WiID: Wi-Fi Based Intelligent Indoor Intrusion Detection with Tensor Decomposition

Yu-Ru Duan, Shaoshi Yang, *Senior Member, IEEE*, Hou-Yu Zhai, Xiao-Yang Wang, Jing-Sheng Tan, Yu-Song Luo and Sheng Chen, *Life Fellow, IEEE*

***Abstract*—Wi-Fi sensing has emerged as a promising paradigm for indoor intrusion detection, as it offers a robust and high-accuracy solution without the need for extra hardware deployment. However, existing schemes often compromise the inherent structure of channel state information (CSI) during feature extraction through lossy preprocessing, causing high false alarm rates and poor generalization. As a remedy, we propose a novel tensor-based framework for indoor intrusion detection, which enables reliable perception of fine-grained human activities through structured feature extraction, even in motion-ambiguous scenarios. Our approach integrates tensor-based feature extraction, multi-dimensional feature consolidation, and a modified deep learning (DL) network for accurate intrusion recognition. To validate our framework, we collected a comprehensive through-wall CSI dataset under the IEEE 802.11n standard, encompassing five common human activities in realistic scenarios. Extensive experimental results demonstrate the superior performance of our method compared to existing state-of-the-art schemes.**

***Index Terms*—Wi-Fi sensing, channel state information (CSI), indoor intrusion detection, tensor decomposition.**

I. Introduction

THE proliferation of ubiquitous sensing technologies presents a significant opportunity to reimagine indoor safety and security, particularly through the development of more intelligent and nuanced intrusion detection systems. Despite this potential, the existing intrusion detection field has been slow to move beyond conventional approaches, which primarily rely on devices such as cameras and passive infrared (PIR) sensors and exhibit significant drawbacks [1]. Vision-based systems [2], for instance, are limited to line-of-sight (LoS) scenarios, constrained by variable lighting conditions, and raise substantial privacy concerns. Similarly, sensor-based networks are expensive to deploy and scale, particularly in complex indoor layouts [3]. These limitations highlight the urgent need for a robust and cost-effective alternative.

Recent advancements in Wi-Fi sensing offer a promising solution to these challenges. This approach can utilize existing Wi-Fi infrastructure, eliminating the need to deploy dedicated devices or for individuals to carry specific devices. By analyzing the subtle perturbations in channel state information (CSI) induced by human movement, Wi-Fi based systems can achieve device-free human activity recognition (HAR). This capability facilitates passive, real-time monitoring in both LoS and non-line-of-sight (NLoS) scenarios, thus offering a robust method for modern indoor intrusion detection systems [4].

The development of Wi-Fi based intrusion detection shows a clear trajectory from employing simple statistical metrics to utilizing multifaceted features extracted and optimized by deep learning (DL) models. Unlike general Wi-Fi sensing tasks, intrusion detection requires extracting fine-grained features from subtle CSI variations, and mapping them to specific anomalous human activities. Early approaches primarily leveraged statistical features extracted from CSI, such as the mean and variance, to train classical machine learning (ML) classifiers. For instance, the authors of [5] employed these metrics with support vector machines (SVMs) to distinguish between static and dynamic states, while others first employed principal component analysis (PCA) to reduce CSI dimensionality [5]. However, these handcrafted features are inherently sensitive to environmental variations, and lack the capability to isolate subtle motion-induced changes from ambient interference.

As a remedy, research has shifted towards DL-based Wi-Fi sensing approaches, which can automatically learn hierarchical and more discriminative features directly from CSI. Early schemes primarily focused on learning temporal features from raw one-dimensional (1D) CSI sequences using recurrent neural networks (RNNs), such as long short-term memory (LSTM) [5], gated recurrent unit (GRU) [6] and attention-based bidirectional LSTM (ABLSTM) [7]. To further enhance feature richness, subsequent works employed convolutional neural network (CNN)-based methods to convert CSI into two-dimensional (2D) image representations to achieve HAR by extracting spatio-temporal features [8]. However, these methods inherently cause feature information loss by compromising the native multi-dimensional structure of CSI. Such loss significantly weakens their ability to capture subtle motion patterns, which are crucial for reliable intrusion detection.

Motivated by limitations of existing solutions and practical challenges, we propose a novel tensor-based framework for indoor intrusion detection. The framework first constructs a third-order tensor from the phase differences between adjacent antennas. Then, tensor decomposition is employed to extract the underlying low-rank feature components. These feature components are then processed by a multi-dimensional feature consolidation module to yield a set of compact features, which are inputted to a modified ABLSTM network for accurate classification. The main contributions are recapped as follows:

- We propose a novel tensor-based framework that enables accurate recognition of subtle human activities through structured feature extraction.

Y.-R. Duan, S. Yang (*Corresponding author*), H.-Y. Zhai, J.-S. Tan and Y.-S. Luo are with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, and the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing 100876, China (e-mail: {duanyuru09, shaoshi.yang, 2hy, jingsheng.tan, yusong.luo}@bupt.edu.cn).

X.-Y Wang is with the Department of Wireless and Terminal Technology, China Mobile Research Institute, Beijing 100053, China (e-mail: wangxiaoyangyjy@chinamobile.com)

S. Chen is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK (e-mail: sqc@ecs.soton.ac.uk).

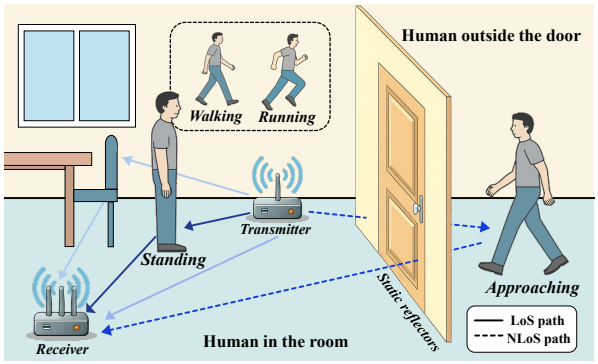

Fig. 1. System model of Wi-Fi based indoor intrusion detection.

- We collected a comprehensive through-wall CSI dataset under the IEEE 802.11n standard, which encompasses five common human activities in realistic scenarios.
- Extensive experimental results demonstrate the superior performance of our proposed scheme over existing state-of-the-art schemes.

II. System Model Characterization

We consider a Wi-Fi based passive intrusion detection system employing multiple-input multiple-output (MIMO) and orthogonal frequency-division multiplexing (OFDM) techniques. As illustrated in Fig. 1, a single-antenna transmitter participates in the wireless communication with the N_{R}-antenna receiver equipped with a uniform linear array (ULA). The presence of a human intruder is detected by monitoring the resultant variations in the wireless channel. To enable intrusion detection through the analysis of Wi-Fi channel variations, we develop a multi-class recognition framework that maps multi-dimensional CSI features to human activity labels.

A. Wi-Fi Based Channel Model

For the gth OFDM symbol, the received signal $Y_m(f_n,g) \in \mathbb{C}$ on the nth subcarrier at the mth antenna is modeled as:

$$Y_m(f_n,g) = H_m(f_n,g)X(f_n,g) + N_m(f_n,g) \quad (1)$$

where $X(f_n,g) \in \mathbb{C}$ represents the transmitted data symbol modulated on the nth subcarrier for the gth OFDM symbol, and $N_m(f_n,g) \sim \mathcal{CN}(0,\sigma^2)$ denotes the additive white Gaussian noise (AWGN). The time-varying channel response, $H_m(f_n,g) \in \mathbb{C}$, captures both quasi-static effects from the environment and dynamic variations from human movement, and can be decomposed as [6]:

$$H_m(f_n,g) = H_m^{\text{static}} + H_m^{\text{dynamic}}(f_n,g) + \Delta H_m^{\text{noise}}(f_n,g) \quad (2)$$

where $H_m^{\text{static}} \in \mathbb{C}$ represents the quasi-static environmental components, $H_m^{\text{dynamic}}(f_n,g) \in \mathbb{C}$ captures the human-induced dynamic variations, and $\Delta H_m^{\text{noise}}(f_n,g) \in \mathbb{C}$ denotes the residual channel estimation error, respectively.

B. Phase Difference Feature Construction

Through the pilot-assisted channel estimation method, we obtain CSI measurements and then extract features from them for accurate intrusion detection. The phase component derived from CSI is leveraged as the core feature due to its high sensitivity to subtle human motion [4]–[6]. While the amplitude component is more robust to noise, it is less sensitive to the slight channel variations caused by human movement. In contrast, the phase component better captures small-scale multipath effects, making it a superior feature for human motion detection [5].[1]

However, the raw phase measurements are corrupted by a combination of deterministic and stochastic distortions. These originate from key imperfections: sampling frequency offset (SFO), carrier frequency offset (CFO), and phase-locked loop (PLL). Formally, the measured phase on the nth subcarrier at the mth antenna for the gth OFDM symbol, denoted as $\tilde{\phi}_m(f_n,g)$, can be expressed as:

$$\tilde{\phi}_m(f_n,g) = \overline{\phi}_m(f_n,g) + n_s n + n_c g + \phi^{\text{PLL}} + \phi^{\text{noise}} \quad (3)$$

where $m \in \{1,\cdots,N_{\mathrm{R}}\}$, $n \in \{0,1,\cdots,N_{\mathrm{c}}-1\}$ with N_{c} denoting the number of subcarriers, $\overline{\phi}_m(f_n,g) \in \mathbb{R}$ denotes the true phase, $\phi^{\text{PLL}} \in \mathbb{R}$ is the initial phase offset caused by the PLL, and $\phi^{\text{noise}} \in \mathbb{R}$ represents environmental noise. The terms $n_s n$ and $n_c g$ denote deterministic phase shifts introduced by SFO and CFO, respectively [5]. Since the phase error components $n_s n$ and $n_c n$ are common across all receiving antennas, we can eliminate them by taking the phase difference between adjacent antennas, which is defined as:

$$\begin{aligned}\Delta\phi_{m'}(f_n,g) &= \tilde{\phi}_{m'+1}(f_n,g) - \tilde{\phi}_{m'}(f_n,g) \\ &= \Delta\overline{\phi}_{m'}(f_n,g) + \Delta\phi^{\text{PLL}} + \Delta\phi^{\text{noise}} + \epsilon_{m'}(f_n,g) \quad (4)\end{aligned}$$

where $m' \in \{1,\cdots,N_{\mathrm{R}}-1\}$, $\Delta\overline{\phi}_{m'}(f_n,g) = \overline{\phi}_{m'+1}(f_n,g) - \overline{\phi}_{m'}(f_n,g)$ is the true phase difference, $\Delta\phi^{\text{PLL}}$ is the constant offset induced by the PLL, $\Delta\phi^{\text{noise}}$ is the noise difference, and $\epsilon_{m'}(f_n,g) \sim \mathcal{N}(0,\sigma_0^2)$ represents phase noise.

To jointly model multi-dimensional dependencies and enhance structured signal patterns, we first convert each time series into a 2D Hankel matrix to preserve the time-shift patterns and reveal the low-rank structure of repetitive human motions. Specifically, for the nth subcarrier, the Hankel matrix constructed from the phase difference sequence between the m'th and $(m'+1)$th antennas over all OFDM symbols, $\mathbf{H}_{n,m'} \in \mathbb{R}^{I\times J}$, where $I \in \mathbb{Z}^+$ and $J \in \mathbb{Z}^+$ denote the window length and embedding dimension, respectively, is given as:

$$\mathbf{H}_{n,m'} = \begin{bmatrix} \Delta\phi_{m'}(f_n,0) & \cdots & \Delta\phi_{m'}\left(f_n,\frac{G-1}{2}\right) \\ \Delta\phi_{m'}(f_n,1) & \cdots & \Delta\phi_{m'}\left(f_n,\frac{G+1}{2}\right) \\ \vdots & \vdots & \vdots \\ \Delta\phi_{m'}\left(f_n,\frac{G-1}{2}\right) & \cdots & \Delta\phi_{m'}(f_n,G-1) \end{bmatrix} \quad (5)$$

Here $g \in \{0,\cdots,G-1\}$, and we set $I = J = (G+1)/2$ to balance the temporal and delay dimensions in Hankel matrices.

To capture joint features across multiple dimensions and separate low-rank structures, we construct a third-order tensor $\mathcal{Y} \in \mathbb{R}^{I\times J\times K}$ by stacking all Hankel matrices $\{\mathbf{H}_{n,m'}\}_{m'=1}^{N_{\mathrm{R}}-1}$ across subcarriers and antenna pairs, where $K = N_{\mathrm{c}} \times (N_{\mathrm{R}}-1)$ [9]. This formulation enables comprehensive modeling of

[1] Other motion-related features, such as Doppler and direction of arrival, are not considered in this paper, as their extraction typically requires specialized hardware or large antenna arrays, limiting their practicality in lightweight sensing systems.

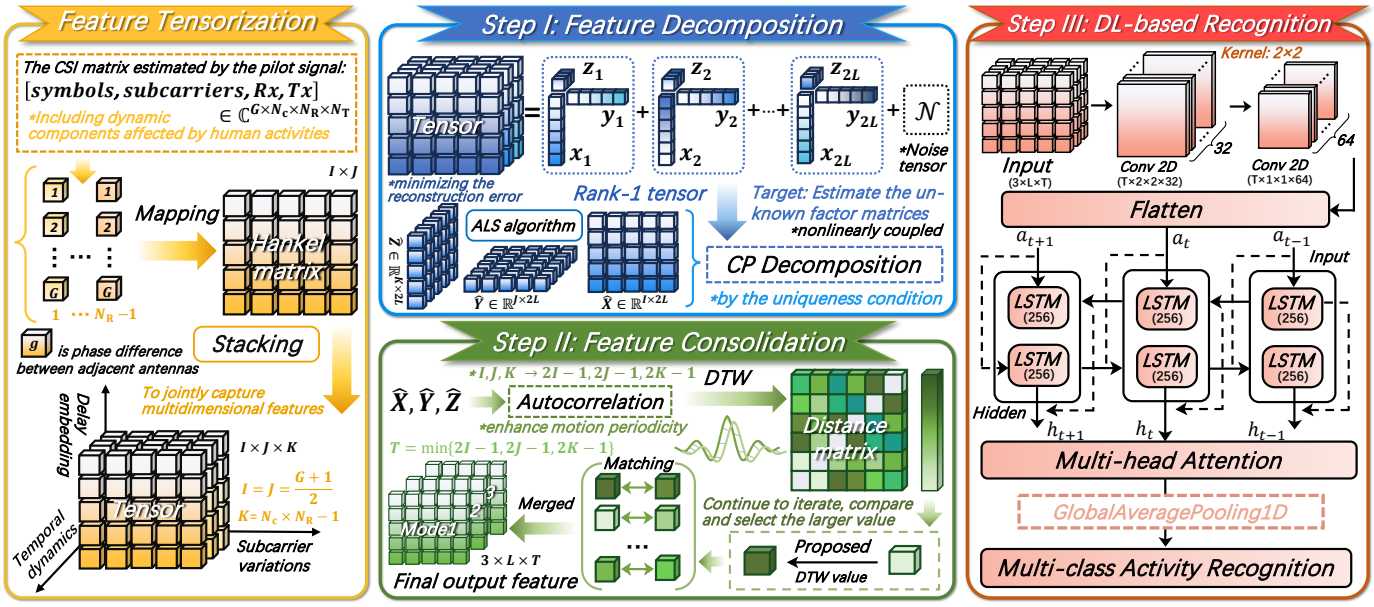

Fig. 2. The proposed tensor-based indoor intrusion detection framework. After collecting CSI measurements, we construct phase differences between adjacent antennas into third-order tensors. In *Step I*, CP decomposition is applied to extract mode-separated and low-rank features. In *Step II*, autocorrelation analysis and dynamic time warping (DTW) based similarity matching are used to consolidate redundant components into motion-representative features. In *Step III*, the selected features are fed into a modified ABLSTM network for fine-grained activity classification.

global temporal evolution (I) and local delay structures (J), and subcarrier-domain variations induced by human motion (via K), thereby providing a structured input for subsequent low-rank decomposition.

III. Tensor-Based Intrusion Detection Framework

The proposed tensor-based framework is shown in Fig. 2. To enable fine-grained human motion detection, we extract human-induced and noise-robust features from $\mathcal{Y}$, which are then fed into a DL network for classification. Specifically, tensor decomposition is employed to extract mode-separated and low-rank features from $\mathcal{Y}$. Then, we refine component-wise features and merge structurally similar components to reduce redundancy. Finally, the extracted features are passed into a modified ABLSTM network for detection.

A. Step I: Tensor-based Feature Decomposition

Directly operating on the high-dimensional tensor $\mathcal{Y}$ to extract features is computationally expensive and may include noise. Thus, to extract compact and mode-separated features, we decompose $\mathcal{Y}$ into a set of low-rank latent components using CANDECOMP-PARAFAC (CP) decomposition. Via CP decomposition, $\mathcal{Y}$ can be factorized into a sum of rank-one tensors, each representing an underlying factor. The uniqueness of this decomposition is guaranteed under mild conditions.

The decomposition rank R must be selected prior to decomposition, as it directly determines the number of resolvable targets. For a signal comprising L dominant propagation paths, its representation via the stacked Hankel matrices in (5) results in an effective rank of approximately $2L$ [10]. Accordingly, to capture these motion-related components, we set the rank to $R = 2L$. Then, the decomposition problem is formulated as the minimization of the following reconstruction error [11]:

$$\textbf{P1:} \quad \min_{\mathbf{X},\mathbf{Y},\mathbf{Z}} \left\| \mathcal{Y} - \sum\nolimits_{l=1}^{2L} \mathbf{x}_l \circ \mathbf{y}_l \circ \mathbf{z}_l \right\|_{\mathrm{F}}^2 \tag{6}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. The terms $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_{2L}] \in \mathbb{R}^{I\times 2L}$, $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_{2L}] \in \mathbb{R}^{J\times 2L}$, and $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_{2L}] \in \mathbb{R}^{K\times 2L}$ denote the factor matrices corresponding to temporal dynamics, delay embeddings, and subcarrier-domain variations, respectively. The operator $\circ$ denotes the vector outer product.

To solve **P1** efficiently, we employ the alternating least squares (ALS) algorithm [12]. ALS is initialized with factor matrices obtained via higher-order singular value decomposition (HOSVD), and it then updates each matrix, $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, in turn repeatedly until the solution converges. We use the update for factor matrix $\mathbf{X}$ as an illustrative example. $\mathcal{Y}$ is first unfolded into its mode-1 matricization, $\mathcal{Y}_{(1)}$:

$$\mathcal{Y}_{(1)} \in \mathbb{R}^{I\times(JK)} = \mathbf{X}(\mathbf{Z} \odot \mathbf{Y})^{\mathrm{T}} \tag{7}$$

where $\odot$ denotes the Khatri-Rao product. This transforms **P1** into the following least-squares problem with respect to $\mathbf{X}$:

$$\textbf{P2:} \quad \min_{\mathbf{X}} \left\| \mathcal{Y}_{(1)} - \mathbf{X}(\mathbf{Z} \odot \mathbf{Y})^{\mathrm{T}} \right\|_{\mathrm{F}}^2 \tag{8}$$

The closed-form solution is given by $\hat{\mathbf{X}} = \mathcal{Y}_{(1)}(\mathbf{Z} \odot \mathbf{Y})(\mathbf{Z}^{\mathrm{T}}\mathbf{Z} * \mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{\dagger} \in \mathbb{R}^{I\times 2L}$, where $\dagger$ denotes the Moore-Penrose pseudo-inverse and $*$ denotes the Hadamard (element-wise) product. The factor matrices $\hat{\mathbf{Y}} \in \mathbb{R}^{J\times 2L}$ and $\hat{\mathbf{Z}} \in \mathbb{R}^{K\times 2L}$ are updated analogously using the mode-2 and mode-3 matricizations of the tensor, namely, $\mathcal{Y}_{(2)}$ and $\mathcal{Y}_{(3)}$, respectively.

By exploiting the structure of the Khatri-Rao product, the pseudo-inverse calculation is simplified from a large-scale $JK \times 2L$ matrix to a $2L \times 2L$ matrix. This reduces the complexity per iteration from $O(JK(2L)^2)$ to $O(L^3)$, given that $L \ll \min\{JK, IK, IJ\}$.

B. Step II: Multi-Dimensional Feature Consolidation

The factor matrices $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, and $\hat{\mathbf{Z}}$ each contain $2L$ components, yet only about L of them capture truly informative

motion patterns. The remaining components are often redundant or noisy, and directly using them introduces unnecessary complexity. To address this, we propose to identify and consolidate similar components into L representative ones for each mode, while preserving the overall multi-dimensional feature consolidation structure.

Let $u \in \{1, 2, 3\}$ denote the mode index corresponding to factor matrices $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, $\hat{\mathbf{Z}}$, where each mode u contains a set of $2L$ components $\{\hat{\mathbf{f}}_l^{(u)}\}_{l=1}^{2L}$, with $\hat{\mathbf{f}}_l^{(u)} \in \mathbb{R}^{d_u}$ and $d_u \in \{I, J, K\}$ denoting the component length in mode u. To enhance periodicity and stabilize motion pattern representation[2], we compute the full autocorrelation sequence of each component, denoted by $\hat{\mathbf{a}}_l^{(u)} \in \mathbb{R}^{2d_u-1}$:

$$\hat{\mathbf{a}}_l^{(u)}[\tau] = \sum\nolimits_{n=0}^{d_u-1-|\tau|} \hat{\mathbf{f}}_l^{(u)}[n] \cdot \hat{\mathbf{f}}_l^{(u)}[n + |\tau|] \tag{9}$$

where $\tau \in \{1-d_u, \cdots, d_u-1\}$ is the lag index.

Then, our goal is to identify and consolidate redundant components characterized by shape similarity. To effectively capture similarity despite temporal misalignment, we adopt dynamic time warping (DTW) [12], which robustly aligns periodic sequences and accurately measures their structural similarity. The DTW-based dissimilarity between components i and j is computed as:

$$D_{ij} = \min_{\mathcal{W}} \sum\nolimits_{(p,q)\in\mathcal{W}} |\hat{\mathbf{a}}_i[p] - \hat{\mathbf{a}}_j[q]| \tag{10}$$

where $\mathcal{W} = \{(p_k, q_k)\}_{k=1}^{K}$ denotes a valid warping path of K-length, ensuring valid alignment. For each sample, based on $\{D_{ij}\}_{i,j=1}^{2L}$, we perform stable matching to iteratively merge each pair into a single representative sequence, resulting in L components per mode.

Due to differing sequence lengths among modes, we truncate the feature matrices to a common length $T = \min_u(2d_u - 1)$ to simplify further processing. The resulting matrices are then stacked along a new axis to form the final input three-dimensional (3D) feature $\mathbf{F} \in \mathbb{R}^{3\times L\times T}$, which is used as input to the downstream DL network.

C. Step III: DL-based Multi-class Activity Recognition

The high inter-class similarity and subtle differences in human motion patterns make it challenging to conduct the multi-class classification task. To fully leverage the structured 3D feature obtained in the previous steps, we design a modified ABLSTM network that effectively models features across all modes. Specifically, to preserve temporal continuity and enable early-stage fusion across modes and components, the input feature $\mathbf{F}$ is reshaped into a time-distributed format. Subsequently, we apply time-distributed 2D convolutional layers to the 3D input features, extracting local spatial patterns across mode and component dimensions at each time step.

Then, we stack multiple bidirectional LSTM (Bi-LSTM) layers to model temporal evolution in both directions, capturing contextual information from both past and future. To better distinguish activities with subtle differences despite similar overall dynamics, residual connections are employed between Bi-LSTM layers to facilitate deeper temporal feature learning. The output of the lth Bi-LSTM layer at time step t is:

$$\mathbf{h}_t^{(l)} = \text{Bi-LSTM}^{(l)}(\mathbf{h}_t^{(l-1)}), \quad \mathbf{h}_t^{(l)} \in \mathbb{R}^{d_h} \tag{11}$$

where $d_h \in \mathbb{Z}^+$ denotes the hidden size.

To enhance the capability to discriminate subtle differences among activities, we apply a multi-head attention mechanism to adaptively emphasize multiple motion-sensitive segments within the temporal sequence. Finally, global average pooling aggregates the weighted temporal features, followed by a fully connected classification layer that outputs activity predictions, enabling robust and effective multi-class activity recognition.

IV. Experiments

This section details the experimental setup and results to validate the effectiveness of the proposed scheme for indoor intrusion detection. Existing public datasets for indoor intrusion detection rarely address fine-grained scenarios and often overlook critical pre-intrusion activities that are essential for early detection. To bridge this gap, we conducted experiments on a CSI dataset collected from a real-world indoor environment, which captures the progressive stages of intrusion[3]. Specifically, the dataset comprises five representative human activities and can be structured into three categories: 1) no-intrusion, where *no activity* is present in the monitored area; 2) pre-intrusion, where a person *approaches* the area but does not enter; and 3) intrusion, where a person remains inside and performs actions such as *standing, walking, or running*.

To comprehensively evaluate the performance, the proposed scheme was benchmarked against several state-of-the-art methods, including WiHGR [6], which employs a modified GRU model, and WiDSAR [8], which is based on a CNN-LSTM architecture; as well as classic DL models, LSTM [5] and ABLSTM [7]. We set L to 3 and the evaluation was performed using a comprehensive suite of metrics, including detection accuracy ↑, precision ↑, recall ↑, F1 score ↑, false positive rate (FPR) ↓, false negative rate (FNR) ↓, and area under the curve (AUC) ↑, where '↓' indicates 'the smaller the better', while '↑' indicates 'the larger the better'.

Firstly, we evaluate the overall intrusion detection performance under the binary classification scenario in Table I. The proposed scheme surpasses all the baseline methods by achieving the highest detection accuracy and F1 score, indicating balanced and reliable classification. Notably, it also achieves the highest precision, AUC and the lowest FPR, reflecting the ability to detect intrusions while suppressing false alarms.

Then, we evaluate the capability of all the methods to discriminate between different types of intrusion-related human activities. As illustrated in Fig. 3, the proposed scheme achieves the highest classification accuracy in the majority of five predefined scenarios. This demonstrates its ability to effectively capture the critical boundary between non-intrusion,

[2]Human motion-induced variations typically exhibit pseudo-periodic fluctuations. Autocorrelation is effective in capturing these patterns and offers more stable features than conventional transforms.

[3]For brevity, detailed information regarding the experimental setup and the CSI acquisition methodology, along with the full dataset, can be found at: https://github.com/ddduan0817/WiFi-CSI-indoor-intrusion-detection-dataset.

TABLE I
BINARY CLASSIFICATION PERFORMANCE COMPARISON (INTRUSION: STANDING, WALKING, RUNNING VS NON-INTRUSION: NO ACTIVITY, APPROACHING)

Model	Detection Accuracy↑ (%)	Precision↑	Recall↑	F1 Score↑	FPR↓	FNR↓	AUC↑
LSTM Model [5]	88.61	0.8829	0.9506	0.9155	0.2331	0.0494	0.9535
ABLSTM Model [7]	89.90	0.8861	0.9671	0.9248	0.2297	0.0329	0.9676
WiHGR Scheme [6]	88.61	0.8631	0.9799	0.9178	0.2872	0.0201	0.9672
WiSDAR Scheme [8]	91.81	0.8970	**0.9872**	0.9399	0.2095	**0.0128**	0.9783
Proposed Scheme	**95.26**	**0.9884**	0.9378	**0.9625**	**0.0203**	0.0622	**0.9863**

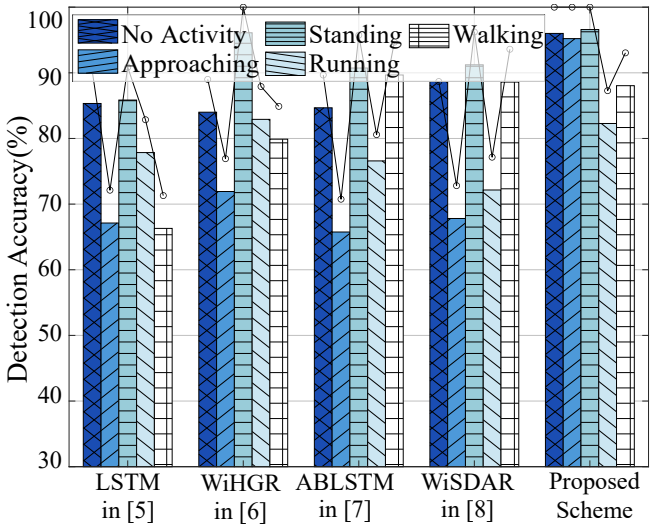

Fig. 3. Accuracy comparison across five intrusion scenarios.

pre-intrusion, and intrusion states. This fine-grained sensitivity to subtle changes highlights the clear boundary discrimination capability of our scheme. These results validate the effectiveness of our design in preserving structured features and leveraging the modified DL model for fine-grained behavior discrimination.

Finally, we evaluate the model robustness against signal disturbances by adding Gaussian phase noise ($\sigma_0^2 \in [0, 1]$) to the phase difference data in (4). As depicted in Fig. 4, the accuracy decreases for all the methods as noise variance increases. Notably, the proposed scheme consistently maintains the highest accuracy across the entire noise range. In contrast, other methods not only start with lower accuracy but also suffer greater degradation as the noise increases. This robustness of our proposed scheme stems from the joint effect of its feature extraction and modified DL network, which effectively suppresses noise while preserving informative patterns.

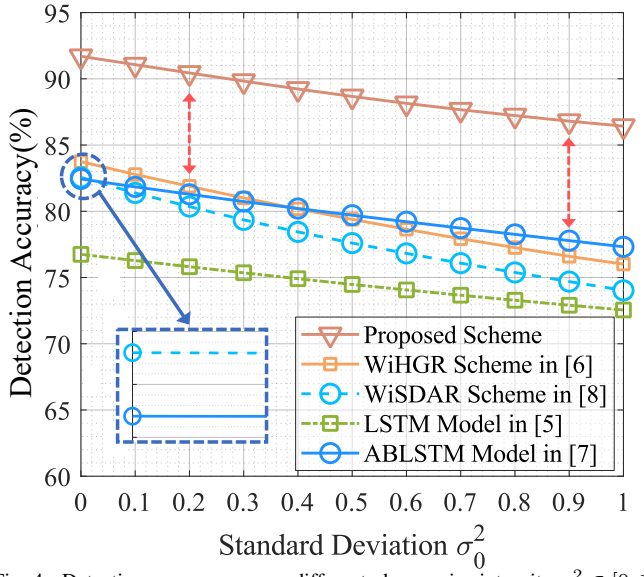

Fig. 4. Detection accuracy versus different phase noise intensity $\sigma_0^2 \in [0, 1]$.

V. CONCLUSION

To achieve high-precision and robust Wi-Fi based intrusion detection, we have proposed a tensor-based framework, which can capture subtle motion-induced features by transforming CSI into multi-dimensional tensors. Our approach integrates tensor-based feature extraction, multi-dimensional feature consolidation, and a modified DL network. Extensive experiments on a comprehensive through-wall CSI dataset have validated the superior performance and generalization of our proposed framework compared to existing state-of-the-art methods.

REFERENCES

[1] I. Ahmad, A. Ullah, and W. Choi, "WiFi-based human sensing with deep learning: Recent advances, challenges, and opportunities," *IEEE Open J. Communi. Society*, vol. 5, pp. 3595–3623, Jul. 2024.
[2] Q. Lei, *et al.*, "A survey of vision-based human action evaluation methods," *Sensors*, vol. 19, no. 19, pp. 4129–4156, Sep. 2019.
[3] T. Choubisa, *et al.*, "An optical-camera complement to a PIR sensor array for intrusion detection and classification in an outdoor environment," *Proc. LCN Workshops 2017* (Singapore), Oct. 9, 2017, pp. 44–52.
[4] T. Ropitault, *et al.*, "IEEE 802.11bf WLAN sensing procedure: Enabling the widespread adoption of WiFi sensing," *IEEE Communi. Standards Mag.*, vol. 8, no. 1, pp. 58–64, Mar. 2024.
[5] S. Yousefi, *et al.*, "A survey on behavior recognition using WiFi channel state information," *IEEE Communi. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.
[6] W. Meng, *et al.*, "WiHGR: A robust WiFi-based human gesture recognition system via sparse recovery and modified attention-based BGRU," *IEEE Internet of Things J.*, vol. 9, no. 12, pp. 10272–10282, Jun. 2022.
[7] Z. Chen, *et al.*, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.
[8] F. Wang, W. Gong, and J. Liu, "On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach," *IEEE Internet of Things J.*, vol. 6, no. 2, pp. 2035–2047, Apr. 2019.
[9] B. Zhao, *et al.*, "A tensor-based joint AoA and ToF estimation method for Wi-Fi systems," *IEEE Wireless Communi. Lett.*, vol. 10, no. 11, pp. 2543–2546, Nov. 2021.
[10] L. D. Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in Rank-(L_r, L_r, 1) terms," *SIAM J. Matrix Analysis and Applications*, vol. 32, no. 4, pp. 1451–1474, 2011.
[11] J. Ding, *et al.*, "Three-dimensional indoor localization and tracking for mobile target based on WiFi sensing," *IEEE Internet of Things J.*, vol. 9, no. 21, pp. 21687–21701, Nov. 2022.
[12] J. C. H. Soto, *et al.*, "Wi-Fi CSI-based human presence detection using DTW features and machine learning," *Proc. LATINCOM 2022* (Rio de Janeiro, Brazil), Nov. 30-Dec. 2, 2022, pp. 1–6.