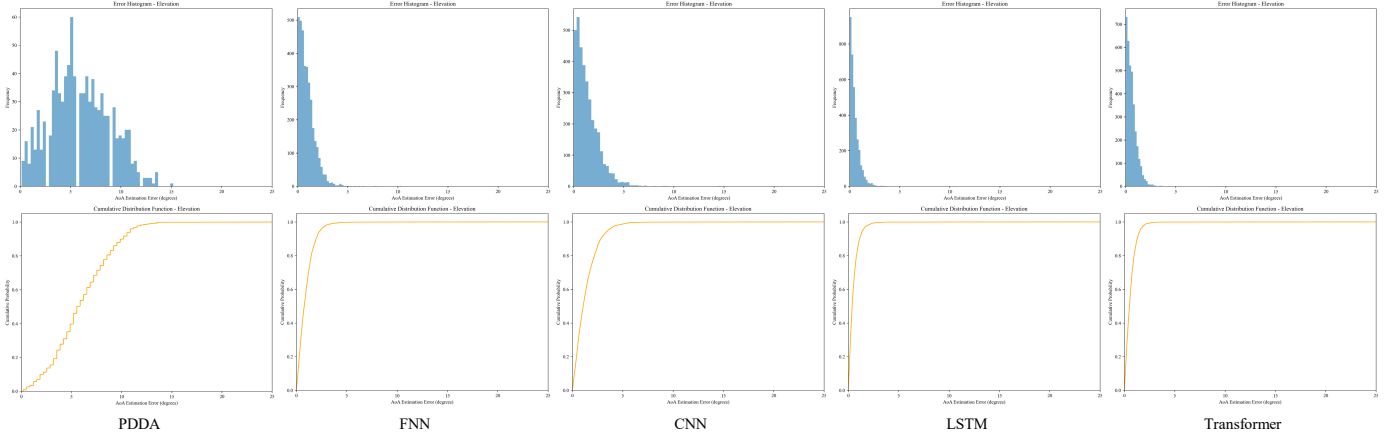


Fig. 6. The error histogram and CDF of elevation estimation.

processes, across metrics of azimuth and elevation estimation accuracy. Evaluation criteria such as MSE, mean absolute error (MAE), training time, prediction time, and the distribution of specific angular errors are utilized to assess model performance. The aggregated findings are presented in Table I. It is evident that the TransAoA surpasses the other methods in most respects. Specifically, the TransAoA achieves an MSE of 3.62 and an MAE of 1.35 for azimuth estimation, and 0.61 MSE and 0.53 MAE for elevation. The LSTM is the next best performing method, with MSE values of 6.51 and 0.80, and MAE values of 1.91 and 0.65 for azimuth and elevation, respectively. The FNN demonstrates fewer errors than the CNN but more than the LSTM, suggesting a lack of graph-like structure inherent in I/Q samples. Conversely, PPDA exhibits the least favorable performance, with significantly larger errors in both azimuth and elevation estimation. In terms of computational efficiency, also considering model size, TransAoA shows the shortest training time per parameter (TPPP) to maintain convergence, averaging 1.53ms, whereas it would sacrifice the prediction time, taking 31ms. Besides, PPDA, despite its simplicity, exhibits a significantly longer prediction time compared to all deep learning techniques. This suggests that the inference efficiency makes learning methods more suitable for real-time localization applications.

In addition, histograms and cumulative distribution functions (CDFs) of the absolute angular errors for azimuth and elevation are depicted in Fig. 5 and 6, respectively. Note that the TransAoA achieves the most precise estimations, with 98.85% of azimuth errors and 99.97% of elevation errors falling within 5 degrees. Furthermore, it attains 99.91% of azimuth errors within 10 degrees and 99.97% within 20 degrees, while maintaining 100% accuracy within 10 degrees for elevation. A particular emphasis is placed on the LSTM that reaches 100%

prediction accuracy within 20 degrees for azimuth, surpassing the TransAoA. This finding highlights the generalization capability of the LSTM in handling deviated data. The LSTM also nearly matches the performance of the TransAoA in elevation estimation.

Furthermore, Fig. 7 illustrates the training loss trajectories of the four deep learning methods throughout the optimization process. The TransAoA exhibits rapid convergence with minimal loss at the start. Although the LSTM experiences notable fluctuations early in training, it maintains a low loss in the end. The FNN shows robust convergence despite an initially high loss, whereas the CNN model undergoes persistent oscillations over several epochs.

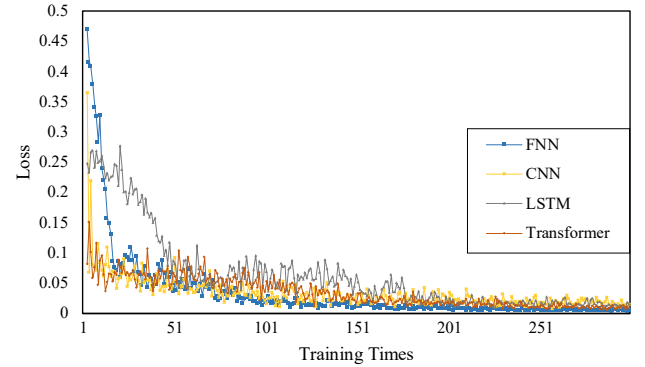


Fig. 7. Comparison of the training loss in the optimization process.

Second, the influence of data filtering in relation to the quality of I/Q samples on AoA estimation accuracy is examined. According to the findings presented in Table II, the absence of quality filtering leads to an increase in MSE and MAE values,

TABLE II
RESULTS OF AOA ESTIMATION WITH ONLY FEATURE EXTRACTION

Method	Azimuth					Elevation				
	MSE	MAE	Within 5°C	Within 10°C	Within 20°C	MSE	MAE	Within 5°C	Within 10°C	Within 20°C
PPDA	128.47	11.04	3.28%	59.65%	92.87%	57.63	6.81	35.46%	87.91%	99.85%
FNN + FE	15.44	2.54	88.34%	98.34%	99.61%	1.92	1.03	99.49%	99.94%	100%
CNN + FE	27.94	3.78	73.52%	93.77%	99.54%	3.15	1.35	98.40%	99.97%	100%
LSTM + FE	7.65	2.08	94.43%	99.34%	99.88%	1.12	0.71	99.66%	99.93%	100%
Trans + FE	6.77	1.87	98.02%	99.83%	99.88%	0.64	0.61	99.85%	99.93%	100%

TABLE III
RESULTS OF AoA ESTIMATION WITH ONLY QUALITY FILTERING

Method	Azimuth					Elevation				
	MSE	MAE	Within 5°C	Within 10°C	Within 20°C	MSE	MAE	Within 5°C	Within 10°C	Within 20°C
FNN + QF	18.74	2.87	84.92%	97.47%	99.52%	1.79	1.01	99.52%	99.97%	100%
CNN + QF	22.41	3.22	79.98%	95.23%	99.57%	1.76	0.97	99.51%	100.00%	100%
LSTM + QF	7.63	1.88	94.37%	99.29%	99.95%	0.94	0.67	99.78%	99.95%	100%
Trans + QF	5.57	1.51	97.33%	99.78%	99.90%	0.67	0.58	99.88%	99.98%	100%

alongside a general decrease in the accuracy of angle error ratios across different ranges. This contrast with the results shown in Table I underscores the critical role of quality filtering in enhancing estimation precision by eliminating I/Q samples distorted by noise. Moreover, this analysis testifies the quality consistency between reference and sampling data, as the quality assessment exclusively involves reference data. Despite these challenges, the TransAoA method consistently outperforms other techniques in angle prediction. Notably, the FNN is the most affected by the absence of quality filtering, experiencing a 30.74% and 27.15% increase in MSE for azimuth and elevation, respectively, illustrating its limited generalization capability.

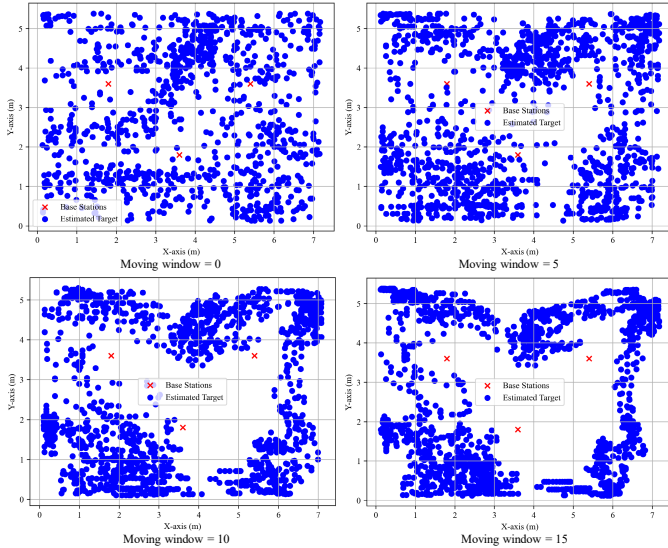


Fig. 8. Comparison of different moving windows for the triangulation of moving targets.

Third, this part explores the effects of a bespoke multi-feature extraction technique on deep learning-based AoA prediction. The Transformer architecture still achieves superior prediction accuracy among the four evaluated methods, as shown in Table III. Distinctively, most methods exhibit an increase in MSE in the absence of prior feature extraction, only the CNN shows a decrease, by 22.41 and 1.76 for azimuth and elevation estimation, respectively. This suggests that the diversity of input data may adversely affect CNN performance. Compared to the results in Table II, the impact of feature extraction is less pronounced than that of quality filtering for the LSTM, whereas it is more critical for both the Transformer and FNN. Without feature extraction, the Transformer demonstrates a more

significant reduction in the accuracy of angle error estimation within 5 degrees than without quality filtering, 1.52% opposed to 0.83% for azimuth. The influence on elevation estimation across all methods is comparatively minor except the CNN, with a maximum deviation of 0.2% for errors within 5 degrees.

Fourth, leveraging the superior angle estimation capabilities of the TransAoA, it is used to triangulate the positions of moving people. An average moving window technique is applied to refine the predicted outcomes. Fig. 8 visualizes the comparative impact of different moving window sizes (0, 5, 10, and 15) on localization performance. The results indicate that the average moving window significantly enhances localization accuracy. Without its application, the predicted locations are dispersed across the map, whereas a more coherent movement path emerges as the window size increases. However, excessively large windows may restrict the range and adaptability of positioning, thereby diminishing accuracy. Based on the analysis, a moving window size of 10 strikes an optimal balance. In conclusion, the localization outcomes validate the effectiveness of the TransAoA as a comprehensive solution for BLE indoor localization on the ground of deep learning-enhanced AoA measurements.

VI. CONCLUSION

This paper presents an innovative BLE-based indoor positioning framework, the TransAoA, that significantly improves the accuracy and robustness of AoA estimation through the application of the Transformer deep learning architecture. The proposed method outperformed traditional and contemporary deep learning techniques in a simulated industrial environment, attaining 98.85% and 99.97% of estimation errors within 5 degrees for azimuth and elevation, respectively. Besides, the integration of a data quality filtering and specialized feature extraction methods enabled the model to readily capture complex signal interactions, leading to rapid convergence and enhanced generalization capabilities. It is anticipated that this technique framework would inspire new ideas for researchers and act as a guide for practitioners to easily replicate in order to meet similar application requirements.

However, a limitation of this work is the reliance on linear triangulation for location estimation. The least square technique might be susceptible to noise and fail to adequately model the inherent non-linearity of localization based on AoAs, potentially harming positioning accuracy. Correspondingly, future research could explore more robust and non-linear approaches to mitigate the impact of noise on AoA-based localization performance. Besides, enhancing the scalability

and adaptability of the TransAoA framework for deployment in diverse industrial environments presents a promising direction. Investigations into the integration of additional environmental and contextual data could further improve the robustness and adaptability of the system. Moreover, extensive field trials in various industrial scenarios would be invaluable to validate the framework's performance and facilitate its adoption in Industry 5.0 applications.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

- [1] Y. Zhu, J. Cheng, Z. Liu, Q. Cheng, X. Zou, H. Xu, Y. Wang, and F. Tao, "Production logistics digital twins: Research profiling, application, challenges and opportunities," *Robot. Cim-int. Manuf.*, vol. 84, pp. 102592, May 2023.
- [2] Z. Zhao, M. Zhang, W. Wu, G. Q. Huang, and L. Wang, "Spatial-temporal traceability for cyber-physical industry 4.0 systems," *J. Manuf. Syst.*, vol. 74, pp. 16-29, Mar. 2024.
- [3] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and Industry 5.0—Inception, conception and perception," *J. Manuf. Syst.*, vol. 61, pp. 530-535, Oct. 2021.
- [4] S. I. Khan, B. R. Ray, and N. C. Karmakar, "RFID localization in construction with IoT and security integration," *Autom. Constr.*, vol. 159, p. 105249, Dec. 2023.
- [5] W. Wu, L. Shen, Z. Zhao, A. R. Harish, R. Y. Zhong, and G. Q. Huang, "Internet of Everything and Digital Twin Enabled Service Platform for Cold Chain Logistics," *J. Ind. Inf. Integr.*, vol. 33, p. 100443, 2023.
- [6] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, and J. A. Castellanos, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Trans. Robot.*, Mar. 2023, doi: 10.1109/TRO.2023.3248510.
- [7] W. Wu, L. Shen, Z. Zhao, M. Li, and G. Q. Huang, "Industrial IoT and long short-term memory network-enabled genetic indoor-tracking for factory logistics," *IEEE Trans. Ind. Inform.*, vol. 18, no. 11, pp. 7537-7548, Nov. Jan. 2022.
- [8] Y. Assayag, H. Oliveira, E. Souto, R. Barreto, and R. Pazzi, "Adaptive path loss model for BLE indoor positioning system," *IEEE Internet Things J.*, Mar. 2023, doi: 10.1109/JIOT.2023.3253660.
- [9] F. J. Aranda, F. Parralejo, F. J. Álvarez, and J. A. Paredes, "Performance analysis of fingerprinting indoor positioning methods with BLE," *Expert Syst. Appl.*, vol. 202, pp. 117095, Sep. 2022.
- [10] M. Comelli, P. Patras, and F. Gringoli, "Dead on arrival: An empirical study of the Bluetooth 5.1 positioning system," in *Proc. 13th Int. Workshop Wireless Netw. Testbeds, Exp. Eval. Charact.*, pp. 13-20, Oct. 2019.
- [11] C. Huang, Y. Zhuang, H. Liu, J. Li, and W. Wang, "A performance evaluation framework for direction finding using BLE AoA/AoD receivers," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3331-3345, Sep. 2020.
- [12] D. Xiao, S. Hu, K. Kang, and H. Qian, "An improved AoA estimation algorithm for BLE system in the presence of phase noise," *IEEE Trans. Consum. Electron.*, vol. 69, Mar. 2023, doi: 10.1109/TCE.2023.3272817.
- [13] Y. Zheng, M. Sheng, J. Liu, and J. Li, "Exploiting AoA estimation accuracy for indoor localization: A weighted AoA-based approach," *IEEE Wirel. Commun. Lett.*, vol. 8, no. 1, pp. 65-68, Jul. 2018.
- [14] L. Cheng and J. Guan, "An indoor localization algorithm based on improved particle filter and directional probabilistic data association for wireless sensor network," *KSII Trans. Internet Inf. Syst.*, vol. 17, no. 11, pp. 3145-3162, Nov. 2023.
- [15] Toasa, F. A., Tello-Oquendo, L., Peñafiel-Ojeda, C. R., & Cuzco, G., "Experimental demonstration for indoor localization based on AoA of Bluetooth 5.1 using software defined radio," in *2021 IEEE 18th Annual Consumer Communications & Networking Conf. (CCNC)*, pp. 1-4, Mar. 2021.
- [16] S. Wielandt, A. Van Nieuwenhuysse, J. P. Goemaere, B. Nauwelaers, and L. De Strycker, "Evaluation of angle of arrival estimation for localization in multiple indoor environments," in *2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS)*, pp. 36-43, IEEE, Nov. 2014.
- [17] X. Yang, Z. Wu, and Q. Zhang, "Bluetooth indoor localization with Gaussian-Bernoulli restricted Boltzmann machine plus liquid state machine," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-8, Jan. 2022.
- [18] G. Guo, R. Chen, K. Yan, Z. Li, L. Qian, S. Xu, ..., and L. Chen, "Large-scale indoor localization solution for pervasive smartphones using corrected acoustic signals and data-driven PDR," *IEEE Internet Things J.*, vol. 10, Mar. 2023, doi: 10.1109/JIOT.2023.3263551.
- [19] Y. Dong, T. Arslan, and Y. Yang, "Magnetic disturbance detection for smartphone-based indoor positioning systems with unsupervised learning," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-11, Mar. 2022.
- [20] L. Li, X. Guo, and N. Ansari, "SmartLoc: Smart wireless indoor localization empowered by machine learning," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6883-6893, Aug. 2020.
- [21] Z. Liu, L. Chen, X. Zhou, Z. Jiao, G. Guo, and R. Chen, "Machine learning for time-of-arrival estimation with 5G signals in indoor positioning," *IEEE Internet Things J.*, vol. 10, no. 11, pp. 9782-9795, Jun. 2023.
- [22] C. Cao, Y. Zhu, L. Zhao, D. Sun, Z. Liu, and Y. Lian, "An accurate positioning method based on time-division strategy for indoor moving target," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1-12, June 2023.
- [23] 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.
- [24] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 3, pp. 2568-2599, Apr. 2019.
- [25] B. Shin, J. H. Lee, C. Yu, H. Kyung, and T. Lee, "Received signal strength-based robust positioning system in corridor environment," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-15, July 2022.
- [26] D. Giovanelli, E. Farella, D. Fontanelli, and D. Macii, "Bluetooth-based indoor positioning through ToF and RSSI data fusion," in *2018 Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1-8, IEEE, Sept. 2018.
- [27] Q. Wan, T. Wu, K. Zhang, X. Liu, K. Cheng, J. Liu, and J. Zhu, "A high precision indoor positioning system of BLE AOA based on ISSS algorithm," *Measurement*, vol. 224, p. 113801, Jan. 2024.
- [28] A. Sonny, A. Kumar, and L. R. Cenkramaddi, "A survey of application of machine learning in wireless indoor positioning systems," *arXiv:2403.04333*, Mar. 2024.
- [29] D. R. Philips, E. Salami, H. Ramiah, and J. Kanesan, "Location accuracy optimization in Bluetooth Low Energy (BLE) 5.1-based indoor positioning system (IPS)—A machine learning approach," *IEEE Access*, vol. 11, pp. 140186-140201, 2023.
- [30] P. Kriz, F. Maly, and T. Kozel, "Improving indoor localization using bluetooth low energy beacons," *Mobile Inf. Syst.* 2016.1: 2083094, 2016.
- [31] A. H. M. Kamal, M. G. R. Alam, M. R. Hassan, T. S. Apon, & M. M. Hassan, "Explainable indoor localization of BLE devices through RSSI using recursive continuous wavelet transformation and XGBoost classifier," *Future Gener. Comput. Syst.*, 141, 230-242, Dec. 2022.
- [32] W. Cui, L. Zhang, B. Li, J. Guo, W. Meng, H. Wang, and L. Xie, "Received Signal Strength Based Indoor Positioning Using a Random Vector Functional Link Network," *IEEE Trans. Ind. Informatics*, vol. 14, no. 5, pp. 1846-1855, May 2018.
- [33] P. Babakhani, T. Merk, M. Mahlig, I. Sarris, D. Kalogiros and P. Karlsson, "Bluetooth Direction Finding using Recurrent Neural Network," 2021 Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN), Lloret de Mar, Spain, pp. 1-7, Dec. 2021.
- [34] Z. Turgut and A. G. Kakisim, "An explainable hybrid deep learning architecture for WiFi-based indoor localization in Internet of Things environment," *Future Gener. Comput. Syst.*, vol. 151, pp. 196-213, Oct. 2023.
- [35] R. Talla-Chumpitaz, M. Castillo-Cara, L. Orozco-Barbosa, and R. Garcia-Castro, "A novel deep learning approach using blurring image techniques for Bluetooth-based indoor localisation," *Inform. Fusion*, vol. 91, pp. 173-186, Oct. 2022.
- [36] Y. H. Ho, Y. Liu, C. Zhang, Y. Sartayeva, and H. C. B. Chan, "Hybrid learning for mobile ad-hoc distancing/positioning using Bluetooth Low Energy," *IEEE Internet Things J.*, vol. 10, no. 14, pp. 12293-12307, Jul. 2023.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

- [38]H. Chefer, S. Gur, and L. Wolf, "Transformer Interpretability Beyond Attention Visualization," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 782–791, 2021.
- [39]N. Kitaev, Ł. Kaiser, and A. Levskaya, "Reformer: The Efficient Transformer," arXiv preprint arXiv:2001.04451, Feb. 2020.
- [40]W. Li, X. Meng, Z. Zhao, Z. Liu, C. Chen, and H. Wang, "LoT: A Transformer-Based Approach Based on Channel State Information for Indoor Localization," *IEEE Sensors J.*, vol. 23, no. 22, pp. 28205–28219, Nov. 2023.
- [41]H. C. Lee, S. C. Wang, and Y. K. Chen, "A BLE-Based Marathon Runner Timing System Using Multi-Directional Antennas to Improve Timing Stability," *IEEE Trans. Instrum. Meas.*, June 2024.
- [42]T. Dag and T. Arsan, "Received signal strength based least squares iteration algorithm for indoor localization," *Comput. Electr. Eng.*, vol. 66, pp. 114–126, Mar. 2018.



Wei Wu (Member, IEEE) received the B.E. degree in vehicle engineering from Hunan University, Hunan, China, in 2017, and the Ph.D. degree in industrial and manufacturing systems engineering from the University of Hong Kong, Hong Kong, China, in 2022.

He is currently an assistant research professor with the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China. His research interests include indoor localization, industrial Internet of Things, artificial intelligence, cyber-physical systems in the field of manufacturing and logistics.



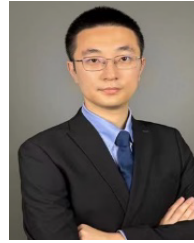
Didi Zhou received the B.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2020, where he is currently pursuing the M.S. degree in electronic and information. He is also a Research Assistant with The Hong Kong Polytechnic University, Hong Kong, China, since 2023. His current research

interests include indoor positioning systems, and industrial IoT.



Leidi Shen received the B.S. degree in Industrial Engineering from School of Mechanical Engineering, Zhejiang University, Hangzhou, China, in 2015, and the Ph.D. degree from Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong, China, in 2022.

He is currently a post-doctoral research fellow with the Department of Industrial and Manufacturing Systems Engineering (IMSE), The University of Hong Kong. His research interests include digital twin, smart manufacturing and logistics, industrial IoT, Blockchain and distributed ledgers.



Zhiheng Zhao (Senior Member, IEEE) received the M.E. and Ph.D. degrees in industrial and manufacturing systems engineering from The University of Hong Kong, Hong Kong, China, in 2013 and 2017, respectively.

He is currently a Research Assistant Professor at The Hong Kong Polytechnic University. His main research areas are the

analysis and optimization of Industry 4.0 manufacturing and spatial-temporal tracking for sustainable logistics. In the past five years, he has published over 30 academic papers and has secured several research funds, including National Natural Science Foundation of China and the Guangdong Provincial Natural Science Foundation, amounting to approximately 5 million HKD. His research achievements have also been applied in the industrial sector, securing over 3 million HKD in corporate collaboration projects.



Congbo Li (Senior Member, IEEE) received the B.S. and Ph.D. degrees from Chongqing University, China, in 2004 and 2009, respectively.

He is currently a Professor of mechanical engineering with Chongqing University. His research interests include modeling and optimization control, and energy-efficient manufacturing. His work has

resulted in one book and over 100 journal and conference proceeding articles. Dr. Li is currently serving as the Co-Chair for the IEEE Robotics and Automation TC on Sustainable Production Automation, a Guest Editor for Special Issue of the IEEE Transactions on Automation Science and Engineering, and Robotics and Computer-Integrated Manufacturing.



George Huang (Follow, IEEE) received the B.E. degree from Southeast University, Nanjing, China, in 1983, and the Ph.D. degree from Cardiff University, Cardiff, U.K., in 1991, both in mechanical engineering.

He is currently a Chair Professor with Department of Industrial and Systems Engineering, The Hong Kong Polytechnic

University, Hong Kong, China. Prior to this appointment, he was Chair Professor of Industrial and Systems Engineering and Head of Department in Department of Industrial and Manufacturing Systems Engineering at The University of Hong Kong. He has conducted research projects in areas of Smart Manufacturing, Logistics, and Construction through IoT-enabled Cyber-Physical Internet and Systems Analytics. His research has been supported with substantial government and industrial grants. He has published extensively and his works have been highly cited by research communities.

He serves as associate editors and editorial members for several international journals. He is Chartered Engineer (CEng), Fellow of IEEE, IISE, ASME, HKIE, IET and CILT.